



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

Dipartimento di FISICA
Corso di Laurea Magistrale in FISICA
Insegnamento di METODI PROBABILISTICI DELLA FISICA

Nicola Cufaro Petroni

LEZIONI DI
PROBABILITÀ
E
PROCESSI STOCASTICI

anno accademico 2016/17

Copyright © 2016 Nicola Cufaro Petroni
Università degli Studi di Bari *Aldo Moro*
Dipartimento di Matematica
via E. Orabona 4, 70125 Bari

Indice

I	Probabilità	7
1	Spazi di probabilità	9
1.1	Campioni	9
1.2	Eventi	13
1.3	Probabilità	18
1.4	Probabilità condizionate	22
1.5	Eventi indipendenti	24
2	Misure di probabilità	27
2.1	Probabilità su \mathbf{N}	27
2.1.1	Spazi finiti e numerabili	27
2.1.2	Modello di Bernoulli	29
2.2	Probabilità su \mathbf{R}	34
2.2.1	Funzioni di distribuzione	34
2.2.2	Distribuzioni discrete	37
2.2.3	Distribuzioni assolutamente continue: densità	38
2.2.4	Distribuzioni singolari	43
2.2.5	Miscele	43
2.3	Probabilità su \mathbf{R}^n	45
2.3.1	Funzioni di distribuzione multivariate	45
2.3.2	Densità multivariate	48
2.3.3	Distribuzioni marginali	49
2.3.4	Copule	52
2.4	Probabilità su \mathbf{R}^∞ e \mathbf{R}^T	55
3	Variabili aleatorie	59
3.1	Variabili aleatorie	59
3.1.1	Misurabilità	59
3.1.2	Leggi e distribuzioni	60
3.1.3	Costruzione di v -a	63
3.2	Vettori aleatori e processi stocastici	65
3.2.1	Elementi aleatori	65
3.2.2	Distribuzioni e densità congiunte e marginali	67

3.2.3	Indipendenza di v - a	71
3.2.4	Decomposizione di v - a binomiali	73
3.3	Valore d'attesa	76
3.3.1	Integrazione e valore d'attesa	76
3.3.2	Cambio di variabili di integrazione	79
3.3.3	Varianza e covarianza	85
3.4	Condizionamento	92
3.4.1	Distribuzioni condizionate	92
3.4.2	Attese condizionate	95
3.4.3	Stima ottimale in media quadratica	101
3.5	Trasformazioni di v - a	102
3.5.1	Funzioni di v - a	102
3.5.2	Somme di v - a indipendenti	106
4	Teoremi limite	109
4.1	Convergenza	109
4.2	Funzioni caratteristiche	112
4.2.1	Definizioni e proprietà	112
4.2.2	Leggi Gaussiane	118
4.2.3	Composizione e decomposizione di leggi	121
4.3	Leggi dei Grandi Numeri	123
4.4	Teoremi Gaussiani	127
4.5	Teoremi di Poisson	129
4.6	Applicabilità dei teoremi limite	134
II	Processi Stocastici	137
5	Concetti generali	139
5.1	Leggi e identità dei p - s	139
5.2	Attese e correlazioni	142
5.3	Convergenza e continuità	143
5.4	Derivazione e integrazione in mq	144
5.5	Stazionarietà ed ergodicità	146
5.6	Spettro di potenza	150
6	Costruzioni euristiche	153
6.1	Processo di Poisson	153
6.1.1	Processi di punto e rinnovi	153
6.1.2	Processo di Poisson	158
6.1.3	Processo di Poisson compensato	164
6.1.4	Processo di Poisson composto	165
6.1.5	Rumore granulare	169
6.2	Processo di Wiener	171

6.2.1	<i>Random walk</i>	171
6.2.2	Processo di Wiener	172
6.2.3	Processo di Wiener geometrico	178
6.3	Rumore bianco	179
6.4	Moto Browniano	183
6.4.1	Einstein (1905)	184
6.4.2	Langevin (1908)	186
7	Processi di Markov	189
7.1	Processi di Markov	189
7.1.1	Proprietà di Markov	189
7.1.2	Equazioni di Chapman-Kolmogorov	193
7.1.3	Processi ad incrementi indipendenti	195
7.1.4	Stazionarietà e omogeneità	198
7.1.5	Ergodicità	200
7.1.6	Processi di Lévy	202
7.1.7	Continuità e salti	203
7.1.8	Processi di Poisson, Wiener e Cauchy	205
7.1.9	Processi di Ornstein-Uhlenbeck	209
7.1.10	Processi Gaussiani non Markoviani	211
7.2	Equazioni di evoluzione	213
7.2.1	Equazioni <i>in avanti</i>	214
7.2.2	Equazioni <i>all'indietro</i>	218
7.2.3	Classi di equazioni <i>in avanti</i>	219
7.2.4	Equazioni di processi particolari	223
8	Elementi di calcolo stocastico	229
8.1	Rumore bianco Wieneriano	229
8.2	Integrali stocastici	232
8.2.1	Integrale di Wiener	233
8.2.2	Integrale di Itô	235
8.3	Calcolo stocastico di Itô	237
8.3.1	Regole elementari di integrazione	237
8.3.2	Attese e covarianze	241
8.3.3	Infinitesimi stocastici	242
8.3.4	Regole di differenziazione e di integrazione per parti	244
8.4	Equazioni differenziali stocastiche (<i>EDS</i>)	247
8.4.1	Differenziali stocastici e formula di Itô	247
8.4.2	<i>EDS</i> e loro soluzioni	248
8.4.3	<i>EDS</i> ed equazioni di Fokker-Planck	249
8.5	Forme particolari di <i>EDS</i>	251
8.5.1	<i>EDS</i> a coefficienti costanti	251
8.5.2	<i>EDS</i> a coefficienti dipendenti solo dal tempo	252

8.5.3	<i>EDS</i> con diffusione lineare in x	253
8.5.4	<i>EDS</i> con trascinamento lineare in x	256
9	Teoria dinamica del moto Browniano	259
9.1	Particella Browniana libera	259
9.2	Confronto con Einstein-Smoluchowski	262
9.3	Markovianità di Ornstein-Uhlenbeck	264
9.4	Particella Browniana in un campo di forze	266
9.5	Distribuzione di Boltzmann	269
III	Appendici	275
A	Consistenza (Sez. 2.3.4)	277
B	Disuguaglianze (Sez. 3.3.2)	285
C	Paradosso di Bertrand (Sez. 3.5.1)	289
D	Spazi L^p di v - a (Sez. 4.1)	293
E	Momenti e cumulanti (Sez. 4.2.1)	295
F	Teoremi limite binomiali (Sez. 4.3)	299
G	Processi di punto con intensità non uniforme (Sez. 6.1.1)	303
H	Paradossi del calcolo stocastico (Sez. 6.4.2)	305
I	Processi pseudo-Markoviani (Sez. 7.1.2)	311
J	Moto Browniano frazionario (Sez. 7.1.10)	315
K	Equazioni del processo di Ornstein-Uhlenbeck (Sez. 7.2.4)	317
L	Integrale di Stratonovich (Sez. 8.2.2)	321
	Indice analitico	323

Parte I
Probabilità

Capitolo 1

Spazi di probabilità

1.1 Campioni

L'origine del calcolo delle probabilità (verso la metà del XVII secolo) è strettamente legata a celebri problemi di gioco d'azzardo dai quali partiremo anche noi per introdurre i primi concetti. Il caso più semplice è quello del lancio di una moneta nel quale si osserva uno dei due possibili risultati: la moneta cade mostrando la faccia con la *testa* (T); oppure la moneta cade mostrando la faccia con la *croce* (C). Dire che la moneta è *equa* significa che non vi è ragione per supporre che uno dei due risultati sia favorito rispetto all'altro. In tal caso T e C sono *equiprobabili*, e per dare una veste quantitativa a questa affermazione si usa attribuire ad ognuno una **probabilità** corrispondente a una frazione dell'unità, per cui nel nostro caso avremo:

$$p = \mathbf{P}\{T\} = \frac{1}{2} \quad q = \mathbf{P}\{C\} = \frac{1}{2}$$

Osserviamo che $p + q = 1$: questo riflette il fatto che *con certezza* (ossia con probabilità eguale ad 1) si verifica uno dei due casi T o C , e che non vi sono altre possibilità. Considerazioni del tutto analoghe possono essere adattate al caso di un dado *equo*, e conducono alla seguente attribuzione di probabilità alle sei facce che qui indicheremo con le cifre romane I, II, \dots, VI :

$$p_1 = \mathbf{P}\{I\} = \frac{1}{6}; \quad \dots \quad ; p_6 = \mathbf{P}\{VI\} = \frac{1}{6}$$

Ovviamente anche in questo caso si ha $p_1 + \dots + p_6 = 1$.

Dagli esempi precedenti si ricava pertanto l'idea che, almeno per i casi elementari, si possano attribuire delle probabilità mediante una semplice *enumerazione*, una procedura che conduce alla cosiddetta **definizione classica** della probabilità di una proposizione A che non riguardi necessariamente un solo risultato possibile: ad esempio per un dado sia A l'affermazione "*esce una faccia contrassegnata da un numero pari*". In tal caso è intuitivo sommare le probabilità dei possibili risultati per i quali A si verifica: in pratica, si enumerano i risultati *possibili* – giudicati equiprobabili in base a qualche ipotesi – e quelli *favorevoli* all'evento in considerazione,

e si attribuisce all'evento A la probabilità

$$P\{A\} = \frac{\text{numero dei casi favorevoli}}{\text{numero dei casi possibili}}$$

Si noti che, come già negli esempi precedenti, la probabilità assegnata ad A è un numero positivo e compreso fra 0 ed 1. Pertanto, se per il lancio di un dado equo consideriamo gli eventi $A =$ “appare una faccia contrassegnata da un numero pari”, $B =$ “appare una faccia contrassegnata da un multiplo di tre”, e $C =$ “appare una faccia diversa da VI”, una semplice enumerazione ci porta a concludere che, essendo 6 i casi equiprobabili possibili, e rispettivamente 3, 2 e 5 i casi favorevoli ad A , B e C , si avrà

$$P\{A\} = \frac{1}{2}, \quad P\{B\} = \frac{1}{3}, \quad P\{C\} = \frac{5}{6}$$

Se invece consideriamo un lancio di *due* dadi non truccati i risultati elementari possibili saranno 36, cioè quante sono le coppie *ordinate* (n, m) dove n ed m possono assumere i 6 valori I, \dots, VI . L'ipotesi che i dadi siano equi vuol dunque dire ora che i 36 eventi elementari $(I, I); (I, II); \dots; (VI, VI)$ sono tutti equiprobabili e pertanto che

$$P\{I, I\} = \frac{1}{36}, \quad P\{I, II\} = \frac{1}{36}, \quad \dots \quad ; \quad P\{VI, VI\} = \frac{1}{36}$$

Sempre per enumerazione si può verificare allora che all'evento $A =$ “non appare la coppia (VI, VI) ” spetta la probabilità

$$P\{A\} = \frac{35}{36}$$

Dalla discussione precedente possiamo dunque trarre una prima conclusione: la probabilità di un risultato sperimentale può essere valutata con un numero compreso tra 0 ed 1; il valore 1 indica la *certezza* del suo verificarsi, il valore 0 la sua *impossibilità* e i valori intermedi rappresentano tutti gli altri casi. Queste assegnazioni di probabilità ai singoli risultati permettono (almeno nei casi più semplici) di calcolare poi le probabilità di eventi più complicati mediante un conteggio di risultati equiprobabili. È evidente quindi l'importanza di stabilire preliminarmente quale è l'*insieme dei possibili risultati dell'esperimento*, ma va detto subito che questo metodo diretto di calcolo diviene rapidamente impraticabile se il numero di tali risultati cresce troppo. Ad esempio le possibili successioni (senza ripetizioni) delle 52 carte di un mazzo di carte francesi sono

$$52 \cdot 51 \cdot 50 \cdot \dots \cdot 2 \cdot 1 = 52! \simeq 8 \cdot 10^{67}$$

un numero enorme che rende vana in questo caso qualunque speranza di risolvere problemi con conteggi espliciti.

Definizione 1.1. Chiameremo *spazio dei campioni* l'insieme Ω (finito o infinito) costituito da tutti i possibili risultati ω di un esperimento.

Si noti che Ω non è necessariamente un insieme di numeri: in generale i suoi elementi possono essere oggetti di varia natura. Nel caso dei nostri esempi precedenti lo spazio dei campioni $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$ era uno spazio *finito* con cardinalità N : così per un lancio di una moneta esso è composto di soli due elementi

$$\Omega = \{T, C\}; \quad N = 2$$

mentre per il lancio di un dado si ha

$$\Omega = \{I, II, \dots, VI\}; \quad N = 6$$

Se invece l'esperimento consiste in due lanci di una moneta si avrebbe

$$\Omega = \{TT, TC, CT, CC\}; \quad N = 4$$

e nel caso di n lanci

$$\Omega = \{\omega = (a_1, \dots, a_n) : a_i = T \text{ oppure } C\}; \quad N = 2^n$$

I casi più noti di spazi dei campioni *infiniti* sono invece l'insieme dei numeri interi \mathbf{N} , l'insieme dei numeri reali \mathbf{R} , l'insieme delle n -ple di numeri reali \mathbf{R}^n , l'insieme \mathbf{R}^∞ delle successioni di numeri reali e l'insieme \mathbf{R}^T delle funzioni da un insieme T in \mathbf{R} . Avremo modo nel seguito di discutere più in dettaglio la natura di questi spazi.

Per gli spazi Ω finiti per i quali possiamo pensare di adottare la *definizione classica* della probabilità è ovviamente cruciale conoscere innanzitutto la cardinalità N , cioè il numero dei *casi possibili*. Introdurremo quindi alcuni esempi:

Esempio 1.2. Consideriamo una scatola contenente M palline numerate (quindi distinguibili) ed estraiamo successivamente n palline rimettendo nella scatola dopo ogni estrazione la pallina estratta: in questo caso parleremo di **campionamento con rimessa**. Se registriamo i numeri segnati sulle palline estratte un possibile risultato dell'esperimento sarà del tipo $\omega = (a_1, a_2, \dots, a_n)$ con $a_i = 1, 2, \dots, M$ ed $i = 1, 2, \dots, n$, con possibili **ripetizioni**. Lo spazio dei campioni – cioè l'insieme di tali n -ple – potrà ora essere costruito in due maniere differenti:

1. Spazio dei campioni **ordinati** (a_1, \dots, a_n) : i campioni sono considerati distinti se differiscono anche solo per l'ordine in cui si presentano le etichette delle palline estratte e prendono il nome di **disposizioni**; in questo caso, ad esempio per $n = 4$ estrazioni, il campione $(4, 1, 2, 1)$ è considerato diverso dal campione $(1, 1, 2, 4)$. È facile verificare allora che il numero N_d di elementi di tale Ω è

$$N_d = M^n$$

2. Spazio dei campioni **non ordinati** $[a_1, \dots, a_n]$: in questo caso i campioni $(4, 1, 2, 1)$ e $(1, 1, 2, 4)$ sono considerati identici, sicché il numero N_r degli

elementi di Ω , che prendono il nome di **ripartizioni**, è inferiore a quello del caso precedente, e si dimostra¹ che

$$N_r = \binom{M+n-1}{n} = \frac{(M+n-1)!}{n!(M-1)!}$$

Se invece si esegue l'estrazione dell'esempio precedente senza rimettere a posto le palline dopo averle estratte si parlerà di **campionamento senza rimessa**. In tal caso è evidente che i campioni ottenuti (a_1, \dots, a_n) saranno composti di elementi tutti diversi (**senza ripetizioni**) e che $n \leq M$ visto che non potremo estrarre un numero di palline superiore a quello contenuto nella scatola. Anche qui ci sono due possibili spazi dei campioni:

1. Spazio dei campioni **ordinati** (a_1, \dots, a_n) : si tratta delle cosiddette **permutazioni** di M oggetti su n posti. Il numero di tali permutazioni (cioè il numero N_p di elementi del nostro spazio Ω) è ora

$$N_p = (M)_n = \frac{M!}{(M-n)!} = M(M-1)\dots(M-n+1)$$

Infatti ogni volta che estraiamo una pallina lasciamo una possibilità in meno per le successive estrazioni. Osserviamo che, nel caso in cui $n = M$, risulta $N = M!$ che è il numero delle permutazioni di M oggetti su M posti.

2. Spazio dei campioni **non ordinati** $[a_1, \dots, a_n]$: si tratta delle cosiddette **combinazioni** di M oggetti su n posti. Il loro numero N_c è

$$N_c = \binom{M}{n} \tag{1.1}$$

Infatti ogni campione non ordinato $[a_1, \dots, a_n]$ ammette $n!$ permutazioni distinte dall'ordine dei suoi simboli (vedi il punto precedente). Ne segue che $N_c \cdot n! = N_p$ da cui discende la relazione richiesta.

All'inizio di questa sezione abbiamo intuitivamente assegnato delle probabilità ai singoli elementi equiprobabili ω di uno spazio dei campioni Ω , con la prospettiva di poter poi utilizzare questa assegnazione per calcolare le probabilità di casi più complicati mediante somme e conteggi (definizione classica). Sarà però fatto chiaro nel seguito non solo che, come già detto, questa procedura è molto scomoda per spazi Ω di grandi cardinalità, ma anche che essa risulta del tutto inapplicabile nel caso di spazi con un numero infinito non numerabile di elementi. Sarà quindi necessario introdurre delle ulteriori strutture formali che ci permettano poi di definire correttamente le nostre probabilità.

¹N. Cufaro Petroni, **CALCOLO DELLE PROBABILITÀ**, Edizioni dal Sud (Bari, 1996)

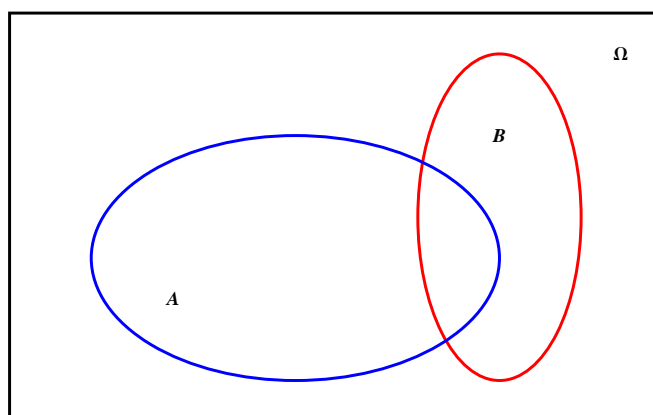


Figura 1.1: Diagrammi di Venn.

1.2 Eventi

Abbiamo già notato che un sottoinsieme $A \subseteq \Omega$ rappresenta una affermazione sui risultati del nostro esperimento. Nel caso di tre lanci di una moneta, ad esempio, lo spazio dei campioni è composto di $N = 2^3 = 8$ elementi

$$\Omega = \{TTT, TTC, \dots, CCC\}$$

e il sottoinsieme

$$A = \{TTT, TTC, TCT, CTT\} \subseteq \Omega$$

rappresenterà la proposizione “*T appare almeno due volte su tre lanci*”. Nel seguito Chiameremo **eventi** i sottoinsiemi $A \subseteq \Omega$, e diremo che un determinato evento A si *verifica* se il risultato ω dell’esperimento cade in A , cioè se $\omega \in A$.

I nostri eventi corrispondono quindi ad una famiglia di *proposizioni logiche* e le corrispondenti operazioni tra eventi (intese nel senso di operazioni tra insiemi) possono essere considerate come un modello per i *connettivi logici* che uniscono tali proposizioni. Così, ad esempio, i connettivi *oppure* (OR) ed *e* (AND) sono rappresentati rispettivamente dalle operazioni di *unione* ed *intersezione*:

$$\begin{aligned} A \cup B &= \{\omega : \omega \in A, \text{ oppure } \omega \in B\} \\ A \cap B = AB &= \{\omega : \omega \in A, \text{ e } \omega \in B\}. \end{aligned}$$

Il significato logico di altre operazioni, come quelle riportate qui di seguito, è facilmente deducibile anche tenendo presenti i *diagrammi di Venn* della Figura 1.1:

$$\begin{aligned} \bar{A} &= \{\omega : \omega \notin A\}; \\ A \setminus B &= A \cap \bar{B} = \{\omega : \omega \in A, \text{ ma } \omega \notin B\}; \\ A \Delta B &= (A \setminus B) \cup (B \setminus A) \quad (\text{differenza simmetrica}) \end{aligned}$$

In questo contesto ovviamente Ω rappresenta l'evento *certo*, dato che Ω contiene tutti i risultati ω , e \emptyset rappresenta l'evento *impossibile* dato che \emptyset non contiene nessun risultato ω . Diremo inoltre che due eventi A e B sono *disgiunti* o *incompatibili* quando $A \cap B = \emptyset$, cioè quando un risultato ω non può mai verificare contemporaneamente gli eventi A e B . Naturalmente le proprietà delle operazioni fra insiemi riproducono proprietà delle operazioni logiche, come ad esempio nel caso delle seguenti identità:

$$\overline{A \cap B} = \overline{A} \cup \overline{B}; \quad \overline{A \cup B} = \overline{A} \cap \overline{B}$$

note anche come *leggi di de Morgan*. Così ad esempio nel caso di due lanci di una moneta, a partire dagli eventi

$$\begin{aligned} A &= \{TT, TC, CT\} = \text{appare } T \text{ almeno una volta} \\ B &= \{TC, CT, CC\} = \text{appare } C \text{ almeno una volta} \end{aligned}$$

è possibile formare altri eventi come

$$A \cup B = \{TT, TC, CT, CC\} = \Omega, \quad A \cap B = \{TC, CT\}, \quad A \setminus B = \{TT\}$$

Va detto subito, però, che in generale la famiglia degli eventi di Ω che si prende in considerazione *non coincide* obbligatoriamente con la famiglia $\wp(\Omega)$ di tutte le parti di Ω . Tipicamente, infatti, si preferisce scegliere una qualche particolare sottofamiglia di parti di Ω , come avviene, ad esempio, quando $\Omega = \mathbf{R}$ perchè in questo caso $\wp(\mathbf{R})$ conterrebbe insiemi particolarmente difficili da trattare e sostanzialmente irrilevanti da un punto di vista pratico. Solo per spazi Ω finiti, invece, la scelta più conveniente è proprio quella della intera famiglia $\wp(\Omega)$. Pertanto nella costruzione di un modello probabilistico dovremo innanzitutto cominciare con lo stabilire quale è la famiglia degli eventi, ed è opportuno domandarsi qui se questa famiglia possa essere del tutto arbitraria o se debba soddisfare particolari requisiti di coerenza. In particolare è evidente che l'applicazione di connettivi logici (congiunzioni, disgiunzioni, negazioni, ...) alle proposizioni del nostro modello deve produrre altre proposizioni ugualmente accettabili, e quindi dovremo richiedere corrispondentemente che la famiglia dei nostri eventi sia chiusa sotto tutte le possibili operazioni insiemistiche. A questo scopo si introducono le seguenti definizioni

Definizione 1.3. Una famiglia non vuota $\mathcal{F} \subseteq \wp(\Omega)$ di parti di un insieme Ω costituisce un'**algebra** quando

$$\begin{aligned} \Omega &\in \mathcal{F} \\ \overline{A} &\in \mathcal{F}, \quad \forall A \in \mathcal{F} \\ A \cap B &\in \mathcal{F}, \quad \forall A, B \in \mathcal{F} \end{aligned}$$

Si dice invece che \mathcal{F} costituisce una σ -**algebra** se è un'algebra e se soddisfa anche l'ulteriore condizione

$$\bigcap_n A_n \in \mathcal{F} \quad \forall (A_n)_{n \in \mathbf{N}} \text{ di elementi di } \mathcal{F}$$

Proposizione 1.4. *Se \mathcal{F} è una σ -algebra, allora*

$$\begin{aligned} \emptyset &\in \mathcal{F} \\ A \cup B &\in \mathcal{F} \quad \forall A, B \in \mathcal{F} \\ A \setminus B &\in \mathcal{F} \quad \forall A, B \in \mathcal{F} \\ \bigcup_n A_n &\in \mathcal{F} \quad \forall (A_n)_{n \in \mathbf{N}} \text{ di elementi di } \mathcal{F} \end{aligned}$$

Dimostrazione: Omessa² □

In pratica una σ -algebra è una famiglia di parti di Ω chiusa sotto tutte le operazioni insiemistiche, anche numerabili: una proprietà, quest'ultima, importante per la costruzione rigorosa di un modello probabilistico. Si noti comunque che queste proprietà delle σ -algebre non si estendono alle operazioni insiemistiche eseguite su famiglie $(A_t)_{t \in T}$ non numerabili di eventi. D'ora in poi supporremo sempre che i nostri eventi costituiscano una σ -algebra \mathcal{F} , e chiameremo inoltre **spazio probabilizzabile** la coppia (Ω, \mathcal{F}) in cui \mathcal{F} è una σ -algebra di eventi di Ω . Dato un insieme Ω gli esempi più evidenti di σ -algebre sono

$$\begin{aligned} \mathcal{F}_* &= \{\emptyset, \Omega\}, \\ \mathcal{F}_A &= \{A, \overline{A}, \emptyset, \Omega\}, \quad (A \subseteq \Omega), \\ \mathcal{F}^* &= \wp(\Omega). \end{aligned}$$

In particolare la σ -algebra \mathcal{F}_A prende il nome di **σ -algebra generata da A** ed è il caso più semplice di una situazione generale: data infatti una famiglia $\mathcal{E} \subseteq \wp(\Omega)$ di parti di Ω , chiameremo **σ -algebra generata da \mathcal{E}** la più piccola σ -algebra $\sigma(\mathcal{E})$ contenente \mathcal{E}

Proposizione 1.5. *Data una famiglia $\mathcal{E} \subseteq \wp(\Omega)$ di parti di Ω esiste sempre la σ -algebra $\sigma(\mathcal{E})$ generata da \mathcal{E}*

Dimostrazione: Omessa³ □

Definizione 1.6. *Diremo che una famiglia (finita o numerabile) di sottoinsiemi $\mathcal{D} = \{D_1, D_2, \dots\}$ è una **decomposizione** di Ω in **atomi** D_k , se le D_k sono parti di Ω non vuote, disgiunte e tali che $\bigcup_k D_k = \Omega$.*

Le decomposizioni sono famiglie di eventi dei quali si verifica sempre uno, e uno solo, e quindi sono il modello per descrivere un insieme di possibilità mutuamente esclusive che esauriscono tutte i casi possibili. È evidente però che una decomposizione non è né un'algebra, né tanto meno una σ -algebra: essa, ad esempio, non contiene le unioni dei suoi atomi. Ciononostante, sulla base della Proposizione 1.5, a partire da

²N. Cufaro Petroni, CALCOLO DELLE PROBABILITÀ, Edizioni dal Sud (Bari, 1996)

³N. Cufaro Petroni, CALCOLO DELLE PROBABILITÀ, Edizioni dal Sud (Bari, 1996)

una data decomposizione \mathcal{D} è sempre possibile costruire la σ -algebra $\mathcal{F} = \sigma(\mathcal{D})$ da essa generata. Il caso più banale è la decomposizione di Ω in soli due sottoinsiemi

$$\mathcal{D}_A = \{A, \bar{A}\}$$

che genera la σ -algebra \mathcal{F}_A . La Proposizione 1.5 si rivela dunque uno strumento essenziale per la costruzione di σ -algre su particolari spazi dei campioni

Esempio 1.7. *Esamineremo ora alcune importanti σ -algre⁴:*

1. *Iniziamo con il caso in cui lo spazio dei campioni Ω è l'insieme \mathbf{R} dei numeri reali, e consideriamo innanzitutto la famiglia \mathcal{I} degli intervalli (limitati o illimitati) chiusi solo a destra*

$$I = (a, b], \quad -\infty \leq a < b \leq +\infty$$

In pratica \mathcal{I} contiene tutte le parti di \mathbf{R} del tipo

$$(a, b], \quad (-\infty, b], \quad (a, +\infty), \quad (-\infty, +\infty)$$

*con $a, b \in \mathbf{R}$ (per convenzione gli intervalli illimitati a destra sono considerati chiusi a destra). Siccome in generale le unioni di intervalli non sono intervalli, \mathcal{I} non è né un'algebra, né tanto meno una σ -algebra. Costruiremo allora (ma noi trascureremo i dettagli) la σ -algebra generata da \mathcal{I} che sarà indicata con il simbolo $\mathcal{B}(\mathbf{R})$, e che prenderà il nome di **σ -algebra di Borel di \mathbf{R}** , mentre i suoi elementi saranno chiamati **boreliani di \mathbf{R}** . Si dimostra allora che $\mathcal{B}(\mathbf{R})$ contiene tutte le parti di \mathbf{R} della forma*

$$\emptyset, \quad \{a\}, \quad [a, b], \quad [a, b), \quad (a, b], \quad (a, b), \quad \mathbf{R}$$

e tutte le loro unioni e intersezioni finite o numerabili. In realtà si potrebbe far vedere che la medesima σ -algebra può essere generata anche partendo da altre famiglie di insiemi, e in particolare dalla famiglia degli aperti di \mathbf{R} . Lo spazio probabilizzabile così ottenuto verrà d'ora in poi indicato con il simbolo

$$(\mathbf{R}, \mathcal{B}(\mathbf{R}))$$

2. *Consideriamo ora il caso in cui $\Omega = \mathbf{R}^n$ sia il prodotto cartesiano di n copie della retta reale: gli elementi di Ω sono quindi le n -ple di numeri reali $\omega = \mathbf{x} = \{x_1, x_2, \dots, x_n\}$. Come nell' esempio precedente, anche qui ci sono varie maniere equivalenti di costruire la σ -algebra necessaria per definire il nostro spazio probabilizzabile. Si può dimostrare che tutte le σ -algre generate a partire da famiglie di rettangoli, o semplicemente di aperti di \mathbf{R}^n sono coincidenti: anche esse portano il nome di **σ -algre di Borel di \mathbf{R}^n** e si indicano con il simbolo $\mathcal{B}(\mathbf{R}^n)$, sicché lo spazio probabilizzabile sarà ora*

$$(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$$

⁴N. Cufaro Petroni, **CALCOLO DELLE PROBABILITÀ**, Edizioni dal Sud (Bari, 1996)

3. Se $(\mathbf{R}_n)_{n \in \mathbf{N}}$ è una successione di copie dell'insieme \mathbf{R} dei numeri reali, con il simbolo $\mathbf{R}^\infty = \mathbf{R}_1 \times \dots \times \mathbf{R}_n \times \dots$ indicheremo l'insieme delle successioni di numeri reali, e avremo $\omega = x = (x_n)_{n \in \mathbf{N}}$. In questo caso la σ -algebra richiesta viene generata partendo da opportuni sottoinsiemi di \mathbf{R}^∞ detti **cilindri**: questi consistono negli insiemi di successioni $(x_n)_{n \in \mathbf{N}}$ individuati dal fatto che un certo numero finito m di loro componenti $(x_{n_1}, x_{n_2}, \dots, x_{n_m})$ cade in un boreliano $B \in \mathcal{B}(\mathbf{R}^m)$ detto **base del cilindro**. Utilizzando il fatto che le basi sono di dimensione finita, e sfruttando i risultati degli esempi precedente, si definiscono famiglie di cilindri che possono essere usate per generare una σ -algebra che viene indicata con il simbolo $\mathcal{B}(\mathbf{R}^\infty)$ e prende ancora il nome di **σ -algebra dei Boreliani di \mathbf{R}^∞** . Il corrispondente spazio probabilizzabile sarà quindi indicato con il simbolo

$$(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$$

e si può mostrare che insiemi di successioni di \mathbf{R}^∞ del tipo

$$\begin{aligned} \{x \in \mathbf{R}^\infty : \sup_n x_n > a\}, & \quad \{x \in \mathbf{R}^\infty : \inf_n x_n < a\}, \\ \{x \in \mathbf{R}^\infty : \underline{\lim}_n x_n \leq a\}, & \quad \{x \in \mathbf{R}^\infty : \overline{\lim}_n x_n > a\}, \\ \{x \in \mathbf{R}^\infty : x \text{ converge}\}, & \quad \{x \in \mathbf{R}^\infty : \lim_n x_n > a\}, \end{aligned}$$

sono tutti elementi di $\mathcal{B}(\mathbf{R}^\infty)$.

4. Detto infine T un arbitrario sottoinsieme di \mathbf{R} (in generale non numerabile come un intervallo $[a, b]$), indicheremo con \mathbf{R}^T l'insieme delle funzioni definite da T in \mathbf{R} . Gli elementi ω di \mathbf{R}^T saranno indicati in uno dei seguenti modi secondo l'opportunità della notazione: x , $x(\cdot)$, $x(t)$, $(x_t)_{t \in T}$. Seguendo lo schema già utilizzato nel caso di \mathbf{R}^∞ , per definire una σ -algebra su \mathbf{R}^T si prendono in considerazione innanzitutto i cilindri con basi B di dimensione finita o numerabile: esse sono sottoinsiemi di funzioni che, in un dato insieme (finito o numerabile) di punti t_j , assumono valori che cadono in B . Si costruisce poi la σ -algebra generata da tali cilindri che sarà indicata con il solito simbolo $\mathcal{B}(\mathbf{R}^T)$, mentre il simbolo

$$(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$$

rappresenterà il corrispondente spazio probabilizzabile. Si dimostra però che $\mathcal{B}(\mathbf{R}^T)$ coincide proprio con la famiglia dei cilindri con basi finite o numerabili, cioè con la famiglia di parti di \mathbf{R}^T che possono essere individuate mediante restrizioni su $(x_t)_{t \in T}$ imposte in un insieme di punti finito o numerabile. Una conseguenza di questo risultato è che alcuni sottoinsiemi di \mathbf{R}^T – tipicamente quelli caratterizzati dal comportamento della funzione $(x_t)_{t \in T}$ in un insieme

non numerabile di punti t – non sono elementi di $\mathcal{B}(\mathbf{R}^T)$: ad esempio, con $T = [0, 1]$, si verifica che gli insiemi

$$\begin{aligned} A_1 &= \{x \in \mathbf{R}^{[0,1]} : \sup_{t \in [0,1]} x_t < a, a \in \mathbf{R}\} \\ A_2 &= \{x \in \mathbf{R}^{[0,1]} : \exists t \in [0, 1] \ni' x_t = 0\} \\ A_3 &= \{x \in \mathbf{R}^{[0,1]} : x_t \text{ è continua in } t_0 \in [0, 1]\} \end{aligned}$$

non sono elementi di $\mathcal{B}(\mathbf{R}^{[0,1]})$. Siccome però tali insiemi rappresentano affermazioni importanti nella teoria dei processi stocastici, questo crea una difficoltà che è necessario superare: un modo è quello di limitarsi a classi di funzioni più ristrette di \mathbf{R}^T . Ad esempio è possibile costruire opportune σ -algebre partendo dall'insieme C delle **funzioni continue** $x(t)$: si può mostrare che in questo modo i suddetti sottoinsiemi di funzioni A_1, A_2 ed A_3 risultano essere inclusi nella σ -algebra. Qui comunque per semplicità trascureremo di elaborare ulteriormente questo punto.

1.3 Probabilità

Nel caso di uno spazio probabilizzabile (Ω, \mathcal{F}) con Ω di cardinalità finita N è possibile definire una *probabilità* assegnandone innanzitutto una quota $p(\omega_k)$ ad ogni singolo $\omega_k \in \Omega$ in modo tale che

$$0 \leq p(\omega_k) \leq 1, \quad k = 1, \dots, N; \quad \sum_{k=1}^N p(\omega_k) = 1$$

Agli eventi $A \in \mathcal{F}$ si assegna poi una probabilità

$$P\{A\} = \sum_{\omega_k \in A} p(\omega_k)$$

e si chiama **spazio finito di probabilità** la terna (Ω, \mathcal{F}, P) . Si prova facilmente che una probabilità così definita gode di varie proprietà (che esamineremo in dettaglio più oltre nel caso più generale) tra le quali ricorderemo

$$\begin{aligned} P\{\emptyset\} &= 0 \\ P\{\Omega\} &= 1 \\ P\{A \cup B\} &= P\{A\} + P\{B\} && \text{se } A \cap B = \emptyset && (\text{additività}) \\ P\{\bar{A}\} &= 1 - P\{A\} \\ P\{A\} &\leq P\{B\} && \text{se } A \subseteq B \end{aligned}$$

Questa procedura è estendibile (con qualche accortezza per garantire la convergenza delle somme) anche al caso in cui Ω è un insieme *infinito numerabile*, ma in tutti

e due i casi l'assegnazione effettiva dei numeri $p(\omega_k)$ ai singoli campioni ω_k non è sempre un compito semplice. Le tecniche per ricavare le $p(\omega_k)$ a partire dalle osservazioni empiriche sono argomento della **statistica** e noi qui ne parleremo solo di sfuggita. Per i nostri primi esempi adotteremo invece la **definizione classica della probabilità** a cui abbiamo già fatto cenno nella Sezione 1.1. Il metodo consiste nel ricondursi ad uno spazio dei campioni finito i cui N elementi possano essere considerati *equiprobabili* per ragioni di simmetria. In tal caso ad ogni ω_k si associa la probabilità $p(\omega_k) = 1/N$, e detto $N(A)$ il numero di elementi di un $A \in \mathcal{F}$, ad A si attribuisce la probabilità

$$\mathbf{P}\{A\} = \frac{N(A)}{N}$$

Esempio 1.8. Problema delle coincidenze: *Mostreremo ora discutendo un problema classico che gli spazi finiti di probabilità (costruiti a partire dalla definizione classica) sono già uno strumento molto versatile che può consentire interessanti risultati. Supponiamo di estrarre con rimessa da una scatola contenente M palline numerate una successione di n palline e di registrare i numeri estratti tenendo conto dell'ordine di estrazione. Abbiamo visto nella Sezione 1.1 che Ω conterrà $N = M^n$ elementi $\omega = (a_1, \dots, a_n)$ che considereremo equiprobabili. Posto allora*

$$A = \{\omega : i \text{ valori delle } a_k \text{ sono tutti diversi}\}$$

cominciamo con il calcolo di $\mathbf{P}\{A\}$ secondo la definizione classica: siccome, come si vede per enumerazione, il numero $N(A)$ dei campioni contenuti in A è

$$N(A) = M(M-1) \dots (M-n+1) = (M)_n = \frac{M!}{(M-n)!}$$

si ha dalla definizione classica

$$\mathbf{P}\{A\} = \frac{(M)_n}{M^n} = \left(1 - \frac{1}{M}\right) \left(1 - \frac{2}{M}\right) \dots \left(1 - \frac{n-1}{M}\right).$$

Questo risultato permette in particolare di discutere il cosiddetto **problema dei compleanni**: date n persone quale è la probabilità P_n che almeno due di esse celebrino il compleanno nello stesso giorno? Si tratta di un problema di conteggi su campioni ordinati – dato che una diversa disposizione delle stesse date di nascita sulle n persone (distinguibili) deve essere ritenuta un'eventualità distinta – e la discussione precedente ci permette di dare una risposta ponendo $M = 365$. Infatti, essendo $\mathbf{P}\{A\}$ la probabilità che tutti i compleanni cadano in giorni differenti, si ha

$$P_n = 1 - \mathbf{P}\{A\} = 1 - \frac{(365)_n}{365^n}$$

e si ottengono dei risultati numerici a prima vista sorprendenti

n	4	16	22	23	40	64
P_n	0.016	0.284	0.467	0.507	0.891	0.997

È notevole infatti che già con $n = 23$ la probabilità di almeno due compleanni coincidenti supera $1/2$, e che con solo 64 persone tale probabilità sfiora la certezza. Si osservi inoltre che se $n \geq 366$ risulta $\mathbf{P}\{A\} = 0$ (e quindi $P_n = 1$) dato che nel prodotto compare un fattore nullo: ciò corrisponde al fatto intuitivo che se ci sono più di 365 persone è certo che ci saranno compleanni coincidenti. Questi risultati sono però meno sorprendenti se si si riflette al fatto che essi cambierebbero molto se la domanda fosse stata posta in modo diverso: “supponendo che io sia una delle persone considerate nel problema precedente, quale è la probabilità P'_n che almeno un'altra celebri il suo compleanno nello stesso giorno in cui lo celebri io?” Infatti in questo caso si ha $N(A) = 365 \cdot 364^{n-1}$ e quindi la probabilità cercata è

$$P'_n = 1 - \left(\frac{364}{365}\right)^{n-1}$$

sicché in particolare questa volta risulta

n	4	16	22	23	40	64	101
P'_n	0.011	0.040	0.056	0.059	0.101	0.159	0.240

Inoltre in questo caso P'_n è sempre diversa da 1 (anche se $n \geq 366$) in quanto, quale che sia il numero delle persone, può sempre capitare che nessuno celebri il suo compleanno nello stesso giorno in cui lo celebri io: un'osservazione che mette bene in luce la differenza fra le due formulazioni del problema.

I modelli finiti o numerabili di probabilità si rivelano però insufficienti perché gli spazi dei campioni sono spesso **insiemi non numerabili**: sarebbe facile verificare, ad esempio, che anche il ben noto insieme delle possibili *successioni infinite* di lanci di moneta non è numerabile. In questi casi la definizione di una probabilità mediante assegnazione di pesi numerici ai singoli elementi di Ω è semplicemente impraticabile. In particolare se attribuiamo pesi $p(\omega) > 0$ a un insieme non numerabile di risultati $\omega \in \Omega$, la condizione $\sum_{\omega \in \Omega} p(\omega) = 1 < +\infty$ non potrebbe mai essere verificata, e pertanto nessuna definizione coerente di probabilità potrebbe essere costruita su questa base. Nel caso generale, quindi, la definizione di $\mathbf{P}\{A\}$ non può più essere costruita per *enumerazione di elementi individuali* come nel caso finito o numerabile, ma deve essere data tramite il concetto di *misura di insiemi*.

Definizione 1.9. Data una σ -algebra \mathcal{F} di parti di un insieme Ω , chiameremo **misura** su \mathcal{F} ogni applicazione $\mu : \mathcal{F} \rightarrow [0, +\infty]$ che sia **σ -additiva**, cioè tale che, comunque scelta una successione $(A_n)_{n \in \mathbf{N}}$ di elementi disgiunti di \mathcal{F} , risulti

$$\mu\left\{\bigcup_n A_n\right\} = \sum_n \mu\{A_n\};$$

diremo inoltre che μ è una **misura finita** se $\mu\{\Omega\} < +\infty$, e che è una **misura σ -finita** se Ω può essere decomposto nell'unione $\Omega = \bigcup_n A_n$, $A_n \in \mathcal{F}$ di insiemi disgiunti tali che $\mu\{A_n\} < +\infty$, $\forall n \in \mathbf{N}$. Una misura finita \mathbf{P} tale che $\mathbf{P}\{\Omega\} = 1$ si dice infine **misura di probabilità**.

Definizione 1.10. Diremo che una proprietà è vera ***P*-quasi ovunque** (***P*-qo**) se essa si verifica per ogni $\omega \in \Omega$, fatta eccezione per un insieme di misura ***P*** nulla.

Si noti che comunque scelto $A \in \mathcal{F}$, risulterà sempre $\mu\{A\} \leq \mu\{\Omega\}$ in quanto μ è additiva, positiva e $\Omega = A \cup \bar{A}$. Ne segue in particolare che se μ è finita avremo anche $\mu\{A\} < +\infty$. Osserviamo infine che, mentre una misura finita risulta sempre anche σ -finita, il viceversa non è vero: ad esempio l'usuale **misura di Lebesgue** sulla retta reale, quella che ad ogni intervallo $[a, b]$ associa come misura la sua ampiezza $|b - a|$, è ovviamente σ -finita, ma non finita.

Proposizione 1.11. Data una misura di probabilità $\mathbf{P} : \mathcal{F} \rightarrow [0, 1]$ le seguenti proprietà sono sempre verificate:

1. $\mathbf{P}\{\emptyset\} = 0$
2. $\mathbf{P}\{A \setminus B\} = \mathbf{P}\{A\} - \mathbf{P}\{AB\}$, $\forall A, B \in \mathcal{F}$
3. $\mathbf{P}\{A \cup B\} = \mathbf{P}\{A\} + \mathbf{P}\{B\} - \mathbf{P}\{AB\}$, $\forall A, B \in \mathcal{F}$
4. $\mathbf{P}\{A \Delta B\} = \mathbf{P}\{A\} + \mathbf{P}\{B\} - 2\mathbf{P}\{AB\}$, $\forall A, B \in \mathcal{F}$
5. $\mathbf{P}\{B\} \leq \mathbf{P}\{A\}$ se $B \subseteq A$, con $A, B \in \mathcal{F}$
6. $\mathbf{P}\{\bigcup_n A_n\} \leq \sum_n \mathbf{P}\{A_n\}$, per ogni successione di eventi $(A_n)_{n \in \mathbf{N}}$.

L'ultima proprietà porta anche il nome di **subadditività**

Dimostrazione: Omessa⁵

□

Definizione 1.12. Assiomi di Kolmogorov: Chiameremo **spazio di probabilità** una terna ordinata $(\Omega, \mathcal{F}, \mathbf{P})$ in cui Ω è un insieme di elementi ω detto anche **spazio dei campioni**, \mathcal{F} è una σ -algebra di eventi di Ω e \mathbf{P} è una **misura di probabilità definita su \mathcal{F}** .

Sarà il caso di notare qui che un *evento di probabilità 0* non è necessariamente coincidente con l'insieme vuoto \emptyset , e che un *evento di probabilità 1* non è necessariamente coincidente con tutto Ω . Questa osservazione è particolarmente importante nel caso di spazi *infiniti non numerabili*, ma anche in spazi finiti può essere utile assegnare pesi probabilistici nulli ad alcuni degli elementi $\omega \in \Omega$. In effetti in questo caso potrebbe sembrare più opportuno modificare la struttura di Ω eliminando tali eventi elementari in quanto ci attendiamo che non si verifichino mai, ma questa operazione potrebbe rivelarsi solo un'apparente razionalizzazione. Ad esempio a volte è importante poter confrontare due o più assegnazioni di probabilità \mathbf{P} allo stesso spazio dei campioni, e può capitare che eventi di probabilità non nulla secondo un'assegnazione siano di probabilità nulla secondo l'altra, o viceversa. In questo caso sarebbe particolarmente scomodo modificare ogni volta la struttura dello spazio Ω e si preferisce piuttosto conservarne tutti gli elementi anche se qualcuno può risultare di probabilità nulla.

⁵N. Cufaro Petroni, CALCOLO DELLE PROBABILITÀ, Edizioni dal Sud (Bari, 1996)

1.4 Probabilità condizionate

Esamineremo ora le procedure mediante le quali della nuova informazione può essere amalgamata con quella già presente in uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ permettendoci di valutare la probabilità in una maniera diversa da quella iniziale.

Definizione 1.13. *Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ e due eventi $A, B \in \mathcal{F}$ con $\mathbf{P}\{B\} \neq 0$, chiameremo **probabilità condizionata** di A rispetto ad B (cioè probabilità che si verifichi A sapendo che si è verificato B) la quantità*

$$\mathbf{P}\{A|B\} \equiv \frac{\mathbf{P}\{A \cap B\}}{\mathbf{P}\{B\}} = \frac{\mathbf{P}\{AB\}}{\mathbf{P}\{B\}}$$

La quantità $\mathbf{P}\{AB\}$ prende invece il nome di **probabilità congiunta** di A e B (cioè probabilità che si verifichino contemporaneamente A e B).

Si noti che la condizione $\mathbf{P}\{B\} \neq 0$ è per il momento cruciale per poter accettare questa definizione: rinviando alla Sezione 3.4 la discussione sull'estensione della definizione anche al caso $\mathbf{P}\{B\} = 0$. È facile mostrare, comunque, che la nuova $\mathbf{P}\{\cdot|B\} : \mathcal{F} \rightarrow [0, 1]$ della definizione gode ancora delle proprietà di una probabilità:

$$\begin{aligned} \mathbf{P}\{\emptyset|B\} &= 0 & \mathbf{P}\{\Omega|B\} &= \mathbf{P}\{B|B\} = 1 \\ \mathbf{P}\{A_1 \cup A_2|B\} &= \mathbf{P}\{A_1|B\} + \mathbf{P}\{A_2|B\} & \text{se } A_1 \cap A_2 &= \emptyset \\ \mathbf{P}\{\bar{A}|B\} &= 1 - \mathbf{P}\{A|B\} \end{aligned}$$

sicché $(\Omega, \mathcal{F}, \mathbf{P}\{\cdot|B\})$ è a tutti gli effetti un nuovo spazio di probabilità.

Proposizione 1.14. Formula della probabilità totale: *Dati in uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ un evento $A \in \mathcal{F}$ e una decomposizione $\mathcal{D} = \{D_1, \dots, D_n\}$ con $\mathbf{P}\{D_j\} \neq 0, j = 1, \dots, n$ si ha*

$$\mathbf{P}\{A\} = \sum_{j=1}^n \mathbf{P}\{A|D_j\} \mathbf{P}\{D_j\}$$

Dimostrazione: Dato che

$$A = A \cap \Omega = A \cap \left(\bigcup_{j=1}^n D_j \right) = \bigcup_{j=1}^n (A \cap D_j)$$

basterà osservare che gli eventi $A \cap D_j$ sono tutti disgiunti per ottenere

$$\mathbf{P}\{A\} = \mathbf{P}\left\{ \bigcup_{j=1}^n (A \cap D_j) \right\} = \sum_{j=1}^n \mathbf{P}\{A \cap D_j\} = \sum_{j=1}^n \mathbf{P}\{A|D_j\} \mathbf{P}\{D_j\}$$

a causa dell'additività di \mathbf{P} □

In particolare, quando $\mathcal{D} = \{B, \overline{B}\}$, la formula della probabilità totale diviene

$$\mathbf{P}\{A\} = \mathbf{P}\{A|B\} \mathbf{P}\{B\} + \mathbf{P}\{A|\overline{B}\} \mathbf{P}\{\overline{B}\} \quad (1.2)$$

espressione particolarmente facile da usare ed interpretare come mostrato nel successivo esempio

Esempio 1.15. Estrazioni successive: *Consideriamo una scatola contenente M palline delle quali m sono bianche ed $M - m$ nere ed eseguiamo due estrazioni successive. Tralasciando una descrizione dettagliata del corrispondente spazio di probabilità, possiamo intuitivamente introdurre gli eventi*

$$\begin{aligned} B &= \text{“la prima pallina estratta è bianca”} \\ A &= \text{“la seconda pallina estratta è bianca”} \end{aligned}$$

e, se le palline sono tutte equiprobabili, potremo usare la definizione classica per attribuire loro delle probabilità. Così, se la prima estrazione è effettuata con rimessa, è facile convincersi del fatto che le probabilità di B e A sono uguali e $\mathbf{P}\{A\} = \mathbf{P}\{B\} = \frac{m}{M}$. Viceversa, se la prima estrazione venisse effettuata senza rimessa e se sapessimo, ad esempio, che la prima pallina estratta è bianca (cioè che si è verificato B), la probabilità di A sarebbe $\frac{m-1}{M-1}$; mentre se in precedenza fosse stata estratta una pallina nera (cioè se si fosse verificato \overline{B}) si avrebbe $\frac{m}{M-1}$. Estraiamo allora, in successione e senza rimessa, due palline e, **senza guardare la prima**, chiediamoci quale è la probabilità che la seconda sia bianca, cioè che si verifichi A . Dalla discussione che precede sappiamo che

$$\begin{aligned} \mathbf{P}\{B\} &= \frac{m}{M} & \mathbf{P}\{\overline{B}\} &= \frac{M-m}{M} \\ \mathbf{P}\{A|B\} &= \frac{m-1}{M-1} & \mathbf{P}\{A|\overline{B}\} &= \frac{m}{M-1} \end{aligned}$$

e quindi tenendo conto del Teorema della probabilità totale (1.2) si ha

$$\mathbf{P}\{A\} = \frac{m-1}{M-1} \frac{m}{M} + \frac{m}{M-1} \frac{M-m}{M} = \frac{m}{M} = \mathbf{P}\{B\}$$

Questa discussione mette in evidenza il fatto che la probabilità di A dipende dalle informazioni effettivamente a disposizione: se le estrazioni sono effettuate senza rimessa il risultato della prima influenza la probabilità della seconda, sicché $\mathbf{P}\{A|B\}$ e $\mathbf{P}\{A|\overline{B}\}$ sono diversi fra loro, e sono anche diversi da $\mathbf{P}\{B\}$. Ma se il risultato della prima estrazione è sconosciuto si ha di nuovo $\mathbf{P}\{A\} = \mathbf{P}\{B\} = \frac{m}{M}$ come nel caso con rimessa.

Proposizione 1.16. Formula di moltiplicazione: *Comunque dati in $(\Omega, \mathcal{F}, \mathbf{P})$ gli eventi A_1, \dots, A_n con $\mathbf{P}\{A_1 \dots A_{n-1}\} \neq 0$, si ha*

$$\mathbf{P}\{A_1 \dots A_n\} = \mathbf{P}\{A_n|A_{n-1} \dots A_1\} \mathbf{P}\{A_{n-1}|A_{n-2} \dots A_1\} \dots \mathbf{P}\{A_2|A_1\} \mathbf{P}\{A_1\}$$

Dimostrazione: Infatti dalla definizione di probabilità condizionate risulta

$$\begin{aligned} & \mathbf{P}\{A_n|A_{n-1}\dots A_1\} \mathbf{P}\{A_{n-1}|A_{n-2}\dots A_1\} \dots \mathbf{P}\{A_2|A_1\} \mathbf{P}\{A_1\} \\ &= \frac{\mathbf{P}\{A_1\dots A_n\}}{\mathbf{P}\{A_1\dots A_{n-1}\}} \frac{\mathbf{P}\{A_1\dots A_{n-1}\}}{\mathbf{P}\{A_1\dots A_{n-2}\}} \dots \frac{\mathbf{P}\{A_1A_2\}}{\mathbf{P}\{A_1\}} \mathbf{P}\{A_1\} = \mathbf{P}\{A_1\dots A_n\} \end{aligned}$$

come volevasi dimostrare. \square

La Formula di Moltiplicazione è una proprietà molto generale della probabilità che giocherà un ruolo importante nella discussione della Markovianità nella seconda parte di queste lezioni

Proposizione 1.17. Teorema di Bayes: *Dati in $(\Omega, \mathcal{F}, \mathbf{P})$ due eventi A, B con $\mathbf{P}\{A\} \neq 0$ e $\mathbf{P}\{B\} \neq 0$, risulta*

$$\mathbf{P}\{A|B\} = \frac{\mathbf{P}\{B|A\} \mathbf{P}\{A\}}{\mathbf{P}\{B\}}$$

Inoltre, se $\mathcal{D} = \{D_1, \dots, D_n\}$ è una decomposizione con $\mathbf{P}\{D_j\} \neq 0$, $j = 1, \dots, n$, risulta anche

$$\mathbf{P}\{D_j|B\} = \frac{\mathbf{P}\{B|D_j\} \mathbf{P}\{D_j\}}{\sum_{k=1}^n \mathbf{P}\{B|D_k\} \mathbf{P}\{D_k\}}$$

Dimostrazione: La dimostrazione della prima relazione (detta anche **Formula di Bayes**) si basa sul fatto che per definizione di probabilità condizionata si ha

$$\mathbf{P}\{B|A\} \mathbf{P}\{A\} = \mathbf{P}\{AB\} = \mathbf{P}\{A|B\} \mathbf{P}\{B\}$$

La seconda relazione si ottiene poi dalla prima tramite il Teorema della Probabilità Totale. \square

Nelle applicazioni statistiche gli eventi D_j del Teorema di Bayes sono spesso chiamati *ipotesi*, e $\mathbf{P}\{D_j\}$ *probabilità a priori* di tali ipotesi, mentre le probabilità condizionate $\mathbf{P}\{D_j|B\}$ si chiamano *probabilità a posteriori*. Questa terminologia deriva, come vedremo con un esempio nella Sezione 2.1.2, dall'osservazione che, in generale, il verificarsi di un certo evento B modifica la probabilità inizialmente assegnata alle ipotesi D_j del nostro problema

1.5 Eventi indipendenti

Due eventi sono indipendenti quando il verificarsi di uno di essi non influenza la probabilità che viene attribuita all'altro. Sulla base della definizione di probabilità condizionata diremo quindi che l'evento A è indipendente dall'evento B se $\mathbf{P}\{A|B\} = \mathbf{P}\{A\}$, e quindi se $\mathbf{P}\{AB\} = \mathbf{P}\{A\} \mathbf{P}\{B\}$. Il vantaggio di questa seconda formulazione è che essa, a differenza di quella basata sulle probabilità condizionate, ha senso anche quando $\mathbf{P}\{B\} = 0$. È facile inoltre, sulla base della simmetria di queste relazioni, convincersi del fatto che se A è indipendente da B , anche B è indipendente da A . Adotteremo pertanto le seguenti definizioni

Definizione 1.18. Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ diremo che A e B sono **eventi indipendenti** quando

$$\mathbf{P}\{AB\} = \mathbf{P}\{A\} \mathbf{P}\{B\}$$

diremo inoltre che due σ -algebre \mathcal{F}_1 e \mathcal{F}_2 di eventi di Ω (più precisamente due sotto- σ -algebre di \mathcal{F}) sono **σ -algebre indipendenti** se ogni evento di \mathcal{F}_1 è indipendente da ogni evento di \mathcal{F}_2 .

Il concetto di indipendenza può essere esteso anche al caso in cui il numero di eventi è maggiore di due, ma bisogna fare molta attenzione al fatto che sarà ora possibile parlare di *indipendenza due a due*, nel senso di $\mathbf{P}\{AB\} = \mathbf{P}\{A\} \mathbf{P}\{B\}$, di *indipendenza tre a tre*, nel senso di $\mathbf{P}\{ABC\} = \mathbf{P}\{A\} \mathbf{P}\{B\} \mathbf{P}\{C\}$, e così via, e che tali livelli di indipendenza non si implicano affatto l'uno con l'altro, nel senso che ad esempio tre eventi possono essere indipendenti due a due senza esserlo tre a tre, e viceversa. Daremo quindi la seguente definizione

Definizione 1.19. Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ diremo che $A_1, \dots, A_n \in \mathcal{F}$ sono **eventi indipendenti** se comunque scelti k indici j_1, \dots, j_k (con $k = 1, \dots, n$) risulta

$$\mathbf{P}\{A_{j_1} \dots A_{j_k}\} = \mathbf{P}\{A_{j_1}\} \dots \mathbf{P}\{A_{j_k}\}$$

cioè quando sono indipendenti due a due, tre a tre, \dots , n a n .

L'indipendenza di due, o più, eventi è determinata dalla probabilità $\mathbf{P}\{\cdot\}$: gli stessi eventi (identici dal punto di vista insiemistico) possono risultare indipendenti o dipendenti secondo la probabilità $\mathbf{P}\{\cdot\}$ assegnata. Questo diviene particolarmente evidente nel concetto di *indipendenza condizionata* che permette di confrontare l'indipendenza di eventi rispetto a $\mathbf{P}\{\cdot\}$ e $\mathbf{P}\{\cdot|D\}$.

Definizione 1.20. Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ diremo che A e B sono **eventi condizionatamente indipendenti rispetto a D** quando

$$\mathbf{P}\{AB|D\} = \mathbf{P}\{A|D\} \mathbf{P}\{B|D\}$$

essendo $D \in \mathcal{F}$ un evento tale che $\mathbf{P}\{D\} \neq 0$.

Potrebbe essere mostrato con semplici esempi – che per brevità trascureremo – che eventi A e B non indipendenti secondo la probabilità \mathbf{P} possono invece risultare condizionatamente indipendenti rispetto a qualche opportuno evento D . Anche il concetto di indipendenza condizionata giocherà un ruolo rilevante nella discussione della Markovianità nella seconda parte delle lezioni

Capitolo 2

Misure di probabilità

2.1 Probabilità su N

2.1.1 Spazi finiti e numerabili

Studieremo in questo capitolo le procedure con le quali è possibile definire una probabilità \mathbf{P} su uno spazio probabilizzabile (Ω, \mathcal{F}) che chiameremo anche **legge** o **distribuzione**, e inizieremo con l'esame di alcuni importanti esempi di spazi Ω finiti o numerabili: in questo caso infatti, come abbiamo visto, sarà possibile introdurre in maniera elementare la probabilità \mathbf{P} attribuendo un opportuno peso probabilistico a ciascun elemento ω dello spazio. I casi più noti riguardano insiemi di numeri interi e sono descritti negli esempi seguenti

Esempio 2.1. Distribuzioni binomiali: *Come primo esempio di spazio di campioni finito consideriamo l'insieme dei primi $n + 1$ numeri interi $\Omega_n = \{0, 1, \dots, n\}$ con la σ -algebra $\wp(\Omega_n)$ di tutte le sue parti: dato un numero $p \in [0, 1]$, e posto $q = 1 - p$, possiamo allora costruire una probabilità \mathbf{P} attribuendo innanzitutto ad ogni $\omega = k$ di Ω_n la probabilità*

$$p_n(k) = \binom{n}{k} p^k q^{n-k} \quad k = 0, 1, \dots, n \quad (2.1)$$

e definendo poi come

$$\mathbf{P}\{B\} = \sum_{k \in B} p_n(k) \quad (2.2)$$

la probabilità di ogni $B \subseteq \Omega_n$. Si verifica facilmente che tale \mathbf{P} è σ -additiva, che i suoi valori sono tutti in $[0, 1]$, e infine che

$$\mathbf{P}\{\Omega_n\} = \sum_{k=0}^n p_n(k) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1 \quad (2.3)$$

L'insieme dei numeri $p_n(k)$, che varia con $n = 1, 2, \dots$ e $p \in [0, 1]$, prende il nome di **distribuzione binomiale**, e nel seguito adotteremo il simbolo $\mathfrak{B}(n; p)$ per indicare

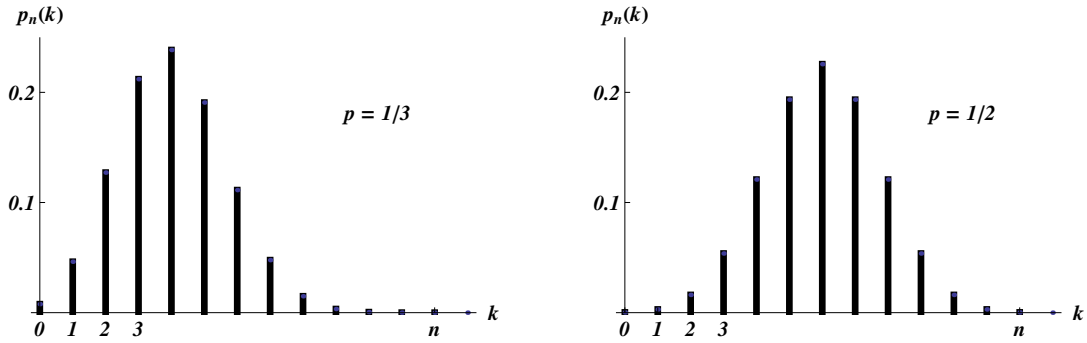


Figura 2.1: Diagrammi a barre di distribuzioni binomiali $\mathfrak{B}(n; p)$.

la famiglia di tali distribuzioni al variare di n e p . Il caso particolare $\mathfrak{B}(1; p)$ su $\Omega_1 = \{0, 1\}$ con

$$p_1(1) = p \quad p_1(0) = q = 1 - p$$

è particolarmente importante e porta il nome di **distribuzione di Bernoulli**. Nella Figura 2.1 è riportato il diagramma a barre di una tipica distribuzione binomiale per due differenti valori di p . Il significato di queste distribuzioni e la loro connessione con esperimenti consistenti, ad esempio, in n estrazioni indipendenti di palline da un'urna sarà discusso nella successiva Sezione 2.1.2.

Esempio 2.2. Distribuzione di Poisson: Generalizzando lo schema introdotto nell'esempio precedente per il caso finito, supponiamo ora che il nostro spazio dei campioni sia l'insieme infinito $\Omega = \mathbf{N} = \{0, 1, 2, \dots\}$, e che $\mathcal{F} = \wp(\mathbf{N})$ sia la σ -algebra degli eventi. Anche in questo caso possiamo costruire una probabilità \mathbf{P} partendo dalla seguente attribuzione di probabilità elementari ai singoli $\omega = k \in \mathbf{N}$

$$\mathbf{P}\{\omega\} = p_\alpha(k) = e^{-\alpha} \frac{\alpha^k}{k!} \quad \alpha > 0 \quad (2.4)$$

e definendo poi la probabilità di un $A \in \mathcal{F}$ tramite

$$\mathbf{P}\{A\} = \sum_{k \in A} p_\alpha(k)$$

Infatti la positività e l'additività si verificano facilmente, mentre la condizione $\mathbf{P}\{\Omega\} = 1$ deriva da

$$\mathbf{P}\{\Omega\} = \sum_{k \in \mathbf{N}} e^{-\alpha} \frac{\alpha^k}{k!} = e^{-\alpha} \sum_{k \in \mathbf{N}} \frac{\alpha^k}{k!} = e^{-\alpha} e^\alpha = 1$$

I valori (2.4) prendono il nome di **distribuzione di Poisson**, si indicano globalmente con il simbolo $\mathfrak{P}(\alpha)$ e assegnano probabilità non nulle a tutti gli infiniti

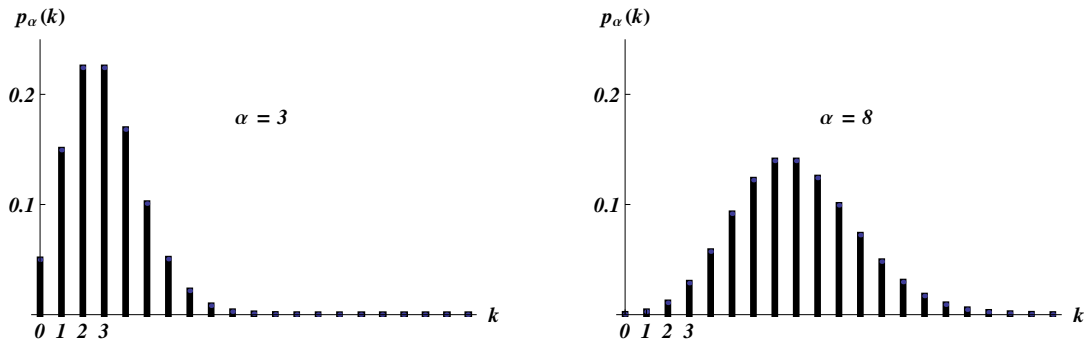


Figura 2.2: Diagrammi a barre di distribuzioni di Poisson $\mathfrak{P}(\alpha)$.

numeri interi $k = 0, 1, 2, \dots$. Il parametro $\alpha > 0$ che classifica le diverse $\mathfrak{P}(\alpha)$ e la formula (2.4) sono per il momento del tutto arbitrari: il loro significato sarà chiarito nella Sezione 4.5, e sarà in quell'occasione mostrato che in genere Ω rappresenta i risultati di conteggi del tipo “numero di particelle α emesse da un campione radioattivo e registrate da un contatore Geiger in un intervallo di 5 minuti”, oppure “numero di telefonate pervenute ad un centralino in un intervallo di un'ora”. Nella Figura 2.2 sono rappresentati alcuni esempi di distribuzioni di Poisson per diversi valori di α .

2.1.2 Modello di Bernoulli

Le distribuzioni binomiali come definite nell'Esempio 2.1 sono prive di un riferimento a situazioni concrete che possono chiarirne il significato: in particolare l'assegnazione iniziale delle $p_n(k)$ agli elementi $\omega = k \in \Omega_n$, pur essendo coerente, appare piuttosto immotivata. Proviamo allora a ricavare $\mathfrak{B}(n; p)$ dalla discussione di qualche modello più empirico, e per far questo consideriamo un esperimento consistente in **n estrazioni con rimessa** da un'urna contenente palline bianche e nere, e prendiamo come nuovo spazio dei campioni l'insieme Ω delle n -ple di estrazioni. Per comodità converremo di codificare i risultati delle estrazioni con un numero: 1 indicherà una pallina bianca, e 0 una nera. In questo modo i risultati dell'esperimento si presenteranno come n -ple ordinate di numeri eguali a 0, o a 1

$$\omega = (a_1, \dots, a_n) \quad a_i = 0, 1; \quad i = 1, \dots, n \quad (2.5)$$

Inoltre, trattandosi di uno spazio di campioni finito, adotteremo come famiglia degli eventi \mathcal{F} tutto $\wp(\Omega)$. Assegnamo ora ad ogni singolo $\omega = (a_1, \dots, a_n)$ la probabilità

$$\mathbf{P}\{\omega\} = p^k q^{n-k} \quad (2.6)$$

dove $k = \sum_i a_i$ è il numero di palline bianche in ω , $p \in [0, 1]$ è un numero arbitrario, e $q = 1 - p$. Questa scelta ci consente di nuovo, in analogia con (2.2), di definire

anche una probabilità su tutti gli eventi $A \in \mathcal{F}$ come

$$\mathbf{P}\{A\} = \sum_{\omega \in A} \mathbf{P}\{\omega\} \quad (2.7)$$

In realtà la definizione di \mathbf{P} su (Ω, \mathcal{F}) data da (2.6) e (2.7) è ancora una volta del tutto arbitraria, ma il suo significato sarà reso chiaro nel seguito della discussione. Innanzitutto si vede facilmente che la positività e l'additività della nostra \mathbf{P} sono garantite dalla (2.7). Per controllare invece che $\mathbf{P}\{\Omega\} = 1$ conviene definire preliminarmente gli $n + 1$ eventi

$$\begin{aligned} D_k &= \text{“trovo } k \text{ palline bianche su } n \text{ estrazioni”} \\ &= \left\{ \omega \in \Omega : \sum_{i=1}^n a_i = k \right\} \quad k = 0, \dots, n \end{aligned} \quad (2.8)$$

che ovviamente costituiscono di una decomposizione \mathcal{D} di Ω , e discutere poi la natura delle probabilità $\mathbf{P}\{D_k\}$

Proposizione 2.3. *Le probabilità $\mathbf{P}\{D_k\}$ degli elementi della decomposizione \mathcal{D} in (2.8) coincidono con le $p_n(k)$ di una distribuzione binomiale $\mathfrak{B}(n; p)$*

Dimostrazione: Siccome i k simboli 1 presenti nelle $\omega \in D_k$ possono collocarsi in vario modo sugli n posti disponibili, ogni D_k sarà costituito da un certo numero n_k di campioni a ciascuno dei quali la (2.6) assegna la stessa probabilità $\mathbf{P}\{\omega\} = p^k q^{n-k}$, e pertanto risulterà

$$\mathbf{P}\{D_k\} = \sum_{\omega \in D_k} \mathbf{P}\{\omega\} = n_k p^k q^{n-k}$$

Per determinare poi il valore di n_k notiamo che ogni risultato $\omega = (a_1 \dots a_n)$ con $k = \sum_i a_i$ fissato è univocamente individuato da un insieme di *numeri d'occupazione* $[b_1, \dots, b_k]$ che indicano la *posizione* dei k numeri 1 sugli n posti di ω : ad esempio, con $n = 7$

$$\omega = (0, 1, 1, 0, 0, 1, 0) \leftrightarrow [2, 3, 6]$$

È inoltre chiaro che per i nostri calcoli l'ordine delle b_j in $[b_1, \dots, b_k]$ è perfettamente irrilevante ($[2, 3, 6]$ e $[3, 6, 2]$ indicheranno nel nostro esempio la medesima 7-pla con il simbolo 1 al 2° , 3° e 6° posto); inoltre i valori delle b_j saranno tutti diversi dato che non indicheremo più di una volta che un posto b_j è occupato da un 1. Pertanto n_k coincide con il numero delle possibili k -ple non ordinate e senza ripetizioni $[b_1, \dots, b_k]$ dove ogni b_j assume i valori $1, 2, \dots, n$. Da (1.1) si ha allora che

$$n_k = \binom{n}{k} \quad \mathbf{P}\{D_k\} = p_n(k) = \binom{n}{k} p^k q^{n-k}$$

Pertanto le $p_n(k)$, introdotte senza motivazione nella definizione della distribuzione binomiale $\mathfrak{B}(n; p)$, coincidono con le probabilità $\mathbf{P}\{D_k\}$ degli eventi D_k in uno spazio Ω di n -ple di estrazioni con una \mathbf{P} definita a partire da (2.6) e (2.7) \square

A questo punto, essendo i D_k tutti disgiunti con $\bigcup_k D_k = \Omega$, dalla additività di \mathbf{P} e da (2.3) si ha che

$$\mathbf{P}\{\Omega\} = \mathbf{P}\left\{\bigcup_{k=0}^n D_k\right\} = \sum_{k=0}^n \mathbf{P}\{D_k\} = \sum_{k=0}^n p_n(k) = 1$$

il che completa la verifica della coerenza dell'assegnazione delle probabilità (2.6)

Ci resta da dare un significato più intuitivo al numero $p \in [0, 1]$ e alle formule (2.6) arbitrariamente introdotti all'inizio della discussione. A questo scopo definiamo innanzitutto, con $j = 1, \dots, n$, gli eventi

$$A_j = \text{"si trova una pallina bianca alla } j^{\text{ma}} \text{ estrazione"} = \{\omega \in \Omega : a_j = 1\}$$

Ovviamente il ritrovamento di una pallina nera alla j^{ma} estrazione corrisponde all'evento \bar{A}_j . Diversamente dagli eventi D_k , gli eventi A_j non sono disgiunti (l'aver trovato, infatti, una pallina bianca alla j^{ma} estrazione non esclude che se ne possano trovare altre in altre estrazioni) e non costituiscono quindi una decomposizione, ma essi sono comunque indispensabili per la discussione seguente

Proposizione 2.4. *Il numero $p \in [0, 1]$ coincide con la probabilità di trovare una pallina bianca in ogni singola estrazione, mentre $q = 1 - p$ è la corrispondente probabilità di trovarne una nera, ovvero si ha*

$$\mathbf{P}\{A_j\} = p, \quad \mathbf{P}\{\bar{A}_j\} = q = 1 - p$$

inoltre, comunque scelto p , gli eventi A_j sono tutti indipendenti fra di loro rispetto alla probabilità \mathbf{P} definita da (2.6), e questo spiega il significato di tali formule

Dimostrazione: Per brevità eviteremo la verifica completa¹ di questi risultati e ci limiteremo solo a qualche osservazione particolare. Per $n = 1$, cioè per una sola estrazione, lo spazio Ω si riduce a $\Omega_1 = \{0, 1\}$ e da (2.6) si ha banalmente

$$\mathbf{P}\{A_1\} = \mathbf{P}\{1\} = p \quad \mathbf{P}\{\bar{A}_1\} = \mathbf{P}\{0\} = q = 1 - p$$

Pertanto p risulta essere la probabilità di trovare una pallina bianca in una singola estrazione isolata, ma questo risultato può essere esteso poi anche al caso di estrazioni multiple con una verifica che ometteremo. D'altra parte, sempre con un calcolo diretto che qui trascuriamo, si prova che per $j \neq \ell$ si ha

$$\mathbf{P}\{A_j A_\ell\} = p^2 \quad \mathbf{P}\{A_j \bar{A}_\ell\} = pq \quad \mathbf{P}\{\bar{A}_j \bar{A}_\ell\} = q^2$$

sicché gli eventi A_j, A_k e le loro negazioni sono tutti indipendenti rispetto a \mathbf{P} di (2.7). Infine, siccome ogni $\omega \in \Omega$ coincide con l'intersezione di k eventi A_j e $n - k$ eventi \bar{A}_ℓ , è evidente che la scelta (2.6) per la probabilità di ω è stata operata proprio per garantire l'indipendenza di tali eventi □

¹N. Cufaro Petroni, CALCOLO DELLE PROBABILITÀ, Edizioni dal Sud (Bari, 1996)

In conclusione potremo dire che lo spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ con \mathbf{P} definita da (2.6) descrive *n tentativi indipendenti di verifica dell'evento*: “viene estratta una pallina bianca”, e che le $p_n(k)$ di una distribuzione binomiale $\mathfrak{B}(n; p)$ sono le *probabilità di trovare k palline bianche su n estrazioni indipendenti e con rimessa da un'urna contenente palline bianche e nere*. Ovviamente quello delle estrazioni dall'urna è solo un esempio particolare, e lo stesso modello potrà essere adattato a *n tentativi indipendenti di verifica di un qualsiasi evento A* che si verifichi con probabilità p in ogni singolo tentativo.

Questo schema prende il nome di **modello di Bernoulli** e costituisce un esempio di spazio di probabilità costruito come **prodotto diretto** di altri spazi di probabilità. Date infatti *n* copie distinte dello spazio di probabilità $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ che descrive una sola estrazione con $\Omega_1 = \{0, 1\}$, $\mathcal{F}_1 = \{1, 0, \Omega_1, \emptyset\}$ e \mathbf{P}_1 di Bernoulli $\mathfrak{B}(1; p)$ tale che

$$\mathbf{P}_1\{1\} = p, \quad \mathbf{P}_1\{0\} = 1 - p$$

lo spazio prodotto diretto sarà definito dallo spazio dei campioni prodotto Cartesiano $\Omega = \Omega_1 \times \dots \times \Omega_1$ delle *n*-ple $\omega = (a_1, \dots, a_n)$ di simboli 0 e 1, dalla σ -algebra \mathcal{F} di tutte le sue parti, e dalla probabilità \mathbf{P} definita sui singoli elementi di Ω dal prodotto

$$\mathbf{P}\{\omega\} = \mathbf{P}_1\{a_1\} \cdot \dots \cdot \mathbf{P}_1\{a_n\} = p^k q^{n-k} \quad k = \sum_{j=1}^n a_j$$

cioè da (2.7). Questa probabilità assegnata come *prodotto* non è l'unica scelta possibile per (Ω, \mathcal{F}) , ma evidentemente è proprio quella che corrisponde all'ipotesi di indipendenza delle estrazioni.

Esempio 2.5. Applicazioni del Teorema di Bayes: *Siamo ora in grado di discutere, come annunciato alla fine della Sezione 1.4, una semplice applicazione statistica del Teorema di Bayes (Proposizione 1.17). Consideriamo – con le notazioni della Sezione 1.4 – due urne D_1 e D_2 , esteriormente indistinguibili, contenenti ambedue palline bianche e nere, ma in proporzioni diverse. Più precisamente, la frazione di palline bianche in D_1 sia $\frac{1}{2}$, mentre quella in D_2 sia $\frac{2}{3}$. Supponiamo inoltre di non poter esaminare direttamente l'interno delle due urne, ma di poter eseguire un numero arbitrario di estrazioni con rimessa. Scegliamo ora una delle due urne, e chiediamoci quale delle due abbiamo preso. È evidente che $\mathcal{D} = \{D_1, D_2\}$ costituisce una decomposizione di Ω , e che, in assenza di altre informazioni, le probabilità a priori di aver preso una o l'altra urna saranno uguali e si avrà*

$$\mathbf{P}\{D_1\} = \mathbf{P}\{D_2\} = \frac{1}{2}$$

È però intuitivo pensare che per saperne di più sull'urna scelta basterà eseguire un certo numero di estrazioni con rimessa: così, ad esempio, l'osservazione di un numero elevato di palline bianche ci farebbe propendere verso l'idea di aver preso D_2 , e viceversa nel caso contrario. Il Teorema di Bayes e le considerazioni svolte

sulla distribuzione binomiale ci permettono ora di dare una veste quantitativa a queste considerazioni rimaste fin qui puramente qualitative. Supponiamo, infatti, per fissare le idee, di eseguire $n = 10$ estrazioni con rimessa dall'urna scelta, e di trovare $k = 4$ volte una pallina bianca, e $n - k = 6$ volte una pallina nera, cioè che si verifichi l'evento

$$B = \text{“su } n = 10 \text{ estrazioni si trovano } k = 4 \text{ palline bianche”}$$

Abbiamo già osservato che, nei due casi possibili di scelta delle nostre urne D_1 e D_2 , le probabilità dell'evento B sono rispettivamente date dalle distribuzioni binomiali $\mathfrak{B}(10; \frac{1}{2})$ e $\mathfrak{B}(10; \frac{2}{3})$, cioè sappiamo che

$$\begin{aligned} P\{B|D_1\} &= \binom{10}{4} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{10-4} = \binom{10}{4} \frac{1}{2^{10}} \\ P\{B|D_2\} &= \binom{10}{4} \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^{10-4} = \binom{10}{4} \frac{2^4}{3^{10}} \end{aligned}$$

e pertanto dal Teorema di Bayes otteniamo

$$\begin{aligned} P\{D_1|B\} &= \frac{P\{B|D_1\} P\{D_1\}}{P\{B|D_1\} P\{D_1\} + P\{B|D_2\} P\{D_2\}} = \frac{\frac{1}{2^{10}}}{\frac{1}{2^{10}} + \frac{2^4}{3^{10}}} \\ &= \frac{3^{10}}{3^{10} + 2^{14}} = 0.783 \\ P\{D_2|B\} &= \frac{2^{14}}{3^{10} + 2^{14}} = 0.217 \end{aligned}$$

Come si noterà, l'aver osservato un numero relativamente scarso di palline bianche favorisce – ma ora con una stima numerica precisa delle probabilità a posteriori – l'ipotesi di aver preso l'urna D_1 che contiene la proporzione più piccola di palline bianche. Naturalmente ulteriori estrazioni produrranno delle modifiche di questa valutazione ma, a lungo andare, ci attendiamo intuitivamente una stabilizzazione del risultato.

Esempio 2.6. Distribuzione multinomiale: La discussione sul Modello di Bernoulli ci permette anche di generalizzare in maniera intuitiva la distribuzione binomiale presentata nell'Esempio 2.1. Supponiamo di avere uno spazio dei campioni Ω composto ancora di n -ple ordinate $\omega = (a_1, \dots, a_n)$ di simboli che però ora possono assumere $r + 1$ (con $r \geq 1$) valori distinti b_0, b_1, \dots, b_r e non solo due. Per continuare con la presentazione adottata in questa sezione, possiamo pensare che si tratti di eseguire n estrazioni con rimessa da un'urna contenente palline di $r + 1$ colori diversi, e anche in questo caso per comodità possiamo indicare gli $r + 1$ possibili colori con i numeri $0, 1, 2, \dots, r$. Supponendo allora che k_i , $i = 0, 1, \dots, r$ sia il numero delle palline che in una data estrazione ω assumono il colore b_i , si parte con l'assegnare a tale ω la probabilità

$$P\{\omega\} = p_0^{k_0} p_1^{k_1} \cdot \dots \cdot p_r^{k_r}$$

dove ovviamente $k_0 + k_1 + \dots + k_r = n$, mentre p_0, p_1, \dots, p_r sono $r + 1$ arbitrari numeri non negativi tali che $p_0 + p_1 + \dots + p_r = 1$. Definiti poi gli eventi

$$D_{k_1 \dots k_r} = \text{“su } n \text{ estrazioni troviamo } k_0 \text{ volte } b_0, k_1 \text{ volte } b_1, \dots, k_r \text{ volte } b_r \text{”}$$

si prova che essi costituiscono una decomposizione di Ω , che ciascuno contiene

$$\binom{n}{k_1, \dots, k_r} = \frac{n!}{k_0! k_1! \dots k_r!}$$

possibili campioni ω equiprobabili, e infine che

$$\mathbf{P}\{D_{k_1 \dots k_r}\} = p_n(k_1, \dots, k_r) = \binom{n}{k_1, \dots, k_r} p_0^{k_0} p_1^{k_1} \cdot \dots \cdot p_r^{k_r} \quad (2.9)$$

L'insieme di tali probabilità prende il nome di **distribuzione multinomiale** e si indica con il simbolo $\mathfrak{B}(n; p_1, \dots, p_r)$. Si tratta quindi anche in questo caso di una famiglia di distribuzioni classificate dal numero n di estrazioni effettuate e dai numeri $p_i \in [0, 1]$, con $p_0 + p_1 + \dots + p_r = 1$, che sono – come si potrebbe verificare con un calcolo diretto – le probabilità di trovare il risultato b_i in ogni singola estrazione. Il corrispondente modello di Bernoulli sarà dunque quello di n estrazioni successive, con rimessa e indipendenti, da un'urna contenente palline di $r + 1$ colori diversi, e $p_n(k_1, \dots, k_r)$ rappresenterà la probabilità di trovare k_1 volte il colore $b_1 \dots k_r$ volte il colore b_r (non è necessario invece citare esplicitamente il colore b_0 trovato k_0 volte). La distribuzione binomiale $\mathfrak{B}(n; p)$ si otterrà come caso particolare per $r = 1$: in genere in questo caso le probabilità p_1 e p_0 vengono rispettivamente denotate con i simboli p e q , mentre $k_1 = k$ e $k_0 = n - k$.

2.2 Probabilità su \mathbf{R}

Passeremo ora ad esaminare gli strumenti con i quali è possibile definire una probabilità su spazi non finiti e non numerabili iniziando, ovviamente, con $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Le distribuzioni introdotte in precedenza in maniera elementare nella Sezione 2.1 su *spazi dei campioni finiti o numerabili* costituiti da numeri interi saranno qui ritrovate come casi particolari nella discussione sulle *distribuzioni discrete*.

2.2.1 Funzioni di distribuzione

Supponiamo prima di tutto che in qualche modo sia stata assegnata su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ una misura di probabilità \mathbf{P} , e consideriamo la funzione

$$F(x) = \mathbf{P}\{(-\infty, x]\}, \quad \forall x \in \mathbf{R} \quad (2.10)$$

Si verifica facilmente allora che $F(x)$ soddisfa alcune importanti proprietà

Proposizione 2.7. *La $F(x)$ definita da (2.10) soddisfa le seguenti proprietà*

1. $F(x)$ è non decrescente;
2. $F(+\infty) = 1$, $F(-\infty) = 0$;
3. $F(x)$ è continua da destra e ammette limite da sinistra $\forall x \in \mathbf{R}$ (cadlag).

Dimostrazione: Mentre 1 e 2 sono piuttosto ovvie se si tiene conto della (2.10), per provare 3 si osservi che $F(x)$, essendo monotona e limitata fra 0 ed 1, ammette sempre i limiti $F(x^+)$ ed $F(x^-)$ da destra e da sinistra in ogni $x \in \mathbf{R}$. Se allora $(x_n)_{n \in \mathbf{N}}$ è una successione numerica monotona con $x_n \downarrow x$ da destra, risulterà $(-\infty, x_n] \rightarrow (-\infty, x]$, e le proprietà di continuità delle probabilità (che qui non discuteremo) permettono allora di scrivere

$$F(x^+) = \lim_n F(x_n) = \lim_n \mathbf{P}\{(-\infty, x_n]\} = \mathbf{P}\{(-\infty, x]\} = F(x)$$

per cui $F(x)$ risulta continua da destra. Non si può invece dire, in generale, che $F(x)$ sia anche continua da sinistra: se infatti $x_n \uparrow x$ da sinistra risulta $(-\infty, x_n] \rightarrow (-\infty, x)$ e quindi

$$F(x^-) = \lim_n F(x_n) = \lim_n \mathbf{P}\{(-\infty, x_n]\} = \mathbf{P}\{(-\infty, x)\} \neq F(x)$$

Infatti, essendo $(-\infty, x] = (-\infty, x) \cup \{x\}$, avremo in generale

$$F(x) = \mathbf{P}\{(-\infty, x]\} = \mathbf{P}\{(-\infty, x)\} + \mathbf{P}\{x\} = F(x^-) + \mathbf{P}\{x\}$$

cioè $F(x^-) = F(x) - \mathbf{P}\{x\}$, e quindi $F(x)$ risulterà continua anche da sinistra (e quindi, in definitiva, continua *tout court*) se e solo se $\mathbf{P}\{x\} = 0$ \square

Dai risultati precedenti discende in particolare che $\mathbf{P}\{x\}$ può essere diversa da zero se e solo se $F(x)$ è discontinua in x , nel qual caso si ha

$$\mathbf{P}\{x\} = F(x) - F(x^-) = F(x^+) - F(x^-) \quad (2.11)$$

Inoltre, siccome $(-\infty, b] = (-\infty, a] \cup (a, b]$, dalla additività di \mathbf{P} risulta anche

$$\mathbf{P}\{(-\infty, b]\} = \mathbf{P}\{(-\infty, a]\} + \mathbf{P}\{(a, b]\}$$

e quindi

$$\mathbf{P}\{(a, b]\} = F(b) - F(a) \quad (2.12)$$

comunque siano scelti $-\infty \leq a < b \leq +\infty$

Definizione 2.8. *Chiameremo **Funzione di Distribuzione** (fdd) su \mathbf{R} ogni funzione $F(x)$ che soddisfi le proprietà 1, 2 e 3.*

La discussione svolta fin qui mostra dunque che ad ogni \mathbf{P} su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ è sempre associata una *fdd* $F(x)$. Il Teorema seguente mostra poi che è vera anche l'affermazione opposta: ad ogni *fdd* definita su \mathbf{R} è sempre possibile associare una misura di probabilità \mathbf{P} su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$

Teorema 2.9. *Data una fdd $F(x)$ su \mathbf{R} , esiste sempre un'unica probabilità \mathbf{P} su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ tale che*

$$\mathbf{P}\{(a, b]\} = F(b) - F(a)$$

comunque scelti a e b con $-\infty \leq a < b \leq +\infty$.

Dimostrazione: Omessa² □

Viene così stabilita una corrispondenza biunivoca fra le leggi \mathbf{P} su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ e le *fdd* $F(x)$ su \mathbf{R} in modo tale che una probabilità su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ può considerarsi assegnata se e solo se viene data la sua *fdd* $F(x)$. Inoltre, siccome nel seguito introdurremo su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ anche delle misure (come quella di Lebesgue) che non sono delle probabilità, sarà utile qui definire delle generalizzazioni dei concetti fin qui incontrati

Definizione 2.10. *Diremo che μ è una **misura di Lebesgue-Stieltjes (L-S)** su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ se essa è σ -additiva, e se $\mu(B) < +\infty$ per ogni intervallo limitato B . Chiameremo poi **Funzione di Distribuzione Generalizzata** su \mathbf{R} (*fddg*) ogni funzione $G(x)$ definita su \mathbf{R} che soddisfi le proprietà 1 e 3, ma in generale non la 2*

È facile mostrare ora che, se μ è una misura di L-S su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, la funzione $G(x)$ definita a meno di una costante additiva da

$$G(y) - G(x) = \mu(x, y], \quad x < y$$

è una *fddg*, mentre la proposizione inversa secondo la quale ad ogni *fddg* $G(x)$ si può sempre associare un'unica misura di L-S è enunciata nel seguente Teorema

Teorema 2.11. *Data una fddg $G(x)$ su \mathbf{R} , esiste sempre un'unica misura di L-S μ su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ tale che*

$$\mu(a, b] = G(b) - G(a)$$

comunque scelti a e b con $-\infty \leq a < b \leq +\infty$.

Dimostrazione: Omessa³ □

Si noti che una *fddg* $G(x)$ gode delle stesse proprietà di una *fdd* con l'eccezione della 2, e che pertanto essa può assumere valori negativi o maggiori di 1 e i suoi limiti per $x \rightarrow \pm\infty$ possono non essere finiti. Un ben noto esempio di questo genere è la **misura di Lebesgue** su \mathbf{R} , cioè la misura σ -finita λ che ad ogni $(a, b] \in \mathcal{B}(\mathbf{R})$ associa $\lambda(a, b] = b - a$: in questo caso la *fddg* è semplicemente

$$G(x) = x$$

²N. Cufaro Petroni, **CALCOLO DELLE PROBABILITÀ**, Edizioni dal Sud (Bari, 1996)

³A.N. Shiryaev, **PROBABILITY**, Springer (New York, 1996)

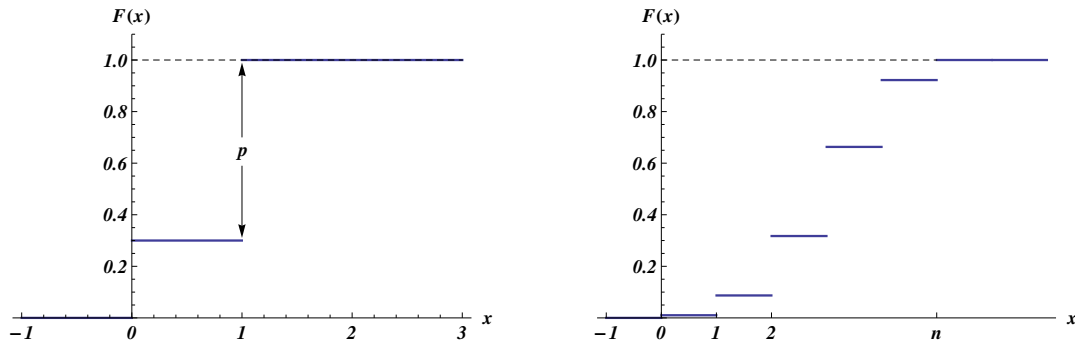


Figura 2.3: *fdd* di una Bernoulli $\mathfrak{B}(1, p)$ e di una binomiale $\mathfrak{B}(n, p)$.

2.2.2 Distribuzioni discrete

Definizione 2.12. Diremo che \mathbf{P} è una **distribuzione discreta** su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ se la sua *fdd* $F(x)$ è costante a tratti e cambia valore in maniera discontinua in un insieme (finito o numerabile) di punti x_1, x_2, \dots nei quali risulta $F(x_i) - F(x_i^-) > 0$.

La *fdd* di una legge discreta, dunque, presenta un tipico andamento a scala (vedi ad esempio la Figura 2.3) dal quale si deduce subito che $\mathbf{P}\{(a, b]\} = 0$ se $(a, b]$ non contiene nessuno dei punti di discontinuità x_i , mentre in generale si ha

$$\mathbf{P}(a, b] = \sum_{x_i \in (a, b]} [F(x_i) - F(x_i^-)] = F(b) - F(a)$$

In particolare risulta $\mathbf{P}\{x\} = 0$ nei punti x nei quali $F(x)$ è continua, e $p_i = \mathbf{P}\{x_i\} = F(x_i) - F(x_i^-)$ nei punti di discontinuità di $F(x)$. La probabilità \mathbf{P} risulta quindi concentrata sui punti x_1, x_2, \dots e può anche essere assegnata elencando i numeri p_1, p_2, \dots : anche essi per estensione prendono il nome di **distribuzione discreta**. Si noti che gli esempi di spazi di probabilità finiti o numerabili presentati nella Sezione 2.1 sono tutti casi particolari di distribuzioni discrete con le x_k coincidenti con numeri interi, come, per semplicità, supporremo anche nel seguito. La differenza principale fra le due trattazioni risiede nel fatto che nella Sezione 2.1 lo spazio Ω era identificato *solo* con l'insieme dei punti x_k , mentre qui Ω è tutto \mathbf{R} e le x_k sono i punti su cui è concentrata la probabilità. Questo implica che ora potremo rappresentare una distribuzione discreta non solo con i diagrammi a barre delle Figure 2.1 e 2.2, ma anche mediante una *fdd* $F(x)$ con x variabile continuamente in \mathbf{R} .

Esempio 2.13. *fdd* di distribuzioni discrete: Come primo esempio possiamo innanzitutto considerare quello delle distribuzioni in cui viene assunto un solo valore $b \in \mathbf{R}$ con probabilità 1, cioè \mathbf{P} -*qo*: la famiglia di tali distribuzione che portano il nome di **distribuzioni degeneri** viene indicata con il simbolo δ_b e sarà caratterizzata da una *fdd* $F(x)$ con un solo gradino di altezza 1 in $x = b$, coincidente quindi

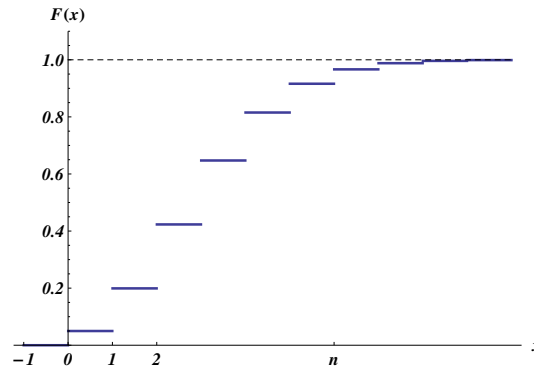


Figura 2.4: fdd della distribuzione di Poisson $\mathfrak{P}(\alpha)$.

con la cosiddetta **funzione di Heaviside**

$$\vartheta(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases} \quad (2.13)$$

Il diagramma a barre delle distribuzioni degeneri presenta ovviamente una sola barra di altezza 1 in $x = b$. Invece nella famiglia $\mathfrak{B}(1;p)$ delle **distribuzioni di Bernoulli** vengono assunti solo i valori 1 e 0 rispettivamente con probabilità p e $q = 1 - p$, mentre nella famiglia $\mathfrak{B}(n;p)$ delle **distribuzioni binomiali** vengono assunti solo i valori $k = 0, \dots, n$ con le probabilità

$$p_n(k) = \binom{n}{k} p^k q^{n-k}, \quad q = 1 - p, \quad 0 \leq p \leq 1$$

Pertanto le fdd delle Bernoulli e delle binomiali hanno l'andamento mostrato nella Figura 2.3. Infine nella famiglia $\mathfrak{P}(\alpha)$ delle **distribuzioni di Poisson** vengono assunti tutti i valori interi $k \in \mathbf{N}$ con le probabilità

$$p_k = \frac{\alpha^k e^{-\alpha}}{k!}, \quad \alpha > 0$$

per cui la corrispondente fdd ha l'aspetto mostrato in Figura 2.4

2.2.3 Distribuzioni assolutamente continue: densità

Definizione 2.14. Date due misure μ e ν su un medesimo spazio probabilizzabile (Ω, \mathcal{F}) , si dice che ν è **assolutamente continua (ac)** rispetto a μ (e si scrive $\nu \ll \mu$) quando $\mu(A) = 0$, con $A \in \mathcal{F}$, implica $\nu(A) = 0$. Nel caso particolare di misure su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, se una misura di probabilità \mathbf{P} è ac rispetto alla misura di Lebesgue si usa anche dire, per brevità, che la fdd $F(x)$ di \mathbf{P} è assolutamente continua

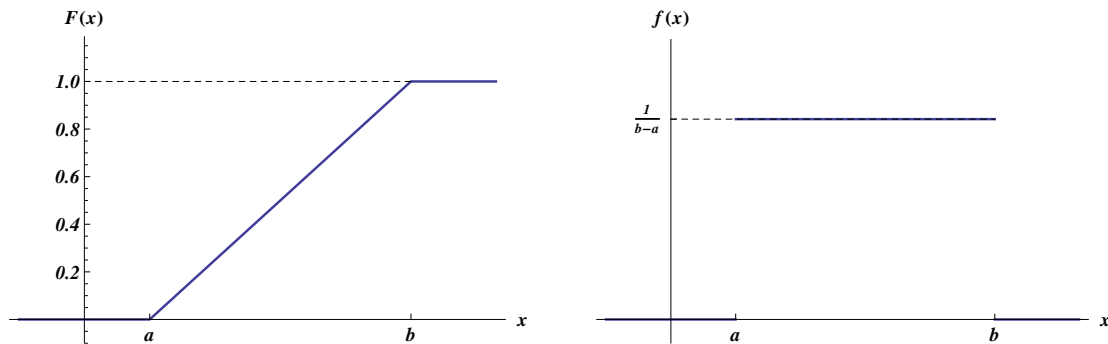


Figura 2.5: *fdd* e *ddp* della distribuzione uniforme $\mathfrak{U}(a, b)$.

Teorema 2.15. Teorema di Radon-Nikodym su \mathbf{R} : Una *fdd* $F(x)$ su \mathbf{R} è *ac* se e solo se esiste una funzione $f(x)$ definita su \mathbf{R} a valori non negativi e tale che

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \quad F(x) = \int_{-\infty}^x f(z) dz$$

La $f(x)$ prende il nome di **densità di probabilità (ddp)** di $F(x)$.

Dimostrazione: Omessa⁴ □

Sarebbe facile mostrare che, data una funzione $f(x)$ integrabile secondo Lebesgue, normalizzata e a valori non negativi, la funzione

$$F(x) = \int_{-\infty}^x f(z) dz$$

risulterebbe sempre essere una *fdd ac*. Il Teorema di Radon-Nikodym stabilisce il fatto notevole che risulta vero anche il viceversa: ogni *fdd ac* risulta essere sempre la primitiva di un'opportuna *ddp* $f(x)$, sicché tutte e sole le *fdd ac* possono essere assegnate tramite una *ddp*. È possibile dimostrare che se una *fdd* $F(x)$ è *ac* essa è anche continua e risulta derivabile dovunque, eventualmente con l'eccezione di un insieme di punti di misura di Lebesgue nulla⁵, e in questo caso la *ddp* non è altro che la sua derivata

$$f(x) = F'(x)$$

Tenendo allora conto di (2.12) e della continuità di $F(x)$, una *ddp* $f(x)$ permette di calcolare la probabilità di intervalli $[a, b]$ secondo la formula

⁴M. Métivier, NOTIONS FONDAMENTALES DE LA THÉORIE DES PROBABILITÉS, Dunod (Paris, 1972)

⁵Viceversa esistono, come vedremo nella Sezione 2.2.4, delle *fdd* $F(x)$ continue, ma non *ac*, sicché l'esistenza della *ddp* non è garantita dalla semplice continuità della *fdd*

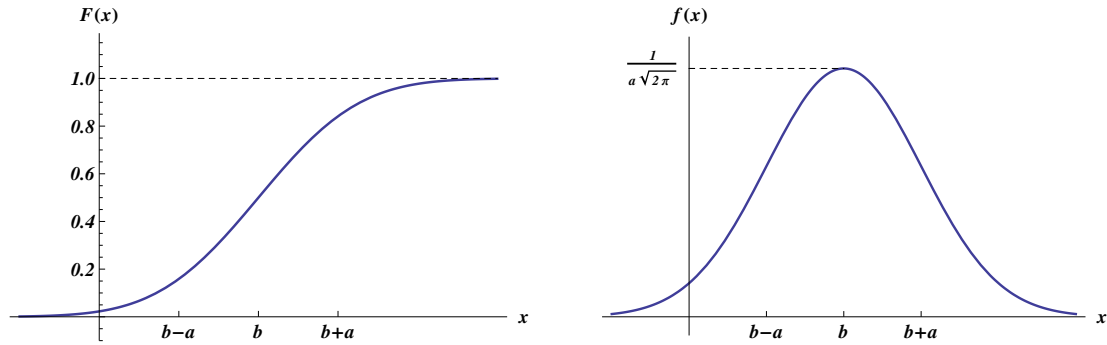


Figura 2.6: *fdd* e *ddp* della distribuzione normale $\mathfrak{N}(b, a^2)$.

$$\mathbf{P}\{[a, b]\} = F(b) - F(a) = \int_a^b f(t) dt$$

Viceversa è evidente che una *fdd* discreta non può mai essere *ac* dal momento che essa non è neanche continua: in questo caso quindi non si può mai assegnare una *ddp*, e bisogna limitarsi all'uso della *fdd*

Esempio 2.16. Distribuzioni uniformi: *Il primo caso è la famiglia dalle leggi uniformi su un intervallo $[a, b]$ che denoteremo con il simbolo $\mathfrak{U}(a, b)$. La fdd è*

$$F(x) = \begin{cases} 0, & \text{se } x < a, \\ \frac{x-a}{b-a}, & \text{se } a \leq x \leq b, \\ 1, & \text{se } b < x, \end{cases}$$

e il suo grafico tipico è presentato nella Figura 2.5. Questa fdd definisce su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ una probabilità \mathbf{P} concentrata su $[a, b]$ che ad ogni intervallo $[x, y] \subseteq [a, b]$ associa la probabilità

$$\mathbf{P}\{[x, y]\} = \frac{y - x}{b - a}$$

Viceversa intervalli esterni ad $[a, b]$ ricevono probabilità nulla, e – data la continuità della $F(x)$ – risulterà anche $\mathbf{P}\{x\} = 0$ per ogni evento ridotto a un solo punto x . La *ddp*, che si ricava facilmente per derivazione, è

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{altrove} \end{cases} \quad (2.14)$$

ed è rappresentata nella Figura 2.5 assieme alla sua *fdd*. Questa *ddp* rende anche conto del nome attribuito a queste distribuzioni in quanto la probabilità di ogni intervallo $[x_1, x_2] \subseteq [a, b]$ dipende solo dalla sua ampiezza $x_2 - x_1$ e non dalla sua posizione all'interno di $[a, b]$: in altri termini tutte le zone di $[a, b]$ sono uniformemente pesate

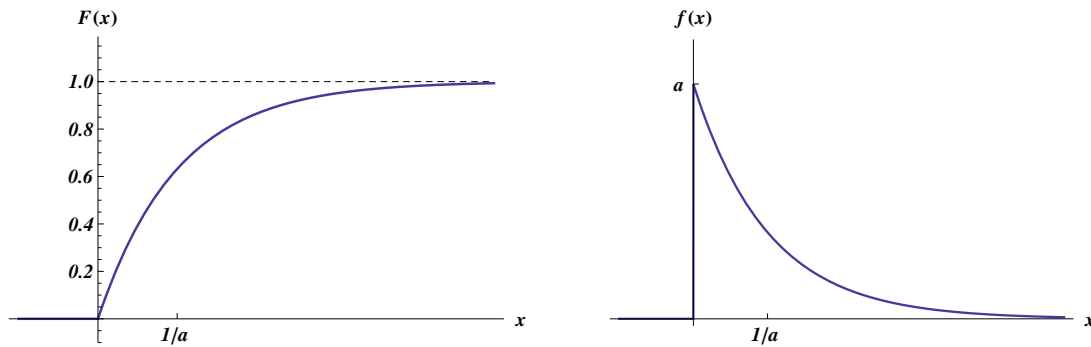


Figura 2.7: fdd e ddp della distribuzione esponenziale $\mathfrak{E}(a)$.

Esempio 2.17. Distribuzioni Gaussiane (normali): La famiglia delle **leggi Gaussiane (normali)** $\mathfrak{N}(b, a^2)$ è caratterizzata dalla ddp

$$f(x) = \frac{1}{a\sqrt{2\pi}} e^{-(x-b)^2/2a^2} \quad a > 0, b \in \mathbf{R} \quad (2.15)$$

nota come *funzione di Gauss* e rappresentata assieme alla corrispondente fdd nella Figura 2.6. Il caso cosiddetto *degenere* di $a = 0$, che qui escludiamo, ha bisogno di una discussione specifica che sarà svolta in seguito. Questa ddp ha una caratteristica forma a campana con il massimo collocato in $x = b$. I due flessi in $x = b \pm a$ danno una misura della larghezza della curva che quindi dipende dal parametro a . Parleremo inoltre di **legge normale standard** se la ddp (2.15) ha $b = 0$ e $a = 1$, cioè è

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Le fdd Gaussiane standard e non standard, dette anche **funzioni errore**, sono rispettivamente

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz \quad F(x) = \frac{1}{a\sqrt{2\pi}} \int_{-\infty}^x e^{-(z-b)^2/2a^2} dz \quad (2.16)$$

e hanno un grafico come quello mostrato in Figura 2.6: esse non sono esprimibili mediante combinazioni finite di funzioni elementari, e possono essere usate solo calcolandole numericamente

Esempio 2.18. Distribuzioni esponenziali: La famiglia delle **leggi esponenziali** $\mathfrak{E}(a)$ è caratterizzata dalla ddp

$$f(x) = a e^{-ax} \vartheta(x) = \begin{cases} a e^{-ax} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases} \quad a > 0 \quad (2.17)$$

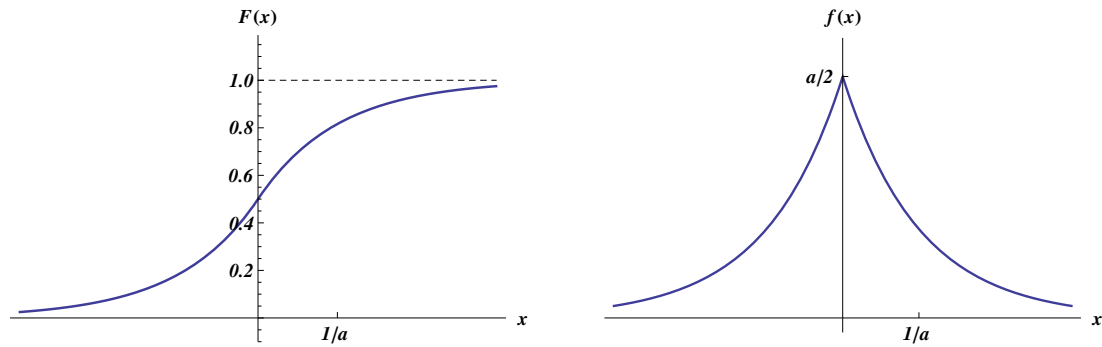


Figura 2.8: fdd e ddp della distribuzione di Laplace $\mathfrak{L}(a)$.

e dalla corrispondente fdd

$$F(x) = (1 - e^{-ax})\vartheta(x) = \begin{cases} 1 - e^{-ax} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$$

ambidue rappresentate nella Figura 2.7.

Esempio 2.19. Distribuzioni di Laplace: Si chiamano invece **leggi di Laplace**, o anche esponenziali bilatere, e si indicano con il simbolo $\mathfrak{L}(a)$, le leggi con ddp

$$f(x) = \frac{a}{2} e^{-a|x|} \quad a > 0 \quad (2.18)$$

e fdd

$$F(x) = \frac{1}{2} + \frac{|x|}{x} \frac{1 - e^{-a|x|}}{2}$$

rappresentate nella Figura 2.8.

Esempio 2.20. Distribuzioni di Cauchy: Infine la famiglia delle **leggi di Cauchy** $\mathfrak{C}(b, a)$ ha come ddp

$$f(x) = \frac{1}{\pi} \frac{a}{a^2 + (x - b)^2} \quad a > 0 \quad (2.19)$$

e come fdd

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{x - b}{a}$$

ambidue rappresentate nella Figura 2.9. Come si può notare da un rapido confronto fra le Figure 2.6 e 2.9, qualitativamente le ddp delle $\mathfrak{N}(b, a^2)$ e $\mathfrak{C}(b, a)$ sono piuttosto simili: in tutti e due i casi si tratta di curve a campana, centrate simmetricamente attorno a $x = b$ e con una larghezza regolata dal parametro $a > 0$. Esse differiscono invece essenzialmente per la velocità con la quale le loro code tendono a zero: mentre

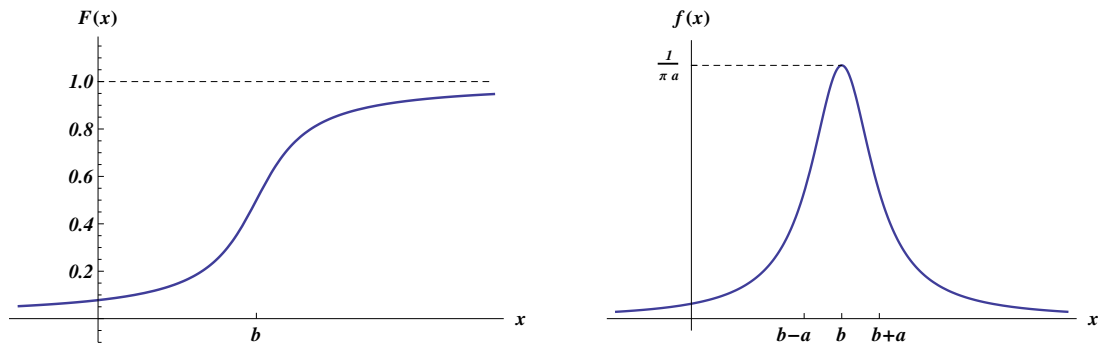


Figura 2.9: *fdd* e *ddp* della distribuzione di Cauchy $\mathfrak{C}(b, a)$.

la *ddp* normale si annulla molto rapidamente, la funzione di Cauchy tende a zero molto più lentamente come x^{-2} . Conseguentemente il corpo centrale della *ddp* di Cauchy è più sottile rispetto a quello della normale, mentre invece le code sono corrispondentemente più robuste

2.2.4 Distribuzioni singolari

Definizione 2.21. Diremo che \mathbf{P} è una **distribuzione singolare** quando la sua *fdd* $F(x)$ è continua, ma non *ac*.

Abbiamo visto che le misure di probabilità *ac* rispetto alla misura di Lebesgue, cioè quelle con una *fdd* $F(x)$ *ac*, sono caratterizzate da una *ddp* $f(x)$. Sappiamo inoltre che una $F(x)$ *ac* è anche certamente continua, mentre in generale il viceversa non è vero: esistono cioè – e se ne potrebbero produrre esempi che qui trascureremo – *fdd* $F(x)$ continue (quindi non discrete) che non sono *ac* (quindi non ricavabili da una *ddp*). Questa osservazione spiega la necessità di introdurre la definizione precedente e mette in evidenza il fatto che una legge singolare non può essere assegnata né tramite una distribuzione discreta di probabilità (cioè tramite un insieme di numeri p_k), né tramite una *ddp* $f(x)$: l'unico modo per definirla, quindi, è quello di assegnare una opportuna *fdd* continua $F(x)$ la cui esistenza è comunque garantita. Nel seguito ci limiteremo però solo ai casi in cui la probabilità può essere data o con una distribuzione discreta, oppure con una *ddp*, per cui le distribuzioni singolari giocheranno per noi solo un ruolo marginale.

2.2.5 Miscela

Definizione 2.22. Diremo che una distribuzione \mathbf{P} è una **miscela** quando la sua *fdd* $F(x)$ è una combinazione convessa di altre *fdd*, cioè quando essa può essere

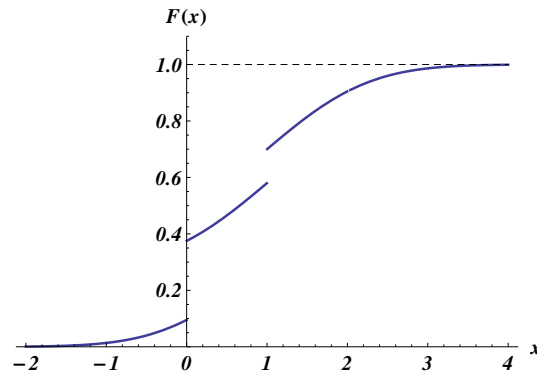


Figura 2.10: *fdd* corrispondente alla miscela di una Bernoulli e di una normale.

rappresentata come

$$F(x) = \sum_{k=1}^n p_k F_k(x), \quad 0 \leq p_k \leq 1, \quad \sum_{k=1}^n p_k = 1$$

dove le $F_k(x)$ sono n *fdd*.

In questa definizione le $F_k(x)$ possono essere di natura varia. Ovviamente se tutte sono *ac* con *ddp* $f_k(x)$ si vede subito che anche $F(x)$ risulta *ac* con *ddp*

$$f(x) = \sum_{k=1}^n p_k f_k(x)$$

Niente vieta, però, che una miscela possa essere composta di tutti i tipi di *fdd*, anzi il seguente importante risultato usa proprio questa eventualità per mettere in evidenza che i casi di distribuzioni fin qui introdotti (discreto, assolutamente continuo e singolare) esauriscono in effetti tutte le possibilità a disposizione

Teorema 2.23. Teorema di Lebesgue-Nikodym: *Ogni distribuzione \mathbf{P} sullo spazio $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ può essere rappresentata come una miscela di probabilità discrete, ac e singolari, cioè la sua *fdd* $F(x)$ è una combinazione convessa*

$$F(x) = p_1 F_1(x) + p_2 F_2(x) + p_3 F_3(x)$$

con F_1 discreta, F_2 *ac*, F_3 singolare, e con p_1, p_2, p_3 numeri positivi tali che $p_1 + p_2 + p_3 = 1$.

Dimostrazione: Omessa⁶

□

⁶M. Métivier, NOTIONS FONDAMENTALES DE LA THÉORIE DES PROBABILITÉS, Dunod (Paris, 1972)

Esempio 2.24. Miscela: *Illustreremo questi concetti con un semplice esempio: nella Figura 2.10 è riportata la fdd di una miscela di una normale $\mathfrak{N}(b, a^2)$ e di una Bernoulli $\mathfrak{B}(1; p)$, con coefficienti p_1, p_2 arbitrari. La $F(x)$ risultante presenterà ovviamente delle discontinuità in $x = 0$ ed $x = 1$ a causa della presenza di una componente di Bernoulli, ma non risulterà costante (come per una distribuzione puramente discreta) nei tratti di continuità a causa della presenza della componente normale. Pertanto in questo semplice caso, anche in assenza di componenti singolari, la probabilità non può essere data né con una distribuzione discreta p_k , né con una ddp $f(x)$: l'unica rappresentazione corretta è quella che fa uso della fdd $F(x)$*

2.3 Probabilità su \mathbf{R}^n

Per spazi dei campioni di tipo $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ possiamo estendere, con solo qualche complicazione formale, le procedure adottate nel caso di $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ nella Sezione 2.2. Compare però in aggiunta l'importante novità delle *distribuzioni marginali* di una *distribuzione multivariata*, e il problema delle loro relazioni reciproche

2.3.1 Funzioni di distribuzione multivariate

In analogia con quanto fatto nella Sezione 2.2.1 considereremo innanzitutto \mathbf{P} come una probabilità assegnata su $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$, e definiremo la seguente funzione di n variabili

$$F(\mathbf{x}) = F(x_1, \dots, x_n) = \mathbf{P}\{(-\infty, x_1] \times \dots \times (-\infty, x_n]\} \quad (2.20)$$

con $\mathbf{x} = (x_1, \dots, x_n)$. Introducendo allora le notazioni sintetiche

$$\begin{aligned} \Delta_k F(\mathbf{x}) &= F(x_1, \dots, x_k + \Delta x_k, \dots, x_n) - F(x_1, \dots, x_k, \dots, x_n) \\ (\mathbf{x}, \mathbf{x} + \Delta \mathbf{x}) &= (x_1, x_1 + \Delta x_1] \times \dots \times (x_n, x_n + \Delta x_n] \end{aligned}$$

con $\Delta x_k \geq 0$, è possibile vedere che

$$\mathbf{P}\{(\mathbf{x}, \mathbf{x} + \Delta \mathbf{x})\} = \Delta_1 \dots \Delta_n F(\mathbf{x})$$

Ad esempio nel caso $n = 2$ si ha

$$\begin{aligned} \mathbf{P}\{(\mathbf{x}, \mathbf{x} + \Delta \mathbf{x})\} &= \Delta_1 \Delta_2 F(\mathbf{x}) \\ &= [F(x_1 + \Delta x_1, x_2 + \Delta x_2) - F(x_1 + \Delta x_1, x_2)] \\ &\quad - [F(x_1, x_2 + \Delta x_2) - F(x_1, x_2)] \end{aligned}$$

come si può controllare facilmente esaminando la Figura 2.11. Si osservi che, diversamente dal caso $n = 1$, la probabilità $\mathbf{P}\{(\mathbf{x}, \mathbf{x} + \Delta \mathbf{x})\}$ del dominio ottenuto come prodotto cartesiano di intervalli non coincide con la semplice differenza $F(\mathbf{x} + \Delta \mathbf{x}) - F(\mathbf{x})$, ma è data da una combinazione di 2^n termini prodotti dall'applicazione ripetuta dell'operazione Δ_k . Le funzioni $F(\mathbf{x})$ così definite godono di alcune proprietà che generalizzano quelle del caso $n = 1$ discusse nella Sezione 2.2.1

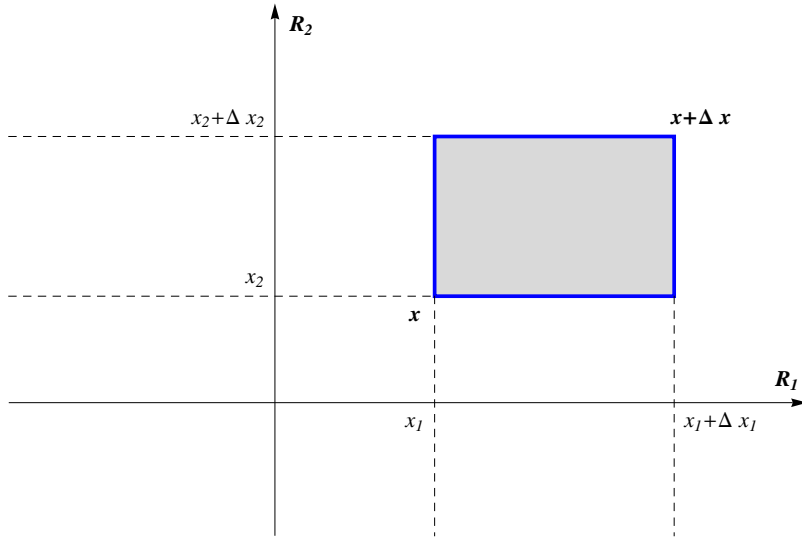


Figura 2.11: Probabilità di prodotti cartesiani di intervalli.

Proposizione 2.25. *La $F(\mathbf{x})$ definita da (2.20) gode delle seguenti proprietà:*

1. *Comunque scelti dei $\Delta x_k \geq 0$ con $k = 1, \dots, n$ risulta sempre*

$$\Delta_1 \dots \Delta_n F(\mathbf{x}) \geq 0$$

e in particolare $F(\mathbf{x})$ è non decrescente in ciascuna delle sue n variabili.

2. *Si ha sempre*

$$\lim_{\mathbf{x} \rightarrow +\infty} F(\mathbf{x}) = 1 \qquad \lim_{\mathbf{x} \rightarrow -\infty} F(\mathbf{x}) = 0$$

*dove il limite per $\mathbf{x} \rightarrow +\infty$ è inteso nel senso che **ciascuna** delle x_k tende verso $+\infty$, mentre quello per $\mathbf{x} \rightarrow -\infty$ è inteso nel senso che **almeno una** delle x_k tende a $-\infty$*

3. *$F(\mathbf{x})$ risulta continua **da sopra**, cioè*

$$\lim_{\mathbf{x}_k \downarrow \mathbf{x}} F(\mathbf{x}_k) = F(\mathbf{x})$$

*dove il limite per $\mathbf{x}_k \downarrow \mathbf{x}$ è inteso nel senso che tutte le componenti degli elementi della successione \mathbf{x}_k devono tendere **decrescendo** verso le omologhe componenti di \mathbf{x} .*

Dimostrazione: La prima deriva immediatamente dalla positività della probabilità; la seconda discende invece dal fatto ovvio che l'insieme

$$(-\infty, +\infty] \times \dots \times (-\infty, +\infty] = \mathbf{R}^n$$

coincide con tutto lo spazio dei campioni, mentre ogni prodotto cartesiano di intervalli nel quale anche un solo fattore coincide con la parte vuota è esso stesso vuoto. Per l'ultima, infine, si può sviluppare un argomento analogo a quello della Proposizione 2.7 che ometteremo per brevità \square

Sulla base di queste osservazioni possiamo ora dare delle definizioni ed enunciare dei risultati che generalizzano quelli dati per il caso $n = 1$.

Definizione 2.26. Chiameremo **funzione di distribuzione multivariata** su \mathbf{R}^n ogni funzione $F(\mathbf{x})$ che soddisfi le proprietà 1, 2 e 3; chiameremo invece **funzione di distribuzione multivariata generalizzata** su \mathbf{R}^n ogni funzione $G(\mathbf{x})$ che soddisfi le proprietà 1, 3 senza essere vincolata da 2 a restare confinata tra 0 ed 1.

Teorema 2.27. Data una fdd multivariata $F(\mathbf{x})$ su \mathbf{R}^n , esiste sempre un'unica probabilità \mathbf{P} su $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ tale che comunque scelti $\Delta x_k \geq 0$ con $k = 1, \dots, n$ risulti

$$\mathbf{P}\{(\mathbf{x}, \mathbf{x} + \Delta\mathbf{x})\} = \Delta_1 \dots \Delta_n F(\mathbf{x})$$

Analogamente, data una fddg multivariata $G(\mathbf{x})$ su \mathbf{R}^n , esiste sempre un'unica misura di Lebesgue-Stieltjes μ su $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ tale che comunque scelti $\Delta x_k \geq 0$ con $k = 1, \dots, n$ risulti

$$\mu(\mathbf{x}, \mathbf{x} + \Delta\mathbf{x}) = \Delta_1 \dots \Delta_n G(\mathbf{x})$$

Dimostrazione: Omessa e analoga a quella della Proposizione 2.9 \square

In sostanza anche nel caso n -dimensionale una probabilità \mathbf{P} o una misura di Lebesgue-Stieltjes μ possono sempre essere assegnate rispettivamente mediante una fdd $F(\mathbf{x})$ o una fddg $G(\mathbf{x})$.

Esempio 2.28. Date la fdd univariata di una legge uniforme su $[0, 1]$, e la fddg univariata della misura di Lebesgue

$$F_1(x) = \begin{cases} 0, & \text{se } x < 0; \\ x, & \text{se } 0 \leq x \leq 1; \\ 1, & \text{se } 1 < x, \end{cases} \quad G_1(x) = x$$

si prova facilmente che le funzioni

$$\begin{aligned} F(\mathbf{x}) &= F_1(x_1) \cdot \dots \cdot F_1(x_n) \\ G(\mathbf{x}) &= G_1(x_1) \cdot \dots \cdot G_1(x_n) = x_1 \cdot \dots \cdot x_n \end{aligned}$$

sono rispettivamente la fdd di una legge uniforme sull'ipercubo $[0, 1]^n$, e la fddg della misura di Lebesgue su \mathbf{R}^n .

Quello precedente è un esempio di una situazione più generale: se $F_1(x), \dots, F_n(x)$ sono n fdd su \mathbf{R} si verifica subito che la funzione di n variabili

$$F(\mathbf{x}) = F_1(x_1) \cdot \dots \cdot F_n(x_n)$$

è sempre una fdd su \mathbf{R}^n . Non è invece vero invece, in generale, il viceversa: una generica fdd su \mathbf{R}^n non si fattorizza sempre nel prodotto di n fdd su \mathbf{R} ; ciò avviene solo in alcuni casi notevoli che saranno discussi nel seguito.

2.3.2 Densità multivariate

Quando \mathbf{P} è *ac* rispetto alla misura di Lebesgue su $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$, e quindi $F(\mathbf{x})$ è *ac*, una semplice generalizzazione del Teorema di Radon–Nikodym 2.15 garantisce l'esistenza di una **funzione di densità multivariata** $f(\mathbf{x})$ non negativa e normalizzata

$$\int_{\mathbf{R}^n} f(x_1, \dots, x_n) dx_1 \dots dx_n = \int_{\mathbf{R}^n} f(\mathbf{x}) d^n \mathbf{x} = 1$$

Date allora una *fdd ac* $F(\mathbf{x})$ e la corrispondente *ddp* $f(\mathbf{x})$, risulta

$$F(\mathbf{x}) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(\mathbf{z}) d^n \mathbf{z} \quad f(\mathbf{x}) = \frac{\partial^n F(\mathbf{x})}{\partial x_1 \dots \partial x_n} \quad (2.21)$$

mentre le probabilità di prodotti cartesiani di intervalli si calcolano come

$$\mathbf{P} \{(a_1, b_1] \times \dots \times (a_n, b_n]\} = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f(\mathbf{x}) d^n \mathbf{x}$$

Esempio 2.29. Leggi Gaussiane (normali) multivariate: *La famiglia delle leggi normali multivariate $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ è caratterizzata dai vettori $\mathbf{b} = (b_1, \dots, b_n)$ di numeri reali, e dalle matrici $\mathbb{A} = \|a_{ij}\|$, simmetriche ($a_{ij} = a_{ji}$) e definite non negative⁷: il significato statistico degli elementi di \mathbf{b} e \mathbb{A} sarà chiarito nella Sezione 3.34. Quando \mathbb{A} è strettamente positiva esiste la matrice inversa \mathbb{A}^{-1} e la $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ è dotata di una *ddp multivariata*:*

$$f(\mathbf{x}) = f(x_1, \dots, x_n) = \sqrt{\frac{|\mathbb{A}^{-1}|}{(2\pi)^n}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{b}) \cdot \mathbb{A}^{-1}(\mathbf{x}-\mathbf{b})} \quad (2.22)$$

dove con $|\mathbb{A}^{-1}|$ indichiamo il determinante di \mathbb{A}^{-1} , con $\mathbf{x} \cdot \mathbf{y} = \sum_k x_k y_k$ indichiamo l'usuale prodotto scalare Euclideo tra in vettori \mathbf{x} e \mathbf{y} , mentre fra vettori e matrici viene adottato l'usuale prodotto righe per colonne sicché, ad esempio

$$\mathbf{x} \cdot \mathbb{A} \mathbf{y} = \sum_{i,j=1}^n a_{ij} x_i y_j$$

⁷La matrice \mathbb{A} si dice **definita non negativa** se, comunque scelto un vettore di numeri reali $\mathbf{x} = (x_1, \dots, x_n)$, risulta sempre

$$\mathbf{x} \cdot \mathbb{A} \mathbf{x} = \sum_{i,j=1}^n a_{ij} x_i x_j \geq 0$$

la matrice si dice, invece, positiva se tale somma è sempre strettamente **positiva** (cioè è sempre diversa da zero). Se \mathbb{A} è positiva essa risulta anche **non singolare** nel senso che il suo determinante $|\mathbb{A}| > 0$ è non nullo, e quindi è dotata di inversa \mathbb{A}^{-1} .

Invece (analogamente al caso $a = 0$ per $n = 1$) le leggi $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ con \mathbb{A} singolare e non invertibile, pur essendo ben definite, non ammettono una ddp e saranno discusse in maniera più dettagliata nella Sezione 4.2.2. La ddp normale multivariata (2.22) è dunque una generalizzazione di quella univariata discussa nella Sezione 2.2.3: nel caso $n = 1$ la ddp della $\mathfrak{N}(b, a^2)$ era caratterizzata solo da due parametri numerici, b e $a \geq 0$; nel caso multivariato, invece, bisogna assegnare un vettore \mathbf{b} e una matrice simmetrica e non negativa \mathbb{A} . Notiamo in particolare che per $n = 2$ l'inversa di \mathbb{A} è

$$\mathbb{A}^{-1} = \frac{1}{(1-r^2)a_1^2a_2^2} \begin{pmatrix} a_2^2 & -a_1a_2r \\ -a_1a_2r & a_1^2 \end{pmatrix} \quad (2.23)$$

e la **ddp normale bivariata** assume la forma

$$f(x_1, x_2) = \frac{e^{-\frac{1}{2(1-r^2)} \left[\frac{(x_1-b_1)^2}{a_1^2} - 2r \frac{(x_1-b_1)(x_2-b_2)}{a_1a_2} + \frac{(x_2-b_2)^2}{a_2^2} \right]}}{2\pi a_1 a_2 \sqrt{1-r^2}} \quad (2.24)$$

dove i tre numeri $a_k = \sqrt{a_{kk}} > 0$, $k = 1, 2$, e $r = a_{12}/\sqrt{a_{11}a_{22}}$ con la limitazione $|r| < 1$, sono derivati dai valori dei tre elementi distinti della matrice \mathbb{A} .

2.3.3 Distribuzioni marginali

Data una *fdd* $F(\mathbf{x})$ su $\mathbf{R}^n = \mathbf{R}_1 \times \dots \times \mathbf{R}_n$ si può dimostrare facilmente che la funzione di $n - 1$ variabili

$$F^{(1)}(x_2, \dots, x_n) = F(+\infty, x_2, \dots, x_n) = \lim_{x_1 \rightarrow +\infty} F(x_1, x_2, \dots, x_n) \quad (2.25)$$

è ancora una *fdd* su $\mathbf{R}^{n-1} = \mathbf{R}_2 \times \dots \times \mathbf{R}_n$ nel senso che essa soddisfa le proprietà 1, 2 e 3 della Sezione 2.3.1. Questo rimane vero quale che sia la coordinata x_i scelta per eseguire il limite; scegliendo però coordinate diverse si ottengono *fdd* $F^{(i)}$ che sono in generale diverse fra loro. Per questo motivo, dovunque necessario per evitare ambiguità, adotteremo una notazione con indici in alto che segnalano *quali coordinate sono state eliminate*. Il limite può inoltre essere anche eseguito su $m < n$ variabili, e anche in questo caso si otterrebbero delle *fdd* su un \mathbf{R}^{n-m} : ad esempio

$$F^{(1,2)}(x_3, \dots, x_n) = F(+\infty, +\infty, x_3, \dots, x_n) \quad (2.26)$$

è una *fdd* su $\mathbf{R}^{n-2} = \mathbf{R}_3 \times \dots \times \mathbf{R}_n$. Proseguendo in questo modo possono infine essere ricavate n *fdd* ad una sola variabile, come ad esempio la *fdd* su \mathbf{R}_1

$$F^{(2,\dots,n)}(x_1) = F(x_1, +\infty, \dots, +\infty)$$

Tutte le *fdd* così ottenute a partire da una *fdd* multivariata F si chiamano **funzioni di distribuzione marginali**, e l'operazione con la quale si ricavano si chiama **marginalizzazione**. Sulla base del Teorema 2.27 queste osservazioni si estendono poi alle misure probabilità: se \mathbf{P} è la probabilità su $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ associata a $F(\mathbf{x})$, potremo definire delle probabilità marginali su spazi $(\mathbf{R}^k, \mathcal{B}(\mathbf{R}^k))$ con $k < n$ associandole

alle *fdd* marginali. Ad esempio possiamo definire una $\mathbf{P}^{(1)}$ su $\mathbf{R}^{n-1} = \mathbf{R}_2 \times \dots \times \mathbf{R}_n$, o una $\mathbf{P}^{(2,\dots,n)}$ su \mathbf{R}_1 associandole rispettivamente alle *fdd* marginali $F^{(1)}$ e $F^{(2,\dots,n)}$ ricavate prima. In questo modo la relazione fra \mathbf{P} e le sue marginali è del tipo

$$\begin{aligned} \mathbf{P}^{(2,\dots,n)}(-\infty, x_1] &= F^{(2,\dots,n)}(x_1) \\ &= F(x_1, +\infty, \dots, +\infty) = \mathbf{P}((-\infty, x_1] \times \mathbf{R}_2 \times \dots \times \mathbf{R}_n) \end{aligned}$$

Se la *fdd* multivariata F è *ac* anche le sue marginali lo saranno e da (2.21) si vede che ad esempio le *ddp* di $F^{(1)}$, $F^{(1,2)}$ e $F^{(2,\dots,n)}$ sono rispettivamente

$$f^{(1)}(x_2, \dots, x_n) = \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n) dx_1 \quad (2.27)$$

$$f^{(1,2)}(x_3, \dots, x_n) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_1, x_2, x_3, \dots, x_n) dx_1 dx_2 \quad (2.28)$$

$$f^{(2,\dots,n)}(x_1) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n) dx_2 \dots dx_n \quad (2.29)$$

Nel caso delle *ddp*, quindi, le marginalizzazioni sono effettuate mediante integrazioni sulle variabili che si vogliono eliminare.

Partendo da una *fdd* (o da una *ddp*) multivariata su \mathbf{R}^n è dunque possibile, in maniera semplice e *univoca*, ricavare tutta una gerarchia di *fdd* marginali su spazi con un numero di dimensioni sempre più piccolo, fino ad arrivare ad ottenere n *fdd* univariate. Ci si domanda allora se non sia possibile percorrere lo stesso cammino in senso inverso: dato un certo numero di *fdd* (univariate o multivariate) è possibile ricavare *in maniera unica* una *fdd* multivariata della quale le iniziali *fdd* siano le marginali? La risposta a questa domanda è in generale *negativa per quel che riguarda l'unicità*, e merita qualche riflessione. Partiamo dal caso in cui siano assegnate n *fdd* univariate $F_1(x), \dots, F_n(x)$: è facile convincersi del fatto che la

$$F(\mathbf{x}) = F_1(x_1) \cdot \dots \cdot F_n(x_n)$$

che come già osservato alla fine della Sezione 2.3.1 è una *fdd* multivariata, ha come marginali proprio le $F_k(x)$. Infatti dalle regole di marginalizzazione appena esposte si avrebbe ad esempio

$$F^{(2,\dots,n)}(x_1) = F_1(x_1)F_2(+\infty) \dots F_n(+\infty) = F_1(x_1)$$

Naturalmente se le *fdd* assegnate sono *ac* con *ddp* $f_k(x)$, anche la F multivariata ottenuta come prodotto sarà *ac* con *ddp*

$$f(\mathbf{x}) = f_1(x_1) \cdot \dots \cdot f_n(x_n)$$

e le *ddp* marginali univariate sono proprio le $f_k(x)$. Si verifica però subito con dei semplici controesempi che la *fdd* costruita come *prodotto* di univariate assegnate *non* è *l'unica fdd* multivariata che ammette le F_k come marginali.

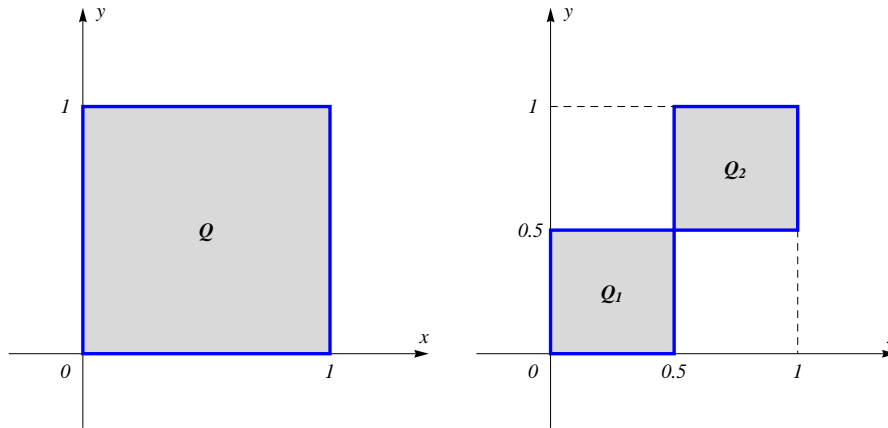


Figura 2.12: Densità uniformi su domini differenti.

Esempio 2.30. Distribuzioni multivariate e marginali: *Supponiamo di considerare le seguenti due ddp bivariate e uniformi sui domini⁸ rappresentati in Figura 2.12*

$$f(x, y) = \begin{cases} 1, & \text{se } (x, y) \in Q, \\ 0, & \text{se } (x, y) \notin Q, \end{cases} \quad Q = [0, 1] \times [0, 1]$$

$$g(x, y) = \begin{cases} 2, & \text{se } (x, y) \in Q_1 \cup Q_2, \\ 0, & \text{se } (x, y) \notin Q_1 \cup Q_2, \end{cases} \quad \begin{aligned} Q_1 &= [0.0, 0.5] \times [0.0, 0.5] \\ Q_2 &= [0.5, 1.0] \times [0.5, 1.0] \end{aligned}$$

È facile mostrare ora con delle integrazioni elementari che le marginali univariate di f sono le due uniformi $\mathfrak{U}(0, 1)$ rispettivamente con ddp

$$f_1(x) = f^{(2)}(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \begin{cases} 1, & \text{se } x \in [0, 1], \\ 0, & \text{se } x \notin [0, 1], \end{cases}$$

$$f_2(y) = f^{(1)}(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \begin{cases} 1, & \text{se } y \in [0, 1], \\ 0, & \text{se } y \notin [0, 1], \end{cases}$$

e che esse coincidono anche esattamente con le corrispondenti marginali g_1 e g_2 di g calcolate in maniera analoga; si ha cioè:

$$f_1(x) = g_1(x) \quad f_2(y) = g_2(y)$$

Ciò mostra esplicitamente che ddp multivariate differenti possono avere le medesime ddp marginali, e quindi che dalla sola conoscenza delle marginali non si può risalire

⁸Siccome le leggi con ddp f e g sono ovviamente ac, le frontiere dei domini scelti saranno sicuramente di misura nulla e quindi potremo sempre considerare chiusi tali domini senza rischio di errori

in maniera unica alla ddp multivariata che le genera. Peraltro, data l'evidente differenza fra le due bivariante f e g , queste osservazioni mostrano anche che

$$f(x, y) = f_1(x)f_2(y) \quad g(x, y) \neq g_1(x)g_2(y) \quad (2.30)$$

cioè che il prodotto delle due marginali univariate è sì una possibile ddp bivariata con le date marginali, ma non l'unica.

Esempio 2.31. Marginali di leggi Gaussiane multivariate: *Un'integrazione laboriosa ma elementare del tipo (2.29) – che ometteremo per brevità – permetterebbe di calcolare esplicitamente le marginali univariate della ddp Gaussiana multivariata $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ (2.22): si scoprirebbe così che tali marginali sono ancora tutte normali $\mathfrak{N}(b_k, a_k^2)$ di tipo (2.15), con $a_k^2 = a_{kk}$ e ddp*

$$f_k(x_k) = \frac{1}{a_k \sqrt{2\pi}} e^{-(x_k - b_k)^2 / 2a_k^2} \quad (2.31)$$

Ci vuol poco però ad accorgersi del fatto che in generale il prodotto di queste ddp – pur essendo ancora una legittima ddp normale multivariata – non coincide con l'originale ddp multivariata (2.22) a meno che \mathbb{A} non sia una matrice diagonale: infatti nel semplice prodotto non si ritroverebbero tutti i termini non diagonali della forma quadratica che forma l'esponente della (2.22). In particolare, dalla discussione sviluppata nell'Esempio 2.29, si ha che la matrice \mathbb{A} di una normale **bivariata** risulta diagonale se e solo se $r = 0$

2.3.4 Copule

Le osservazioni espote nella sezione precedente ci inducono a porci due problemi fra loro strettamente associati:

1. che relazione c'è in generale fra una *fdd* n -variata $F(x_1, \dots, x_n)$ e le sue n marginali univariate $F_k(x)$?
2. assegnate n *fdd* univariate $F_k(x)$, esistono delle *fdd* n -variate $F(x_1, \dots, x_n)$ che abbiano le F_k per marginali? E in caso affermativo, come e in quante maniere possiamo ricostruirle?

Inizieremo con il caso bivariato di una *fdd* $H(x, y)$ con le sue marginali $F(x)$ e $G(y)$

Definizione 2.32. *Diremo che una funzione $C(u, v)$ definita su $[0, 1] \times [0, 1]$ a valori in $[0, 1]$ è una **copula** se essa soddisfa le seguenti proprietà:*

1. $C(u, 0) = C(0, v) = 0, \quad \forall u, v \in [0, 1]$
2. $C(u, 1) = u, \quad C(1, v) = v, \quad \forall u, v \in [0, 1]$
3. $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0, \quad \forall u_1 \leq u_2, v_1 \leq v_2$

Una copula può essere quindi vista come la restrizione al quadrato $[0, 1] \times [0, 1]$ di una *fdd* con marginali uniformi $\mathfrak{U}(0, 1)$. Esempi classici sono

$$\begin{aligned} C_M(u, v) &= u \wedge v = \min\{u, v\} \\ C_m(u, v) &= (u + v - 1)^+ = \max\{u + v - 1, 0\} \\ C_0(u, v) &= uv \\ C_\theta(u, v) &= (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} \quad \theta > 0 \quad (\text{Clayton}) \end{aligned}$$

mentre altri casi possono essere ricavati con varie combinazioni di questi tipi più elementari. Si dimostra inoltre che tutte le copule $C(u, v)$ soddisfano la relazione

$$C_m(u, v) \leq C(u, v) \leq C_M(u, v)$$

Teorema 2.33. Teorema di Sklar:

- Se $H(x, y)$ è una *fdd* bivariata e $F(x) = H(x, +\infty)$, $G(y) = H(+\infty, y)$ sono le sue marginali, esiste sempre una copula $C(u, v)$ tale che

$$H(x, y) = C[F(x), G(y)] \tag{2.32}$$

tale copula è anche unica se F e G sono continue; in caso contrario C è unica solo sui punti (u, v) che sono possibili valori di $(F(x), G(y))$;

- se $F(x)$ e $G(y)$ sono due *fdd* e $C(u, v)$ è una copula, allora la $H(x, y)$ definita da (2.32) è sempre una *fdd* bivariata che ammette F e G come sue marginali.

Dimostrazione: Omessa⁹ □

In sostanza il Teorema di Sklar afferma che ogni *fdd* bivariata si ricava dall'applicazione di una opportuna copula alle sue marginali, e viceversa che il risultato dell'applicazione di un'arbitraria copula a un'arbitraria coppia di *fdd* univariate è sempre una *fdd* bivariata. In particolare si mostra che la costruzione di una bivariata tramite *prodotto* di due *fdd* univariate non è altro che l'applicazione della copula C_0 , e quindi è solo una delle molte possibilità a disposizione.

Esempio 2.34. Copule di Cauchy: Date due distribuzioni di Cauchy $\mathfrak{C}(0, 1)$ rispettivamente con *fdd* *ac* e *ddp*

$$\begin{aligned} F(x) &= \frac{1}{2} + \frac{1}{\pi} \arctan x & f(x) &= \frac{1}{\pi} \frac{1}{1+x^2} \\ G(y) &= \frac{1}{2} + \frac{1}{\pi} \arctan y & g(y) &= \frac{1}{\pi} \frac{1}{1+y^2} \end{aligned}$$

l'applicazione della copula prodotto C_0 produce la *fdd* bivariata

$$H_0(x, y) = \left(\frac{1}{2} + \frac{1}{\pi} \arctan x \right) \left(\frac{1}{2} + \frac{1}{\pi} \arctan y \right)$$

⁹R.B. Nelsen, AN INTRODUCTION TO COPULAS, Springer (New York, 1999)

che è ovviamente ancora ac con ddp

$$h_0(x, y) = \frac{1}{\pi} \frac{1}{1+x^2} \cdot \frac{1}{\pi} \frac{1}{1+y^2} = f(x)g(y)$$

Se invece adottassimo la più semplice copula di Clayton con $\theta = 1$

$$C_1(u, v) = (u^{-1} + v^{-1} - 1)^{-1} = \frac{uv}{u + v - uv}$$

otterremmo una fdd $H_1(x, y)$ diversa (che non riportiamo per brevità) ma ancora ac con una ddp

$$h_1(x, y) = \frac{32\pi^2(\pi + 2 \arctan x)(\pi + 2 \arctan y)}{(1+x^2)(1+y^2)[2 \arctan x(\pi - 2 \arctan y) + \pi(3\pi + 2 \arctan y)]^3}$$

che ha ancora per marginali le due distribuzioni di Cauchy iniziali, pur non essendone più il semplice prodotto. Si noti comunque che in generale la fdd risultante dall'applicazione di una copula può non essere ac anche se le due fdd univariate lo sono.

Si può dimostrare che il Teorema di Sklar 2.33 ammette una generalizzazione secondo la quale tutte le fdd multivariate $H(\mathbf{x}) = H(x_1, \dots, x_n)$ possono essere ricavate dall'applicazione di opportune copule multivariate $C(\mathbf{u}) = C(u_1, \dots, u_n)$ a fdd univariate $F_1(x_1), \dots, F_n(x_n)$. In particolare si dimostra che anche in questo caso n fdd univariate arbitrarie possono sempre – e in vario modo – essere combinate in fdd multivariate.

Diverso è invece il problema di determinare la connessione con fdd marginali multivariate. Non entreremo nei dettagli, ma ci limiteremo ad osservare che, data ad esempio una fdd trivariata $F(x, y, z)$, potremo sempre calcolare le sue tre marginali bivariate

$$F^{(1)}(y, z) = F(+\infty, y, z), \quad F^{(2)}(x, z) = F(x, +\infty, z), \quad F^{(3)}(x, y) = F(x, y, +\infty)$$

e che è anche possibile – ancorché non banale – pensare di determinare dei metodi generali per ricostruire F a partire da tali marginali con delle opportune copule. Il problema inverso, però, a differenza dal caso del Teorema di Sklar, non ammette sempre una soluzione, nel senso che non possiamo pensare di trovare sempre una fdd $F(x, y, z)$ che abbia come marginali bivariate tre fdd $F_1(y, z)$, $F_2(x, z)$ e $F_3(x, y)$ assegnate arbitrariamente. Diversamente dal caso delle marginali univariate, si pone infatti ora un problema di compatibilità fra le fdd assegnate. Ad esempio è evidente che – per poter essere tutte marginali della medesima trivariata $F(x, y, z)$ – esse dovranno per lo meno concordare sulle marginali univariate da esse deducibili, cioè dovrà risultare

$$F_1(+\infty, z) = F_2(+\infty, z), \quad F_1(y, +\infty) = F_3(+\infty, y), \quad F_2(x, +\infty) = F_3(x, +\infty)$$

e questa condizione (peraltro solo necessaria) non è garantita se F_1 , F_2 e F_3 sono del tutto arbitrarie. L'assegnazione di *fdd* marginali multivariate dovrà quindi rispettare dei criteri di *consistenza* che peraltro non si limitano solo a quelli indicati in questa sezione: una breve discussione su questo punto è riportata nell'Appendice A. Le precedenti osservazioni sulla compatibilità di famiglie di *fdd* torneranno inoltre utili nella discussione della sezione successiva

2.4 Probabilità su \mathbf{R}^∞ e \mathbf{R}^T

L'estensione delle precedenti osservazioni al caso dello spazio $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ delle successioni di numeri reali è meno immediata e richiede qualche cautela. C'è infatti una importante differenza fra gli spazi $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ e quelli con un numero finito di dimensioni $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$: mentre su questi ultimi possiamo assegnare la probabilità tramite opportune *fdd* con n variabili, è evidente che non è possibile estendere tale costruzione anche al caso di $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ in quanto non avrebbe alcun senso assegnare *fdd* con un numero *infinito* di variabili. Il problema di assegnare una probabilità su $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ dovrà quindi essere risolto con strumenti differenti.

A questo scopo osserveremo innanzitutto che, se su $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ fosse assegnata una probabilità \mathbf{P} , sarebbe anche possibile ricavare induttivamente un'intera famiglia di probabilità sugli spazi di dimensione finita ottenuti selezionando un numero arbitrario, ma finito, di componenti. Infatti n arbitrarie componenti delle successioni di $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ costituiranno un elemento di uno spazio $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$: per definire una probabilità su tale $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ partendo dalla \mathbf{P} assegnata basterà allora considerare un Boreliano $B \in \mathcal{B}(\mathbf{R}^n)$ come base di un *cilindro* in $\mathcal{B}(\mathbf{R}^\infty)$, e poi attribuire alla base B la probabilità che \mathbf{P} assegna a tutto il cilindro. Si ottiene in questo modo un'intera famiglia di spazi di probabilità di dimensioni finite che gode di una semplice, ma importante proprietà di **consistenza** (si vedano anche le considerazioni riportate nell'Appendice A): le *fdd* di spazi $(\mathbf{R}^k, \mathcal{B}(\mathbf{R}^k))$ che siano sottospazi di un $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ con $k \leq n$ si ottengono per marginalizzazione delle componenti in eccesso usando le solite relazioni (2.25) e (2.26).

Questo fatto suggerisce il tentativo di definire una probabilità su $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ seguendo un *percorso inverso*: si assegnano prima delle probabilità su tutti i sottospazi di dimensione finita $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$, e poi le si estende a tutto $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$. Sulla base di osservazioni analoghe a quelle svolte alla fine della sezione precedente sappiamo però che tali probabilità non possono essere assegnate in maniera completamente arbitraria: in effetti per risultare estendibili esse devono costituire una **famiglia consistente di probabilità**, nel senso della *consistenza* discussa prima. Il teorema seguente garantisce allora che, una volta assegnata una famiglia consistente di probabilità sugli spazi di dimensioni finite, essa definisce per estensione e in maniera univoca una probabilità su tutto $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$

Teorema 2.35. Teorema di Kolmogorov su \mathbf{R}^∞ : *Data una famiglia consistente di spazi di probabilità $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n), \mathbf{P}_n)$ esiste sempre un'unica probabilità \mathbf{P} su $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ che risulta essere un prolungamento della famiglia assegnata.*

Dimostrazione: Omessa¹⁰ □

Esempio 2.36. Successioni di Bernoulli: *La maniera più semplice di soddisfare i requisiti del Teorema 2.35 consiste nel prendere una successione di fdd univariate $G_k(x)$, $k \in \mathbf{N}$ e nel definire poi la successione di fdd multivariate*

$$F_n(x_1, \dots, x_n) = G_1(x_1) \cdot \dots \cdot G_n(x_n) \quad n \in \mathbf{N}$$

Sulla base del Teorema 2.27 potremo allora definire sugli spazi $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ le probabilità \mathbf{P}_n associate alle F_n , e potremo verificare facilmente che tali \mathbf{P}_n costituiscono una famiglia consistente di probabilità: esiste pertanto, in base al Teorema di Kolmogorov 2.35, un'unica misura di probabilità \mathbf{P} sullo spazio delle successioni numeriche $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ tale che

$$\mathbf{P}\{x \in \mathbf{R}^\infty : (x_1, \dots, x_n) \in B\} = \mathbf{P}_n\{B\} \quad \forall B \in \mathcal{B}(\mathbf{R}^n)$$

e in particolare

$$\mathbf{P}\{x \in \mathbf{R}^\infty : x_1 \leq a_1, \dots, x_n \leq a_n\} = F_n(a_1, \dots, a_n) = G_1(a_1) \cdot \dots \cdot G_n(a_n)$$

Ad esempio se le $G_n(x)$ sono tutte distribuzioni di Bernoulli $\mathfrak{B}(1; p)$ identiche fra loro in modo che

$$G_n(x) = \begin{cases} 0, & \text{se } x < 0; \\ 1 - p, & \text{se } 0 \leq x < 1; \\ 1, & \text{se } 1 \leq x, \end{cases}$$

potremo definire una \mathbf{P} sullo spazio delle successioni i cui elementi x_j assumono solo valori $a_j = 0, 1$, in modo tale che comunque scelto $k = 0, 1, \dots, n$

$$\mathbf{P}\left\{x \in \mathbf{R}^\infty : x_1 = a_1, \dots, x_n = a_n, \text{ con } \sum_{j=1}^n a_j = k\right\} = p^k q^{n-k}$$

*Tale \mathbf{P} è l'estensione allo spazio (non numerabile) delle successioni infinite di estrazioni (**successioni di Bernoulli**), delle distribuzioni binomiali definite con le (2.6) e (2.7) sugli spazi finiti delle n -ple di estrazioni come esposto nella Sezione 2.1.2: questa estensione è essenziale per poter introdurre limiti per un numero infinito di estrazioni, come vedremo nella Appendice F.*

¹⁰A.N. Shiryaev, **PROBABILITY**, Springer (New York, 1996)

Consideriamo infine lo spazio $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$ e supponiamo innanzitutto, come nel caso di $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$, che su di esso sia stata assegnata una probabilità \mathbf{P} . Anche in questo caso la conoscenza di \mathbf{P} ci permette di ricavare tutta una famiglia di probabilità sui sottospazi di dimensione finita usando degli opportuni cilindri secondo lo schema dell'Esempio 1.7. Si verifica che la famiglia così costruita è consistente in un senso analogo a quello del caso di \mathbf{R}^∞ . Ci si chiede allora se viceversa, partendo da una famiglia consistente di spazi di probabilità di dimensione finita, non sia possibile definire su tutto $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$ una \mathbf{P} che risulti esserne il prolungamento: ciò consentirebbe ancora una volta di garantire l'esistenza di una probabilità sullo spazio $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$ con *infinite dimensioni* (non numerabili), mediante l'assegnazione di una famiglia infinita di probabilità su spazi di *dimensione finita*. La risposta affermativa a questa domanda è contenuta nel seguente teorema.

Teorema 2.37. Teorema di Kolmogorov su \mathbf{R}^T : *Detto $S = \{t_1, \dots, t_n\}$ un arbitrario sottoinsieme finito di T , e data una famiglia consistente di spazi di probabilità $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n), \mathbf{P}_S)$ esiste sempre un'unica probabilità \mathbf{P} sullo spazio $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$ che risulta essere un prolungamento della famiglia assegnata.*

Dimostrazione: Omessa¹¹ □

Esempio 2.38. Misura di Wiener: *Consideriamo il caso in cui $T = [0, +\infty)$, sicché \mathbf{R}^T risulterà essere l'insieme delle funzioni $(x_t)_{t \geq 0}$, e prendiamo la seguente famiglia di ddp normali $\mathfrak{N}(0, t)$*

$$\varphi_t(x) = \frac{e^{-x^2/2t}}{\sqrt{2\pi t}} \quad t > 0$$

Dato allora $S = \{t_1, \dots, t_n\}$ con $0 < t_1 < \dots < t_n$, e un Boreliano di \mathbf{R}^n , ad esempio $B = A_1 \times \dots \times A_n \in \mathcal{B}(\mathbf{R}^n)$, definiremo \mathbf{P}_S assegnando a B la probabilità

$$\mathbf{P}_S\{B\} = \int_{A_n} \dots \int_{A_1} \varphi_{t_n-t_{n-1}}(x_n - x_{n-1}) \dots \varphi_{t_2-t_1}(x_2 - x_1) \varphi_{t_1}(x_1) dx_1 \dots dx_n$$

Si verifica poi che la famiglia $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n), \mathbf{P}_S)$ così costruita al variare di S è consistente, e quindi il Teorema 2.37 ci garantisce l'esistenza di una probabilità \mathbf{P} definita su tutto $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$ che risulta essere il prolungamento di quella assegnata. La misura di probabilità così ottenuta prende il nome di **misura di Wiener** e gioca un ruolo molto importante nella teoria dei processi stocastici. Il suo significato può essere esposto intuitivamente nel modo seguente: se interpretiamo $(x_t)_{t \geq 0}$ come la traiettoria (aleatoria) di un punto materiale, il cilindro con base $B = A_1 \times \dots \times A_n$ rappresenterà il fascio di traiettorie che partono da $x = 0$, e che agli istanti $t_1 < \dots < t_n$ passano attraverso le finestre A_1, \dots, A_n . Inoltre le $\varphi_{t_k-t_{k-1}}(x_k - x_{k-1}) dx_k$ rappresentano le probabilità (gaussiane) che la particella, partendo da x_{k-1} all'istante t_{k-1} ,

¹¹A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

si trovi in $[x_k, x_k + dx_k]$ dopo un tempo $t_k - t_{k-1}$, mentre il prodotto di tali ddp che compare nella definizione esprime l'ipotesi di indipendenza degli spostamenti negli intervalli di tempo $[0, t_1], [t_1, t_2], \dots, [t_{n-1}, t_n]$. Infine l'integrale multiplo introdotto nella definizione permette di calcolare la probabilità assegnata al fascio di traiettorie che, agli istanti $t_1 < \dots < t_n$, passano attraverso le finestre A_1, \dots, A_n .

Capitolo 3

Variabili aleatorie

3.1 Variabili aleatorie

3.1.1 Misurabilità

Definizione 3.1. *Dati gli spazi probabilizzabili (Ω, \mathcal{F}) e $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, una funzione $X : \Omega \rightarrow \mathbf{R}$ si dice \mathcal{F} -**misurabile** – o, se non ci sono ambiguità, semplicemente **misurabile** – quando (vedi la Figura 3.1 per una descrizione grafica)*

$$X^{-1}(B) = \{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbf{R}).$$

e, per richiamare le σ -algebre rispetto alle quali X è misurabile, scriveremo anche

$$X : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$$

In probabilità una X misurabile prende il nome di **variabile aleatoria** (*v-a*), e se (Ω, \mathcal{F}) coincide con $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ viene chiamata **funzione di Borel**.

Si noti subito che la definizione di *v-a* è del tutto slegata dalle eventuali misure di probabilità definibili sugli spazi probabilizzabili citati: che X sia una *v-a* dipende solo dalla sua misurabilità. Viceversa è essenziale dichiarare esplicitamente quali sono le σ -algebre coinvolte: senza di esse, infatti, la definizione non avrebbe senso.

Esempio 3.2. Indicatori e *v-a* semplici e degeneri: *Il caso più elementare di *v-a* è rappresentato dai cosiddetti indicatori $I_A(\omega)$ di un evento $A \in \mathcal{F}$. Essi sono definiti come*

$$I_A(\omega) = \begin{cases} 1, & \text{se } \omega \in A, \\ 0, & \text{se } \omega \notin A, \end{cases}$$

e risultano ovviamente funzioni misurabili rispetto ad \mathcal{F} dato che $A \in \mathcal{F}$. Gli indicatori sono *v-a* che godono di un certo numero di proprietà; in particolare, ad esempio, si verifica facilmente che $\forall \omega \in \Omega$

$$\begin{aligned} I_{\emptyset}(\omega) &= 0, & I_{\Omega}(\omega) &= 1, & I_A(\omega) + I_{\bar{A}}(\omega) &= 1, \\ I_{AB}(\omega) &= I_A(\omega) I_B(\omega), & I_{A \cup B}(\omega) &= I_A(\omega) + I_B(\omega) - I_A(\omega) I_B(\omega) \end{aligned}$$

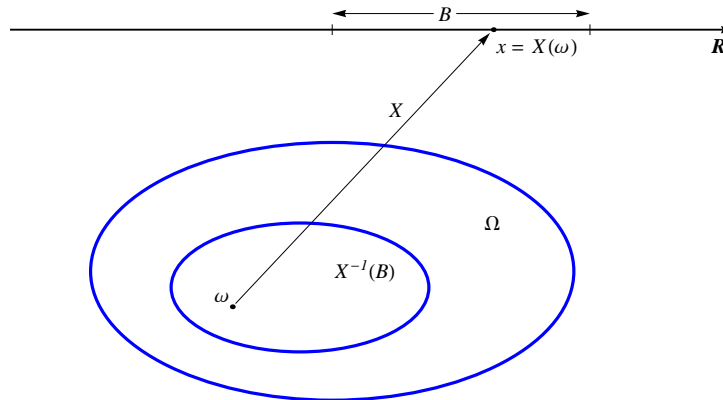


Figura 3.1: Illustrazione grafica della definizione di variabile aleatoria.

Oltre gli indicatori giocano poi un ruolo importante le cosiddette **v-a semplici**, cioè le v-a che prendono solo un numero finito di valori x_k , $k = 1, \dots, n$ e sono quindi della forma

$$X(\omega) = \sum_{k=1}^n x_k I_{D_k}(\omega)$$

dove le $D_k \in \mathcal{F}$ costituiscono una decomposizione finita di Ω : praticamente la v-a semplice X assumerà il valore x_k su tutte le ω di D_k con $k = 1, \dots, n$. Infine va ricordato che non è escluso il caso limite di una **v-a degenera (o costante)** caratterizzata dal fatto di assumere un solo valore b su tutto Ω .

In definitiva qualunque funzione misurabile $X : \Omega \rightarrow \mathbf{R}$ è una v-a e potrà essere innanzitutto considerata come una semplice maniera di associare valori numerici in \mathbf{R} agli elementi ω di Ω . Qualunque regola che ad esempio associ una vincita in denaro alle facce di un dado (la misurabilità in questi casi semplici è ovviamente garantita) potrà dunque essere considerata una v-a. La motivazione per richiedere anche che tali funzioni numeriche siano misurabili sarà resa evidente nella sezione successiva.

3.1.2 Leggi e distribuzioni

Il ruolo giocato dalla misurabilità risulta chiaro quando si assegna una probabilità \mathbf{P} su (Ω, \mathcal{F}) : essa consente infatti di indurre anche sullo spazio di arrivo $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ una nuova probabilità associata alla v-a X , proiezione resa possibile, come si vede nella seguente definizione, proprio dal fatto che $\{X \in B\} \in \mathcal{F}$, $\forall B \in \mathcal{B}(\mathbf{R})$

Definizione 3.3. Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ ed una v-a $X : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$, chiameremo **legge o distribuzione di X** la misura di probabilità \mathbf{P}_X indotta da X su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ dalla relazione

$$\mathbf{P}_X\{B\} = \mathbf{P}\{X \in B\}, \quad B \in \mathcal{B}(\mathbf{R})$$

mentre la fdd $F_X(x)$ di \mathbf{P}_X , cioè la funzione

$$F_X(x) = \mathbf{P}_X\{(-\infty, x]\} = \mathbf{P}\{X \leq x\} = \mathbf{P}\{\omega \in \Omega : X(\omega) \leq x\}, \quad x \in \mathbf{R}$$

prenderà il nome di **funzione di distribuzione (fdd) di X** , ed è la probabilità che la v -a X risulti minore o uguale al numero x . Infine adotteremo la notazione

$$X \sim \mathbf{P}_X$$

per indicare che \mathbf{P}_X è la distribuzione della v -a X

È piuttosto evidente che v -a differenti (nel senso della Definizione 3.1) X e Y , definite su un unico spazio $(\Omega, \mathcal{F}, \mathbf{P})$, saranno in genere dotate di leggi differenti \mathbf{P}_X e \mathbf{P}_Y , ma va anche detto subito che non è affatto escluso il caso in cui esse siano invece dotate proprio della medesima legge e siano quindi *identicamente distribuite*. Ad esempio sullo spazio dei campioni delle facce di un dado equo possiamo definire le seguenti due diverse v -a: X che assume valore 1 in corrispondenza delle prime quattro facce (e 0 sulle altre due) ed Y che assume valore 0 sulle prime due facce (e 1 sulle altre quattro). È facile controllare che queste due v -a, sebbene diverse come funzioni di ω , assumono gli stessi due valori (0 e 1) con le medesime probabilità ($\frac{1}{3}$ e $\frac{2}{3}$), cioè hanno la medesima distribuzione. Viceversa, una stessa v -a X può essere dotata di molte leggi differenti secondo le probabilità \mathbf{P} definite su (Ω, \mathcal{F}) : si ricordi a questo proposito che i condizionamenti modificano la probabilità su (Ω, \mathcal{F}) . Tutto ciò mostra che la legge – pur essendo quello che in definitiva è praticamente accessibile alle nostre osservazioni – non determina affatto la v -a, ma ne fissa solo, per così dire, le caratteristiche statistiche. Diventa importante a questo punto stabilire cosa intendiamo esattamente per uguaglianza di v -a.

Definizione 3.4. Due v -a X ed Y definite su un medesimo $(\Omega, \mathcal{F}, \mathbf{P})$ sono

- **indistinguibili**, e scriveremo semplicemente $X = Y$, quando

$$X(\omega) = Y(\omega) \quad \forall \omega \in \Omega$$

- **uguali \mathbf{P} -qo**, e scriveremo $X \stackrel{qo}{=} Y$, se

$$\mathbf{P}\{X \neq Y\} = \mathbf{P}\{\omega \in \Omega : X(\omega) \neq Y(\omega)\} = 0$$

- **identicamente distribuite (id)**, e scriveremo $X \stackrel{d}{=} Y$, se le loro leggi coincidono cioè se $\mathbf{P}_X = \mathbf{P}_Y$ e quindi

$$F_X(x) = F_Y(x) \quad \forall x \in \mathbf{R}$$

Si vede facilmente che v -a indistinguibili sono anche uguali \mathbf{P} -qo, e che v -a uguali \mathbf{P} -qo sono anche *id*: il viceversa invece in generale non è vero come è stato mostrato poc'anzi con alcuni semplici controesempi. Procederemo ora ad una prima classificazione delle v -a sulla base delle loro leggi

v-a discrete

Si dicono *v-a discrete* quelle del tipo

$$X(\omega) = \sum_k x_k I_{D_k}(\omega)$$

dove k varia in un insieme finito o numerabile di numeri interi, mentre gli eventi $D_k = \{X = x_k\}$ costituiscono una decomposizione di Ω . È facile vedere che la distribuzione \mathbf{P}_X di queste *v-a* è una misura di probabilità discreta concentrata sull'insieme – al più numerabile – di numeri $x_k \in \mathbf{R}$, per cui la sua *fdd* è *discreta* nel senso discusso nella Sezione 2.2.1. In questo caso, posto

$$p_k = \mathbf{P}_X\{x_k\} = \mathbf{P}\{X = x_k\} = F_X(x_k) - F_X(x_k^-)$$

risulta

$$\mathbf{P}_X\{B\} = \sum_{k: x_k \in B} p_k, \quad B \in \mathcal{B}(\mathbf{R})$$

Ovviamente le *v-a semplici* introdotte in precedenza sono *v-a discrete* che assumono solo un numero finito di valori, e in particolare lo sarà anche una *v-a P-go degenera* uguale a un solo valore b con probabilità 1, cioè con $\mathbf{P}_X\{b\} = 1$. Le *v-a discrete* saranno denominate in base alle loro leggi, e in questo modo avremo ad esempio *v-a degeneri* δ_b , di Bernoulli $\mathfrak{B}(1; p)$, binomiali $\mathfrak{B}(n; p)$ e di Poisson $\mathfrak{P}(\alpha)$ le cui distribuzioni sono state già introdotte nella Sezione 2.2.2.

v-a continue e ac

Una *v-a* si dice invece *continua* se la *fdd* $F_X(x)$ è una funzione continua di x , e in particolare si chiama *assolutamente continua (ac)* se la sua F_X è *ac*, cioè se esiste una *densità di probabilità (ddp)* $f_X(x)$ non negativa, normalizzata e tale che

$$F_X(x) = \int_{-\infty}^x f_X(y) dy$$

Ricordiamo qui che non tutte le *v-a continue* sono anche *ac*: esistono infatti *v-a con fdd continue* ma singolari. Nel seguito, però, non avremo quasi mai a che fare con questi casi, anzi va segnalato che nella letteratura di carattere applicativo le *v-a ac* sono spesso indicate semplicemente con il nome di *v-a continue*. Ovviamente tutte le osservazioni e gli esempi relativi alle *ddp* riportati nella Sezione 2.2.3 possono essere estesi in modo naturale alle *v-a ac* che saranno denominate in base alle loro leggi: avremo così *v-a uniformi* $\mathfrak{U}(a, b)$, normali $\mathfrak{N}(b, a^2)$, esponenziali $\mathfrak{E}(a)$, di Laplace $\mathfrak{L}(a)$, di Cauchy $\mathfrak{C}(b, a)$ e via dicendo. Sarà infine utile ricordare qui – ma questo punto sarà ripreso più tardi con maggiore dettaglio – come si calcola $\mathbf{P}_X(B)$ a partire dalla *ddp* $f_X(x)$ quando questa esiste: per $B = (-\infty, x]$ dalle definizioni date finora avremo

$$\mathbf{P}_X\{(-\infty, x]\} = \mathbf{P}\{X \leq x\} = F_X(x) = \int_{-\infty}^x f_X(t) dt = \int_{(-\infty, x]} f_X(t) dt$$

mentre per $B = (a, b]$ si ha invece

$$\mathbf{P}_X\{(a, b]\} = \mathbf{P}\{a < X \leq b\} = F_X(b) - F_X(a) = \int_a^b f_X(t) dt = \int_{(a,b]} f_X(t) dt$$

Con una generalizzazione piuttosto naturale potremo allora pensare che nel caso generale di un $B \in \mathcal{B}(\mathbf{R})$ generico avremo per analogia

$$\mathbf{P}_X\{B\} = \int_B dF_X(x) = \int_B f_X(x) dx$$

il cui significato preciso sarà comunque precisato nella discussione del Corollario 3.23.

3.1.3 Costruzione di v -a

Abbiamo visto che dato una v -a X su uno spazio $(\Omega, \mathcal{F}, \mathbf{P})$ essa è sempre dotata di una distribuzione assegnata mediante una *fdd* $F_X(x)$: sarà bene osservare allora prima di tutto che è vero anche il viceversa. Se infatti è assegnata solo una *fdd* $F(x)$ (ma non uno spazio di probabilità o una v -a) potremo prendere $\Omega = \mathbf{R}$ come spazio dei campioni, e definire una v -a X come l'*applicazione identica* $X : \mathbf{R} \rightarrow \mathbf{R}$ che ad ogni $\omega = x$ associa lo stesso $x \in \mathbf{R}$. Se allora su $(\Omega, \mathcal{F}) = (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ si definisce una \mathbf{P} tramite la *fdd* assegnata $F(x)$, è facile convincersi del fatto che anche la *fdd* di X sarà proprio $F(x)$. Queste osservazioni conducono alla seguente definizione

Definizione 3.5. *Data una fdd $F(x)$, e detta \mathbf{P} la corrispondente probabilità su $(\Omega, \mathcal{F}) = (\mathbf{R}, \mathcal{B}(\mathbf{R}))$, si chiama **v -a canonica** X l'applicazione identica dallo spazio $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mathbf{P})$ in $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$: la sua *fdd* coinciderà con la $F(x)$ assegnata*

Esamineremo ora alcune importanti procedure mediante le quali è possibile costruire nuove v -a a partire da v -a date: in particolare ci soffermeremo sulle v -a definite come *funzioni di v -a*, e su quelle ottenute come *limiti di successioni di v -a*

Funzioni di v -a

Proposizione 3.6. *Data una v -a $X(\omega)$ ($\omega \in \Omega$) e una funzione di Borel $\phi(x)$ ($x \in \mathbf{R}$) definita su tutti i valori assunti da X , la funzione composta*

$$\phi[X(\omega)] = Y(\omega)$$

è ancora misurabile, e quindi è una v -a

Dimostrazione: Omessa¹ □

La Proposizione precedente consente di dire che, se X è una v -a, anche $Y = \phi(X)$ è una v -a purché ϕ sia misurabile. Così ad esempio X^n , $|X|$, $\cos X$, e^X sono nuove v -a.

¹A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

Viceversa va detto subito che, date due generiche v -a X e Y , non è detto che si possa trovare una funzione di Borel ϕ tale che Y coincida con $\phi(X)$: si potrebbero mostrare a questo proposito dei semplici esempi che qui trascureremo per brevità. Questa osservazione permette però di mettere ancora una volta in evidenza l'importanza del concetto di misurabilità. A questo scopo introdurremo un nuovo concetto: se $X : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ è una v -a si può provare che anche la seguente famiglia di parti di Ω

$$\mathcal{F}_X = (X^{-1}(B))_{B \in \mathcal{B}(\mathbf{R})} \subseteq \mathcal{F}$$

è una σ -algebra che prenderà il nome di **σ -algebra generata** da X . È evidente che \mathcal{F}_X risulta essere anche la più piccola σ -algebra di parti di Ω rispetto alla quale X è misurabile. Ora, date due v -a X e Y , le rispettive σ -algebre \mathcal{F}_X e \mathcal{F}_Y in generale non solo non coincidono, ma non sono neanche contenute una nell'altra. Assume quindi particolare rilevanza il seguente risultato

Teorema 3.7. *Date due v -a X e Y , se Y è \mathcal{F}_X -misurabile (cioè se $\mathcal{F}_Y \subseteq \mathcal{F}_X$) allora esiste una funzione di Borel ϕ tale che $Y(\omega) = \phi[X(\omega)]$, $\forall \omega \in \Omega$.*

Dimostrazione: Omessa² □

In pratica Y è una funzione di X quando tutti gli eventi della sua σ -algebra \mathcal{F}_Y sono anche eventi di \mathcal{F}_X , ovvero quando tutte le affermazioni che si possono formulare attorno a Y (gli eventi di \mathcal{F}_Y) possono essere equivalentemente riformulate come affermazioni su X (cioè sono anche eventi di \mathcal{F}_X).

Limiti di successioni di v -a

Una seconda procedura per generare v -a consiste nel costruirle come *limiti di successioni* di v -a $(X_n)_{n \in \mathbf{N}}$. Dovremo pertanto introdurre preliminarmente (ma la questione sarà ripresa con maggiori dettagli nella Sezione 4.1) un'opportuna definizione di **convergenza**. Osserviamo innanzitutto che con l'espressione *convergente* intenderemo d'ora in poi indicare sia successioni convergenti verso limiti finiti che divergenti verso $+\infty$ o $-\infty$; l'espressione *non convergente* sarà invece riservata al caso di successioni che non ammettono limite. Ricorderemo inoltre che il concetto di convergenza che introdurremo qui è solo il primo fra quelli che tratteremo in queste lezioni: infatti, siccome si tratta di limiti di funzioni, esistono diversi, e non equivalenti, tipi di convergenza che saranno discussi in seguito.

Si osservi innanzitutto che una successione di v -a $(X_n)_{n \in \mathbf{N}}$ non è una successione di numeri, ma di funzioni di $\omega \in \Omega$. Solo quando si fissa un arbitrario ω le X_n della successione assumono i particolari valori x_n e si ottiene la successione numerica $(x_n)_{n \in \mathbf{N}}$ come *campione* della successione di v -a. Naturalmente scegliendo un altro ω' si otterrà un'altra successione numerica $(x'_n)_{n \in \mathbf{N}}$, e così di seguito con ω'' ... Da questo punto di vista possiamo allora pensare la successione di v -a $(X_n)_{n \in \mathbf{N}}$ come l'insieme di tutte le sue possibili realizzazioni $(x_n)_{n \in \mathbf{N}}$ al variare di $\omega \in \Omega$.

²A.N. Shiryaev, **PROBABILITY**, Springer (New York, 1996)

Definizione 3.8. Diremo che una successione di v -a $(X_n)_{n \in \mathbf{N}}$ **converge puntualmente** quando tutte le successioni numeriche $(x_n)_{n \in \mathbf{N}}$, con $x_n = X_n(\omega)$, convergono comunque sia scelto $\omega \in \Omega$.

Naturalmente quando la nostra successione $(X_n)_{n \in \mathbf{N}}$ converge nel senso appena indicato ogni campione $(x_n)_{n \in \mathbf{N}}$ converge, al variare di $\omega \in \Omega$, verso un diverso numero x , per cui il limite è una nuova funzione $X(\omega)$ che ad ogni $\omega \in \Omega$ associa un numero $x \in \mathbf{R}$: si prova allora che tale funzione è ancora una v -a, e che viceversa ogni v -a X si può sempre ottenere come limite di opportune successioni di v -a semplici

Proposizione 3.9. Se $(X_n)_{n \in \mathbf{N}}$ è una successione di v -a che converge puntualmente verso $X(\omega)$, allora anche X è misurabile, cioè è una v -a.

Dimostrazione: Omessa³ □

Teorema 3.10. Teorema di Lebesgue: Se X è una v -a non negativa ($X \geq 0$), è sempre possibile determinare una successione $(X_n)_{n \in \mathbf{N}}$ di v -a semplici non decrescenti

$$0 \leq X_n \leq X_{n+1} \leq X, \quad \forall \omega \in \Omega, \quad \forall n \in \mathbf{N}$$

che converge puntualmente (da sotto) verso X , e scriveremo anche $X_n \uparrow X$. Se invece X è una generica v -a è sempre possibile determinare una successione $(X_n)_{n \in \mathbf{N}}$ di v -a semplici che soddisfa

$$|X_n| \leq |X|, \quad \forall \omega \in \Omega, \quad \forall n \in \mathbf{N}$$

e che converge puntualmente verso X .

Dimostrazione: Omessa⁴ □

Sulla base dei teoremi precedenti si può provare che, se X e Y sono v -a, allora anche $X \pm Y$, XY , X/Y ... sono variabili aleatorie, sempre che non assumano valori indeterminati del tipo $\infty - \infty$, ∞/∞ , $0/0$. Si noti infine che talora, per tenere conto di eventuali successioni divergenti (positivamente o negativamente) per qualche ω , dovremo supporre che le v -a possano assumere anche i valori $+\infty$ e $-\infty$. In questo caso parleremo di **v -a estese** per le quali, con qualche dovuta accortezza, valgono i medesimi risultati che valgono per le v -a ordinarie.

3.2 Vettori aleatori e processi stocastici

3.2.1 Elementi aleatori

Il concetto di v -a come funzione misurabile $X : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ può essere generalizzato supponendo che tale funzione prenda valori in spazi diversi da $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$.

³A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

⁴A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

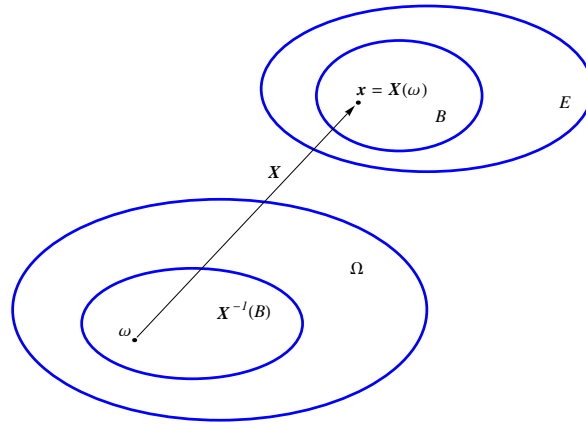


Figura 3.2: Illustrazione grafica della definizione di elemento aleatorio.

In effetti l'unica proprietà di $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ indispensabile per la definizione di una funzione misurabile è che esso sia uno spazio probabilizzabile: le altre proprietà dei numeri reali sono usate solo nella discussione di esempi particolari. Pertanto supporremo nel seguito che la nostra funzione prenda valori in spazi più generali, e parleremo di *elementi aleatori* che, se non si riducono ad una *v-a*, indicheremo con il simbolo \mathbf{X} .

Definizione 3.11. *Dati due spazi probabilizzabili (Ω, \mathcal{F}) ed (E, \mathcal{E}) diremo che una funzione $\mathbf{X} : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ è un **elemento aleatorio** se essa risulta \mathcal{F}/\mathcal{E} -misurabile, cioè se (vedi anche Figura 3.2)*

$$\mathbf{X}^{-1}(B) \in \mathcal{F}, \quad \forall B \in \mathcal{E}$$

Vettori aleatori

Supponiamo innanzitutto che $(E, \mathcal{E}) = (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ con $\mathbf{R}^n = \mathbf{R}_1 \times \dots \times \mathbf{R}_n$ prodotto cartesiano di n copie dell'insieme dei numeri reali. In questo caso i valori assunti dall'elemento aleatorio \mathbf{X} in corrispondenza di un $\omega \in \Omega$ sono le n -ple di numeri reali

$$\mathbf{x} = (x_1, \dots, x_n) = (x_k)_{k=1, \dots, n} \in \mathbf{R}^n$$

e si usa dire che \mathbf{X} è un **vettore aleatorio** (*vett-a*) a valori in \mathbf{R}^n . Si verifica subito che esso è equivalentemente rappresentabile come una n -pla $(X_k)_{k=1, \dots, n}$ in cui le X_k sono delle *v-a* in $(\mathbf{R}_k, \mathcal{B}(\mathbf{R}_k))$ che prendono il nome di **componenti** del *vett-a*. In pratica considerare complessivamente un *vett-a* $\mathbf{X} = (X_k)_{k=1, \dots, n}$ equivale a prendere in esame le n *v-a* che costituiscono le sue componenti. In particolare possiamo considerare il caso in cui $(E, \mathcal{E}) = (\mathbf{C}, \mathcal{B}(\mathbf{C}))$ dove \mathbf{C} è l'insieme dei numeri complessi $z = x + iy$: data la nota isomorfia fra \mathbf{C} e \mathbf{R}^2 potremo definire una ***v-a* complessa** \mathbf{Z} come un *vett-a* con $n = 2$ dato tramite le sue componenti

(le sue parti reale e immaginaria) nella forma $\mathbf{Z}(\omega) = X(\omega) + iY(\omega)$, dove X ed Y sono ordinarie v -a reali.

Successioni aleatorie

Quando invece $(E, \mathcal{E}) = (\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ i valori dell'elemento aleatorio sono delle successioni di numeri reali

$$\mathbf{x} = (x_1, \dots, x_n, \dots) = (x_n)_{n \in \mathbf{N}} \in \mathbf{R}^\infty$$

e \mathbf{X} prende il nome di **successione aleatoria (s-a)**. In questo caso si può dire equivalentemente che \mathbf{X} coincide con una **successione di v-a** $(X_n)_{n \in \mathbf{N}}$ analoga a quelle già introdotte nella sezione precedente quando abbiamo discusso le prime idee di convergenza.

Processi stocastici

Infine se $(E, \mathcal{E}) = (\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$ con T sottoinsieme (in generale non finito e non numerabile) di \mathbf{R} , il nostro elemento aleatorio \mathbf{X} associerà ad ogni $\omega \in \Omega$ una intera funzione $(x_t)_{t \in T}$, o anche $x(t)$, detta **traiettoria**, mentre l'elemento aleatorio stesso prenderà il nome di **processo stocastico (p-s)**. Anche in questo caso si può mostrare che un p -s può sempre essere visto come una famiglia $(X_t)_{t \in T}$ di v -a X_t , ovvero $X(t)$. Pertanto \mathbf{X} può essere considerato equivalentemente o come un'applicazione che associa ad ogni $\omega \in \Omega$ una intera funzione (traiettoria) $(x_t)_{t \in T} \in \mathbf{R}^T$, oppure come un'applicazione che associa ad ogni $t \in T$ una v -a $X_t : (\Omega, \mathcal{F}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$. In pratica le componenti $X_t(\omega) = X(\omega; t)$ di un p -s \mathbf{X} sono funzioni di due variabili t e ω , e le diverse notazioni adottate serviranno a mettere in evidenza alternativamente le due dipendenze. Maggiori dettagli sui p -s saranno introdotti nella seconda parte di queste lezioni. Osserviamo che quando in particolare $T = \{1, 2, \dots\} = \mathbf{N}$ è l'insieme dei numeri naturali il p -s si riduce ad una successione di v -a $(X_n)_{n \in \mathbf{N}}$ e per questo viene anche chiamata *processo stocastico a tempi discreti (interi)*.

3.2.2 Distribuzioni e densità congiunte e marginali

Se supponiamo che sullo spazio probabilizzabile (Ω, \mathcal{F}) sia anche definita una probabilità \mathbf{P} potremo indurre, con le usuali procedure, delle probabilità anche sui vari tipi di spazio di arrivo (E, \mathcal{E}) e quindi leggi e distribuzioni dei diversi elementi aleatori.

Leggi di vettori aleatori

Dato un *vett-a* $\mathbf{X} = (X_k)_{k=1, \dots, n}$ chiameremo **legge o distribuzione congiunta** delle sue componenti X_k la $\mathbf{P}_{\mathbf{X}}$ definita su $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ da

$$\mathbf{P}_{\mathbf{X}}\{B\} = \mathbf{P}\{\mathbf{X}^{-1}(B)\} = \mathbf{P}\{\mathbf{X} \in B\} = \mathbf{P}\{(X_1, \dots, X_n) \in B\}$$

dove B è un generico elemento di $\mathcal{B}(\mathbf{R}^n)$; chiameremo invece **leggi o distribuzioni marginali** delle componenti X_k le \mathbf{P}_{X_k} definite su $(\mathbf{R}_k, \mathcal{B}(\mathbf{R}_k))$ da

$$\mathbf{P}_{X_k}\{A\} = \mathbf{P}\{X_k^{-1}(A)\} = \mathbf{P}\{X_k \in A\}$$

dove A è un generico elemento di $\mathcal{B}(\mathbf{R}_k)$. Corrispondentemente chiameremo **funzione di distribuzione congiunta** (*fdd* congiunta) del *vett-a* \mathbf{X} la *fdd* di $\mathbf{P}_{\mathbf{X}}$, cioè la funzione

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= F_{\mathbf{X}}(x_1, \dots, x_n) = \mathbf{P}_{\mathbf{X}}\{(-\infty, x_1] \times \dots \times (-\infty, x_n]\} \\ &= \mathbf{P}\{X_1 \leq x_1, \dots, X_n \leq x_n\} \end{aligned}$$

con $\mathbf{x} = (x_1, \dots, x_n) \in \mathbf{R}^n$, e **funzioni di distribuzione marginali** (*fdd* marginali) delle componenti X_k le *fdd* delle \mathbf{P}_{X_k} , cioè le funzioni

$$F_{X_k}(x_k) = \mathbf{P}\{X_k \leq x_k\} = \mathbf{P}_{X_k}\{(-\infty, x_k]\} = \mathbf{P}_{\mathbf{X}}\{\mathbf{R}_1 \times \dots \times (-\infty, x_k] \times \dots \times \mathbf{R}_n\}$$

con $x_k \in \mathbf{R}_k$ e $k = 1, \dots, n$.

Leggi di successioni aleatorie

I concetti relativi ai *vett-a* si estendono facilmente anche al caso delle *s-a* $\mathbf{X} = (X_n)_{n \in \mathbf{N}}$ selezionando dei sottoinsiemi finiti di componenti che possono essere trattati come *vett-a*: se infatti $\{k_1, \dots, k_m\}$ è uno di tali sottoinsiemi finiti di indici, chiameremo **legge o distribuzione congiunta finito-dimensionale** della *s-a* \mathbf{X} la distribuzione congiunta del *vett-a* $(X_{k_1}, \dots, X_{k_m})$. Naturalmente scegliendo in tutte le maniere possibili il sottoinsieme finito di indici $\{k_1, \dots, k_m\}$ potremo costruire una intera famiglia (consistente, nel senso indicato nella Sezione 2.4) di distribuzioni finito-dimensionali di \mathbf{X} dotate delle corrispondenti **funzioni di distribuzione congiunte finito-dimensionali**

$$F(x_1, k_1; \dots; x_m, k_m) = \mathbf{P}\{X_{k_1} \leq x_1; \dots; X_{k_m} \leq x_m\}$$

Le distribuzioni (e le funzioni di distribuzione) delle singole componenti X_k assumono invece la qualifica di **leggi o distribuzioni marginali**. Ricorderemo infine che viceversa, in base al Teorema 2.35, l'assegnazione di una famiglia consistente di distribuzioni finito-dimensionali è sufficiente per definire la misura di probabilità $\mathbf{P}_{\mathbf{X}}$ su tutto lo spazio $(\mathbf{R}^\infty, \mathcal{B}(\mathbf{R}^\infty))$ nel quale la nostra *s-a* prende valori.

Leggi di processi stocastici

Analogamente può essere infine trattato il caso dei *p-s* $\mathbf{X} = (X_t)_{t \in T}$: detto infatti $\{t_1, \dots, t_m\}$ un insieme finito di punti di T chiameremo **legge o distribuzione congiunta finito-dimensionale** del *p-s* \mathbf{X} la distribuzione congiunta del *vett-a*

$(X_{t_1}, \dots, X_{t_m})$ e **funzione di distribuzione congiunta finito-dimensionale** la corrispondente *fdd*

$$F(x_1, t_1; \dots; x_m, t_m) = \mathbf{P}\{X_{t_1} \leq x_1; \dots; X_{t_m} \leq x_m\}$$

Le distribuzioni delle singole componenti X_{t_i} , invece, saranno chiamate **leggi o distribuzioni marginali**. Anche qui, scegliendo in tutti i modi possibili i punti $\{t_1, \dots, t_m\}$ potremo definire una intera famiglia consistente di distribuzioni e *fdd*, e viceversa, in base al Teorema 2.37, l'assegnazione di una famiglia consistente di *fdd* finito-dimensionali è tutto ciò che è necessario per definire $\mathbf{P}_{\mathbf{X}}$ su $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$.

Marginalizzazione

Le osservazioni sulle relazioni fra distribuzioni congiunte e marginali discusse nella Sezione 2.3.3 trovano applicazione anche qui. In particolare, data la *fdd* congiunta $F_{\mathbf{X}}(\mathbf{x})$ di un *vett-a* \mathbf{X} , sarà sempre possibile ricavare in maniera univoca le *fdd* marginali seguendo l'usuale procedura, cioè

$$F_{X_k}(x_k) = F_{\mathbf{X}}(+\infty, \dots, x_k, \dots, +\infty)$$

Infatti si vede facilmente che

$$\begin{aligned} F_{X_k}(x_k) &= \mathbf{P}\{X_k \leq x_k\} = \mathbf{P}\{X_1 < +\infty, \dots, X_k \leq x_k, \dots, X_n < +\infty\} \\ &= F_{\mathbf{X}}(+\infty, \dots, x_k, \dots, +\infty) \end{aligned}$$

Invece, come è stato mostrato sempre nella Sezione 2.3.3, date le *fdd* marginali F_{X_k} non si può, in generale, ricostruire la *fdd* congiunta $F_{\mathbf{X}}$ *in maniera univoca*.

Leggi e distribuzioni *ac*

Quando invece la *fdd* congiunta $F_{\mathbf{X}}(\mathbf{x})$ di un *vett-a* \mathbf{X} è *ac* esiste anche una **densità congiunta** (*ddp* congiunta) $f_{\mathbf{X}}(\mathbf{x}) \geq 0$ per la quale avremo

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{\mathbf{X}}(y_1, \dots, y_n) dy_1 \dots dy_n$$

Naturalmente sussiste anche la relazione inversa

$$f_{\mathbf{X}}(x_1, \dots, x_n) = \frac{\partial^n F_{\mathbf{X}}(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n}$$

Si mostra che in questo caso anche le singole componenti X_k hanno *fdd* marginali $F_{X_k}(x_k)$ *ac* con *ddp* $f_{X_k}(x_k)$ che si ricavano dalla *ddp* congiunta $f_{\mathbf{X}}(\mathbf{x})$ con la solita regola di **marginalizzazione**

$$f_{X_k}(x_k) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x_1, \dots, x_k, \dots, x_n) dx_1 \dots dx_{k-1} dx_{k+1} \dots dx_n$$

dove le $n - 1$ integrazioni sono effettuate su tutte le variabili tranne la k -ma. Ovviamente risulta anche

$$f_{X_k}(x_k) = F'_{X_k}(x_k)$$

Queste $f_{X_k}(x_k)$ prendono il nome di **densità marginali** ed anche qui vale l'osservazione che, mentre dalla ddp congiunta $f_{\mathbf{X}}(\mathbf{x})$ è sempre possibile ricavare con un'integrazione le ddp marginali $f_{X_k}(x_k)$, la ricostruzione della ddp congiunta a partire dalle marginali in generale non è unica. La conoscenza delle ddp congiunte consente infine di calcolare $\mathbf{P}_{\mathbf{X}}\{B\}$ con $B \in \mathcal{B}(\mathbf{R}^n)$ mediante un'integrazione: ad esempio se $B = (a_1, b_1] \times \dots \times (a_n, b_n]$ si ha, in analogia con il caso unidimensionale,

$$\begin{aligned} \mathbf{P}_{\mathbf{X}}\{B\} &= \mathbf{P}\{\mathbf{X} \in B\} = \mathbf{P}\{a_1 < X_1 \leq b_1, \dots, a_n < X_n \leq b_n\} \\ &= \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n \end{aligned}$$

una formula che sarà ripresa più oltre con maggiori dettagli.

Esempio 3.12. Vettori aleatori discreti: *I casi più semplici sono quelli di **vett-a con componenti discrete**: in questo caso infatti la $F_{\mathbf{X}}$ è discreta, e sarà sufficiente assegnare delle distribuzioni discrete congiunte per definirne la legge. Come esempio esplicito considereremo quello di un **vett-a multinomiale** $\mathbf{X} = (X_1, \dots, X_r)$ con $r = 1, 2, \dots$, e con legge $\mathfrak{B}(n; p_1, \dots, p_r)$ definita nell'Esempio 2.6. Si ricorderà che questa legge descrive un esperimento aleatorio con $r+1$ possibili risultati. Le singole componenti X_j – che rappresentano il numero di volte in cui si ottiene il j -mo possibile risultato su n tentativi – potranno assumere valori da 0 ad n con legge congiunta multinomiale, mentre è sottintesa la v -a X_0 tale che $X_0 + X_1 + \dots + X_r = n$. Dalla (2.9) si avrà allora*

$$\mathbf{P}_{\mathbf{X}}\{\mathbf{k}\} = \mathbf{P}\{X_1 = k_1, \dots, X_r = k_r\} = \binom{n}{k_1, \dots, k_r} p_0^{k_0} p_1^{k_1} \cdot \dots \cdot p_r^{k_r} \quad (3.1)$$

con $k_0 + k_1 + \dots + k_r = n$, e $p_0 + p_1 + \dots + p_r = 1$. Naturalmente se $r = 1$ \mathbf{X} si riduce a una sola componente X_1 con legge binomiale $\mathfrak{B}(n; p_1)$.

Riprendendo la discussione della Sezione 2.1.2, potremo poi considerare come ulteriore esempio il vett-a $\mathbf{X} = (X_1, \dots, X_n)$ le cui componenti X_k rappresentano i singoli **esiti di n lanci di moneta**, e assumono quindi solo i valori 0 e 1. Pertanto i valori di \mathbf{X} sono le n -ple ordinate del tipo (2.5) che nel Modello di Bernoulli si realizzano con probabilità (2.6), sicché la distribuzione congiunta del vett-a \mathbf{X} è ora

$$\mathbf{P}_{\mathbf{X}}\{a_1, \dots, a_n\} = \mathbf{P}\{X_1 = a_1, \dots, X_n = a_n\} = p^k q^{n-k} \quad a_j = 0, 1 \quad (3.2)$$

dove $k = \sum_j a_j$ e $q = 1 - p$. Un semplice calcolo mostrerebbe infine che le marginali delle singole componenti X_k sono tutte id con legge di Bernoulli $\mathfrak{B}(1; p)$. Per evitare inutili complicazioni formali non daremo invece la forma esplicita delle fdd congiunte $F_{\mathbf{X}}$ di questi due esempi

Esempio 3.13. Vettori aleatori Gaussiani $\mathbf{X} \sim \mathfrak{N}(\mathbf{b}, \mathbb{A})$: Un caso importante di vett-*a* ac dotato di ddp è quello dei **vett-*a* Gaussiani (normali)** $\mathbf{X} = (X_1, \dots, X_n)$ caratterizzati da ddp Gaussiane multivariate $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ del tipo (2.22) già discusse nella Sezione 2.3.2. Naturalmente ogni vett-*a* Gaussiano ha poi come ddp marginali le Gaussiane univariate $\mathfrak{N}(b_k, a_k^2)$ del tipo (2.31) come può essere visto con un calcolo diretto marginalizzando la $\mathfrak{N}(\mathbf{b}, \mathbb{A})$. Non bisogna invece pensare che un vett-*a* con marginali Gaussiane univariate sia, solo per questo, anche un vett-*a* Gaussiano: la ddp congiunta, infatti, non è univocamente definita a partire dalle marginali, e pertanto il nostro vett-*a* potrebbe non avere una ddp congiunta della famiglia $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ pur avendo delle marginali tutte Gaussiane $\mathfrak{N}(b_k, a_k^2)$

3.2.3 Indipendenza di *v-a*

Definizione 3.14. Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ ed una famiglia $\mathbf{X} = (X_s)_{s \in S}$ di *v-a*, con S insieme di indici arbitrario (finito, numerabile o non numerabile), diremo che le componenti X_s di \mathbf{X} sono ***v-a* indipendenti** se, comunque scelte m componenti con indici s_1, \dots, s_m e comunque presi i sottoinsiemi $B_k \in \mathcal{B}(\mathbf{R})$, con $k = 1, \dots, m$ risulta

$$\mathbf{P}\{X_{s_1} \in B_1, \dots, X_{s_m} \in B_m\} = \mathbf{P}\{X_{s_1} \in B_1\} \cdot \dots \cdot \mathbf{P}\{X_{s_m} \in B_m\}$$

In pratica l'indipendenza delle *v-a* di \mathbf{X} si riconduce alla indipendenza (nel senso usuale della Sezione 1.5) di tutti gli eventi che possono essere costruiti a partire da un arbitrario sottoinsieme (finito) di *v-a* di \mathbf{X} , e in linea di principio essa richiede la verifica di un numero molto grande di relazioni. Il caso più semplice è quello in cui $S = \{1, \dots, n\}$ e \mathbf{X} è un vett-*a* con n componenti, ma la definizione si adatta anche al caso delle *s-a* e dei *p-s*. Infine, nello spirito delle osservazioni svolte nella Sezione 3.1.2, ricorderemo che possono esistere *v-a* identicamente distribuite, ma diverse ed anche indipendenti: in questi casi parleremo di ***v-a* indipendenti e identicamente distribuite (iid)**.

Teorema 3.15. Dato un vett-*a* $\mathbf{X} = (X_k)_{k=1, \dots, n}$ su $(\Omega, \mathcal{F}, \mathbf{P})$ le seguenti due affermazioni sono equivalenti:

- (a) le componenti X_k sono indipendenti;
- (b) $F_{\mathbf{X}}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n)$;

se poi $F_{\mathbf{X}}(\mathbf{x})$ è dotata di ddp, tali due affermazioni sono anche equivalenti a

- (c) $f_{\mathbf{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n)$.

Dimostrazione: Omessa⁵

□

⁵A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

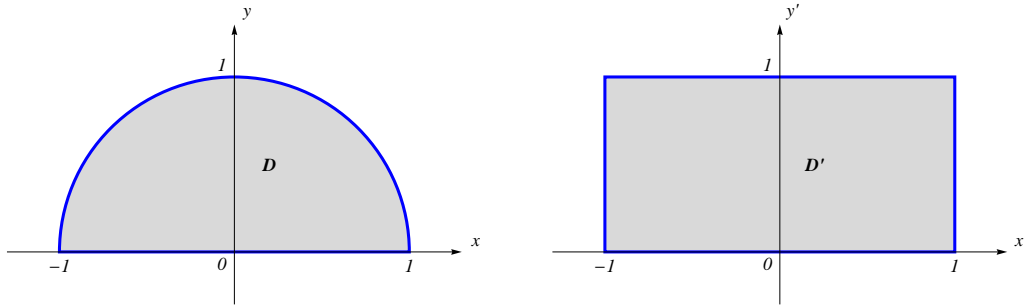


Figura 3.3: Domini D e D' dell'Esempio 3.16

Esempio 3.16. *Un primo esempio di applicazione del Teorema 3.15 si ha osservando che il vett-a \mathbf{X} definito nell'Esempio 3.12 con distribuzione (3.2), e che descrive l'esito di n lanci di moneta nel Modello di Bernoulli, risulta evidentemente composto di v -a iid. Infatti la distribuzione congiunta (3.2) (e quindi anche la corrispondente fdd congiunta) risulta essere proprio il prodotto delle n identiche leggi di Bernoulli $\mathfrak{B}(1;p)$ che rappresentano le leggi marginali delle X_k . Possiamo quindi dire che il nostro \mathbf{X} , con la legge (3.2) che gli è stata assegnata, è un vett-a composto di n **v-a di Bernoulli iid** e scriveremo anche simbolicamente $\mathbf{X} \sim [\mathfrak{B}(1;p)]^n$*

Riprendendo poi l'Esempio 2.30 associato alla Figura 2.12, supponiamo che f e g siano rispettivamente le due diverse ddp congiunte di due vett-a \mathbf{U} e \mathbf{V} con due componenti ciascuno. Abbiamo mostrato che le marginali di questi vett-a coincidono, ma che i loro rapporti con le rispettive distribuzioni congiunte riassunti in (2.30) sono differenti: in particolare la ddp congiunta di \mathbf{U} coincide con il prodotto delle sue marginali, mentre questo non avviene per \mathbf{V} . Possiamo dunque concludere in base al Teorema 3.15 che \mathbf{U} e \mathbf{V} , pur avendo identiche marginali, differiscono sostanzialmente per il fatto che le componenti di \mathbf{U} sono indipendenti, mentre quelle di \mathbf{V} non lo sono.

In un terzo esempio analogo al precedente supporremo ancora di considerare due vett-a $\mathbf{U} = (X, Y)$ e $\mathbf{V} = (X', Y')$ dotati di ddp costanti (leggi uniformi) in due diverse regioni del piano \mathbf{R}^2 :

$$f_{\mathbf{U}}(x, y) = \begin{cases} \frac{2}{\pi}, & \text{se } (x, y) \in D, \\ 0, & \text{se } (x, y) \notin D, \end{cases} \quad f_{\mathbf{V}}(x', y') = \begin{cases} \frac{1}{2}, & \text{se } (x', y') \in D', \\ 0, & \text{se } (x', y') \notin D', \end{cases}$$

dove le regioni D e D' rappresentate in Figura 3.3 sono

$$D = \{(x, y) \in \mathbf{R}^2 : -1 \leq x \leq +1, 0 \leq y \leq \sqrt{1-x^2}\}$$

$$D' = \{(x', y') \in \mathbf{R}^2 : -1 \leq x' \leq +1, 0 \leq y' \leq 1\}$$

Si vede ora con un'integrazione elementare che le ddp marginali di \mathbf{U} sono

$$f_X(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, & \text{se } x \in [-1, 1] \\ 0, & \text{se } x \notin [-1, 1] \end{cases} \quad f_Y(y) = \begin{cases} \frac{4}{\pi}\sqrt{1-y^2}, & \text{se } y \in [0, 1] \\ 0, & \text{se } y \notin [0, 1] \end{cases}$$

mentre le marginali di \mathbf{V} sono

$$f_{X'}(x') = \begin{cases} \frac{1}{2}, & \text{se } x' \in [-1, 1] \\ 0, & \text{se } x' \notin [-1, 1] \end{cases} \quad f_{Y'}(y') = \begin{cases} 1, & \text{se } y' \in [0, 1] \\ 0, & \text{se } y' \notin [0, 1] \end{cases}$$

e pertanto

$$f_{\mathbf{U}}(x, y) \neq f_X(x)f_Y(y) \quad f_{\mathbf{V}}(x', y') = f_{X'}(x')f_{Y'}(y')$$

sicch  le componenti di \mathbf{V} sono *v-a* indipendenti mentre quelle di \mathbf{U} non lo sono.

3.2.4 Decomposizione di *v-a* binomiali

Le somme di *v-a* e le loro leggi giocano un ruolo molto importante e saranno discusse pi  compiutamente nella successiva Sezione 3.5.2: qui ci limiteremo solo ad una breve trattazione preliminare per particolari *v-a* discrete che ci consentir  tuttavia di ottenere un importante risultato sulle *v-a* binomiali.

Sia $\mathbf{U} = (X, Y)$ un *vett-a* le cui componenti assumano solo valori interi, e siano

$$\mathbf{P}_{\mathbf{U}}\{j, k\} = \mathbf{P}\{X = j, Y = k\} \quad \mathbf{P}_X\{j\} = \mathbf{P}\{X = j\} \quad \mathbf{P}_Y\{k\} = \mathbf{P}\{Y = k\}$$

le rispettive distribuzioni congiunta e marginali. Nel seguito sar  sufficiente che X e Y assumano solo un *numero finito* di valori

$$X = j \in \{0, 1, \dots, m\} \quad Y = k \in \{0, 1, \dots, n\} \quad (3.3)$$

ma per semplificare le notazioni supporremo che i loro possibili valori comprendano tutti i numeri relativi \mathbf{Z} (interi positivi, negativi e nulli), precisando per  che le probabilit  sono nulle su tutti i valori diversi da (3.3). Definiamo poi la nuova *v-a* $W = X + Y$ che ovviamente assumer  i valori

$$W = \ell \in \{0, 1, \dots, n + m\}$$

e poniamoci il problema di calcolare la sua distribuzione \mathbf{P}_W a partire dalle distribuzioni $\mathbf{P}_{\mathbf{U}}$, \mathbf{P}_X e \mathbf{P}_Y

Proposizione 3.17. *Con le notazioni fin qui introdotte si ha*

$$\mathbf{P}_W\{\ell\} = \sum_{k \in \mathbf{Z}} \mathbf{P}_{\mathbf{U}}\{\ell - k, k\}$$

Se poi X e Y sono anche **indipendenti** risulta

$$\mathbf{P}_W\{\ell\} = \sum_{k \in \mathbf{Z}} \mathbf{P}_X\{\ell - k\} \mathbf{P}_Y\{k\} = (\mathbf{P}_X * \mathbf{P}_Y)\{\ell\} \quad (3.4)$$

e diremo che la legge \mathbf{P}_W   la **convoluzione (discreta)** di \mathbf{P}_X e \mathbf{P}_Y

Dimostrazione: Si ha infatti

$$\begin{aligned} P_W\{\ell\} &= P\{W = \ell\} = P\{X + Y = \ell\} = \sum_{j+k=\ell} P\{X = j, Y = k\} \\ &= \sum_{j+k=\ell} P_U\{j, k\} = \sum_{k \in \mathbf{Z}} P_U\{\ell - k, k\} \end{aligned}$$

Se inoltre X e Y sono indipendenti dal Teorema 3.15 si ha $P_U\{j, k\} = P_X\{j\}P_Y\{k\}$ e quindi si ricava facilmente il risultato (3.4) \square

La formula (3.4) in realtà si riferisce solo alle leggi delle nostre v -a X, Y e W ed è quindi importante sottolineare che il concetto di *convoluzione* (per il momento *discreta*) si applica prima di tutto alle distribuzioni e non direttamente alle v -a

Definizione 3.18. Se \mathfrak{P} e \mathfrak{Q} sono due leggi discrete con distribuzioni rispettivamente $p(j)$ e $q(k)$ diremo che la legge \mathfrak{R} è la loro **convoluzione (discreta)** $\mathfrak{P} * \mathfrak{Q}$ se la sua distribuzione $r(\ell)$ si ottiene come

$$r(\ell) = \sum_{k \in \mathbf{Z}} p(\ell - k)q(k) \quad (3.5)$$

Nel *modello di Bernoulli* introdotto nella Sezione 2.1.2 lo spazio Ω era l'insieme delle n -ple di estrazioni ω di tipo (2.5), dove $a_j = 0, 1$ rappresenta l'esito della j -ma estrazione, e $k = \sum_j a_j$ è il numero di palline bianche in una n -pla di estrazioni. Si definivano poi gli eventi

$$\begin{aligned} A_j &= \{\omega \in \Omega : a_j = 1\} \quad j = 1, \dots, n \\ D_k &= \left\{ \omega \in \Omega : \sum_{j=0}^n a_j = k \right\} \quad k = 0, 1, \dots, n \end{aligned}$$

e, assegnata un'opportuna probabilità su Ω mediante la (2.6), abbiamo mostrato che le $P\{D_k\}$ seguono una distribuzione binomiale $\mathfrak{B}(n; p)$ mentre le A_j sono eventi indipendenti con $P\{A_j\} = p$. Questo modello può ora essere ricostruito in termini di v -a: consideriamo prima di tutto il *vett-a* $\mathbf{X} = (X_1, \dots, X_n)$ (già introdotto nelle Sezioni 3.2.2 e 3.2.3) le cui componenti indipendenti sono le v -a (indicatori)

$$X_j = I_{A_j}$$

che rappresentano gli esiti delle singole estrazioni, e poi la v -a semplice

$$S_n = \sum_{k=0}^n k I_{D_k}$$

che rappresenta il numero di palline bianche in n estrazioni. Ovviamente avremo ora che $D_k = \{S_n = k\}$ e $A_j = \{X_j = 1\}$, ed è chiaro che fra le nostre nuove v -a sussiste la relazione

$$S_n = \sum_{j=1}^n X_j \quad (3.6)$$

Infatti, dalla definizione, per ogni $\omega \in \Omega$ il valore di S_n (numero di palline bianche su n estrazioni) coincide proprio con la somma di tutti i valori delle X_j . Pertanto le v -a dei due membri dell'uguaglianza (3.6) sono per definizione indistinguibili nel senso della Definizione 3.4, e quindi sono anche *id*. Nella Sezione 2.1.2 abbiamo poi mostrato in maniera diretta che

$$X_j \sim \mathfrak{B}(1; p) \quad S_n \sim \mathfrak{B}(n; p)$$

e quindi potremo ora affermare che la somma di n v -a di Bernoulli *indipendenti* $X_j \sim \mathfrak{B}(1; p)$ è una v -a binomiale $S_n \sim \mathfrak{B}(n; p)$. Vogliamo ora rivedere ed estendere questo risultato alla luce della Proposizione 3.17

Proposizione 3.19. *Ogni legge binomiale $\mathfrak{B}(n; p)$ è convoluzione di n leggi di Bernoulli $\mathfrak{B}(1; p)$ nel senso che*

$$\mathfrak{B}(n; p) = [\mathfrak{B}(1; p)]^{*n} = \underbrace{\mathfrak{B}(1; p) * \dots * \mathfrak{B}(1; p)}_{n \text{ volte}} \quad (3.7)$$

e pertanto:

- se il vett-a $\mathbf{X} \sim [\mathfrak{B}(1; p)]^n$ ha componenti iid di Bernoulli, la somma delle sue componenti $S_n = X_1 + \dots + X_n$ seguirà una legge binomiale $\mathfrak{B}(n; p)$
- ogni v -a binomiale $S_n \sim \mathfrak{B}(n; p)$ è (in distribuzione) somma delle componenti di un vett-a $\mathbf{X} \sim [\mathfrak{B}(1; p)]^n$ con componenti iid di Bernoulli

Diremo quindi che una v -a binomiale S_n si **decompone** sempre (in distribuzione) nella somma di n v -a di Bernoulli iid X_j

$$S_n \stackrel{d}{=} X_1 + \dots + X_n \quad (3.8)$$

Dimostrazione: La legge di Bernoulli $\mathfrak{B}(1; p)$ ha distribuzione

$$p_1(k) = \binom{1}{k} p^k q^{1-k} = \begin{cases} p & \text{se } k = 1 \\ q & \text{se } k = 0 \end{cases}$$

sicché da (3.5) la legge $\mathfrak{B}(1; p) * \mathfrak{B}(1; p)$ avrà probabilità non nulle solo su 0, 1, 2

$$\begin{aligned} r(0) &= p_1(0)p_1(0) = q^2 \\ r(1) &= p_1(0)p_1(1) + p_1(1)p_1(0) = 2pq \\ r(2) &= p_1(1)p_1(1) = p^2 \end{aligned}$$

che ovviamente coincidono con la distribuzione

$$p_2(k) = \binom{2}{k} p^k q^{2-k}$$

della legge binomiale $\mathfrak{B}(2; p)$. Il risultato completo (3.7) segue per induzione, ma noi ne trascureremo la verifica. Si capisce facilmente allora che, se X_1, \dots, X_n sono n v -a di Bernoulli $\mathfrak{B}(1; p)$ *iid*, a norma della Proposizione 3.17 la loro somma S_n avrà come distribuzione la convoluzione delle loro leggi e quindi, secondo (3.7), sarà una v -a binomiale $\mathfrak{B}(n; p)$. Viceversa se $S_n \sim \mathfrak{B}(n; p)$ sappiamo da (3.7) che la sua legge binomiale è convoluzione di n Bernoulli $\mathfrak{B}(1; p)$; ma dalla Definizione 3.5 sappiamo anche che alle n leggi $\mathfrak{B}(1; p)$ possiamo associare n v -a di Bernoulli X_1, \dots, X_n *iid* in modo che, dalla Proposizione 3.17, risulti $S_n \stackrel{d}{=} X_1 + \dots + X_n$ \square

3.3 Valore d'attesa

3.3.1 Integrazione e valore d'attesa

Il valore d'attesa di una v -a X è un importante indicatore numerico che caratterizza la posizione del baricentro della sua distribuzione \mathbf{P}_X , e che trae origine dal concetto di media pesata di un campione di numeri. Nel caso di v -a semplici

$$X = \sum_{k=1}^n x_k I_{D_k}$$

definite su uno spazio $(\Omega, \mathcal{F}, \mathbf{P})$ con una decomposizione $D_k = \{X = x_k\} \in \mathcal{F}$, il valore d'attesa viene definito in modo elementare proprio come *media pesata*

$$\mathbf{E}[X] = \sum_{k=1}^n x_k \mathbf{P}_X\{x_k\} = \sum_{k=1}^n x_k \mathbf{P}\{X = x_k\} = \sum_{k=1}^n x_k \mathbf{P}\{D_k\} \quad (3.9)$$

Questa definizione si estende in maniera naturale anche al caso di v -a discrete non semplici, cioè con un insieme infinito numerabile di valori, con la serie

$$\mathbf{E}[X] = \sum_{k=1}^{\infty} x_k \mathbf{P}_X\{x_k\}$$

purché essa converga. In particolare si noti che, comunque dato un evento $A \in \mathcal{F}$, risulta sempre

$$\mathbf{E}[I_A] = \mathbf{P}\{A\} \quad (3.10)$$

un'equazione che, pur nella sua semplicità, mette in evidenza l'importante relazione che intercorre fra i concetti di valore d'attesa e di probabilità.

Quando però X è una generica v -a queste definizioni elementari non sono più applicabili. In questo caso si osserva che se X è una v -a *non negativa* il Teorema 3.10 garantisce l'esistenza di una successione (non decrescente) di v -a semplici e non negative $(X_n)_{n \in \mathbf{N}}$ tale che $X_n(\omega) \uparrow X(\omega)$ comunque scelto $\omega \in \Omega$. Le definizioni elementari precedenti consentono allora di definire la successione numerica

a termini non negativi $(\mathbf{E}[X_n])_{n \in \mathbf{N}}$ che risulta monotona non decrescente, e quindi ammette sicuramente un limite (eventualmente $+\infty$) che potremo prendere come definizione del valore d'attesa di X . Per estendere poi la procedura anche a v -a del tutto generali, ricorderemo che un'arbitraria v -a X è sempre suscettibile di essere rappresentata come differenza di v -a non negative nella forma

$$X = X^+ - X^-$$

dove $X^+ = \max\{X, 0\}$ e $X^- = -\min\{X, 0\}$ prendono rispettivamente i nomi di *parte positiva* e *parte negativa* di X . Conseguentemente se X è una generica v -a basterà considerare separatamente le due v -a non negative X^+ e X^- , definirne i valore d'attesa e finalmente ricostituire quello di X per differenza. Evitando allora di entrare in maggiori dettagli tecnici, queste considerazioni ci consentono di introdurre le seguenti definizioni che si riducono a quelle già introdotta per le v -a semplici nel caso di spazi Ω finiti

Definizione 3.20. *Per definire il **valore d'attesa** di una v -a si introducono i seguenti concetti:*

- se X è una v -a non negativa chiameremo **valore d'attesa** il limite (eventualmente anche $+\infty$)

$$\mathbf{E}[X] = \lim_n \mathbf{E}[X_n]$$

dove $(X_n)_{n \in \mathbf{N}}$ è una successione monotona non decrescente di v -a semplici non negative tale che $X_n \uparrow X$ per ogni $\omega \in \Omega$, e le $\mathbf{E}[X_n]$ sono definite in maniera elementare; l'esistenza delle successioni $(X_n)_{n \in \mathbf{N}}$ è garantita dal Teorema 3.10, e si dimostra che il risultato non dipende dalla particolare successione scelta;

- se X è una generica v -a diremo che il suo valore d'attesa **esiste** (o che è definito) quando almeno uno dei due numeri non negativi $\mathbf{E}[X^+]$, $\mathbf{E}[X^-]$ è finito, e in tal caso si pone

$$\mathbf{E}[X] = \mathbf{E}[X^+] - \mathbf{E}[X^-]$$

se invece $\mathbf{E}[X^+]$ e $\mathbf{E}[X^-]$ sono ambedue $+\infty$ si dice che il valore d'attesa di X **non esiste**;

- quando ambedue i numeri $\mathbf{E}[X^+]$, $\mathbf{E}[X^-]$ sono finiti anche $\mathbf{E}[X]$ risulta finito e si dice che X è **integrabile**, cioè ha valore d'attesa finito; siccome però

$$|X| = X^+ + X^-$$

se X è integrabile essa risulta anche **assolutamente integrabile**, cioè

$$\mathbf{E}[|X|] < +\infty$$

ed è facile vedere che vale anche il viceversa, sicché potremo dire in tutta generalità che una v -a è integrabile se e solo se essa è assolutamente integrabile.

Siccome le procedure delineate nella Definizione 3.20 riproducono quelle usate per la definizione dell'**integrale di Lebesgue** (del quale il valore d'attesa non è che un caso particolare), useremo nel seguito anche la notazione

$$\mathbf{E}[X] = \int_{\Omega} X d\mathbf{P} = \int_{\Omega} X(\omega) \mathbf{P}\{d\omega\}$$

In generale, infatti, se g è una funzione misurabile da (Ω, \mathcal{F}) in $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, e se μ è una misura (ma in generale non una probabilità) su (Ω, \mathcal{F}) , l'integrale di Lebesgue si definisce ripercorrendo i concetti esposti nella Definizione 3.20 e si adotta la notazione

$$\int_{\Omega} g d\mu = \int_{\Omega} g(\omega) \mu\{d\omega\}$$

Ovviamente quando μ non è una probabilità tale integrale non potrà essere interpretato come un valore d'attesa.

È anche possibile definire integrali estesi a sottoinsiemi di Ω : se $A \in \mathcal{F}$ è un sottoinsieme di Ω , si chiama **integrale di Lebesgue esteso all'insieme A** l'integrale (rispettivamente nei casi di una probabilità \mathbf{P} o di una generica misura μ)

$$\int_A X d\mathbf{P} = \int_{\Omega} X I_A d\mathbf{P} = \mathbf{E}[X I_A] \qquad \int_A g d\mu = \int_{\Omega} g I_A d\mu$$

Si noti inoltre che la (3.10) può ora essere scritta nella forma più suggestiva

$$\mathbf{P}\{A\} = \mathbf{E}[I_A] = \int_{\Omega} I_A d\mathbf{P} = \int_A d\mathbf{P} \tag{3.11}$$

per sottolineare che la probabilità di un evento A non è altro che l'integrale della misura \mathbf{P} su A .

Definizione 3.21. *Si chiama **momento di ordine k** di una v -a X il valore d'attesa (quando esso esiste)*

$$\mathbf{E}[X^k] = \int_{\Omega} X^k d\mathbf{P} \qquad k = 0, 1, 2, \dots$$

e **momento assoluto di ordine r** il valore d'attesa (quando esso esiste)

$$\mathbf{E}[|X|^r] = \int_{\Omega} |X|^r d\mathbf{P} \qquad r \geq 0$$

Sarà importante infine richiamare le notazioni usualmente adottate per le integrazioni nel caso particolare in cui $(\Omega, \mathcal{F}) = (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$, g è una funzione di Borel e $G(\mathbf{x})$ è la *fdd* generalizzata di una misura di Lebesgue-Stieltjes μ : parleremo allora di **integrale di Lebesgue-Stieltjes** e scriveremo

$$\int_{\Omega} g d\mu = \int_{\mathbf{R}^n} g dG = \int_{\mathbf{R}^n} g(\mathbf{x}) G(d\mathbf{x}) = \int_{\mathbf{R}^n} g(x_1, \dots, x_n) G(dx_1, \dots, dx_n)$$

In particolare, per $n = 1$ si scrive anche

$$\int_{\mathbf{R}} g dG = \int_{\mathbf{R}} g(x) G(dx) = \int_{-\infty}^{+\infty} g(x) G(dx)$$

mentre l'integrale esteso a un intervallo $(a, b]$ sarà

$$\int_{(a,b]} g(x) G(dx) = \int_{\mathbf{R}} I_{(a,b]}(x)g(x) G(dx) = \int_a^b g(x) G(dx)$$

Se μ è la misura di Lebesgue, $G(dx)$ sarà sostituito da $d\mathbf{x}$ e, nel caso $n = 1$, $G(dx)$ da dx . Se invece μ è una probabilità \mathbf{P} , la G sarà sostituita da una fdd F e gli integrali su \mathbf{R}^n assumeranno il significato di valori d'attesa.

3.3.2 Cambio di variabili di integrazione

Teorema 3.22. (Cambio di variabili di integrazione): Dato su $(\Omega, \mathcal{F}, \mathbf{P})$ il vett- a $\mathbf{X} = (X_1, \dots, X_n)$ con distribuzione congiunta $\mathbf{P}_{\mathbf{X}}$, data la funzione di Borel $g : (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n)) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$, e posto $Y = g(\mathbf{X})$ (vedi Figura 3.4), risulta

$$\begin{aligned} \mathbf{E}[Y] &= \int_{\Omega} Y(\omega) d\mathbf{P}(\omega) = \mathbf{E}[g(\mathbf{X})] = \int_{\Omega} g(\mathbf{X}(\omega)) d\mathbf{P}(\omega) \\ &= \int_{\mathbf{R}^n} g(\mathbf{x}) \mathbf{P}_{\mathbf{X}}\{d\mathbf{x}\} = \int_{\mathbf{R}^n} g(x_1, \dots, x_n) \mathbf{P}_{\mathbf{X}}\{dx_1, \dots, dx_n\} \end{aligned}$$

Dimostrazione: Omessa⁶. Si osservi dalla Figura 3.4 che in generale, con $n \geq 2$, \mathbf{X} è un vett- a e $g(\mathbf{x}) = g(x_1, \dots, x_n)$ una funzione di n variabili secondo lo schema

$$(\Omega, \mathcal{F}) \xrightarrow{X} (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n)) \xrightarrow{g} (\mathbf{R}, \mathcal{B}(\mathbf{R}))$$

mentre $Y = g(\mathbf{X})$ è una $v-a$. Nel caso più semplice $n = 1$ invece il vett- a \mathbf{X} ha una sola componente X e lo schema del Teorema 3.22 si riduce a

$$(\Omega, \mathcal{F}) \xrightarrow{X} (\mathbf{R}, \mathcal{B}(\mathbf{R})) \xrightarrow{g} (\mathbf{R}, \mathcal{B}(\mathbf{R}))$$

con $Y = g(X)$ □

Siccome la Definizione 3.20 di valore d'attesa si presenta sotto una forma astratta e poco adatta al calcolo, il precedente risultato, che – assieme alle notazioni e alle regole di calcolo che ne conseguono – fa essenzialmente uso di formule di integrazione ordinaria, assume una notevole importanza pratica. Il Teorema 3.22 afferma infatti che si può calcolare il valore d'attesa di $Y = g(\mathbf{X})$ (cioè l'integrale astratto nella misura \mathbf{P} di una funzione da Ω in \mathbf{R}) tramite integrazioni nella misura $\mathbf{P}_{\mathbf{X}}$ di

⁶A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

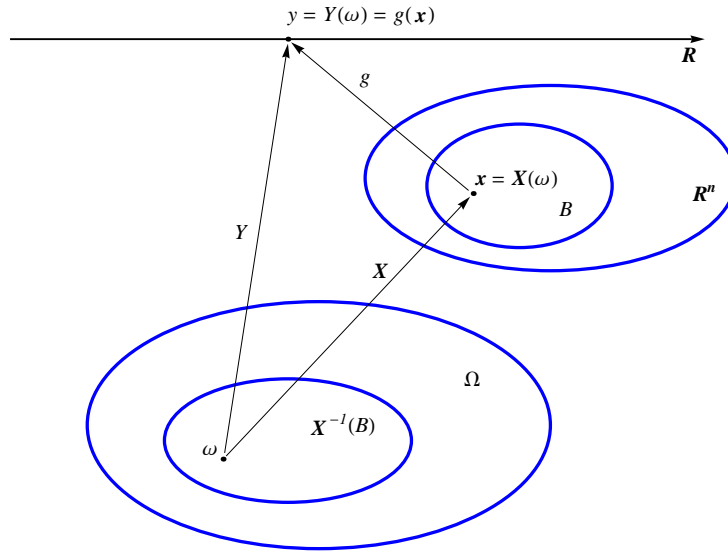


Figura 3.4: Illustrazione grafica del Teorema 3.22.

funzioni di Borel $g(\mathbf{x})$ da \mathbf{R}^n in \mathbf{R} . Siccome le distribuzioni \mathbf{P}_X su $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ sono descritte mediante *fdd* F – o *ddp* f , se esistono – questo risultato permette di fornire alcune familiari regole di calcolo che utilizzano integrazioni ordinarie. A questo scopo si tenga presente che se F_X e f_X sono la *fdd* e la *ddp* di \mathbf{P}_X , si prova preliminarmente (ma noi eviteremo di farlo) che

$$\int_A g(x) F_X(dx) = \int_A g(x) f_X(x) dx \quad \forall A \in \mathcal{B}(\mathbf{R}) \quad (3.12)$$

sicché, se una *ddp* esiste, potremo sempre sostituire $F_X(dx)$ con $f_X(x) dx$ in tutte le integrazioni presentate nelle formule seguenti

Corollario 3.23. *Nel caso $n = 1$, quando il vett-a \mathbf{X} si riduce ad una sola componente X , se \mathbf{P}_X ha una *fdd* $F_X(x)$ discreta, dal Teorema 3.22 avremo*

$$\mathbf{E}[Y] = \mathbf{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) F_X(dx) = \sum_k g(x_k) \mathbf{P}_X\{x_k\} \quad (3.13)$$

e se inoltre $g(x) = x$ (cioè se $Y = X$) otterremo

$$\mathbf{E}[X] = \sum_k x_k \mathbf{P}_X\{x_k\} \quad (3.14)$$

sicché potremo anche scrivere

$$\mathbf{E}[Y] = \sum_\ell y_\ell \mathbf{P}_Y\{y_\ell\} = \sum_k g(x_k) \mathbf{P}_X\{x_k\} \quad (3.15)$$

Se invece $F_X(x)$ è ac con $ddp f_X(x)$, risulterà

$$\mathbf{E}[Y] = \mathbf{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)F_X(dx) = \int_{-\infty}^{+\infty} g(x)f_X(x) dx \quad (3.16)$$

e se $g(x) = x$ si otterrà l'usuale relazione

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} xF_X(dx) = \int_{-\infty}^{+\infty} xf_X(x) dx \quad (3.17)$$

in base alla quale potremo anche scrivere

$$\mathbf{E}[Y] = \int_{-\infty}^{+\infty} yf_Y(y) dy = \int_{-\infty}^{+\infty} g(x)f_X(x) dx \quad (3.18)$$

Nel caso $n > 1$, dette $F_{\mathbf{X}}(\mathbf{x})$ e $f_{\mathbf{X}}(\mathbf{x})$ le fdd e ddp congiunte di \mathbf{X} , e $F_Y(y)$ e $f_Y(y)$ le fdd e ddp di $Y = g(\mathbf{X})$, le (3.16) e (3.18) divengono

$$\begin{aligned} \mathbf{E}[Y] &= \mathbf{E}[g(\mathbf{X})] = \int_{\mathbf{R}^n} yf_Y(y) dy = \int_{\mathbf{R}^n} g(\mathbf{x})f_{\mathbf{X}}(\mathbf{x}) d^n \mathbf{x} \\ &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(x_1, \dots, x_n)f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n \end{aligned} \quad (3.19)$$

Infine probabilità e fdd si calcolano rispettivamente come

$$\mathbf{P}_{\mathbf{X}}\{[a_1, b_1] \times \dots \times [a_n, b_n]\} = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f_{\mathbf{X}}(\mathbf{x}) d^n \mathbf{x} \quad (3.20)$$

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{\mathbf{X}}(\mathbf{y}) d^n \mathbf{y} \quad (3.21)$$

e si ritrovano anche le regole di marginalizzazione

$$f_{X_k}(x_k) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_{k-1} dx_{k+1} \dots dx_n \quad (3.22)$$

Dimostrazione: Trascurando per brevità la verifica completa dei risultati, osserveremo solo che nel caso $n = 1$ con $F_X(x)$ ac e $ddp f_X(x)$ l'enunciato del Teorema 3.22 diventa

$$\begin{aligned} \mathbf{E}[Y] = \mathbf{E}[g(X)] &= \int_{\Omega} Y d\mathbf{P} = \int_{\Omega} g(X) \mathbf{P}\{d\omega\} = \int_{\mathbf{R}} g(x) \mathbf{P}_{\mathbf{X}}\{dx\} \\ &= \int_{\mathbf{R}} g(x) F_X(dx) = \int_{\mathbf{R}} g(x) f_X(x) dx \end{aligned}$$

e fornisce immediatamente (3.16) e (3.17). Inoltre, essendo la (3.17) valida per qualsiasi $v-a$, se introduciamo la $ddp f_Y$ della $v-a$ Y , potremo sicuramente scrivere anche

$$\mathbf{E}[Y] = \int_{-\infty}^{+\infty} yF_Y(dy) = \int_{-\infty}^{+\infty} yf_Y(y) dy \quad (3.23)$$

e quindi confrontando (3.16) con (3.23) si ottiene (3.18): in pratica il valore d'attesa di $Y = g(X)$ può essere calcolato in due modi equivalenti secondo la ddp , f_X o f_Y , che si vuole usare, e l'equazione (3.18) altro non è che l'usuale regola per il cambio di variabile di integrazione $y = g(x)$.

Per ottenere le formule sulle probabilità (sempre per $n = 1$) si considera poi in particolare il caso in cui $g(x)$ è un indicatore $\chi_B(x)$ su $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$

$$g(x) = \chi_B(x) = \begin{cases} 1, & \text{se } x \in B \\ 0, & \text{altrimenti} \end{cases} \quad B \in \mathcal{B}(\mathbf{R})$$

una funzione di Borel che è in relazione con il corrispondente indicatore su (Ω, \mathcal{F}) tramite la relazione

$$\chi_B(X(\omega)) = I_{X^{-1}(B)}(\omega) \quad \forall \omega \in \Omega$$

dato che $\omega \in X^{-1}(B)$ equivale a $X(\omega) \in B$. Dal Teorema 3.22 si ha allora

$$\begin{aligned} \mathbf{P}\{X \in B\} &= \mathbf{P}_X\{B\} = \mathbf{P}\{X^{-1}(B)\} = \mathbf{E} [I_{X^{-1}(B)}] = \int_{\Omega} I_{X^{-1}(B)} d\mathbf{P} \\ &= \int_{\Omega} \chi_B(X) d\mathbf{P} = \int_{\mathbf{R}} \chi_B(x) \mathbf{P}_X\{dx\} = \int_{\mathbf{R}} \chi_B(x) F_X(dx) \\ &= \int_B F_X(dx) = \int_B f_X(x) dx \end{aligned}$$

un risultato già anticipato alla fine della Sezione 3.1.2. In particolare, quando $B = [a, b]$, scriveremo

$$\mathbf{P}_X\{[a, b]\} = \mathbf{P}\{a \leq X \leq b\} = \int_a^b f_X(x) dx \quad (3.24)$$

mentre per $B = (-\infty, x]$ si ha

$$\mathbf{P}_X\{(-\infty, x]\} = \mathbf{P}\{X \leq x\} = F_X(x) = \int_{-\infty}^x f_X(t) dt \quad (3.25)$$

formule che, come altre di questa sezione, sono in generale prese come punti di partenza di trattazioni più elementari. Si noti infine che, quando supponiamo l'esistenza di una ddp f , le fdd delle leggi discusse nel corollario sono evidentemente ac , e quindi continue: pertanto la probabilità assegnata ai singoli punti è rigorosamente nulla. Conseguentemente l'inclusione o l'esclusione degli estremi degli intervalli considerati è perfettamente irrilevante: per questo motivo, ad esempio, il fatto che nella formula (3.24) l'intervallo sia chiuso in ambedue gli estremi è una pura questione di convenienza di notazione \square

Esempio 3.24. Se $X \sim \delta_b$ è una **v-a degenera** il suo valore d'attesa è banalmente

$$\mathbf{E}[X] = b \cdot 1 = b \quad (3.26)$$

mentre se $X \sim \mathfrak{B}(1; p)$ è una **v-a di Bernoulli** avremo

$$\mathbf{E}[X] = 1 \cdot p + 0 \cdot (1 - p) = p \quad (3.27)$$

Se poi $X \sim \mathfrak{B}(n; p)$ è una **v-a binomiale** con un riscaldamento di indici si ha

$$\begin{aligned} \mathbf{E}[X] &= \sum_{k=1}^n k \binom{n}{k} p^k q^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} q^{n-k} \\ &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} = np \end{aligned} \quad (3.28)$$

e infine, in maniera analoga, se $X \sim \mathfrak{P}(\alpha)$ è una **v-a di Poisson** si ha

$$\mathbf{E}[X] = e^{-\alpha} \sum_{k=0}^{\infty} k \frac{\alpha^k}{k!} = e^{-\alpha} \sum_{k=1}^{\infty} \frac{\alpha^k}{(k-1)!} = \alpha e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = \alpha \quad (3.29)$$

Esempio 3.25. La formula (3.17) ci permette ora di calcolare il valore d'attesa delle v-a X dotate di ddp f_X . Per una **v-a uniforme** $X \sim \mathfrak{U}(a, b)$ con ddp (2.14) si ha

$$\mathbf{E}[X] = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{a+b}{2} \quad (3.30)$$

mentre se X è una **v-a Gaussiana** $X \sim \mathfrak{N}(b, a^2)$ con ddp (2.15), posto $y = (x-b)/a$, e tenendo conto dei noti integrali Gaussiani

$$\int_{-\infty}^{\infty} e^{-y^2/2} dy = \sqrt{2\pi} \quad \int_{-\infty}^{\infty} y e^{-y^2/2} dy = 0 \quad \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy = \sqrt{2\pi} \quad (3.31)$$

si ha

$$\begin{aligned} \mathbf{E}[X] &= \frac{1}{a\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-b)^2/2a^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (ay + b) e^{-y^2/2} dy \\ &= \frac{b}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = b \end{aligned} \quad (3.32)$$

Se poi X è una **v-a esponenziale** $X \sim \mathfrak{E}(a)$ con ddp (2.17) il valore d'attesa è

$$\mathbf{E}[X] = \int_0^{+\infty} ax e^{-ax} = \frac{1}{a} [-(1+ax)e^{-ax}]_0^{+\infty} = \frac{1}{a} \quad (3.33)$$

e se è una **v-a di Laplace** $X \sim \mathfrak{L}(a)$ con ddp (2.18) il valore d'attesa è

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} \frac{a}{2} x e^{-a|x|} = 0 \quad (3.34)$$

Sarà importante osservare, invece, che se X è una **v-a di Cauchy** $X \sim \mathfrak{C}(a, b)$ con ddp (2.19) il valore d'attesa non esiste nel senso della Definizione 3.20, cioè né converge, né diverge: semplicemente non è definito. Infatti, usando la funzione di Heaviside (2.13), è facile rendersi conto del fatto che $X^+ = X\vartheta(X)$ e $X^- = -X\vartheta(-X)$. Prendendo allora per semplicità una legge di Cauchy $\mathfrak{C}(a, 0)$ con $b = 0$ in modo che la ddp (2.19) risulti simmetrica con $f_X(-x) = f_X(x)$, si ha

$$\mathbf{E}[X^+] = \mathbf{E}[X^-] = \frac{a}{\pi} \int_0^{+\infty} \frac{x}{a^2 + x^2} dx = \frac{a}{\pi} \left[\frac{1}{2} \ln(a^2 + x^2) \right]_0^{+\infty} = +\infty$$

per cui $\mathbf{E}[X]$ è del tipo $\infty - \infty$, e quindi il valore d'attesa non è definito come integrale di Lebesgue. Non è rilevante invece il fatto che il valor principale

$$\lim_{M \rightarrow +\infty} \frac{a}{\pi} \int_{-M}^{+M} \frac{x}{a^2 + x^2} dx = 0$$

esista finito: il valore d'attesa è un integrale di Lebesgue, e non uno di Riemann, né tanto meno un valor principale. Come vedremo più avanti nella Sezione 4.6, questa differenza è molto più di una sottigliezza matematica e ha delle implicazioni piuttosto profonde.

Proposizione 3.26. *Siano X e Y v-a integrabili definite su $(\Omega, \mathcal{F}, \mathbf{P})$:*

1. $\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]$ con $a, b \in \mathbf{R}$
2. $|\mathbf{E}[X]| \leq \mathbf{E}[|X|]$
3. se $X = 0$, \mathbf{P} -qo, allora $\mathbf{E}[X] = 0$; se poi X è una generica v-a e A un evento

$$\mathbf{E}[XI_A] = \int_A X d\mathbf{P} = 0 \quad \text{se } \mathbf{P}\{A\} = 0$$

4. se $X \leq Y$, \mathbf{P} -qo allora $\mathbf{E}[X] \leq \mathbf{E}[Y]$, e se $X = Y$, \mathbf{P} -qo allora $\mathbf{E}[X] = \mathbf{E}[Y]$
5. se $X \geq 0$ e $\mathbf{E}[X] = 0$, allora $X = 0$, \mathbf{P} -qo, cioè X è degenere δ_0
6. se $\mathbf{E}[XI_A] \leq \mathbf{E}[YI_A]$, $\forall A \in \mathcal{F}$, allora $X \leq Y$, \mathbf{P} -qo, e in particolare se $\mathbf{E}[XI_A] = \mathbf{E}[YI_A]$, $\forall A \in \mathcal{F}$, allora $X = Y$, \mathbf{P} -qo
7. se X e Y sono indipendenti, allora anche XY è integrabile e

$$\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$$

Dimostrazione: Omessa⁷. Questi risultati sono, in una diversa notazione, le solite proprietà dell'integrale di Lebesgue □

⁷N. Cufaro Petroni, **CALCOLO DELLE PROBABILITÀ**, Edizioni dal Sud (Bari, 1996)

Esempio 3.27. Come particolare applicazione osserviamo che dalla proprietà 1. il valore d'attesa $\mathbf{E}[\cdot]$ risulta essere un funzionale lineare, e che questo consente alcune semplificazioni di calcolo. In particolare ricorderemo che per calcolare il valore d'attesa (3.28) di una v -a binomiale S_n con legge $\mathfrak{B}(n; p)$ abbiamo utilizzato nell'Esempio 3.24 la definizione elementare (3.9): sebbene in questo modo il risultato sia stato ottenuto direttamente e con poco sforzo, è opportuno osservare che un metodo ancora più rapido consiste nel ricordare l'equazione (3.8) in base alla quale ogni v -a binomiale coincide in distribuzione con la somma di n v -a X_1, \dots, X_n iid di Bernoulli $\mathfrak{B}(1; p)$. Siccome il valore d'attesa dipende solo dalla distribuzione di una v -a, e siccome da (3.27) sappiamo che $\mathbf{E}[X_j] = p$, dalla linearità del valore d'attesa abbiamo allora immediatamente che

$$\mathbf{E}[S_n] = \mathbf{E}\left[\sum_{j=1}^n X_j\right] = \sum_{j=1}^n \mathbf{E}[X_j] = np \quad (3.35)$$

risultato che ovviamente coincide con (3.28)

I valori d'attesa soddisfano ovviamente anche alcune importanti *disuguaglianze* che sono tipiche della teoria dell'integrazione: esse sono ricapitolate nella loro formulazione probabilistica nell'Appendice B. Nella successiva Sezione 3.3.3 introdurremo invece separatamente la Disuguaglianza di Chebyshev (Proposizione 3.41) per l'importanza che essa rivestirà nella discussione della Legge dei Grandi Numeri

3.3.3 Varianza e covarianza

Il valore d'attesa di una v -a X è un numero che individua per così dire il *baricentro* della sua distribuzione. Esso però da un lato non è l'unico indicatore di centralità, e dall'altro non esaurisce tutta l'informazione contenuta nella legge di X . Ricordiamo infatti che il valore d'attesa può non essere neanche uno dei possibili valori di X : ad esempio il valore d'attesa p di una v -a di Bernoulli $\mathfrak{B}(1; p)$ (che assume solo i valori 0 e 1) non è in generale né 0, né 1. Ci sono inoltre anche altri parametri numerici che possono altrettanto legittimamente individuare il *centro* di una distribuzione: ad esempio la **moda** (o le mode) che, per v -a dotate di *ddp*, è semplicemente il valore (o i valori) di x in cui $f_X(x)$ ha un massimo. Si vede facilmente che – tranne casi molto particolari – il valore d'attesa non coincide con la moda della distribuzione: in particolare quando la distribuzione non è simmetrica essa ha il suo massimo ed il suo baricentro (se esso esiste) in posizioni diverse.

È dunque importante esaminare anche altri parametri numerici che contribuiscano a caratterizzare la v -a X e la sua legge. Uno particolarmente rilevante sarà costituito da una qualche misura della *dispersione* con cui i valori della v -a si distribuiscono attorno al valore d'attesa. Le *deviazioni* dei valori di X rispetto a $\mathbf{E}[X]$ sono descritte dallo scarto $X - \mathbf{E}[X]$ e si potrebbe pensare a prima vista che una misura significativa della dispersione potrebbe essere data dal suo valore d'attesa

$\mathbf{E}[X - \mathbf{E}[X]]$. Si vede subito però che il suo valore è identicamente nullo

$$\mathbf{E}[X - \mathbf{E}[X]] = \mathbf{E}[X] - \mathbf{E}[X] = 0$$

e che quindi esso non può essere assunto come una buona misura della dispersione di X . Siccome è intuitivamente chiaro che questa situazione dipende essenzialmente dal fatto che $X - \mathbf{E}[X]$ assume valori positivi e negativi, si usa prendere in considerazione piuttosto lo scarto quadratico $(X - \mathbf{E}[X])^2$ che invece assume solo valori positivi e che avrà quindi un valore d'attesa in generale diverso da zero

Definizione 3.28. *Se X, Y sono v-a integrabili, e $\mathbf{X} = (X_1, \dots, X_n)$ è un vett-a con componenti integrabili*

- si chiama **varianza** di X il numero non negativo (finito o infinito)

$$\mathbf{V}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \sigma_X^2$$

e **deviazione standard** la sua radice quadrata positiva $\sigma_X = +\sqrt{\mathbf{V}[X]}$

- si chiama **covarianza** di X e Y la quantità (se esiste)

$$\mathbf{cov}[X, Y] = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \quad (\mathbf{cov}[X, X] = \mathbf{V}[X])$$

e (se $\mathbf{V}[X] \neq 0, \mathbf{V}[Y] \neq 0$) **coefficiente di correlazione** la quantità

$$\rho[X, Y] = \frac{\mathbf{cov}[X, Y]}{\sqrt{\mathbf{V}[X]} \sqrt{\mathbf{V}[Y]}}$$

- se $\mathbf{cov}[X, Y] = 0$, ovvero $\rho[X, Y] = 0$, X e Y si dicono **v-a non correlate**
- si chiama **matrice delle covarianze** (rispettivamente **delle correlazioni**) del vett-a \mathbf{X} la matrice $n \times n$, $\mathbb{R} = \|r_{ij}\|$ (rispettivamente $\mathbb{P} = \|\rho_{ij}\|$) i cui elementi sono $r_{ij} = \mathbf{cov}[X_i, X_j]$ (rispettivamente $\rho_{ij} = \rho[X_i, X_j]$).

Proposizione 3.29. *Se X ed Y sono v-a, e a e b sono numeri si ha:*

1. $\mathbf{cov}[X, Y] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$, e in particolare $\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$
2. se $\mathbf{V}[X] = 0$, allora $X = \mathbf{E}[X]$, **P-qo**
3. $\mathbf{V}[a + bX] = b^2 \mathbf{V}[X]$
4. $\mathbf{V}[X + Y] = \mathbf{V}[X] + \mathbf{V}[Y] + 2\mathbf{cov}[X, Y]$
5. se X ed Y sono indipendenti, allora sono anche non correlate

Dimostrazione:

1. Sviluppando il prodotto nella definizione si ha infatti:

$$\mathbf{cov}[X, Y] = \mathbf{E}[XY] - 2\mathbf{E}[X]\mathbf{E}[Y] + \mathbf{E}[X]\mathbf{E}[Y] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$$

Ponendo poi $X = Y$ si ottiene anche il risultato sulla varianza

2. Siccome $(X - \mathbf{E}[X])^2 \geq 0$, il risultato segue direttamente dalla 5 di Proposizione 3.26: infatti se $\mathbf{V}[X] = 0$ allora $X - \mathbf{E}[X] = 0$, \mathbf{P} -qo (cioè è una v -a degenera δ_0) e quindi $X = \mathbf{E}[X]$, \mathbf{P} -qo
3. Per la linearità del valore d'attesa si ha $\mathbf{E}[a + bX] = a + b\mathbf{E}[X]$, e quindi

$$\mathbf{V}[a + bX] = \mathbf{E}[(a + bX - a - b\mathbf{E}[X])^2] = b^2\mathbf{E}[(X - \mathbf{E}[X])^2] = b^2\mathbf{V}[X]$$

4. Si ha infatti

$$\begin{aligned} \mathbf{V}[X + Y] &= \mathbf{E}[(X + Y - \mathbf{E}[X + Y])^2] = \mathbf{E}[(X - \mathbf{E}[X] + Y - \mathbf{E}[Y])^2] \\ &= \mathbf{V}[X] + \mathbf{V}[Y] + 2\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{V}[X] + \mathbf{V}[Y] + 2\mathbf{cov}[X, Y] \end{aligned}$$

Si noti che, diversamente da quel che avviene per il valore d'attesa \mathbf{E} , la varianza \mathbf{V} non è un funzionale lineare. In particolare abbiamo appena mostrato che in generale $\mathbf{V}[X + Y]$ non coincide con la somma $\mathbf{V}[X] + \mathbf{V}[Y]$; ciò avviene solo se le due v -a X e Y sono anche **non correlate**: solo in questo caso infatti si ha

$$\mathbf{V}[X + Y] = \mathbf{V}[X] + \mathbf{V}[Y] \quad \text{se } \mathbf{cov}[X, Y] = 0$$

5. A causa dell'indipendenza di X ed Y dalla definizione si ha

$$\mathbf{cov}[X, Y] = \mathbf{E}[X - \mathbf{E}[X]]\mathbf{E}[Y - \mathbf{E}[Y]] = 0$$

cioè X e Y sono anche non correlate □

Proposizione 3.30. *Comunque date due v -a X e Y , risulta sempre*

$$|\rho[X, Y]| \leq 1$$

Inoltre si ha $|\rho[X, Y]| = 1$ se e solo se esistono due numeri $a \neq 0$ e b tali che $Y = aX + b$, \mathbf{P} -qo; in particolare risulta $a > 0$ se $\rho[X, Y] = +1$, e $a < 0$ se $\rho[X, Y] = -1$

Dimostrazione: Omessa⁸. Queste importanti proprietà del coefficiente di correlazione ρ non si estendono anche alla covarianza che può infatti assumere qualunque

⁸**P.L. Meyer**, INTRODUCTORY PROBABILITY AND STATISTICAL APPLICATIONS, Addison-Wesley (Reading, 1980)

valore reale positivo o negativo. La precedente proposizione mostra in particolare che ρ è una misura della linearità della dipendenza fra X e Y : infatti quando ρ assume i suoi valori estremi ± 1 , Y dipende linearmente da X , addirittura in maniera funzionale. \square

Esempio 3.31. Indipendenza e non correlazione: *Il punto 5 della Proposizione 3.29 afferma che se due v-a sono indipendenti esse sono anche non correlate. È importante però osservare che in generale il viceversa non è vero: due v-a possono non essere indipendenti, pur essendo non correlate. Ad esempio sia α una v-a che assume i tre valori $0, \frac{\pi}{2}, \pi$ con uguali probabilità $\frac{1}{3}$, e siano poi definite le v-a $X = \sin \alpha$ e $Y = \cos \alpha$. Si verifica subito che X ed Y sono non correlate: infatti*

$$\begin{aligned} \mathbf{E}[X] &= \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0 = \frac{1}{3} & \mathbf{E}[Y] &= \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot (-1) = 0 \\ \mathbf{E}[XY] &= \frac{1}{3} \cdot (1 \cdot 0) + \frac{1}{3} \cdot (0 \cdot 1) + \frac{1}{3} \cdot (-1 \cdot 0) = 0 \end{aligned}$$

per cui

$$\mathbf{E}[X] \mathbf{E}[Y] = 0 = \mathbf{E}[XY]$$

Ma si verifica altrettanto facilmente che esse non sono indipendenti: infatti

$$\mathbf{P}\{X = 1\} = \frac{1}{3} \quad \mathbf{P}\{Y = 1\} = \frac{1}{3} \quad \mathbf{P}\{X = 1, Y = 1\} = 0$$

e quindi

$$\mathbf{P}\{X = 1, Y = 1\} = 0 \neq \frac{1}{9} = \mathbf{P}\{X = 1\} \mathbf{P}\{Y = 1\}$$

Proposizione 3.32. *Condizione necessaria e sufficiente affinché una matrice $n \times n$, \mathbb{R} sia matrice delle covarianze di un vett-a $\mathbf{X} = (X_1, \dots, X_n)$, è che tale matrice sia simmetrica e definita non negativa; ovvero, equivalentemente, che esista una matrice $n \times n$, \mathbb{C} tale che $\mathbb{R} = \mathbb{C}\mathbb{C}^T$, dove \mathbb{C}^T è la trasposta della matrice \mathbb{C} .*

Dimostrazione: Si vede facilmente dalla definizione che la matrice delle covarianze \mathbb{R} di un vett-a \mathbf{X} è simmetrica ($r_{ij} = r_{ji}$), e si verifica anche subito che essa è definita non negativa: infatti, comunque scelti n numeri reali $\lambda_1, \dots, \lambda_n$, si ha

$$\begin{aligned} \sum_{i,j=1}^n r_{ij} \lambda_i \lambda_j &= \sum_{i,j=1}^n \lambda_i \lambda_j \mathbf{E}[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])] \\ &= \mathbf{E} \left[\left(\sum_{i=1}^n \lambda_i (X_i - \mathbf{E}[X_i]) \right)^2 \right] \geq 0 \end{aligned}$$

La proposizione afferma che queste proprietà sono caratteristiche delle matrici delle covarianze, nel senso che è vero anche l'inverso: ogni matrice simmetrica e definita

non negativa è una buona matrice delle covarianze di qualche vett-a \mathbf{X} . Omettiamo⁹ la dimostrazione di questa affermazione e del resto della proposizione \square

Esempio 3.33. *Esamineremo innanzitutto alcuni esempi di varianze di leggi discrete: se $X \sim \delta_b$ è una v-a degenera ovviamente si ha $\mathbf{V}[X] = 0$; se invece $X \sim \mathfrak{B}(1; p)$ è una v-a di Bernoulli si vede subito che, essendo $\mathbf{E}[X] = p$, si ha*

$$\mathbf{V}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] = (1 - p)^2 p + (0 - p)^2 (1 - p) = p(1 - p)$$

Se poi $S_n \sim \mathfrak{B}(n; p)$ è una v-a binomiale, siccome da (3.8) sappiamo che in distribuzione essa è la somma di n v-a X_1, \dots, X_n di Bernoulli iid $\mathfrak{B}(1; p)$, a causa della indipendenza delle X_j avremo

$$\mathbf{V}[S_n] = \mathbf{V}\left[\sum_{j=1}^n X_j\right] = \sum_{j=1}^n \mathbf{V}[X_j] = \sum_{j=1}^n p(1 - p) = np(1 - p) \quad (3.36)$$

Se $X \sim \mathfrak{P}(\alpha)$ è una v-a di Poisson, per la quale da (6.12) già sappiamo che $\mathbf{E}[X] = \alpha$, conviene partire dal calcolo di

$$\mathbf{E}[X(X - 1)] = \sum_{k=0}^{\infty} k(k - 1) \frac{e^{-\alpha} \alpha^k}{k!} = e^{-\alpha} \sum_{k=2}^{\infty} \frac{\alpha^k}{(k - 2)!} = \alpha^2 e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = \alpha^2$$

da cui si ricava

$$\mathbf{E}[X^2] = \alpha^2 + \mathbf{E}[X] = \alpha^2 + \alpha$$

e in definitiva

$$\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \alpha^2 + \alpha - \alpha^2 = \alpha$$

Si noti la proprietà delle leggi di Poisson per le quali $\mathbf{E}[X] = \mathbf{V}[X] = \alpha$. Passando poi alle leggi dotate di ddp abbiamo per una v-a uniforme $X \sim \mathfrak{U}(a, b)$

$$\mathbf{E}[X^2] = \int_a^b \frac{x^2}{b - a} dx = \frac{1}{b - a} \left[\frac{x^3}{3} \right]_a^b = \frac{b^3 - a^3}{3(b - a)}$$

e tenendo conto di (3.30) si ottiene

$$\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{b^3 - a^3}{3(b - a)} - \frac{(a + b)^2}{4} = \frac{(b - a)^2}{12} \quad (3.37)$$

Analogamente per una v-a esponenziale $X \sim \mathfrak{E}(a)$ si ha con $y = ax$

$$\mathbf{E}[X^2] = \int_0^{+\infty} x^2 a e^{-ax} dx = \frac{1}{a^2} \int_0^{+\infty} y^2 e^{-y} dy = \frac{1}{a^2} [-(2 + 2y + y^2)e^{-y}]_0^{+\infty} = \frac{2}{a^2}$$

⁹N. Cufaro Petroni, CALCOLO DELLE PROBABILITÀ, Edizioni dal Sud (Bari, 1996)

e quindi tenendo conto di (3.33)

$$\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{2}{a^2} - \frac{1}{a^2} = \frac{1}{a^2} \quad (3.38)$$

In modo simile si dimostra che per una **v-a di Laplace** $X \sim \mathfrak{L}(a)$ si ha

$$\mathbf{V}[X] = \frac{2}{a^2} \quad (3.39)$$

Se infine $X \sim \mathfrak{N}(b, a^2)$ è una **v-a Gaussiana**, tenendo conto degli integrali Gaussiani (3.31) si ha con $y = (x - b)/a$

$$\mathbf{E}[X^2] = \int_{-\infty}^{+\infty} x^2 \frac{e^{-(x-b)^2/2a^2}}{a\sqrt{2\pi}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (ay + b)^2 e^{-y^2/2} dy = a^2 + b^2$$

e quindi da (3.32)

$$\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = (a^2 + b^2) - b^2 = a^2 \quad (3.40)$$

Per una **v-a di Cauchy** $X \sim \mathfrak{C}(a, b)$ invece la varianza non può essere definita perché il suo valore d'attesa non esiste come abbiamo visto alla fine della Sezione 3.3.2, e comunque anche il suo secondo momento è divergente come può essere visto più facilmente nel caso $b = 0$, con $y = x/a$

$$\mathbf{E}[X^2] = \frac{a}{\pi} \int_{-\infty}^{+\infty} \frac{x^2}{a^2 + x^2} dx = \frac{a^2}{\pi} \int_{-\infty}^{+\infty} \frac{y^2}{1 + y^2} dy = \frac{a^2}{\pi} [y - \arctan y]_{-\infty}^{+\infty} = +\infty$$

Esempio 3.34. Vettori Gaussiani bivariati: Se $\mathbf{X} = (X, Y) \sim \mathfrak{N}(\mathbf{b}, \mathbb{A})$ è un vett-a Gaussiano (normale) bivariato sappiamo che la sua ddp congiunta (2.24) è caratterizzata da cinque parametri: le due componenti di $\mathbf{b} = (b_1, b_2) \in \mathbf{R}^2$, e i tre numeri $a_1 > 0$, $a_2 > 0$, $|r| \leq 1$ associati agli elementi della matrice simmetrica e definita positiva \mathbb{A} dalle relazioni

$$a_k = \sqrt{a_{kk}} \quad r = \frac{a_{12}}{\sqrt{a_{11}a_{22}}} = \frac{a_{21}}{\sqrt{a_{11}a_{22}}}$$

Come già indicato nell'Esempio 2.31, le marginali sono a loro volta delle v-a normali $\mathfrak{N}(b_k, a_k^2)$ con ddp

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x, y) dy = \frac{1}{a_1\sqrt{2\pi}} e^{-(x-b_1)^2/2a_1^2}, \\ f_Y(y) &= \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x, y) dx = \frac{1}{a_2\sqrt{2\pi}} e^{-(y-b_2)^2/2a_2^2}, \end{aligned}$$

Un calcolo diretto mostrerebbe ora che l'interpretazione probabilistica dei cinque parametri che compaiono in una $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ bivariata è

$$\begin{aligned} b_1 &= \mathbf{E}[X] & b_2 &= \mathbf{E}[Y] \\ a_1^2 &= a_{11} = \mathbf{V}[X] & a_2^2 &= a_{22} = \mathbf{V}[Y] & r &= \frac{a_{12}}{\sqrt{a_{11}a_{22}}} = \frac{a_{21}}{\sqrt{a_{11}a_{22}}} = \rho[X, Y] \end{aligned}$$

e pertanto il vettore delle medie \mathbf{b} e la matrice delle covarianze \mathbb{A} sono

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \mathbf{E}[X] \\ \mathbf{E}[Y] \end{pmatrix}$$

$$\mathbb{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{V}[X] & \mathbf{cov}[X, Y] \\ \mathbf{cov}[X, Y] & \mathbf{V}[Y] \end{pmatrix} = \begin{pmatrix} a_1^2 & a_1 a_2 r \\ a_1 a_2 r & a_2^2 \end{pmatrix}$$

Abbiamo già messo in evidenza che due v-a X e Y indipendenti sono anche non correlate, ma che in generale il viceversa non è vero. È importante sottolineare allora che invece, se le componenti X_k di un vett-a \mathbf{X} congiuntamente gaussiano sono non correlate, esse risultano anche indipendenti. In altre parole: **le componenti di un vett-a $\mathbf{X} \sim \mathcal{N}(\mathbf{b}, \mathbb{A})$ Gaussiano multivariato sono indipendenti se e solo se esse sono non correlate.** Che ciò sia vero si deduce – ad esempio nel caso bivariato – dal fatto che se le componenti X e Y sono non correlate si ha $r = \rho[X, Y] = 0$, e in tal caso la ddp congiunta (2.24) si riduce a

$$f_X(x, y) = \frac{1}{2\pi a_1 a_2} e^{-(x-b_1)^2/a_1^2} e^{-(y-b_2)^2/a_2^2}$$

e si vede subito che $f_X(x, y) = f_X(x) \cdot f_Y(y)$, sicché in base al Teorema 3.15 X e Y risultano anche indipendenti.

Proposizione 3.35. Disuguaglianza di Chebyshev: Se X è una v-a integrabile e non negativa si ha

$$\mathbf{P}\{X \geq \epsilon\} \leq \frac{\mathbf{E}[X]}{\epsilon} \quad \forall \epsilon > 0 \quad (3.41)$$

Dimostrazione: Il risultato discende immediatamente dalle relazioni

$$\mathbf{E}[X] \geq \mathbf{E}[X I_{\{X \geq \epsilon\}}] \geq \epsilon \mathbf{E}[I_{\{X \geq \epsilon\}}] = \epsilon \mathbf{P}\{X \geq \epsilon\}$$

dove $\epsilon > 0$ è ovviamente arbitrario. □

Corollario 3.36. Se X è una v-a di quadrato integrabile, per ogni $\epsilon > 0$ risulta

$$\mathbf{P}\{|X| \geq \epsilon\} = \mathbf{P}\{X^2 \geq \epsilon^2\} \leq \frac{\mathbf{E}[X^2]}{\epsilon^2}$$

$$\mathbf{P}\{|X - \mathbf{E}[X]| \geq \epsilon\} \leq \frac{\mathbf{E}[(X - \mathbf{E}[X])^2]}{\epsilon^2} = \frac{\mathbf{V}[X]}{\epsilon^2} \quad (3.42)$$

Dimostrazione: Basta applicare la (3.41) alle particolari v-a proposte □

3.4 Condizionamento

3.4.1 Distribuzioni condizionate

Nella Sezione 1.4 la probabilità condizionata è stata definita solo nei casi in cui la probabilità dell'evento condizionante è diversa da zero. Questa limitazione diventa problematica quando l'evento condizionante è $\{Y = y\}$, con Y *v-a ac*: è noto infatti che in questo caso $\mathbf{P}\{Y = y\} = 0$. Sarà quindi necessario assegnare una definizione coerente e una notazione adeguata per questo tipo di condizionamento

Definizione 3.37. *Se X e Y sono due v-a con fdd congiunta $F_{XY}(x, y)$ derivabile in y , e se Y è ac con ddp $f_Y(y)$, chiameremo **fdd di X condizionata dall'evento $\{Y = y\}$** la*

$$F_{X|Y}(x|y) = F_X(x|Y = y) \equiv \frac{\partial_y F_{XY}(x, y)}{f_Y(y)} \quad (3.43)$$

per tutte le y tali che $f_Y(y) \neq 0$ (e quindi \mathbf{P}_Y -q.o.); per le y in cui $f_Y(y) = 0$ (un insieme di misura \mathbf{P}_Y nulla) la $F_X(x|Y = y)$ assume invece un valore arbitrario: ad esempio zero. Se poi anche X è ac e se la ddp congiunta è $f_{XY}(x, y)$ **la ddp di X condizionata dall'evento $\{Y = y\}$ sarà**

$$f_{X|Y}(x|y) = f_X(x|Y = y) \equiv \frac{f_{XY}(x, y)}{f_Y(y)} \quad (3.44)$$

per le y tali che $f_Y(y) \neq 0$, e zero per le y in cui $f_Y(y) = 0$

Per giustificare in maniera intuitiva queste definizioni, dette $F_{XY}(x, y)$, $F_X(x)$ e $F_Y(y)$ rispettivamente le fdd congiunta e marginali di X e Y , e $f_Y(y) = F'_Y(y)$ la ddp di Y , e supponendo che la $F_{XY}(x, y)$ sia derivabile rispetto a y , si considera inizialmente l'evento condizionante *modificato* nella forma $\{y < Y \leq y + \Delta y\}$, in modo che abbia una probabilità diversa da zero: la Definizione 3.37 si ritrova poi passando al limite per $\Delta y \rightarrow 0$. Infatti dalla definizione elementare di probabilità condizionata si ha

$$\begin{aligned} F_X(x|y < Y \leq y + \Delta y) &= \mathbf{P}\{X \leq x | y < Y \leq y + \Delta y\} \\ &= \frac{\mathbf{P}\{X \leq x, y < Y \leq y + \Delta y\}}{\mathbf{P}\{y < Y \leq y + \Delta y\}} \\ &= \frac{F_{XY}(x, y + \Delta y) - F_{XY}(x, y)}{F_Y(y + \Delta y) - F_Y(y)} = \frac{\frac{F_{XY}(x, y + \Delta y) - F_{XY}(x, y)}{\Delta y}}{\frac{F_Y(y + \Delta y) - F_Y(y)}{\Delta y}} \end{aligned}$$

e quindi al limite per $\Delta y \rightarrow 0$ si ritrova la (3.43)

$$F_X(x | Y = y) \equiv \lim_{\Delta y \rightarrow 0} F_X(x | y < Y \leq y + \Delta y) = \frac{\partial_y F_{XY}(x, y)}{F'_Y(y)} = \frac{\partial_y F_{XY}(x, y)}{f_Y(y)}$$

Se poi supponiamo che anche X sia dotata di *ddp* $f_X(x)$, e che $f_{XY}(x, y)$ sia la *ddp* congiunta, con un'ulteriore derivazione di (3.43) rispetto a x si ottiene facilmente la *ddp* condizionata (3.44). Le (3.43) e (3.44) definiscono le distribuzioni condizionate per tutti i valore di y per i quali $f_Y(y) > 0$, e quindi \mathbf{P}_Y -*qo*. Nei punti in cui invece $f_Y(y) = 0$, il valore di $F_X(x | Y = y)$ (o della *ddp*) può essere scelto arbitrariamente (ad esempio uguale a zero) visto che esso comunque non può influenzare il risultato dei calcoli. Si noti inoltre che, se X ed Y sono **indipendenti**, dal Teorema 3.15 segue facilmente che

$$f_{X|Y}(x|y) = f_X(x) \quad (3.45)$$

Formule analoghe valgono ovviamente anche nel caso di *v-a* discrete, ma in questo caso le *ddp* condizionate sono sostituite da probabilità condizionate sulla base delle definizioni elementari della Sezione 1.4.

Proposizione 3.38. *Se X, Y sono due v-a con *ddp* congiunta $f_{XY}(x, y)$, allora*

$$\mathbf{P}_X\{A|Y = y\} = \int_A f_{X|Y}(x|y) dx = \frac{1}{f_Y(y)} \int_A f_{XY}(x, y) dx \quad (3.46)$$

$$\mathbf{P}_{XY}\{A \times B\} = \int_B \mathbf{P}_X\{A|Y = y\} f_Y(y) dy \quad (3.47)$$

$$\mathbf{P}_X\{A\} = \int_{-\infty}^{+\infty} \mathbf{P}_X\{A|Y = y\} f_Y(y) dy \quad (3.48)$$

Dimostrazione: Dalla (3.44) si ottiene innanzitutto la (3.46)

$$\mathbf{P}_X\{A|Y = y\} = \mathbf{P}\{X \in A|Y = y\} = \int_A f_{X|Y}(x|y) dx = \frac{1}{f_Y(y)} \int_A f_{XY}(x, y) dx$$

e poi da questa anche le formule (3.47) e (3.48)

$$\begin{aligned} \mathbf{P}_{XY}\{A \times B\} &= \mathbf{P}\{X \in A, Y \in B\} = \int_{A \times B} f_{XY}(x, y) dx dy \\ &= \int_A dx \int_B dy f_{X|Y}(x|y) f_Y(y) \\ &= \int_B \mathbf{P}_X\{A|Y = y\} f_Y(y) dy \\ \mathbf{P}_X\{A\} &= \mathbf{P}\{X \in A\} = \int_{-\infty}^{+\infty} \mathbf{P}_X\{A|Y = y\} f_Y(y) dy \end{aligned}$$

La (3.48) mostra come si ritrova \mathbf{P}_X dalla sua distribuzione condizionata (3.46) \square

Proposizione 3.39. *Se X, Y sono due v-a con ddp congiunta $f_{XY}(x, y)$, allora*

$$f_{XY}(x, y|a \leq Y \leq b) = \frac{f_{XY}(x, y)}{\int_a^b f_Y(y') dy'} \chi_{[a,b]}(y) \quad (3.49)$$

$$f_X(x|a \leq Y \leq b) = \frac{\int_a^b f_{XY}(x, y') dy'}{\int_a^b f_Y(y') dy'} \quad (3.50)$$

$$f_Y(y|a \leq Y \leq b) = \frac{f_Y(y)}{\int_a^b f_Y(y') dy'} \chi_{[a,b]}(y) \quad (3.51)$$

dove abbiamo introdotto l'indicatore dell'evento B in $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$

$$\chi_B(x) = \begin{cases} 1, & \text{se } x \in B \\ 0, & \text{altrimenti} \end{cases} \quad B \in \mathcal{B}(\mathbf{R})$$

Dimostrazione: Infatti dalle definizioni si ha innanzitutto che

$$\begin{aligned} F_{XY}(x, y|a \leq Y \leq b) &= \frac{\mathbf{P}\{X \leq x, Y \leq y, a \leq Y \leq b\}}{\mathbf{P}\{a \leq Y \leq b\}} \\ &= \begin{cases} 0 & \text{se } y \leq a \\ \frac{\mathbf{P}\{X \leq x, a \leq Y \leq y\}}{\mathbf{P}\{a \leq Y \leq b\}} = \frac{F_{XY}(x, y) - F_{XY}(x, a)}{F_Y(b) - F_Y(a)} & \text{se } a \leq y \leq b \\ \frac{\mathbf{P}\{X \leq x, a \leq Y \leq b\}}{\mathbf{P}\{a \leq Y \leq b\}} = \frac{F_{XY}(x, b) - F_{XY}(x, a)}{F_Y(b) - F_Y(a)} & \text{se } b \leq y \end{cases} \end{aligned}$$

e quindi la (3.49) segue ricordando che

$$\begin{aligned} f_{XY}(x, y|a \leq Y \leq b) &= \partial_x \partial_y F_{XY}(x, y|a \leq Y \leq b) \\ F_Y(b) - F_Y(a) &= \int_a^b f_Y(y') dy' \end{aligned}$$

Dalla (3.49) poi si ricavano (3.50) e (3.51) con una semplice marginalizzazione \square

Proposizione 3.40. *Se $\mathbf{X} = (X_1, X_2) \sim \mathfrak{N}(\mathbf{b}, \mathbb{A})$ è un vett-a Gaussiano bivariato con ddp (2.24), la legge di X_2 condizionata da $X_1 = x_1$ è ancora Gaussiana con*

$$\mathfrak{N}\left(b_2 + r(x_1 - b_1)\frac{a_2}{a_1}, (1 - r^2)a_2^2\right) \quad (3.52)$$

Dimostrazione: Sappiamo che la ddp bivariata di \mathbf{X} ha la forma (2.24), e che le sue due marginali sono $\mathfrak{N}(b_k, a_k^2)$ con ddp (2.31). Un'applicazione diretta della (3.44)

conduce allora alla seguente *ddp* condizionata

$$\begin{aligned}
 f_{X_2|X_1}(x_2|x_1) &= \frac{e^{-\frac{1}{2(1-r^2)}\left[\frac{(x_1-b_1)^2}{a_1^2}-2r\frac{(x_1-b_1)(x_2-b_2)}{a_1a_2}+\frac{(x_2-b_2)^2}{a_2^2}\right]}}{2\pi a_1 a_2 \sqrt{1-r^2}} a_1 \sqrt{2\pi} e^{-\frac{(x_1-b_1)^2}{2a_1^2}} \\
 &= \frac{e^{-\frac{1}{2(1-r^2)}\left[r^2\frac{(x_1-b_1)^2}{a_1^2}-2r\frac{(x_1-b_1)(x_2-b_2)}{a_1a_2}+\frac{(x_2-b_2)^2}{a_2^2}\right]}}{\sqrt{2\pi a_2^2(1-r^2)}} \\
 &= \frac{e^{-\frac{1}{2a_2^2(1-r^2)}\left[(x_2-b_2)-r(x_1-b_1)\frac{a_2}{a_1}\right]^2}}{\sqrt{2\pi a_2^2(1-r^2)}}
 \end{aligned}$$

e quindi al risultato (3.52) □

Fin qui abbiamo trattato solo il caso del condizionamento reciproco fra due *v-a*, ma questa limitazione è dovuta solo al desiderio di semplificare la notazione. Per completezza, quindi, ricorderemo che dati due *vett-a* $\mathbf{X} = (X_1, \dots, X_n)$ e $\mathbf{Y} = (Y_1, \dots, Y_m)$ dotati di *ddp* congiunta $f_{\mathbf{XY}}(x_1, \dots, x_n, y_1, \dots, y_m)$, con una procedura identica a quella adottata nel caso di due sole *v-a* è possibile introdurre anche la *ddp* di \mathbf{X} condizionata dall'evento $\{Y_1 = y_1, \dots, Y_m = y_m\}$ come

$$f_{\mathbf{X}|\mathbf{Y}}(x_1, \dots, x_n | y_1, \dots, y_m) = \frac{f_{\mathbf{XY}}(x_1, \dots, x_n, y_1, \dots, y_m)}{f_{\mathbf{Y}}(y_1, \dots, y_m)} \quad (3.53)$$

3.4.2 Attese condizionate

Sappiamo che se B è un evento di probabilità diversa da zero la probabilità condizionata $\mathbf{P}\{\cdot | B\}$ si definisce in maniera elementare (Sezione 1.4) ed è a tutti gli effetti una misura di probabilità su (Ω, \mathcal{F}) . Se allora X è una *v-a* essa sarà anche dotata in maniera naturale di una distribuzione condizionata $\mathbf{P}_X\{\cdot | B\}$, di una *fdd* condizionata $F_X(\cdot | B)$ ed eventualmente di una *ddp* condizionata $f_X(\cdot | B)$. Se invece l'evento condizionante è $\{Y = y\}$ dove Y è una *v-a ac* abbiamo visto nella sezione precedente in che maniera è possibile introdurre le analoghe quantità $\mathbf{P}_X\{\cdot | Y = y\}$, $F_X(\cdot | Y = y)$ e $f_X(\cdot | Y = y)$. È pertanto possibile ripercorrere le procedure sviluppate nella Sezione 3.3.1 utilizzando queste nuove misure per definire le attese condizionate nei vari casi. Noi supporremo nel seguito che le nostre *v-a* siano sempre dotate di *ddp*

Definizione 3.41. Diremo *valore d'attesa condizionato di X rispetto ad un evento B di misura non nulla il numero*

$$\mathbf{E}[X|B] = \int_{-\infty}^{+\infty} x f_X(x|B) dx \quad (3.54)$$

e *valore d'attesa condizionato di X rispetto a $\{Y = y\}$ la funzione*

$$\mathbf{E}[X|Y = y] = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y) dx \quad (3.55)$$

o più in generale valore d'attesa condizionato di $g(X)$ la funzione

$$\mathbf{E} [g(X)|Y = y] = \int_{-\infty}^{+\infty} g(x) f_{X|Y}(x|y) dx \quad (3.56)$$

Chiameremo infine **valore d'attesa condizionato di X rispetto ad una v -a Y** la nuova v -a $\mathbf{E} [X|Y]$ definita da

$$\mathbf{E} [X|Y] (\omega) \equiv m (Y(\omega)) \quad (3.57)$$

dove abbiamo posto

$$m(y) \equiv \mathbf{E} [X|Y = y] \quad (3.58)$$

Sarà bene ribadire che nell'ultima definizione (3.57) di valore d'attesa condizionato rispetto ad una v -a si parte dall'osservazione che il valore d'attesa (3.55) può essere considerato come una funzione $m(y)$ del valore y della v -a condizionante, e che è quindi anche possibile introdurre – sulla base del Teorema 3.6 – una nuova v -a $m(Y)$ per la quale viene generalmente adottata la nuova notazione sintetica $\mathbf{E} [X|Y]$. È importante ricordare a questo proposito che questa espressione denoterà da ora in poi una v -a, e non più solo un numero o una funzione come negli ordinari valori d'attesa, e che tale nuova v -a giocherà un ruolo particolarmente rilevante negli sviluppi dei capitoli seguenti

Proposizione 3.42. *Date due v -a X e Y su $(\Omega, \mathcal{F}, \mathbf{P})$ si hanno le seguenti proprietà delle attese condizionate rispetto a una v -a*

1. $\mathbf{E} [\mathbf{E} [X|Y]] = \mathbf{E} [X]$
2. $\mathbf{E} [X|Y] = \mathbf{E} [X]$ \mathbf{P} -qo se X e Y sono indipendenti
3. $\mathbf{E} [\varphi(X, Y)|Y = y] = \mathbf{E} [\varphi(X, y)|Y = y]$ \mathbf{P}_Y -qo
4. $\mathbf{E} [\varphi(X, Y)|Y = y] = \mathbf{E} [\varphi(X, y)]$ \mathbf{P}_Y -qo se X e Y sono indipendenti
5. $\mathbf{E} [X g(Y)|Y] = g(Y) \mathbf{E} [X|Y]$ \mathbf{P} -qo

Dimostrazione:

1. Da (3.44), (3.55) e (3.58) si ha

$$\begin{aligned} \mathbf{E} [\mathbf{E} [X|Y]] &= \mathbf{E} [m(Y)] = \int_{\mathbf{R}} m(y) f_Y(y) dy = \int_{\mathbf{R}} \mathbf{E} [X|Y = y] f_Y(y) dy \\ &= \int_{\mathbf{R}} \left[\int_{\mathbf{R}} x f_{X|Y}(x | y) dx \right] f_Y(y) dy \\ &= \int_{\mathbf{R}} \left[\int_{\mathbf{R}} x \frac{f_{XY}(x, y)}{f_Y(y)} dx \right] f_Y(y) dy \\ &= \int_{\mathbf{R}} x \left[\int_{\mathbf{R}} f_{XY}(x, y) dy \right] dx = \int_{\mathbf{R}} x f_X(x) dx = \mathbf{E} [X] \end{aligned}$$

2. Tenendo conto dell'indipendenza e di (3.45) abbiamo facilmente che

$$m(y) = \mathbf{E}[X|Y = y] = \int_{\mathbf{R}} x f_{X|Y}(x|y) dx = \int_{\mathbf{R}} x f_X(x) dx = \mathbf{E}[X]$$

per cui $\mathbf{E}[X|Y] = m(Y) = \mathbf{E}[X]$

3. Ricordando la relazione (3.49) possiamo scrivere

$$\begin{aligned} & \mathbf{E}[\varphi(X, Y)|y \leq Y \leq y + \Delta y] \\ &= \int_{\mathbf{R}} dx \int_{\mathbf{R}} dz \varphi(x, z) f_{XY}(x, z|y \leq Y \leq y + \Delta y) \\ &= \int_{-\infty}^{+\infty} dx \int_y^{y+\Delta y} dz \varphi(x, z) \frac{f_{XY}(x, z)}{F_Y(y + \Delta y) - F_Y(y)} \end{aligned}$$

D'altra parte siccome

$$F_Y(y + \Delta y) - F_Y(y) = F_Y'(y)\Delta y + o(\Delta y) = f_Y(y)\Delta y + o(\Delta y)$$

risulta anche

$$\begin{aligned} \lim_{\Delta y \rightarrow 0} \int_y^{y+\Delta y} dz \varphi(x, z) \frac{f_{XY}(x, z)}{F_Y(y + \Delta y) - F_Y(y)} &= \varphi(x, y) \frac{f_{XY}(x, y)}{f_Y(y)} \\ &= \varphi(x, y) f_{X|Y}(x|y) \end{aligned}$$

e pertanto potremo scrivere

$$\begin{aligned} \mathbf{E}[\varphi(X, Y)|Y = y] &= \lim_{\Delta y \rightarrow 0} \mathbf{E}[\varphi(X, Y)|y \leq Y \leq y + \Delta y] \\ &= \int_{-\infty}^{+\infty} \varphi(x, y) f_{X|Y}(x|y) dx = \mathbf{E}[\varphi(X, y)|Y = y] \end{aligned}$$

4. Se X e Y sono indipendenti il risultato segue dal punto precedente e da (3.45)

5. Dal punto 3 segue in particolare che

$$\mathbf{E}[X g(Y)|Y = y] = \mathbf{E}[X g(y)|Y = y] = g(y) \mathbf{E}[X|Y = y]$$

e l'enunciato si ricava allora applicando queste funzioni di y alla v -a Y . \square

Utilizzando le *ddp* condizionate (3.53) è infine anche possibile introdurre le attese condizionate rispetto a eventi di probabilità nulla del tipo $\{Y_1 = y_1, \dots, Y_m = y_m\}$

$$\begin{aligned} m(y_1, \dots, y_m) &= \mathbf{E}[X|Y_1 = y_1, \dots, Y_m = y_m] \\ &= \int_{\mathbf{R}} x f_{X|Y}(x|y_1, \dots, y_m) dx, \end{aligned}$$

e quindi anche le **attese condizionate rispetto a un vett- \mathbf{a}**

$$\mathbf{E}[X|\mathbf{Y}] = \mathbf{E}[X|Y_1, \dots, Y_m] = m(Y_1, \dots, Y_m) = m(\mathbf{Y}) \quad (3.59)$$

Le proprietà di queste nuove v -a sono simili a quelle delle attese condizionate da una sola v -a introdotte all'inizio di questa sezione e non saranno qui elencate

Esempio 3.43. Vita media: *Supponiamo che il tempo durante il quale opera, senza rompersi, un componente di un dato apparato sia descritto da una v -a Y con ddp $f_Y(y)$, con la precisazione che, se l'apparato comincia a funzionare all'istante $y = 0$, la $f_Y(y)$ sarà definita come identicamente nulla per $y < 0$. Poniamoci allora il problema di determinare*

$$f_{Y-y_0}(y|Y \geq y_0) \quad e \quad \mathbf{E}[Y - y_0|Y \geq y_0]$$

cioè la ddp e il valore d'attesa (**vita media**) del tempo di vita residuo del nostro componente, sotto l'ipotesi che esso sia ancora funzionante al tempo $y_0 > 0$. Supponendo allora, per rendere significativo il nostro problema, che $\mathbf{P}\{Y \geq y_0\} > 0$, da (3.51) con $a = y_0$ e $b = +\infty$ si ha innanzitutto

$$\mathbf{E}[Y - y_0|Y \geq y_0] = \int_{\mathbf{R}} (y - y_0) f_Y(y|Y \geq y_0) dy = \frac{\int_{y_0}^{+\infty} (y - y_0) f_Y(y) dy}{\int_{y_0}^{+\infty} f_Y(y) dy} \quad (3.60)$$

D'altra parte per calcolare la ddp del tempo residuo $Y - y_0$ osserveremo prima che

$$\begin{aligned} F_{Y-y_0}(y|Y \geq y_0) &= \mathbf{P}\{Y - y_0 \leq y|Y \geq y_0\} \\ &= \mathbf{P}\{Y \leq y + y_0|Y \geq y_0\} = F_Y(y + y_0|Y \geq y_0) \end{aligned}$$

e quindi da (3.51) otterremo

$$\begin{aligned} f_{Y-y_0}(y|Y \geq y_0) &= \partial_y F_{Y-y_0}(y|Y \geq y_0) = \partial_y F_Y(y_0 + y|Y \geq y_0) \\ &= f_Y(y_0 + y|Y \geq y_0) = \frac{f_Y(y_0 + y) \chi_{(y_0, +\infty)}(y_0 + y)}{\int_{y_0}^{+\infty} f_Y(y') dy'} \\ &= \frac{f_Y(y_0 + y) \chi_{(0, +\infty)}(y)}{\int_{y_0}^{+\infty} f_Y(y') dy'} \end{aligned} \quad (3.61)$$

Naturalmente questo risultato permette di ricavare di nuovo la vita media (3.60) con un calcolo esplicito

$$\begin{aligned} \mathbf{E}[Y - y_0|Y \geq y_0] &= \int_{-\infty}^{+\infty} y f_{Y-y_0}(y|Y \geq y_0) dy = \frac{\int_0^{+\infty} y f_Y(y_0 + y) dy}{\int_{y_0}^{+\infty} f_Y(y') dy'} \\ &= \frac{\int_{y_0}^{+\infty} (y - y_0) f_Y(y) dy}{\int_{y_0}^{+\infty} f_Y(y) dy} \end{aligned}$$

È interessante esaminare ora questi risultati nel caso particolare in cui $Y \sim \mathfrak{E}(a)$ sia una v -a esponenziale. In tal caso sappiamo infatti da (2.17) e (3.33) che

$$f_Y(y) = ae^{-ay}\vartheta(y) \quad \mathbf{E}[Y] = \frac{1}{a}$$

e siccome con $z = y - y_0$ si ha

$$\begin{aligned} \int_{y_0}^{+\infty} f_Y(y) dy &= \int_{y_0}^{+\infty} ae^{-ay} dy = e^{-ay_0} \\ \int_{y_0}^{+\infty} (y - y_0)f_Y(y) dy &= \int_0^{+\infty} zf_Y(z + y_0) dz = \int_0^{+\infty} za e^{-a(z+y_0)} dz = \frac{e^{-ay_0}}{a} \\ f_Y(y_0 + y) \chi_{(0,+\infty)}(y) &= ae^{-a(y_0+y)} \chi_{(0,+\infty)}(y) = e^{-ay_0} f_Y(y) \end{aligned}$$

si vede da (3.60) e da (3.61) che

$$\mathbf{E}[Y - y_0 | Y \geq y_0] = \frac{1}{a} = \mathbf{E}[Y] \quad f_{Y-y_0}(y | Y \geq y_0) = f_Y(y)$$

Quindi, non solo la vita media del nostro componente – condizionata dal fatto che esso ha funzionato fino all'istante $y = y_0$ – non dipende da y_0 ed è sempre eguale a $\mathbf{E}[Y]$, ma anche la ddp di $Y - y_0$ condizionata da $Y \geq y_0$ non dipende da y_0 e resta sempre eguale alla ddp non condizionata. Questo comportamento (detto anche **assenza di memoria o di usura**) è caratteristico delle v -a esponenziali nel senso che non vi sono altre distribuzioni diverse da quella esponenziale che siano dotate di queste proprietà

Esempio 3.44. Ago di Buffon: Un ago di lunghezza unitaria viene lanciato a caso su un piano sul quale sono tracciate delle linee parallele a distanza unitaria fra loro: quale è la probabilità che l'ago intersechi una di queste linee? Innanzitutto, data la regolarità della ripetizione delle linee sul tavolo, sarà sufficiente esaminare il problema con due sole linee facendo l'ipotesi che il centro dell'ago cada fra queste ultime. Inoltre la posizione di tale centro lungo le rette parallele alle due linee è ininfluyente: potremmo tenerne conto aggiungendo un'opportuna variabile al nostro problema, ma potremmo poi eliminarla facilmente senza cambiare il risultato. La posizione dell'ago sarà allora caratterizzata mediante due sole v -a: la distanza X del suo punto di mezzo dalla linea sinistra, e l'angolo Θ che esso forma con una perpendicolare alle linee parallele (vedi Figura 3.5). Che l'ago sia **lanciato a caso** qui vuol dire che la coppia di v -a X, Θ è distribuita in maniera uniforme in $[0, 1] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$, cioè che

$$f_{X\Theta}(x, \theta) = \frac{1}{\pi} \chi_{[0,1]}(x) \chi_{[-\frac{\pi}{2}, \frac{\pi}{2}]}(\theta)$$

È facile vedere allora che le ddp marginali sono

$$f_X(x) = \chi_{[0,1]}(x) \quad f_\Theta(\theta) = \frac{1}{\pi} \chi_{[-\frac{\pi}{2}, \frac{\pi}{2}]}(\theta)$$

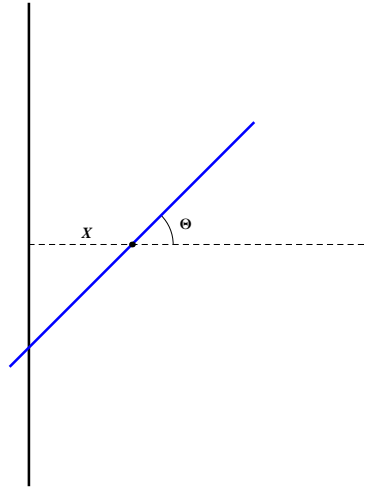


Figura 3.5: Problema dell'ago di Buffon.

e che le due v-a X e Θ sono indipendenti. Posto ora

$$B = \left\{ (x, \theta) : x \leq \frac{1}{2} \cos \theta, \text{ oppure } x \geq 1 - \frac{1}{2} \cos \theta, \text{ con } -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2} \right\}$$

è evidente che l'evento del quale vogliamo calcolare la probabilità è

$$A = \{ \text{l'ago interseca una linea} \} = \{ \omega \in \Omega : (X, \Theta) \in B \}$$

e che $I_A = \chi_B(X, \Theta)$, con $\chi_B(x, \theta)$ indicatore dell'insieme B in \mathbf{R}^2 . Il risultato può essere ottenuto in diverse maniere equivalenti, e noi ne sceglieremo una che fa uso, in successione, di (3.10), di 1 della Proposizione 3.42, dell'uniformità della distribuzione di Θ , e di 4 della Proposizione 3.42:

$$\begin{aligned} \mathbf{P}\{A\} &= \mathbf{E}[I_A] = \mathbf{E}[\chi_B(X, \Theta)] = \mathbf{E}[\mathbf{E}[\chi_B(X, \Theta) | \Theta]] \\ &= \int_{-\pi/2}^{\pi/2} \mathbf{E}[\chi_B(X, \Theta) | \Theta = \theta] \frac{d\theta}{\pi} = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \mathbf{E}[\chi_B(X, \theta)] d\theta \end{aligned}$$

A questo punto basterà ricordare che a causa dell'uniformità di X si ha

$$\mathbf{E}[\chi_B(X, \theta)] = \mathbf{P}\left\{ \left\{ X \leq \frac{1}{2} \cos \theta \right\} \cup \left\{ X \geq 1 - \frac{1}{2} \cos \theta \right\} \right\} = \frac{1}{2} \cos \theta + \frac{1}{2} \cos \theta = \cos \theta$$

per ottenere

$$\mathbf{P}\{A\} = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos \theta d\theta = \frac{2}{\pi}$$

Sarà interessante osservare che questo risultato può essere (ed è stato) usato per dare una **stima empirica** di π : se lanciamo l'ago n volte, se definiamo n v-a iid Y_k (con $k = 1, \dots, n$) di Bernoulli che siano eguali ad 1 se l'ago interseca una linea al

k -mo lancio e a 0 nel caso contrario, se supponiamo che p sia la probabilità che l'ago intersechi una linea in ogni singolo lancio, e se definiamo la v -a $\nu_n = Y_1 + \dots + Y_n$ per indicare il numero di volte in cui l'ago interseca una linea su n lanci, allora è intuitivo capire (e la **Legge dei Grandi Numeri** che discuteremo nella successiva Sezione 4.3 lo confermerà) che, con n abbastanza grande, la v -a ν_n/n approssimerà bene il numero p . Sapendo allora, dalla discussione precedente, che $p = 2/\pi$, una buona approssimazione del valore di π sarà data da un valore misurato della v -a $2n/\nu_n$ con n abbastanza grande. Questo metodo per stimare π è stato usato varie volte nel corso della storia¹⁰ e rappresenta il primo caso conosciuto di applicazioni delle regolarità statistiche alla soluzione di problemi di analisi numerica: un metodo divenuto successivamente noto come **metodo di Monte Carlo** del quale avremo modo di riparlare nel corso di queste lezioni

3.4.3 Stima ottimale in media quadratica

Per mettere meglio in evidenza le relazioni che legano le v -a X, Y e $\mathbf{E}[X|Y]$ definita in (3.57) e (3.58), bisogna ricordare che in generale X e Y possono essere statisticamente dipendenti senza che esista una qualche funzione di Borel $h(y)$ tale che $X = h(Y)$: la *dipendenza statistica*, cioè, non si traduce necessariamente in una *dipendenza funzionale* (vedi anche la discussione nella Sezione 3.1.3). D'altra parte, date X e Y statisticamente dipendenti, può capitare che si vogliano usare le misure di Y per ottenere informazioni sui (*stimare i*) valori della X . In pratica ciò viene fatto considerando delle v -a della forma $h(Y)$ che vengono usate come *stimatori* di X . Siccome però abbiamo osservato che in generale non potremo sperare di trovare una h tale che $X = h(Y)$, sarà necessario fissare un criterio che permetta di scegliere in maniera ottimale uno stimatore di X : il più noto consiste nel cercare di determinare il *miglior stimatore in media quadratica* (in *mq*). Si definisce prima l'errore quadratico medio (*eqm*) che si commette stimando X con un generico stimatore $h(Y)$

$$\mathbf{E} [(X - h(Y))^2]$$

e poi si elegge come miglior stimatore quello ottenuto con la $h^*(Y)$ che rende minimo tale *eqm*

$$\mathbf{E} [(X - h^*(Y))^2] = \inf_h \mathbf{E} [(X - h(Y))^2]$$

La determinazione di tale stimatore ottimale, cioè della funzione di Borel h^* che rende minimo l'*eqm*, costituisce un tipico problema variazionale che comunque, anche nella sua veste più generale, ammette sempre una soluzione formale esatta: è importante sottolineare allora – come espresso dall'enunciato della successiva proposizione – che tale soluzione $h^*(y)$ coincide proprio con la $m(y) = \mathbf{E}[X|Y = y]$ definita in (3.58)

¹⁰La stima fu tentata per la prima volta nel 1850 dall'astronomo svizzero R. Wolf (1816 - 1893) che, lanciando l'ago 5 000 volte ottenne come risultato 3.1596.

Proposizione 3.45. *Il miglior stimatore in mq di X tramite Y è $\mathbf{E}[X|Y]$, ed è quindi definito dalla funzione di Borel*

$$h^*(y) = m(y) = \mathbf{E}[X|Y = y]$$

Dimostrazione: Se $h(Y)$ è il generico stimatore e $h^*(Y) = \mathbf{E}[X|Y]$, si ha:

$$\begin{aligned} \mathbf{E}[(X - h(Y))^2] &= \mathbf{E}[(X - h^*(Y) + h^*(Y) - h(Y))^2] \\ &= \mathbf{E}[(X - h^*(Y))^2] + \mathbf{E}[(h^*(Y) - h(Y))^2] \\ &\quad + 2\mathbf{E}[(X - h^*(Y))(h^*(Y) - h(Y))] \end{aligned}$$

Da 1 e 5 della Proposizione 3.42 si ha però che

$$\begin{aligned} \mathbf{E}[(X - h^*(Y))(h^*(Y) - h(Y))] &= \mathbf{E}\left[\mathbf{E}[(X - h^*(Y))(h^*(Y) - h(Y)) | Y]\right] \\ &= \mathbf{E}\left[(h^*(Y) - h(Y)) \mathbf{E}[X - h^*(Y) | Y]\right] \end{aligned}$$

e siccome

$$\mathbf{E}[X - h^*(Y) | Y] = \mathbf{E}[X|Y] - \mathbf{E}[h^*(Y)|Y] = \mathbf{E}[X|Y] - h^*(Y) = 0$$

in definitiva risulta

$$\mathbf{E}[(X - h(Y))^2] = \mathbf{E}[(X - h^*(Y))^2] + \mathbf{E}[(h^*(Y) - h(Y))^2]$$

D'altra parte $\mathbf{E}[(h^*(Y) - h(Y))^2] \geq 0$, e quindi

$$\mathbf{E}[(X - h(Y))^2] \geq \mathbf{E}[(X - h^*(Y))^2]$$

comunque scelta la funzione di Borel h □

La funzione $m(y) = \mathbf{E}[X|Y = y]$ prende anche il nome di *curva di regressione* di X su Y : tale nome ha la sua origine negli studi di sir F. Galton (1822 - 1911) sulla relazione che intercorre fra le altezze delle generazioni dei genitori e dei figli in una data popolazione. I suoi risultati – esprimibili in termini di attese condizionate – mostravano che quando i genitori sono più alti della media, i figli tendono ad essere più bassi dei genitori; quando invece i genitori sono più bassi, i figli tendono ad essere più alti di loro: in ambedue i casi l'altezza dei figli *regredisce* verso il valore medio dell'altezza della popolazione

3.5 Trasformazioni di v -a

3.5.1 Funzioni di v -a

Proposizione 3.46. *Se X è una v -a con ddp $f_X(x)$, e se $y = \varphi(x)$ è una funzione continua e regolare, il cui intervallo di definizione è decomponibile in n intervalli*

disgiunti $[a_k, b_k]$ all'interno dei quali φ è derivabile e strettamente monotona, con derivata non nulla in ogni punto, allora la ddp $f_Y(y)$ della v-a $Y = \varphi(X)$ è

$$f_Y(y) = \sum_{k=1}^n \frac{f_X(x_k(y))}{|\varphi'(x_k(y))|} \chi_{[\alpha_k, \beta_k]}(y) \quad (3.62)$$

dove $[\alpha_k, \beta_k]$ sono gli intervalli dei valori assunti da φ per $x \in [a_k, b_k]$, e per un dato y le $x_k(y)$ rappresentano le (al più n) soluzioni dell'equazione $\varphi(x) = y$

Dimostrazione: Per cominciare supponiamo che X prenda valori solo nell'intervallo $[a, b]$ (cioè che $f_X(x)$ sia nulla al di fuori di tale intervallo), e che la funzione $\varphi(x)$ sia definita, derivabile e strettamente crescente ($\varphi'(x) > 0$) su tutto $[a, b]$. Detto allora $[\alpha, \beta]$ l'insieme dei valori assunti da $\varphi(x)$, indichiamo con $x_1(y) = \varphi^{-1}(y)$ la soluzione dell'equazione $\varphi(x) = y$ che esiste ed è unica (e monotona come funzione di y) se $y \in [\alpha, \beta]$. È evidente allora che

$$F_Y(y) = \mathbf{P}\{Y \leq y\} = \begin{cases} 0 & \text{per } y < \alpha \\ 1 & \text{per } y > \beta \end{cases}$$

mentre per $y \in [\alpha, \beta]$, con il cambiamento di variabile di integrazione $z = \varphi(x)$, $x = \varphi^{-1}(z) = x_1(z)$, si ha

$$\begin{aligned} F_Y(y) &= \mathbf{P}\{Y \leq y\} = \mathbf{P}\{\varphi(X) \leq y\} = \mathbf{P}\{X \leq \varphi^{-1}(y)\} = \mathbf{P}\{X \leq x_1(y)\} \\ &= \int_a^{x_1(y)} f_X(x) dx = \int_\alpha^y f_X(x_1(z)) x_1'(z) dz = \int_\alpha^y f_Y(z) dz \end{aligned}$$

Pertanto in definitiva avremo

$$f_Y(y) = f_X(x_1(y)) x_1'(y) \chi_{[\alpha, \beta]}(y) = \begin{cases} f_X(x_1(y)) x_1'(y) & \text{per } \alpha \leq y \leq \beta \\ 0 & \text{altrove} \end{cases}$$

D'altra parte, se φ fosse strettamente decrescente un analogo ragionamento condurrebbe alla conclusione

$$f_Y(y) = -f_X(x_1(y)) x_1'(y) \chi_{[\alpha, \beta]}(y)$$

per cui in ogni caso, quando φ è strettamente monotona su $[a, b]$, potremo scrivere

$$f_Y(y) = f_X(x_1(y)) |x_1'(y)| \chi_{[\alpha, \beta]}(y) \quad (3.63)$$

Siccome inoltre per un ben noto risultato dell'analisi

$$x_1'(y) = \frac{1}{\varphi'(x_1(y))}$$

la nostra trasformazione per funzioni φ monotone si potrà enunciare nella forma

$$f_Y(y) = \frac{f_X(x_1(y))}{|\varphi'(x_1(y))|} \chi_{[\alpha, \beta]}(y) \quad (3.64)$$

cioè la (3.62) con un solo termine nella somma. Quando invece φ non è strettamente monotona su tutto l'insieme dei valori assunti da X , in molti casi di utilità pratica il suo insieme di definizione si potrà decomporre nell'unione di n intervalli disgiunti $[a_k, b_k]$ all'interno dei quali essa è derivabile e strettamente monotona, con derivata non nulla in ogni punto. Se allora $[\alpha_k, \beta_k]$ sono gli intervalli dei valori assunti da φ per $x \in [a_k, b_k]$, e se per un dato y le $x_k(y)$ rappresentano le soluzioni dell'equazione $\varphi(x) = y$, in modo analogo al caso monotono si prova¹¹ la (3.62). Occorre precisare comunque che il numero degli addendi di questa somma dipende da y , nel senso che per un dato y saranno presenti solo gli $m \leq n$ termini corrispondenti alle soluzioni di $\varphi(x) = y$ per i quali $\chi_{[\alpha_k, \beta_k]}(y) = 1$. Naturalmente, nel caso in cui l'equazione $\varphi(x) = y$ non ha nessuna soluzione il numero di addendi della somma è nullo e $f_Y(y) = 0$ \square

Nel caso più generale $\mathbf{Y} = \varphi(\mathbf{X})$ è una trasformazione di un *vett-a* \mathbf{X} con n componenti X_j in un *vett-a* \mathbf{Y} con m componenti mediante le funzioni

$$Y_k = \varphi_k(X_1, \dots, X_n), \quad k = 1, \dots, m$$

Non si perde però in generalità se ci si limita al caso $n = m$. Infatti:

- se $m < n$, è sempre possibile aggiungere ad \mathbf{Y} altre $n - m$ componenti ausiliarie tutte eguali a X_{m+1}, \dots, X_n ; una volta risolto il problema in questa forma e determinata la *ddp* congiunta $f_{\mathbf{Y}}(y_1, \dots, y_n)$, si eliminano le variabili y_{m+1}, \dots, y_n in eccesso mediante una marginalizzazione;
- se $m > n$, $m - n$ fra le Y_k risulteranno funzioni delle altre n componenti; in questo caso si risolve il problema per le prime n *v-a* Y_k , e poi la distribuzione delle rimanenti viene ricavata in funzione delle precedenti.

Posto quindi $n = m$, ci limiteremo a ricordare il risultato principale: se per un dato \mathbf{y} , le $\mathbf{x}^j(\mathbf{y})$ sono le (al più n) soluzioni del sistema di n equazioni $y_k = \varphi_k(x_1, \dots, x_n)$, allora la *ddp* congiunta del *vett-a* \mathbf{Y} sarà

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{j=1}^n \frac{f_{\mathbf{X}}(\mathbf{x}^j(\mathbf{y}))}{|J(\mathbf{x}^j(\mathbf{y}))|} \chi_j(\mathbf{y}) \quad (3.66)$$

¹¹Osserviamo per completezza che l'equazione (3.62) si può anche scrivere nella forma

$$f_Y(y) = \int_{-\infty}^{+\infty} f_X(x) \delta[y - \varphi(x)] dx \quad (3.65)$$

introducendo la distribuzione $\delta(x)$ di Dirac per la quale vale, con le nostre notazioni, la relazione

$$\delta[y - \varphi(x)] = \sum_{k=1}^n \frac{\delta[x - x_k(y)]}{|\varphi'(x_k(y))|} \chi_{[\alpha_k, \beta_k]}(y)$$

Per maggiori dettagli vedi **V.S. Vladimirov**, *LE DISTRIBUZIONI NELLA FISICA MATEMATICA*, Mir (Mosca, 1981)

dove $J(\mathbf{x})$ è il determinante Jacobiano della trasformazione con elementi del tipo $\partial\varphi_k/\partial x_l$, mentre le $\chi_j(\mathbf{y})$ sono funzioni che valgono 1 se in corrispondenza di \mathbf{y} esiste la j -ma soluzione, e 0 altrimenti. Questa formula pertanto generalizza la (3.62) con analoghe precisazioni sul numero di addendi della somma in funzione del numero (eventualmente nullo) di soluzioni del dato sistema di equazioni.

Esempio 3.47. Funzioni lineari: Se $Y = aX + b$, cioè $\varphi(x) = ax + b$, con $a \neq 0$, l'equazione $y = \varphi(x)$ ammette sempre una ed una sola soluzione: $x_1(y) = (y - b)/a$. Ne segue che

$$f_{aX+b}(y) = \frac{1}{|a|} f_X \left(\frac{y - b}{a} \right)$$

In particolare, se $X \sim \mathfrak{N}(0, 1)$ è una v-a normale standard, allora $Y = aX + b \sim \mathfrak{N}(b, a^2)$; viceversa, se $X \sim \mathfrak{N}(b, a^2)$, allora $Y = (X - b)/a \sim \mathfrak{N}(0, 1)$ è normale standard.

Funzioni quadratiche: Se invece $Y = X^2$, cioè $\varphi(x) = x^2$, l'equazione $y = \varphi(x)$ ammetterà soluzioni $x_1(y) = -\sqrt{y}$ e $x_2(y) = +\sqrt{y}$ solo se $y \geq 0$. Con $\vartheta(y)$ funzione di Heaviside (2.13) avremo allora

$$f_{X^2}(y) = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})] \vartheta(y)$$

In particolare, se $X \sim \mathfrak{N}(0, 1)$ si ha

$$f_{X^2}(y) = \frac{e^{-y/2}}{\sqrt{2\pi y}} \vartheta(y) \quad (3.67)$$

che, come vedremo nella sezione seguente, è una legge del χ^2 con un grado di libertà.

Funzioni esponenziali: Se infine $Y = e^X$ e $X \sim \mathfrak{N}(b, a^2)$ da (3.62) si ha

$$f_{e^X}(y) = \frac{e^{-(\ln y - b)^2/2a^2}}{ay\sqrt{2\pi}} \vartheta(y)$$

una legge che prende anche il nome di **log-normale**, e che indicheremo con $\ln\mathfrak{N}(b, a^2)$.

Valore di attesa e varianza di una v-a $Y \sim \ln\mathfrak{N}(b, a^2)$ sono

$$\mathbf{E}[Y] = e^{b+a^2/2} \quad \mathbf{V}[Y] = e^{2b+a^2}(e^{a^2} - 1) \quad (3.68)$$

Infatti se $X \sim \mathfrak{N}(b, a^2)$ si ha innanzitutto con $z = \frac{x-b}{a}$

$$\mathbf{E}[Y] = \mathbf{E}[e^X] = \int_{-\infty}^{+\infty} e^x \frac{e^{-\frac{(x-b)^2}{2a^2}}}{a\sqrt{2\pi}} dx = e^{b+a^2/2} \int_{-\infty}^{+\infty} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz = e^{b+a^2/2}$$

Siccome d'altra parte $2X \sim \mathfrak{N}(2b, 4a^2)$ dal risultato precedente segue anche

$$\mathbf{E}[Y^2] = \mathbf{E}[e^{2X}] = e^{2b+2a^2}$$

e quindi

$$\mathbf{V}[Y] = \mathbf{E}[Y^2] - \mathbf{E}[Y]^2 = e^{2b+2a^2} - e^{2b+a^2} = e^{2b+a^2}(e^{a^2} - 1)$$

Un ultimo esempio di applicazione di (3.62) che va sotto il nome di **Paradosso di Bertrand** è rinviato all'Appendice C

3.5.2 Somme di v -a indipendenti

Definizione 3.48. Chiameremo **convoluzione** di due ddp f e g la funzione

$$\begin{aligned} (f * g)(x) &= (g * f)(x) \\ &= \int_{-\infty}^{\infty} f(x-y)g(y) dy = \int_{-\infty}^{\infty} g(x-y)f(y) dy \end{aligned}$$

Si verifica che la convoluzione di due ddp è ancora una ddp

Proposizione 3.49. Date due v -a indipendenti X e Y rispettivamente con ddp $f_X(x)$ e $f_Y(y)$, la ddp della loro somma $Z = X + Y$ è la convoluzione delle loro ddp

$$f_Z(x) = (f_X * f_Y)(x) = (f_Y * f_X)(x)$$

Dimostrazione: Se due v -a X e Y hanno ddp congiunta $f_{XY}(x, y)$ e consideriamo $Z = \varphi(X, Y)$, dove $z = \varphi(x, y)$ è una funzione di Borel, dato z e definito per brevità

$$\{\varphi \leq z\} = \{(x, y) \in \mathbf{R}^2 : \varphi(x, y) \leq z\}$$

la fdd della v -a Z si calcola facilmente come

$$F_Z(z) = \mathbf{P}\{Z \leq z\} = \mathbf{P}\{\varphi(X, Y) \leq z\} = \int_{\{\varphi \leq z\}} f_{XY}(x, y) dx dy$$

Se poi in particolare $\varphi(x, y) = x + y$, e se le v -a X e Y sono indipendenti, cioè se $f_{XY}(x, y) = f_X(x)f_Y(y)$, con il cambiamento di variabile $u = x + y$ avremo che

$$\begin{aligned} F_Z(z) &= \int_{\{x+y \leq z\}} f_X(x)f_Y(y) dx dy = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-x} f_Y(y) dy \right] f_X(x) dx \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^z f_Y(u-x) du \right] f_X(x) dx \\ &= \int_{-\infty}^z \left[\int_{-\infty}^{\infty} f_Y(u-x)f_X(x) dx \right] du \end{aligned}$$

Equivalentemente, invertendo l'ordine delle integrazioni, si ha anche

$$F_Z(z) = \int_{-\infty}^z \left[\int_{-\infty}^{\infty} f_X(u-y)f_Y(y) dy \right] du$$

e quindi potremo dire che $Z = X + Y$ ha come ddp

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(z-x)f_X(x) dx = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y) dy$$

cioè $f_Z = f_X * f_Y = f_Y * f_X$ □

Le considerazioni precedenti si adattano facilmente anche al caso di più di due v -a: se X_1, \dots, X_n sono n v -a indipendenti dotate di ddp , la ddp della loro somma $Z = X_1 + \dots + X_n$ è la convoluzione delle rispettive ddp

$$f_Z(x) = (f_{X_1} * \dots * f_{X_n})(x) \tag{3.69}$$

Esempio 3.50. Somme di v-a uniformi: Se X_1, \dots, X_n sono v-a iid $\mathfrak{U}(-1, 1)$ la loro ddp può anche essere espressa mediante l'uso della funzione $\vartheta(x)$ di Heaviside (2.13)

$$f_{X_k}(x) = f(x) = \frac{1}{2} \vartheta(1 - |x|) \quad k = 1, \dots, n$$

Si mostra allora con un calcolo diretto che

$$\begin{aligned} f_{X_1+X_2}(x) &= \frac{2 - |x|}{4} \vartheta(2 - |x|), \\ f_{X_1+X_2+X_3}(x) &= \left[\vartheta(1 - |x|) \frac{3 - x^2}{8} + \vartheta(|x| - 1) \frac{(3 - |x|)^2}{16} \right] \vartheta(3 - |x|) \end{aligned}$$

e per induzione si prova per $Y = X_1 + \dots + X_n$ la formula

$$f_Y(x) = \frac{\vartheta(n - |x|)}{2^n (n - 1)!} \sum_{k=0}^{[(n+x)/2]} (-1)^k \binom{n}{k} (n + x - 2k)^{n-1}$$

nella quale con $[\alpha]$ indichiamo la parte intera del numero reale α . Pertanto il primo risultato rilevante è che le somme di v-a indipendenti e uniformi non sono affatto uniformi: ad esempio la $f_{X_1+X_2}$ è una distribuzione triangolare su $[-2, 2]$, mentre la $f_{X_1+X_2+X_3}$ è composta di tre rami di parabola connessi con continuità su $[-3, 3]$.

Somme di v-a Gaussiane: Alla luce dell'esempio precedente acquistano interesse le leggi che combinate tramite convoluzione con altre leggi della stessa famiglia producono ancora leggi della medesima famiglia. Così, se $X \sim \mathfrak{N}(b_1, a_1^2)$ ed $Y \sim \mathfrak{N}(b_2, a_2^2)$ sono Gaussiane e indipendenti, si ottiene da un calcolo diretto che $X + Y \sim \mathfrak{N}(b_1 + b_2, a_1^2 + a_2^2)$, ovvero simbolicamente

$$\mathfrak{N}(b_1, a_1^2) * \mathfrak{N}(b_2, a_2^2) = \mathfrak{N}(b_1 + b_2, a_1^2 + a_2^2) \quad (3.70)$$

Questo importante risultato, che si estende anche ad un numero arbitrario di v-a, prende il nome di **proprietà riproduttiva delle v-a normali**, e può essere più facilmente provato (assieme ad altri analoghi risultati per altre famiglie di leggi) con l'uso delle funzioni caratteristiche come vedremo più avanti nella Sezione 4.2.3.

Distribuzione χ_n^2 : Se X_1, \dots, X_n sono v-a iid tutte $\mathfrak{N}(0, 1)$, tenendo conto di (3.67) e applicando ripetutamente la regola delle convoluzioni alle $f_{X_k^2}(x)$, si ottiene per la v-a $Z = X_1^2 + \dots + X_n^2$ la ddp

$$f_Z(x) = \frac{x^{n/2-1} e^{-x/2}}{2^{n/2} \Gamma(n/2)} \vartheta(x)$$

nota con il nome di distribuzione del χ^2 con n gradi di libertà e indicata con il simbolo χ_n^2 . Qui $\Gamma(x)$ è la funzione Gamma definita da

$$\Gamma(x) = \int_0^{+\infty} z^{x-1} e^{-z} dz \quad (3.71)$$

che soddisfa le proprietà

$$\Gamma(x) = (x-1)\Gamma(x-1) \quad \Gamma(1) = 1 \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{2}$$

sicché si ha

$$\Gamma\left(\frac{n}{2}\right) = \begin{cases} (n-2)!! 2^{-n/2} & \text{se } n \text{ è pari} \\ (n-2)!! 2^{-(n-1)/2} & \text{se } n \text{ è dispari} \end{cases}$$

Si dimostra che il valore d'attesa e la varianza di una v -a Z con legge χ_n^2 sono

$$\mathbf{E}[Z] = n \quad \mathbf{V}[Z] = 2n$$

Distribuzione di Student \mathfrak{T}_n : Se X_0, X_1, \dots, X_n sono $n+1$ v -a iid normali $\mathfrak{N}(0, a^2)$, posto

$$T = \frac{X_0}{\sqrt{(X_1^2 + \dots + X_n^2)/n}} = \frac{X_0/a}{\sqrt{(X_1^2 + \dots + X_n^2)/na^2}} = \frac{X_0/a}{\sqrt{Z/n}}$$

dagli esempi precedenti si ha che X_k/a sono tutte $\mathfrak{N}(0, 1)$, mentre $Z = (X_1^2 + \dots + X_n^2)/a^2$ è χ_n^2 . Si dimostra allora che la ddp di T è

$$f_T(t) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$

prende il nome di distribuzione t di Student e si indica con il simbolo \mathfrak{T}_n . Per $n=1$ la distribuzione di Student coincide con quella di Cauchy $\mathfrak{C}(1, 0)$. Si dimostra che valore d'attesa e varianza di una v -a T con legge \mathfrak{T}_n sono

$$\mathbf{E}[T] = 0 \quad \text{per } n \geq 2 \quad \mathbf{V}[T] = \frac{n}{n-2} \quad \text{per } n \geq 3$$

mentre non esistono per gli altri valori di n .

Capitolo 4

Teoremi limite

4.1 Convergenza

I Teoremi Limite sono affermazioni relative ai limiti di successioni di somme di v -a al crescere del numero degli addendi. Siccome al concetto di convergenza di una successione di v -a $(X_n)_{n \in \mathbf{N}}$ si possono dare molti significati non equivalenti fra loro, sarà importante prima di tutto chiarire questo punto elencando i tipi di convergenza più comuni e le relazioni che intercorrono fra loro

Definizione 4.1. *Data una successione di v.a. $(X_n)_{n \in \mathbf{N}}$ su $(\Omega, \mathcal{F}, \mathbf{P})$, essa*

- **converge in probabilità** alla v -a X , e scriveremo $X_n \xrightarrow{\mathbf{P}} X$, se

$$\mathbf{P}\{|X_n - X| > \epsilon\} \xrightarrow{n} 0, \quad \forall \epsilon > 0$$

- **converge quasi ovunque (P-qo) o con probabilità 1** alla v -a X , e scriveremo $X_n \xrightarrow{qo} X$ o anche $X_n \xrightarrow{n} X$ **P-qo**, se

$$\mathbf{P}\{X_n \rightarrow X\} = 1 \quad \text{ovvero} \quad \mathbf{P}\{X_n \not\rightarrow X\} = 0$$

dove $\{X_n \not\rightarrow X\}$ è l'insieme degli ω per i quali $(X_n)_{n \in \mathbf{N}}$ non converge a X

- **converge in L^p** (con $0 < p < +\infty$) alla v -a X e scriveremo $X_n \xrightarrow{L^p} X$, se

$$\mathbf{E}[|X_n - X|^p] \xrightarrow{n} 0$$

se poi $p = 2$ si dice anche che $(X_n)_{n \in \mathbf{N}}$ **converge in media quadratica (mq)** e si usa la notazione $X_n \xrightarrow{mq} X$. Il significato preciso dei simboli $L^p = L^p(\Omega, \mathcal{F}, \mathbf{P})$ è chiarito nell'Appendice D

- **converge in distribuzione**, e scriveremo $X_n \xrightarrow{d} X$, se

$$\mathbf{E}[f(X_n)] \xrightarrow{n} \mathbf{E}[f(X)], \quad \forall f \in \mathcal{C}(\mathbf{R})$$

dove $\mathcal{C}(\mathbf{R})$ indica l'insieme delle funzioni limitate e continue su \mathbf{R}

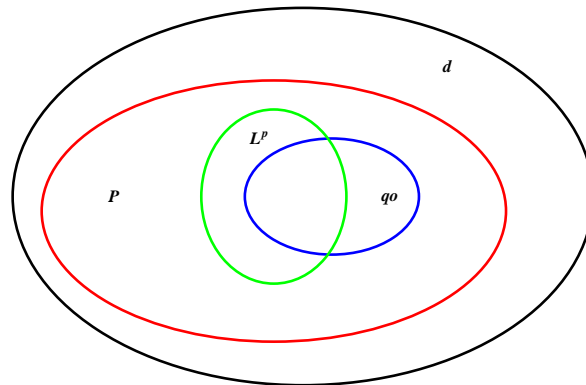


Figura 4.1: Relazioni fra i quattro tipi di convergenza secondo il Teorema 4.4

Definizione 4.2. *Data una successione di fdd $(F_n(x))_{n \in \mathbf{N}}$, essa*

- **converge debolmente** alla fdd $F(x)$, e scriveremo $F_n \xrightarrow{w} F$, se

$$\int_{\mathbf{R}} f(x)F_n(dx) \xrightarrow{n} \int_{\mathbf{R}} f(x)F(dx), \quad \forall f \in \mathcal{C}(\mathbf{R})$$

dove $\mathcal{C}(\mathbf{R})$ indica l'insieme delle funzioni limitate e continue

- **converge in generale** alla fdd $F(x)$, e scriveremo $F_n \xrightarrow{g} F$, se

$$F_n(x) \xrightarrow{n} F(x), \quad \forall x \in P_C(F)$$

dove $P_C(F)$ è l'insieme dei punti $x \in \mathbf{R}$ per i quali $F(x)$ è continua.

Proposizione 4.3. *Una successione di fdd $(F_n(x))_{n \in \mathbf{N}}$ converge debolmente alla fdd $F(x)$ se e solo se essa vi converge in generale*

Dimostrazione: Omessa¹ □

Data ora la successione di *v-a* $(X_n)_{n \in \mathbf{N}}$, e dette F_{X_n} le fdd di X_n , si vede immediatamente che $(X_n)_{n \in \mathbf{N}}$ converge in distribuzione alla *v-a* X con fdd F_X se e solo se $F_{X_n} \xrightarrow{w} F_X$, ovvero equivalentemente se e solo se $F_{X_n} \xrightarrow{g} F_X$. In pratica – con le dovute precisazioni – le convergenze in distribuzione, debole e in generale sono equivalenti. Questo implica che una convergenza in distribuzione di una successione di *v-a* può essere anche dimostrata esaminando semplicemente il comportamento della successione delle loro fdd, ossia delle loro leggi. In particolare basta verificare che $F_{X_n}(x) \xrightarrow{n} F_X(x)$ in tutti i punti x in cui la fdd limite $F_X(x)$ è continua.

I quattro tipi di convergenza elencati nella Definizione 4.1, invece, non sono equivalenti tra loro e le loro reciproche relazioni sono elencate nel seguente Teorema e riassunte graficamente nella Figura 4.1

¹A.N. Shiryaev, **PROBABILITY**, Springer (New York, 1996)

Teorema 4.4. *Data una successione di v-a $(X_n)_{n \in \mathbf{N}}$ e una v-a X , risulta*

1. $X_n \xrightarrow{qo} X \implies X_n \xrightarrow{\mathbf{P}} X$;
2. $X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{\mathbf{P}} X$, $p > 0$;
3. $X_n \xrightarrow{\mathbf{P}} X \implies X_n \xrightarrow{d} X$;
4. $X_n \xrightarrow{d} c \implies X_n \xrightarrow{\mathbf{P}} c$, *se c è un numero (convergenza degenera)*

Dimostrazione: Omessa² □

Implicazioni diverse da quelle qui elencate, invece, non sono in generale assicurate come si potrebbe far vedere con dei semplici controesempi. Ciononostante è possibile individuare delle ipotesi supplementari che garantiscono anche altre implicazioni diverse da quelle espresse nel Teorema 4.4: alcune fra le più note condizioni supplementari sono raccolte nel successivo teorema.

Teorema 4.5. *Data una successione di v-a $(X_n)_{n \in \mathbf{N}}$ e una v-a X*

1. *se $X_n \xrightarrow{\mathbf{P}} X$, allora esiste sempre una successione estratta $(X_{n_k})_{k \in \mathbf{N}}$ tale che $X_{n_k} \xrightarrow{qo} X$;*
2. *se $X_n \xrightarrow{L^p} X$, allora esiste sempre una successione estratta $(X_{n_k})_{k \in \mathbf{N}}$ tale che $X_{n_k} \xrightarrow{qo} X$;*
3. *se $X_n \xrightarrow{qo} X$, e se esiste una v.a. $Y \geq 0$ con $\mathbf{E}[|Y|] < +\infty$ tale che $|X_n - X| < Y$, allora risulta anche $X_n \xrightarrow{L^p} X$.*

Dimostrazione: Omessa³ □

Teorema 4.6. Criterio di convergenza degenera in mq : *Una successione di v-a $(X_n)_{n \in \mathbf{N}}$ converge in mq verso il numero m (convergenza degenera) se e solo se*

$$\mathbf{E}[X_n] \xrightarrow{n} m \quad \mathbf{V}[X_n] \xrightarrow{n} 0 \quad (4.1)$$

Dimostrazione: Infatti si ha

$$\begin{aligned} (X_n - m)^2 &= [(X_n - \mathbf{E}[X_n]) + (\mathbf{E}[X_n] - m)]^2 \\ &= (X_n - \mathbf{E}[X_n])^2 + (\mathbf{E}[X_n] - m)^2 + 2(X_n - \mathbf{E}[X_n])(\mathbf{E}[X_n] - m) \end{aligned}$$

e siccome ovviamente

$$\mathbf{E}[(X_n - \mathbf{E}[X_n])(\mathbf{E}[X_n] - m)] = (\mathbf{E}[X_n] - m)\mathbf{E}[X_n - \mathbf{E}[X_n]] = 0$$

risulta

$$\mathbf{E}[(X_n - m)^2] = \mathbf{V}[X_n] + (\mathbf{E}[X_n] - m)^2$$

e quindi la convergenza degenera in mq è equivalente alle condizioni (4.1) □

²N. Cufaro Petroni, CALCOLO DELLE PROBABILITÀ, Edizioni dal Sud (Bari, 1996)

³A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

4.2 Funzioni caratteristiche

4.2.1 Definizioni e proprietà

Definizione 4.7. Chiameremo **funzione caratteristica (fc)** del vett-a $\mathbf{X} = (X_1, \dots, X_n)$ la funzione

$$\varphi_{\mathbf{X}}(\mathbf{u}) = \varphi_{\mathbf{X}}(u_1, \dots, u_n) = \mathbf{E} [e^{i\mathbf{u} \cdot \mathbf{X}}] \quad \mathbf{u} \in \mathbf{R}^n \quad (4.2)$$

dove $\mathbf{u} \cdot \mathbf{X} = \sum_k u_k X_k$. Se \mathbf{X} è dotato di ddp la fc $\varphi_{\mathbf{X}}(\mathbf{u})$ può essere scritta come

$$\varphi_{\mathbf{X}}(\mathbf{u}) = \int_{\mathbf{R}^n} e^{i\mathbf{u} \cdot \mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) d^n \mathbf{x} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} e^{i\mathbf{u} \cdot \mathbf{x}} f_{\mathbf{X}}(x_1 \dots x_n) dx_1 \dots dx_n$$

e se il vett-a \mathbf{X} si riduce ad una sola componente X la fc diviene

$$\varphi_X(u) = \int_{-\infty}^{+\infty} e^{iux} f_X(x) dx \quad u \in \mathbf{R}$$

per cui una fc non è altro che la **trasformata di Fourier** di una ddp

Proposizione 4.8. Se $\varphi_X(u)$ è la fc della v-a X per ogni $u \in \mathbf{R}$ risulta

$$\varphi_X(u) = \overline{\varphi_X(-u)} \quad |\varphi_X(u)| \leq \varphi_X(0) = 1$$

dove \bar{z} è il complesso coniugato del numero complesso z . Inoltre $\varphi_X(u)$ è uniformemente continua in \mathbf{R} , ed è reale e pari se e solo se $f_X(x)$ è pari

Dimostrazione: La prima affermazione discende immediatamente dalla definizione, mentre per la seconda si ha

$$|\varphi_X(u)| = \left| \int_{-\infty}^{+\infty} e^{iux} f(x) dx \right| \leq \int_{-\infty}^{+\infty} f(x) dx = \varphi_X(0) = 1$$

Inoltre se $f_X(x)$ è pari la parte immaginaria di $\varphi_X(u)$ si annulla per ragioni di simmetria e la parte reale risulta evidentemente pari. Omettiamo⁴ invece la dimostrazione dell'uniforme continuità □

Proposizione 4.9. Se X è una v-a con fc $\varphi_X(u)$, se a, b sono numeri e se $Y = aX + b$ si ha

$$\varphi_Y(u) = e^{ibu} \varphi_X(au) \quad (4.3)$$

Se $\mathbf{X} = (X_1, \dots, X_n)$ è un vett-a, dette $\varphi_{\mathbf{X}}(u_1, \dots, u_n)$ e $\varphi_{X_k}(u_k)$ rispettivamente le fc congiunta e marginali, si ha

$$\varphi_{X_k}(u_k) = \varphi_{\mathbf{X}}(0, \dots, u_k, \dots, 0) \quad (4.4)$$

Se poi le componenti X_k sono indipendenti e $S_n = X_1 + \dots + X_n$, risulta

$$\varphi_{S_n}(u) = \varphi_{X_1}(u) \cdot \dots \cdot \varphi_{X_n}(u) \quad (4.5)$$

⁴N. Cufaro Petroni, **CALCOLO DELLE PROBABILITÀ**, Edizioni dal Sud (Bari, 1996)

Dimostrazione: Se $Y = aX + b$, dalla definizione (4.2) si ha la (4.3)

$$\varphi_Y(u) = \mathbf{E} [e^{iuY}] = e^{iub} \mathbf{E} [e^{i(au)X}] = e^{ibu} \varphi_X(au)$$

Anche la (4.4) deriva immediatamente dalla definizione (4.2), e infine se le X_k sono anche indipendenti si ha la (4.5)

$$\begin{aligned} \varphi_{S_n}(u) &= \mathbf{E} [e^{iuS_n}] = \mathbf{E} [e^{iuX_1} \cdot \dots \cdot e^{iuX_n}] \\ &= \mathbf{E} [e^{iuX_1}] \cdot \dots \cdot \mathbf{E} [e^{iuX_n}] = \varphi_{X_1}(u) \cdot \dots \cdot \varphi_{X_n}(u) \end{aligned}$$

Quest'ultima proprietà sarà particolarmente importante nella discussione dei Teoremi limite e delle proprietà riproduttive: infatti, mentre in base all'equazione (3.69) la *ddp* $f_{S_n}(x)$ della somma di n *v-a* indipendenti è il *prodotto di convoluzione* $(f_{X_1} * \dots * f_{X_n})(x)$ delle rispettive *ddp*, la sua *fc* $\varphi_{S_n}(u)$ è il *prodotto ordinario* $\varphi_{X_1}(u) \cdot \dots \cdot \varphi_{X_n}(u)$ delle corrispondenti *fc* \square

Esempio 4.10. Per le leggi discrete la *fc* si trova calcolando il valore d'attesa (4.2) mediante una somma: in questo modo la *fc* di una **v-a degenera** $X \sim \delta_b$ è

$$\varphi_X(u) = e^{ibu} \quad (4.6)$$

mentre per una **v-a di Bernoulli** $X \sim \mathfrak{B}(1; p)$ si ha

$$\varphi_X(u) = p e^{iu} + q \quad (4.7)$$

Per una **v-a binomiale** $S_n \sim \mathfrak{B}(n; p)$ conviene invece ricordare che in base alla (3.8) risulta $S_n \stackrel{d}{=} X_1 + \dots + X_n$ con X_k Bernoulli iid e pertanto da (4.5) e (4.7) si ha

$$\varphi_X(u) = (p e^{iu} + q)^n \quad (4.8)$$

Infine per una **v-a di Poisson** $X \sim \mathfrak{P}(\alpha)$ la *fc* è

$$\varphi_X(u) = \sum_{k=0}^{\infty} e^{iuk} e^{-\alpha} \frac{\alpha^k}{k!} = e^{-\alpha} \sum_{k=0}^{\infty} \frac{(\alpha e^{iu})^k}{k!} = e^{\alpha(e^{iu}-1)} \quad (4.9)$$

Per le leggi dotate di *ddp* invece il calcolo della *fc* si esegue mediante opportune integrazioni: così per una **v-a uniforme** $X \sim \mathfrak{U}(a, b)$ si ha

$$\varphi_X(u) = \int_a^b \frac{e^{iux}}{b-a} dx = \frac{e^{ibu} - e^{iau}}{i(b-a)u} \quad (4.10)$$

e in particolare per $X \sim \mathfrak{U}(-1, 1)$ risulta

$$\varphi_X(u) = \frac{\sin u}{u} \quad (4.11)$$

Per una **v-a Gaussiana** $X \sim \mathfrak{N}(b, a^2)$ conviene ricordare dall'Esempio 3.47 che $Y = (X - b)/a \sim \mathfrak{N}(0, 1)$ e che da (4.3) si ha

$$\varphi_X(u) = e^{ibu} \varphi_Y(au)$$

sicché sarà sufficiente calcolare la fc φ_Y di una normale standard $\mathfrak{N}(0, 1)$. Usando allora le proprietà di convergenza dello sviluppo in serie di potenze degli esponenziali avremo che

$$\begin{aligned} \varphi_Y(u) &= \mathbf{E} [e^{iuY}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{iux} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} \sum_{n=0}^{\infty} \frac{(iux)^n}{n!} dx = \sum_{n=0}^{\infty} \frac{(iu)^n}{n!} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n e^{-x^2/2} dx \end{aligned}$$

e siccome

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n e^{-x^2/2} dx = \begin{cases} 0 & \text{per } n = 2k + 1 \\ (2k - 1)!! & \text{per } n = 2k \end{cases}$$

si ottiene

$$\varphi_Y(u) = \sum_{k=0}^{\infty} \frac{(iu)^{2k}}{(2k)!} (2k - 1)!! = \sum_{k=0}^{\infty} \left(-\frac{u^2}{2}\right)^k \frac{1}{k!} = e^{-u^2/2} \quad (4.12)$$

da cui in definitiva

$$\varphi_X(u) = e^{ibu - a^2 u^2/2} \quad (4.13)$$

In particolare se $X \sim \mathfrak{N}(0, a^2)$ la ddp e la fc sono rispettivamente

$$f_X(x) = \frac{1}{a\sqrt{2\pi}} e^{-x^2/2a^2} \quad \varphi_X(u) = e^{-a^2 u^2/2}$$

caso che mette bene in evidenza sia il fatto che la fc di una ddp Gaussiana è ancora una funzione Gaussiana, sia la relazione inversa che sussiste fra la larghezza (varianza) a^2 della ddp e la larghezza $1/a^2$ della fc. Con integrazioni elementari si verifica poi che la fc di una **v-a esponenziale** $X \sim \mathfrak{E}(a)$ è

$$\varphi_X(u) = \int_0^{+\infty} a e^{-ax} e^{ixu} dx = \frac{a}{a - iu} = \frac{a^2 + iau}{a^2 + u^2} \quad (4.14)$$

e quella di una **v-a di Laplace** $X \sim \mathfrak{L}(a)$

$$\varphi_X(u) = \int_{-\infty}^{+\infty} \frac{a}{2} e^{-a|x|} e^{ixu} dx = \frac{a^2}{a^2 + u^2} \quad (4.15)$$

Infine per una **v-a di Cauchy** $X \sim \mathfrak{C}(a, b)$ la fc

$$\varphi_X(u) = \int_{-\infty}^{+\infty} \frac{a}{\pi} \frac{e^{ixu}}{a^2 + (x - b)^2} dx = e^{-a|u| + ibu} \quad (4.16)$$

si ricava dal teorema dei residui con un'integrazione in campo complesso

Teorema 4.11. Se X è una v -a con fc $\varphi(u)$, se $\mathbf{E}[|X|^n] < +\infty, \forall n \in \mathbf{N}$, e se

$$\lim_n \frac{\mathbf{E}[|X|^n]^{1/n}}{n} = \frac{1}{R} < +\infty$$

allora $\varphi(u)$ è derivabile ad ogni ordine $n \in \mathbf{N}$ con

$$\varphi^{(n)}(u) = \mathbf{E}[(iX)^n e^{iuX}] \quad \varphi^{(n)}(0) = i^n \mathbf{E}[X^n] \quad (4.17)$$

e inoltre per $|u| < R/3$ si ha lo sviluppo in Serie di Taylor

$$\varphi(u) = \sum_{n=0}^{\infty} \frac{(iu)^n}{n!} \mathbf{E}[X^n] = \sum_{n=0}^{\infty} \frac{u^n}{n!} \varphi^{(n)}(0) \quad (4.18)$$

Se invece $\mathbf{E}[|X|^k] < +\infty$ solo per un numero finito n di esponenti $k = 1, \dots, n$, allora la $\varphi(u)$ risulta derivabile solo fino all'ordine n , e vale la Formula di Taylor

$$\varphi(u) = \sum_{k=0}^n \frac{(iu)^k}{k!} \mathbf{E}[X^k] + o(u^n) = \sum_{k=0}^n \frac{u^k}{k!} \varphi^{(k)}(0) + o(u^n) \quad (4.19)$$

con un resto infinitesimo di ordine superiore a n per $u \rightarrow 0$.

Dimostrazione: Omessa⁵. Osserveremo solo che, una volta verificate le condizioni per il passaggio al limite sotto segno di integrale, la relazione (4.17) non è altro che il risultato di una derivazione multipla sotto segno di integrale. Quanto allo sviluppo (4.18) esso essenzialmente discende dallo sviluppo in serie di Taylor dell'esponenziale secondo lo schema euristico

$$\varphi(u) = \mathbf{E}[e^{iuX}] = \mathbf{E}\left[\sum_{n=0}^{\infty} \frac{(iu)^n}{n!} X^n\right] = \sum_{n=0}^{\infty} \frac{(iu)^n}{n!} \mathbf{E}[X^n] = \sum_{n=0}^{\infty} \frac{u^n}{n!} \varphi^{(n)}(0)$$

Questi risultati mettono in evidenza l'importante relazione che intercorre fra fc $\varphi(u)$ e momenti $\mathbf{E}[X^n]$ di una v -a X : ulteriori dettagli sul cosiddetto **problema dei momenti** e sui **cumulanti** sono riportati nell'Appendice E □

Con una dimostrazione del tutto analoga il precedente teorema può essere esteso anche al caso di sviluppi attorno al punto $u = v$ invece che $u = 0$, e in tal caso – con un opportuno raggio di convergenza – si avrà lo sviluppo

$$\varphi(u) = \sum_{n=0}^{\infty} \frac{i^n (u - v)^n}{n!} \mathbf{E}[X^n e^{ivX}]$$

⁵N. Cufaro Petroni, CALCOLO DELLE PROBABILITÀ, Edizioni dal Sud (Bari, 1996)

Teorema 4.12. Teorema di unicità: *Se $f(x)$ e $g(x)$ sono due ddp dotate della stessa fc, cioè se*

$$\int_{-\infty}^{+\infty} e^{iux} f(x) dx = \int_{-\infty}^{+\infty} e^{iux} g(x) dx, \quad \forall u \in \mathbf{R}$$

allora risulta anche $f(x) = g(x)$ in ogni $x \in \mathbf{R}$, con la possibile eccezione di un insieme di punti di misura di Lebesgue nulla

Dimostrazione: Omessa⁶ □

Teorema 4.13. Formula di inversione: *Se $\varphi(u)$ è la fc di una legge ac, allora la corrispondente ddp è*

$$f(x) = \frac{1}{2\pi} \lim_{T \rightarrow +\infty} \int_{-T}^T e^{-iux} \varphi(u) du = \frac{1}{2\pi} VP \int_{-\infty}^{+\infty} e^{-iux} \varphi(u) du \quad (4.20)$$

Dimostrazione: Omessa⁷ □

Teorema 4.14. Condizione necessaria e sufficiente affinché le componenti di un vett-a $\mathbf{X} = (X_1, \dots, X_n)$ siano indipendenti è che

$$\varphi_{\mathbf{X}}(u_1, \dots, u_n) = \varphi_{X_1}(u_1) \cdot \dots \cdot \varphi_{X_n}(u_n)$$

cioè che la fc congiunta $\varphi_{\mathbf{X}}(\mathbf{u})$ sia il prodotto delle fc marginali $\varphi_{X_k}(u_k)$ delle componenti

Dimostrazione: Omessa⁸ □

In pratica tutti questi risultati suggeriscono che la legge di una v -a può essere equivalentemente rappresentata sia dalla sua ddp (o dalla fdd quando essa non è ac), sia anche dalla sua fc: infatti la conoscenza dell'una consente di ricavare l'altra in maniera univoca, e viceversa. Inoltre tutte le informazioni rilevanti (come attese e momenti) possono essere ottenute, in maniera separata, sia dalla ddp che dalla fc. Prima però di accettare definitivamente l'idea che la legge di una v -a possa essere ben rappresentata dalla sola fc, dobbiamo far notare un punto piuttosto delicato: la difficoltà con la quale a volte si può stabilire se una data funzione $\varphi(u)$ è la fc di qualche legge o meno. Che una funzione $f(x)$ sia una possibile ddp è piuttosto facile da controllare: basta che sia una funzione reale, non negativa e normalizzata. Per una fc invece non basta sapere che $\varphi(u)$ ammette una trasformata di Fourier inversa secondo la formula (4.20): bisogna anche essere certi (e senza eseguire direttamente calcoli che spesso sono complicati) che la trasformata inversa sia una buona ddp. Insomma abbiamo bisogno di una qualche caratterizzazione intrinseca di $\varphi(u)$ che ci permetta di essere sicuri di poterla considerare una buona fc.

⁶A.N. Shiryaev, **PROBABILITY**, Springer (New York, 1996)

⁷A.N. Shiryaev, **PROBABILITY**, Springer (New York, 1996)

⁸N. Cufaro Petroni, **CALCOLO DELLE PROBABILITÀ**, Edizioni dal Sud (Bari, 1996)

Teorema 4.15. Teorema di Bochner: Una funzione continua $\varphi(u)$ è una *fc* se e solo se essa è definita non negativa⁹, e $\varphi(0) = 1$

Dimostrazione: Omessa¹⁰. A questo proposito osserveremo solo che, se $\varphi(u)$ è la *fc* di una *v-a* X , sappiamo che essa è (uniformemente) continua e che $\varphi(0) = 1$. Inoltre è facile controllare che, comunque scelti u_1, \dots, u_n , e comunque presi n numeri complessi z_1, \dots, z_n , risulta

$$\begin{aligned} \sum_{j,k=1}^n z_j \bar{z}_k \varphi(u_j - u_k) &= \sum_{j,k=1}^n z_j \bar{z}_k \mathbf{E} [e^{i(u_j - u_k)X}] = \mathbf{E} \left[\sum_{j,k=1}^n z_j \bar{z}_k e^{iu_j X} e^{-iu_k X} \right] \\ &= \mathbf{E} \left[\sum_{j=1}^n z_j e^{iu_j X} \overline{\sum_{k=1}^n z_k e^{iu_k X}} \right] = \mathbf{E} \left[\left| \sum_{j=1}^n z_j e^{iu_j X} \right|^2 \right] \geq 0 \end{aligned}$$

cioè che $\varphi(u)$ è *definita non negativa*. Il Teorema di Bochner afferma che è vero anche l'inverso: qualunque funzione di variabile reale e a valori complessi $\varphi(u)$ che goda di queste tre proprietà è una buona *fc* \square

La stretta relazione che sussiste fra la *fdd* $F(x)$ (o la *ddp* $f(x)$) e la *fc* $\varphi(t)$ di una legge suggerisce che la convergenza debole di una successione di *fdd* $(F_n(x))_{n \in \mathbf{N}}$ possa essere analizzata mediante lo studio della convergenza puntuale della corrispondente successione di *fc* $(\varphi_n(u))_{n \in \mathbf{N}}$. I risultati del teorema seguente individuano le condizioni sotto le quali la convergenza debole $F_n \xrightarrow{w} F$ è equivalente alla convergenza punto a punto $\varphi_n(u) \rightarrow \varphi(u)$ delle corrispondenti *fc*

Teorema 4.16. Teorema di continuità di Paul Lévy: Data la successione di *fdd* $(F_n(x))_{n \in \mathbf{N}}$ e la corrispondente successione di *fc* $(\varphi_n(u))_{n \in \mathbf{N}}$

1. se $F_n \xrightarrow{w} F$ e se $F(x)$ risulta essere una *fdd*, allora anche $\varphi_n(u) \xrightarrow{n} \varphi(u)$ in ogni $u \in \mathbf{R}$, e $\varphi(u)$ risulta essere la *fc* di $F(x)$;
2. se il limite $\varphi(u) = \lim_n \varphi_n(u)$ esiste per ogni $u \in \mathbf{R}$, e se $\varphi(u)$ è continua in $u = 0$, allora $\varphi(u)$ è la *fc* di una *fdd* $F(x)$ e risulta $F_n \xrightarrow{w} F$
3. se in particolare sappiamo a priori che $F(x)$ è una *fdd* e $\varphi(u)$ la sua *fc*, allora $F_n \xrightarrow{w} F$ se e solo se $\varphi_n(u) \xrightarrow{n} \varphi(u)$ in ogni $u \in \mathbf{R}$

Dimostrazione: Omessa¹¹ \square

⁹Si dice che una funzione $\varphi(u)$ è *definita non negativa* quando, comunque scelti n punti u_1, \dots, u_n , la matrice $\|\varphi(u_j - u_k)\|$ risulta definita non negativa, cioè quando, comunque presi n numeri complessi z_1, \dots, z_n , risulta

$$\sum_{j,k=1}^n z_j \bar{z}_k \varphi(u_j - u_k) \geq 0 \tag{4.21}$$

¹⁰A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

¹¹A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

4.2.2 Leggi Gaussiane

I *vett-a* con legge congiunta Gaussiana $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ giocano un ruolo estremamente importante in probabilità e statistica. Questo è dovuto innanzitutto, come vedremo nella Sezione 4.4, al cosiddetto *Teorema Limite Centrale* secondo il quale somme di un gran numero di *v-a* indipendenti, ma con leggi arbitrarie e soggette a condizioni non troppo restrittive, tendono a distribuirsi in maniera Gaussiana. È questa la base teorica della cosiddetta *Legge degli errori* secondo la quale gli errori casuali commessi nelle misure fisiche – errori che derivano dalla somma di un gran numero di piccoli disturbi elementari indipendenti e incontrollabili – si distribuiscono in maniera Gaussiana. In secondo luogo le *v-a* Gaussiane godono di altre proprietà caratteristiche:

- hanno leggi $\mathfrak{N}(b, a^2)$ determinate solo da un numero piccolo (due) di parametri
- sono caratterizzate dall'equivalenza fra *indipendenza e non correlazione*, una proprietà non posseduta dalle altre *v-a* (come visto nella Sezione 3.3.3)
- possiedono *momenti finiti* di ogni ordine e possono quindi essere analizzate con le tecniche di analisi funzionale introdotte nell'Appendice D

Data la loro importanza sarà quindi utile determinare la maniera più efficace e completa per rappresentare la famiglia di leggi Gaussiane $\mathfrak{N}(b, a^2)$, e qui di seguito esamineremo questo problema alla luce delle proprietà delle loro *fc*.

Come è noto, quando $a > 0$, la *ddp* di una *v-a* $X \sim \mathfrak{N}(b, a^2)$ è

$$f_X(x) = \frac{1}{a\sqrt{2\pi}} e^{-(x-b)^2/2a^2}$$

Siccome $a^2 = \mathbf{V}[X]$, quando $a \downarrow 0$ la legge di X converge intuitivamente verso quella di una *v-a* degenera che prende solo il valore $X = b$, \mathbf{P} -qo. D'altra parte è noto che una *v-a* degenera in b è caratterizzata da una legge δ_b tipicamente non continua, e quindi non dotata di *ddp*. Per questo motivo – nella misura in cui ci limitiamo a rappresentare le leggi solo mediante le loro *ddp* – siamo obbligati a tenere separato il caso $a > 0$ (in cui X è dotata di *ddp* Gaussiana) dal caso $a = 0$ (in cui X degenera attorno al valore b e non può essere descritta da una *ddp*), e ad ammettere che le due descrizioni non passano con continuità una nell'altra quando $a \downarrow 0$. Per aggirare questa difficoltà ricorderemo allora che una *v-a*, oltre che mediante le sue *fdd* o *ddp*, può essere descritta altrettanto bene tramite la sua *fc*, e che per le *v-a* qui considerate abbiamo da (4.6) e (4.13)

$$\varphi_X(u) = \begin{cases} e^{ibu} & \text{se } a = 0, \quad \text{legge } \delta_b \\ e^{ibu - u^2 a^2 / 2} & \text{se } a > 0, \quad \text{legge } \mathfrak{N}(b, a^2) \end{cases}$$

È allora evidente che – diversamente dalla rispettiva *ddp* – la *fc* con $a = 0$ si ottiene con continuità da quella per $a > 0$ al limite per $a \downarrow 0$, e che potremo parlare della

famiglia di leggi $\mathfrak{N}(b, a^2)$ per $a \geq 0$, con $\mathfrak{N}(b, 0) = \delta_b$, nel senso che ora tutte queste leggi saranno rappresentate dalla famiglia di *fc*

$$\varphi_X(u) = e^{ibu - u^2 a^2 / 2} \quad a \geq 0$$

in modo che il caso *degenere* non sia altro (come è anche intuitivamente chiaro) che il caso limite del caso *non degenere*.

Queste osservazioni possono essere estese anche alla definizione dei *vett-a* Gaussiani $\mathbf{X} \sim \mathfrak{N}(\mathbf{b}, \mathbb{A})$ per i quali – con una descrizione in termini di *ddp* – andrebbero distinti i casi in cui la matrice definita non negativa delle covarianze \mathbb{A} è non singolare (definita positiva), da quelli in cui essa è singolare ($|\mathbb{A}| = 0$). Va notato, peraltro, a questo proposito che la difficoltà di dare una descrizione unitaria per *vett-a* con più di una sola componente è accresciuta dal fatto che nei casi singolari le varie componenti possono esibire comportamenti differenti: non è esclusa, infatti, l'eventualità che alcune componenti siano degeneri e altre no, e che quindi la distribuzione sia intrinsecamente mista. L'uso delle *fc* ci permette invece ancora una volta di dare una descrizione unitaria e coerente

Definizione 4.17. Diremo che $\mathbf{X} = (X_1, \dots, X_n) \sim \mathfrak{N}(\mathbf{b}, \mathbb{A})$ è un **vett-a Gaussiano (o normale)** con vettore delle medie $\mathbf{b} = (b_1, \dots, b_n) \in \mathbf{R}^n$ e matrice delle covarianze simmetrica e definita non negativa $\mathbb{A} = \|a_{kl}\|$, se la sua *fc* ha la forma

$$\varphi_{\mathbf{X}}(\mathbf{u}) = \varphi_{\mathbf{X}}(u_1, \dots, u_n) = e^{i\mathbf{b} \cdot \mathbf{u}} e^{-\mathbf{u} \cdot \mathbb{A} \mathbf{u} / 2} \quad \mathbf{u} \in \mathbf{R}^n \quad (4.22)$$

dove con $\mathbf{b} \cdot \mathbf{u} = \sum_k b_k u_k$ indichiamo l'usuale prodotto scalare tra vettori in \mathbf{R}^n

La *fc* (4.22) è ovviamente una diretta generalizzazione della *fc* (4.13) per *v-a* Gaussiane con una sola componente, e riassume quest'ultima forma se b è un numero e la matrice delle covarianze si riduce a un solo elemento a^2 . Si noti che – rispetto alla formulazione (2.22) in termini di *ddp* – la *fc* (4.22) fa uso solo della matrice \mathbb{A} , e non della sua inversa \mathbb{A}^{-1} , e quindi non risente della sua eventuale singolarità. Siccome però il caso singolare costituisce un'estensione di quello che finora avevamo inteso per *vett-a* Gaussiano, senza peraltro essere riducibile nel caso singolare ad una semplice legge degenere, dovremo verificare nella successiva proposizione che la Definizione 4.17 sia completamente accettabile e coerente

Proposizione 4.18. Nel caso non singolare $|\mathbb{A}| \neq 0$ la (4.22) è la *fc* di una *ddp* Gaussiana (2.22); nel caso singolare $|\mathbb{A}| = 0$ essa è comunque la *fc* di una legge che continueremo a chiamare Gaussiana $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ pur non essendo dotata di *ddp*

Dimostrazione: Se \mathbb{A} è una matrice non singolare (cioè se $|\mathbb{A}| \neq 0$) si può definire la sua inversa \mathbb{A}^{-1} e si può mostrare con un calcolo diretto (qui omesso) della trasformata di Fourier inversa che la $\varphi_{\mathbf{X}}(\mathbf{u})$ della Definizione 4.17 è proprio la *fc* di un *vett-a* $\mathbf{X} \sim \mathfrak{N}(\mathbf{b}, \mathbb{A})$ dotato di *ddp* normale multivariata (2.22)

$$f_{\mathbf{X}}(\mathbf{x}) = \sqrt{\frac{|\mathbb{A}^{-1}|}{(2\pi)^n}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{b}) \cdot \mathbb{A}^{-1}(\mathbf{x}-\mathbf{b})}$$

Se invece \mathbb{A} è una matrice singolare (se, cioè, $|\mathbb{A}| = 0$), \mathbb{A}^{-1} non esiste e la (4.22) non potrà più essere considerata come la trasformata di Fourier di una qualche *ddp*. Ciononostante si può mostrare che essa continua ad essere la *fc* della legge di un opportuno *vett-a* non dotato di *ddp*. Infatti, preso $n \in \mathbf{N}$, si considera la matrice $\mathbb{A}_n = \mathbb{A} + \frac{1}{n} \mathbb{I}$, dove \mathbb{I} è la matrice identità, e si osserva che essa è ancora simmetrica e definita non negativa, ma che – a differenza di \mathbb{A} – risulta anche non singolare per ogni $n \in \mathbf{N}$. Ne segue allora che per ogni $n \in \mathbf{N}$ l'inversa \mathbb{A}_n^{-1} esiste e la funzione

$$\varphi_n(\mathbf{u}) = e^{i\mathbf{b} \cdot \mathbf{u}} e^{-\mathbf{u} \cdot \mathbb{A}_n \mathbf{u} / 2}$$

è la *fc* di un *vett-a* con legge $\mathfrak{N}(\mathbf{b}, \mathbb{A}_n)$ dotato di opportuna *ddp* Gaussiana. Siccome però per ogni \mathbf{u} fissato risulta ovviamente

$$\lim_n \varphi_n(\mathbf{u}) = e^{i\mathbf{b} \cdot \mathbf{u}} e^{-\mathbf{u} \cdot \mathbb{A} \mathbf{u} / 2} = \varphi_{\mathbf{X}}(\mathbf{u})$$

e la funzione limite (4.22) è continua in $\mathbf{u} = (0, \dots, 0)$, il Teorema di continuità 4.16 ci garantisce che $\varphi_{\mathbf{X}}(\mathbf{u})$ è la *fc* di una legge, anche se quest'ultima non è dotata di *ddp*. I *vett-a* distribuiti secondo questa legge ottenuta come limite di *fc* Gaussiane, saranno quindi considerati a pieno titolo come dei *vett-a* Gaussiani con legge $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ nel caso singolare $|\mathbb{A}| = 0$ □

Proposizione 4.19. *Dato un vett-a Gaussiano $\mathbf{X} = (X_1, \dots, X_n) \sim \mathfrak{N}(\mathbf{b}, \mathbb{A})$ si ha*

$$b_k = \mathbf{E}[X_k] ; \quad a_{kl} = \mathbf{cov}[X_k, X_l] \quad a_{kk} = a_k^2 = \mathbf{V}[X_k]$$

e le componenti $X_k \sim \mathfrak{N}(b_k, a_k^2)$ sono indipendenti se e solo se sono non correlate

Dimostrazione: Il significato probabilistico degli elementi di \mathbf{b} e di $\mathbb{A} = \|a_{kl}\|$ (già discusso nell'Esempio 3.34 per il caso bivariato non degenere) si ricava con un calcolo diretto, qui omesso, facendo uso della *fc* (4.22). Che le componenti X_k siano ancora Gaussiane $\mathfrak{N}(b_k, a_k^2)$ (come anticipato senza prova nell'Esempio 3.13) può poi essere dimostrato evitando una laboriosa marginalizzazione tramite integrali: infatti da (4.4) si ricava subito che le *fc* marginali del nostro *vett-a* Gaussiano sono

$$\varphi_{X_k}(u_k) = e^{ib_k u_k} e^{-u_k^2 a_k^2 / 2}$$

e quindi sono a loro volta Gaussiane $\mathfrak{N}(b_k, a_k^2)$. Infine l'equivalenza fra indipendenza e non correlazione delle componenti (già discussa nell'Esempio 3.34 per il caso bivariato non degenere) può ora essere dimostrata del tutto in generale: innanzitutto è ovvio che se le X_k sono indipendenti esse sono anche non correlate. Viceversa, se le componenti di un *vett-a* Gaussiano $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ sono non correlate la matrice delle covarianze \mathbb{A} risulta diagonale con $a_{kl} = \delta_{kl} a_k^2$ e pertanto la sua *fc* è

$$\varphi_{\mathbf{X}}(u_1, \dots, u_n) = e^{i\mathbf{b} \cdot \mathbf{u}} e^{-\sum_k a_k^2 u_k^2 / 2} = \prod_{k=1}^n (e^{ib_k u_k} e^{-a_k^2 u_k^2 / 2}) = \prod_{k=1}^n \varphi_{X_k}(u_k)$$

dove le $\varphi_{X_k}(u_k)$ sono le *fc* delle singole componenti. Ne segue in base al Teorema 4.14 che le componenti di \mathbf{X} sono indipendenti. □

Proposizione 4.20. *Dato il vett-a $\mathbf{X} = (X_1, \dots, X_n)$, le seguenti affermazioni sono equivalenti*

1. $\mathbf{X} \sim \mathfrak{N}(\mathbf{b}, \mathbb{A})$
2. $\mathbf{c} \cdot \mathbf{X} \sim \mathfrak{N}(\mathbf{c} \cdot \mathbf{b}, \mathbf{c} \cdot \mathbb{A} \mathbf{c})$ *comunque scelto $\mathbf{c} \in \mathbf{R}^n$*
3. $\mathbf{X} = \mathbb{C} \mathbf{Y} + \mathbf{b}$ *dove $\mathbf{Y} \sim \mathfrak{N}(0, \mathbb{I})$, \mathbb{C} è non singolare e $\mathbb{A} = \mathbb{C} \mathbb{C}^T$*

Dimostrazione: Omessa¹². Si noti in particolare che nel punto 3 il vett-a \mathbf{Y} è Gaussiano con componenti Y_k normali standard $\mathfrak{N}(0, 1)$ e *indipendenti* dato che la sua matrice delle covarianze è δ_{jk} . Pertanto la proposizione afferma tra l'altro che le componenti di un generico vett-a Gaussiano $\mathbf{X} \sim \mathfrak{N}(\mathbf{b}, \mathbb{A})$ risultano sempre essere combinazioni lineari delle componenti normali standard e indipendenti del vett-a $\mathbf{Y} \sim \mathfrak{N}(0, \mathbb{I})$, e quindi che, viceversa, opportune combinazioni lineari delle componenti di un qualunque vett-a $\mathbf{X} \sim \mathfrak{N}(\mathbf{b}, \mathbb{A})$ possono sempre essere rese standard e indipendenti \square

4.2.3 Composizione e decomposizione di leggi

Con il termine di **proprietà riproduttiva di una famiglia di leggi** si suole indicare il fatto che tale famiglia è chiusa sotto l'operazione di convoluzione, nel senso che la **composizione** tramite convoluzione delle *ddp* di due o più leggi della famiglia produce ancora una legge della stessa famiglia. Abbiamo già incontrato le proprietà riproduttive (3.70) delle *v-a* normali $\mathfrak{N}(b, a^2)$ nella Sezione 3.5.2, ma ne abbiamo rinviato la verifica esplicita per evitare delle integrazioni che in questo, come in molti altri casi, potrebbero non essere di facile esecuzione. L'uso delle *fc* permette invece anche in questa occasione una grande semplificazione in quanto, come sappiamo dalla Proposizione 4.9, alla composizione tramite convoluzione delle *ddp* corrisponde la banale composizione tramite prodotto semplice delle *fc*

Esempio 4.21. *Le proprietà riproduttive (3.70) delle leggi Gaussiane $\mathfrak{N}(b, a^2)$*

$$\mathfrak{N}(b_1, a_1^2) * \mathfrak{N}(b_2, a_2^2) = \mathfrak{N}(b_1 + b_2, a_1^2 + a_2^2)$$

*si giustificano ricordando semplicemente la Proposizione 4.9 e (4.13), e osservando che il prodotto delle *fc* $\varphi_1(u)$ e $\varphi_2(u)$ delle leggi $\mathfrak{N}(b_1, a_1^2)$ e $\mathfrak{N}(b_2, a_2^2)$ è*

$$\varphi(u) = \varphi_1(u)\varphi_2(u) = e^{ib_1u - a_1^2u^2/2} e^{ib_2u - a_2^2u^2/2} = e^{i(b_1+b_2)u - (a_1^2+a_2^2)u^2/2}$$

*cioè è la *fc* di una legge $\mathfrak{N}(b_1 + b_2, a_1^2 + a_2^2)$. Pertanto la famiglia di leggi $\mathfrak{N}(b, a^2)$ al variare dei parametri a e b è chiusa sotto operazioni di composizione mediante*

¹²N. Cufaro Petroni, CALCOLO DELLE PROBABILITÀ, Edizioni dal Sud (Bari, 1996)

convoluzione. Come caso particolare con $a_1 = a_2 = 0$ si ritrovano poi anche le proprietà riproduttive delle **leggi degeneri** δ_b

$$\delta_{b_1} * \delta_{b_2} = \delta_{b_1+b_2} \quad (4.23)$$

Con lo stesso metodo si può dimostrare che anche le **leggi di Poisson** $\mathfrak{P}(\alpha)$ godono delle loro proprietà riproduttive nel senso che

$$\mathfrak{P}(\alpha_1) * \mathfrak{P}(\alpha_2) = \mathfrak{P}(\alpha_1 + \alpha_2) \quad (4.24)$$

Infatti da (4.9) vediamo che il prodotto delle fc delle leggi $\mathfrak{P}(\alpha_1)$ e $\mathfrak{P}(\alpha_2)$ è

$$\varphi(u) = e^{\alpha_1(e^{iu}-1)} e^{\alpha_2(e^{iu}-1)} = e^{(\alpha_1+\alpha_2)(e^{iu}-1)}$$

che coincide con la fc della legge $\mathfrak{P}(\alpha_1 + \alpha_2)$.

Le famiglie δ_b , $\mathfrak{N}(b, a^2)$ e $\mathfrak{P}(\alpha)$ – la cui importanza sarà messa ben in evidenza dalla successiva discussione sui teoremi limite – godono anche di un’ulteriore proprietà: esse sono chiuse non solo sotto composizione, ma anche sotto **decomposizione** mediante convoluzioni. Così, ad esempio, se una legge Gaussiana $\mathfrak{N}(b, a^2)$ si decompone nella convoluzione di altre due leggi, queste ultime non possono che essere anche loro due leggi Gaussiane della famiglia $\mathfrak{N}(b, a^2)$. In altri termini, non solo le leggi ottenute componendo due Gaussiane sono ancora Gaussiane, ma *solo* componendo leggi Gaussiane si può ottenere tale risultato. In questo senso si dice che la famiglia $\mathfrak{N}(b, a^2)$ è *chiusa sotto composizione e decomposizione* mediante convoluzioni. Lo stesso risultato è vero anche per le famiglie δ_b e $\mathfrak{P}(\alpha)$, ma, diversamente da quelli sulle composizioni, i risultati sulle decomposizioni sono piuttosto difficili da dimostrare¹³

Esempio 4.22. *Ci sono ancora altre famiglie di leggi chiuse sotto composizione mediante convoluzione: con gli stessi semplici metodi si provano infatti da (4.16) le proprietà riproduttive delle **leggi di Cauchy** $\mathfrak{C}(a, b)$*

$$\mathfrak{C}(a_1, b_1) * \mathfrak{C}(a_2, b_2) = \mathfrak{C}(a_1 + a_2, b_1 + b_2) \quad (4.25)$$

Non bisogna però pensare che si tratti di proprietà del tutto generali: ad esempio una combinazione di **leggi esponenziali** $\mathfrak{E}(a)$ non riproduce un’altra esponenziale con un nuovo parametro, ma produce invece una legge di un genere differente. Infatti si vede facilmente da (4.14) che se $\varphi_1(u)$ e $\varphi_2(u)$ sono le fc delle leggi $\mathfrak{E}(a_1)$ e $\mathfrak{E}(a_2)$, il loro prodotto

$$\varphi(u) = \varphi_1(u)\varphi_2(u) = \frac{a_1}{a_1 - iu} \frac{a_2}{a_2 - iu}$$

non è più la fc di una legge esponenziale. Tralasciando uno studio di quest’altro tipo di leggi, osserveremo invece che limitandoci a comporre solo leggi esponenziali

¹³M. Loève, **PROBABILITY THEORY - I**, Springer (New York, 1977)

con il medesimo parametro a si producono altre famiglie di leggi che saranno usate nel seguito: infatti il prodotto delle fc $\varphi_a(u)$ di n leggi esponenziali $\mathfrak{E}(a)$ tutte con lo stesso a è

$$\varphi(u) = \varphi_a^n(u) = \left(\frac{a}{a - iu} \right)^n \quad (4.26)$$

e si verifica facilmente con un calcolo diretto che la corrispondente ddp è

$$f_Z(x) = \frac{(ax)^{n-1}}{(n-1)!} ae^{-ax} \vartheta(x) \quad n = 1, 2, \dots \quad (4.27)$$

dove $\vartheta(x)$ è la funzione di Heaviside. Le leggi di questa nuova famiglia portano il nome di **leggi di Erlang** $\mathfrak{E}_n(a)$, ed è facile verificare da (3.33) e (3.38) che se $X \sim \mathfrak{E}_n(a)$, allora

$$\mathbf{E}[X] = \frac{n}{a} \quad \mathbf{V}[X] = \frac{n}{a^2} \quad (4.28)$$

Si noti la particolare reciprocità che sussiste fra le distribuzioni di Erlang (4.27) e quelle di Poisson: l'espressione

$$\frac{x^k e^{-x}}{k!} \vartheta(x) \quad k = 0, 1, 2, \dots$$

può infatti essere considerata sia come una distribuzione discreta di Poisson $\mathfrak{P}(x)$ in k (con parametro $x > 0$), sia come una ddp di Erlang $\mathfrak{E}_{k+1}(1)$ in x (con $a = 1$)

4.3 Leggi dei Grandi Numeri

I Teoremi Limite classici sono proposizioni riguardanti limiti per $n \rightarrow \infty$ di somme $S_n = X_1 + \dots + X_n$ di successioni $(X_n)_{n \in \mathbf{N}}$ di v - a , e in essi, come vedremo, giocano un ruolo preminente le famiglie di leggi δ_b , $\mathfrak{N}(b, a^2)$ e $\mathfrak{P}(\alpha)$. Va detto subito però che tali Teoremi Limite non sono una conseguenza delle proprietà di composizione e decomposizione tipiche di queste tre famiglie che abbiamo richiamato nella Sezione 4.2.3. Si tratta invece di risultati più profondi che travalicano i confini delle precedenti discussioni. Innanzitutto, infatti, mentre le proprietà di composizione e decomposizione si riferiscono solo a *somme finite* di v - a , i teoremi che ora esamineremo sono risultati su *limiti di successioni di somme* di v - a . Inoltre, mentre nelle proprietà di composizione e decomposizione una v - a , ad esempio Gaussiana, è somma di un numero finito di altre v - a indipendenti e ancora *Gaussiane*, nel Teorema Limite Centrale la legge normale emerge come limite in distribuzione di somme di v - a indipendenti con *legge arbitraria* all'interno di condizioni molto larghe. Cominceremo discutendo le Leggi dei Grandi Numeri che, diversamente dai Teoremi Gaussiani che saranno discussi nella Sezione 4.4, sono un caso di *convergenza degenera*: la successione S_n convergerà infatti verso un *numero*, ovvero verso una v - a che assume un solo valore \mathbf{P} -qo. La versione più antica di questo primo importante risultato, il Teorema di Bernoulli (1713), è richiamata nell'Appendice F

Teorema 4.23. Legge debole dei Grandi Numeri: *Data una successione $(X_n)_{n \in \mathbf{N}}$ di v-a iid con $\mathbf{E}[|X_n|] < +\infty$, e posto $S_n = X_1 + \dots + X_n$ e $\mathbf{E}[X_n] = m$, risulta*

$$\frac{S_n}{n} \xrightarrow{\mathbf{P}} m$$

Dimostrazione: In questa formulazione le X_k non sono in generale v-a di Bernoulli, e quindi S_n non è più binomiale, sicché sarà necessario introdurre un metodo di dimostrazione diverso da quello utilizzato da Bernoulli. Osserviamo allora innanzitutto che, in base al punto 4 del Teorema 4.4, la convergenza *degenere* in probabilità (nel nostro caso verso il numero m) è *equivalente* alla convergenza in distribuzione verso la medesima costante e quindi per la dimostrazione potremo legittimamente usare il Teorema 4.16. Se allora $\varphi(u)$ è la fc delle X_n , le fc delle S_n/n saranno

$$\varphi_n(u) = \mathbf{E}[e^{iuS_n/n}] = \prod_{k=1}^n \mathbf{E}[e^{iuX_k/n}] = \left[\varphi\left(\frac{u}{n}\right) \right]^n$$

D'altra parte, dato che le nostre v-a sono integrabili, dalla (4.19) si ha

$$\varphi(u) = 1 + ium + o(u) \quad u \rightarrow 0$$

e quindi, con u arbitrario ma fissato,

$$\varphi\left(\frac{u}{n}\right) = 1 + i\frac{u}{n}m + o\left(\frac{1}{n}\right) \quad n \rightarrow \infty$$

Ne segue che, per ogni $u \in \mathbf{R}$, avremo

$$\varphi_n(u) = \left[1 + i\frac{u}{n}m + o\left(\frac{1}{n}\right) \right]^n \xrightarrow{n} e^{imu}$$

e siccome e^{imu} è la fc di una v-a degenere attorno al valore m , il risultato segue dal Teorema 4.16. \square

Esiste anche una versione della *Legge debole dei Grandi Numeri* adatta a successioni di v-a indipendenti, ma *non identicamente distribuite*. Converrà osservare a questo scopo che la tesi del Teorema 4.23 può essere anche posta nella forma

$$\frac{S_n - \mathbf{E}[S_n]}{n} \xrightarrow{\mathbf{P}} 0 \tag{4.29}$$

che, non facendo più richiamo a un comune valore delle attese, è più adatta a trattare il caso delle X_n indipendenti, ma non identicamente distribuite. Nel teorema seguente mostreremo che l'ipotesi di identica distribuzione può essere sostituita da opportune ipotesi sulle varianze $\mathbf{V}[X_n]$ che ora devono essere supposte finite

Teorema 4.24. *Se le v-a di $(X_n)_{n \in \mathbf{N}}$ sono indipendenti, se $\mathbf{E} [|X_n|^2] < +\infty$ e posto $S_n = X_1 + \dots + X_n$, se esiste un numero $C > 0$ tale che*

$$\mathbf{V} [X_n] < C \quad \forall n \in \mathbf{N}$$

allora risulta

$$\frac{S_n - \mathbf{E} [S_n]}{n} \xrightarrow{\mathbf{P}} 0$$

Dimostrazione: Dalla disuguaglianza di Chebyshev (3.42), e per $\epsilon > 0$ arbitrario

$$\begin{aligned} \mathbf{P} \left\{ \left| \frac{S_n - \mathbf{E} [S_n]}{n} \right| \geq \epsilon \right\} &\leq \frac{1}{\epsilon^2} \mathbf{V} \left[\frac{S_n - \mathbf{E} [S_n]}{n} \right] = \frac{1}{n^2 \epsilon^2} \mathbf{V} \left[\sum_{k=1}^n (X_k - \mathbf{E} [X_k]) \right] \\ &= \frac{1}{n^2 \epsilon^2} \sum_{k=1}^n \mathbf{V} [X_k - \mathbf{E} [X_k]] = \frac{1}{n^2 \epsilon^2} \sum_{k=1}^n \mathbf{V} [X_k] \leq \frac{nC}{n^2 \epsilon^2} \\ &= \frac{C}{n \epsilon^2} \xrightarrow{n} 0 \end{aligned}$$

e il teorema è dimostrato per definizione di convergenza in probabilità. □

La Legge dei Grandi Numeri gioca un ruolo fondamentale nel calcolo delle probabilità in quanto, tra l'altro, consente di stimare il valore d'attesa di una v-a X con la media aritmetica dei risultati di un gran numero di osservazioni indipendenti. In pratica si considera una successione $(X_n)_{n \in \mathbf{N}}$ di misure indipendenti di X (le X_n sono quindi *iid*) e si calcola la media aritmetica S_n/n . In base al Teorema 4.23 potremo allora affermare *con grande confidenza* che il valore osservato della media S_n/n si discosta dal numero $\mathbf{E} [X]$ di una quantità infinitesima con n . L'espressione *con grande confidenza*, però, resta per il momento problematica: le Leggi dei Grandi Numeri formulate finora, infatti, assicurano la convergenza di S_n/n verso $\mathbf{E} [X]$ solamente in probabilità, e non \mathbf{P} -qo. Pertanto, la probabilità che la successione S_n/n non converga verso $\mathbf{E} [X]$ potrebbe, a rigore, essere diversa da zero. Se non disponessimo di risultati più forti dei Teoremi 4.23 e 4.24, potremmo sospettare che, con probabilità non nulla, la media di una successione di misure non converga verso $\mathbf{E} [X]$, e questo creerebbe una situazione particolarmente delicata in tutte le applicazioni. Per questo motivo grandi sforzi sono stati dedicati alla ricerca di Leggi dei Grandi Numeri in forma forte (vere cioè \mathbf{P} -qo) in modo da assicurare la correttezza delle procedure empiriche con probabilità 1

Teorema 4.25. Legge forte dei Grandi Numeri: *Data una successione $(X_n)_{n \in \mathbf{N}}$ di v-a iid con $\mathbf{E} [|X_n|] < +\infty$, e posto $S_n = X_1 + \dots + X_n$ e $\mathbf{E} [X_n] = m$, risulta*

$$\frac{S_n}{n} \xrightarrow{qo} m$$

Dimostrazione: Omessa¹⁴. Si noti che le ipotesi di questo Teorema forte coincidono con quelle del Teorema debole 4.23: la differenza dell'enunciato (convergenza

¹⁴A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

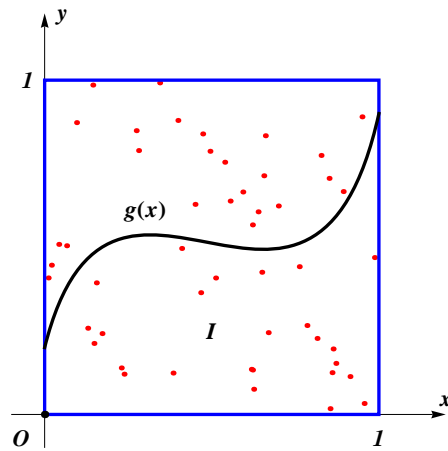


Figura 4.2: Calcolo dell'integrale definito (4.30) con il metodo di Monte Carlo

\mathbf{P} -qo, invece di convergenza in probabilità) è giustificata solo dall'uso di tecniche di dimostrazione più evolute. Si prova poi che è possibile fare a meno dell'identica distribuzione delle X_n a prezzo di qualche ipotesi aggiuntiva sulle varianze, e che i risultati restano validi anche se il valore d'attesa $\mathbf{E}[X_n]$ esiste ma non è finito. Noi però eviteremo di entrare nei dettagli tecnici di questi importanti risultati, e mostreremo invece alcuni esempi di applicazione pratica della Legge forte dei Grandi Numeri □

Esempio 4.26. *Supponiamo di considerare una funzione continua $g(x) : [0, 1] \rightarrow [0, 1]$ e di voler calcolare*

$$I = \int_0^1 g(x) dx \tag{4.30}$$

*in modo numerico, senza cioè trovare la primitiva di $g(x)$. Mostriamo come è possibile far questo sfruttando delle regolarità statistiche, secondo un metodo noto come **Monte Carlo** del quale presenteremo due possibili versioni.*

Consideriamo prima di tutto il vett-a $\mathbf{U} = (X, Y)$ a valori in $(\mathbf{R}^2, \mathcal{B}(\mathbf{R}^2))$ con componenti indipendenti ed uniformemente distribuite in $[0, 1]$ con

$$f_{\mathbf{U}}(x, y) = \begin{cases} 1 & (x, y) \in [0, 1] \times [0, 1] \\ 0 & \text{altrove} \end{cases}$$

in modo che \mathbf{U} risulti uniformemente distribuito in $[0, 1] \times [0, 1]$. Posto allora

$$A = \{(x, y) \in [0, 1] \times [0, 1] : g(x) \geq y\} \in \mathcal{B}(\mathbf{R}^2) \qquad B = \{(X, Y) \in A\} \in \mathcal{F}$$

$$\chi_A(x, y) = \begin{cases} 1 & (x, y) \in A \\ 0 & \text{altrove} \end{cases} \qquad I_B(\omega) = \begin{cases} 1 & \omega \in B \\ 0 & \text{altrove} \end{cases}$$

la v -a $Z = I_B = \chi_A(X, Y)$ sarà una Bernoulli $\mathfrak{B}(1; p)$ con

$$\begin{aligned} p &= \mathbf{E}[Z] = \mathbf{P}\{B\} = \mathbf{P}\{(X, Y) \in A\} = \mathbf{P}\{Y \leq g(X)\} \\ &= \int_A f_{\mathcal{U}}(x, y) dx dy = \int_0^1 \left[\int_0^{g(x)} dy \right] dx = \int_0^1 g(x) dx = I \end{aligned}$$

In pratica il valore di I coincide con la probabilità che un punto aleatorio X, Y verifichi l'evento $Y \leq g(X)$, ovvero con il valore d'attesa di Z . Pertanto l'integrale I può essere calcolato stimando $\mathbf{E}[Z]$ tramite la Legge forte dei Grandi Numeri. Scelti n punti (X, Y) uniformemente distribuiti in $[0, 1] \times [0, 1]$, sia $(Z_n)_{n \in \mathbf{N}}$ una successione di v -a iid distribuite come Z , e $S_n = Z_1 + \dots + Z_n$: il valore $I = \mathbf{E}[Z]$ è allora ben approssimato da S_n/n per grandi valori di n . Concretamente questo vuol dire stimare $I = p = \mathbf{P}\{Y \leq g(X)\}$ contando il numero di punti casuali e uniformi in $[0, 1] \times [0, 1]$ che cadono sotto la curva $y = g(x)$ nella Figura 4.2, e dividendo il risultato per il numero totale di punti estratti.

Il calcolo numerico dell'integrale I può essere eseguito anche con una procedura alternativa, osservando che se X è una v -a uniforme $\mathfrak{U}(0, 1)$, e se $Y = g(X)$, risulta

$$\mathbf{E}Y = \mathbf{E}[g(X)] = \int_0^1 g(x) dx = I$$

Pertanto si potrà anche valutare I stimando il valore d'attesa di $Y = g(X)$ con una media aritmetica: detta $(X_n)_{n \in \mathbf{N}}$ una successione di v -a iid uniformi $\mathfrak{U}(0, 1)$, la Legge dei grandi numeri ci dice che con probabilità 1 avremo

$$\frac{1}{n} \sum_{k=1}^n g(X_k) \xrightarrow{n} \mathbf{E}[g(X)] = I$$

sicché con un numero adeguato di misure si potrà sempre avere una approssimazione del valore di I con l'accuratezza voluta.

4.4 Teoremi Gaussiani

I Teoremi Gaussiani riguardano convergenze in distribuzione verso leggi normali standard $\mathfrak{N}(0, 1)$, e siccome saremo interessati alla forma delle distribuzioni limite più che ai valori d'attesa o alle varianze, sarà utile standardizzare le successioni in questione. Ricorderemo allora che una v -a X si dice standardizzata quando $\mathbf{E}[X] = 0$ e $\mathbf{V}[X] = 1$, per cui nel seguito studieremo il comportamento di somme S_n standardizzate

$$S_n^* = \frac{S_n - \mathbf{E}[S_n]}{\sqrt{\mathbf{V}[S_n]}} \quad (4.31)$$

Nelle versioni più arcaiche di questi teoremi le somme S_n in questione erano v -a binomiali (vedi Appendice F), ma ne esistono ormai delle formulazioni molto più generali che possono essere dimostrate sotto una grande varietà di ipotesi

Teorema 4.27. Teorema Limite Centrale per v-a iid: *Data una successione $(X_n)_{n \in \mathbf{N}}$ di v-a iid con $\mathbf{E}[X_n^2] < +\infty$ e $\mathbf{V}[X_n] > 0$, e definite $S_n = X_1 + \dots + X_n$ e S_n^* come in (4.31), risulta*

$$S_n^* \xrightarrow{d} \mathfrak{N}(0, 1)$$

Dimostrazione: Siccome la convergenza in distribuzione di una successione di v-a equivale alla convergenza in generale della corrispondente successione di fdd (vedi Sezione 4.1), la tesi del nostro teorema afferma che

$$\mathbf{P}\{S_n^* \leq x\} \xrightarrow{n} \Phi(x), \quad \forall x \in \mathbf{R}$$

dove

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz$$

è la funzione errore standard (2.16) che è anche continua in ogni x . Per la dimostrazione faremo allora uso del Teorema 4.16: siccome le X_n sono iid, poniamo

$$m = \mathbf{E}[X_n] \quad \sigma^2 = \mathbf{V}[X_n] \quad \varphi(u) = \mathbf{E}[e^{iu(X_n - m)}]$$

e osserviamo prima di tutto che

$$\mathbf{E}[S_n] = nm \quad \mathbf{V}[S_n] = n\sigma^2 \quad S_n^* = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^n (X_k - m)$$

Dall'indipendenza delle X_n si ha allora

$$\varphi_n(u) = \mathbf{E}[e^{iuS_n^*}] = \mathbf{E}\left[\prod_{k=1}^n e^{iu(X_k - m)/\sigma\sqrt{n}}\right] = \prod_{k=1}^n \mathbf{E}\left[e^{iu(X_k - m)/\sigma\sqrt{n}}\right] = \left[\varphi\left(\frac{u}{\sigma\sqrt{n}}\right)\right]^n$$

Ora $\varphi(u)$ è la fc delle $X_n - m$, con momenti finiti almeno fino al secondo ordine e

$$\mathbf{E}[X_n - m] = 0 \quad \mathbf{E}[(X_n - m)^2] = \sigma^2$$

Quindi dalla (4.19) sappiamo che

$$\varphi(u) = 1 - \frac{\sigma^2 u^2}{2} + o(u^2) \quad u \rightarrow 0$$

e pertanto, con u arbitrario ma fissato e $n \rightarrow \infty$, risulterà

$$\varphi_n(u) = \left[1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right)\right]^n \xrightarrow{n} e^{-u^2/2}$$

Siccome da (4.13) sappiamo che $e^{-u^2/2}$ è la fc di $\mathfrak{N}(0, 1)$, il teorema resta provato sulla base del Teorema 4.16. \square

Si osservi che, siccome ora $\mathbf{V}[S_n] = n\sigma^2$, la formulazione (4.29) della Legge dei Grandi Numeri (tenendo conto della particolare equivalenza fra convergenze in probabilità e in distribuzione verso 0) si potrebbe riscrivere come

$$\frac{S_n - \mathbf{E}[S_n]}{\mathbf{V}[S_n]} \xrightarrow{d} \delta_0 = \mathfrak{N}(0, 0)$$

con un denominatore che cresce come n , mentre la tesi del Teorema Limite Centrale 4.27 è

$$S_n^* = \frac{S_n - \mathbf{E}[S_n]}{\sqrt{\mathbf{V}[S_n]}} \xrightarrow{d} \mathfrak{N}(0, 1)$$

Queste due formulazioni mettono bene in luce le analogie e le differenze fra i due enunciati: nel Teorema Limite Centrale la varianza compare sotto radice per cui il denominatore cresce solo come \sqrt{n} , e questo giustifica in maniera intuitiva che la convergenza non sia più *degenere*. Ricorderemo infine l'enunciato di un'altra versione del Teorema Limite Centrale che, introducendo opportune ulteriori condizioni tecniche, permette di fare a meno dell'ipotesi di identica distribuzione delle X_n

Teorema 4.28. Teorema Limite Centrale per v-a indipendenti: *Data una successione $(X_n)_{n \in \mathbf{N}}$ di v-a indipendenti con $\mathbf{E}[X_n^2] < +\infty$ e $\mathbf{V}[X_n] > 0$, e definite $S_n = X_1 + \dots + X_n$ e S_n^* come in (4.31), posto*

$$m_n = \mathbf{E}[X_n] \quad V_n = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$$

se esiste $\delta > 0$ tale che (**condizioni di Lyapunov**)

$$\frac{1}{V_n^{2+\delta}} \sum_{k=1}^n \mathbf{E}[|X_k - m_k|^{2+\delta}] \xrightarrow{n} 0$$

allora si ha

$$S_n^* \xrightarrow{d} \mathfrak{N}(0, 1)$$

Dimostrazione: Omessa¹⁵

□

4.5 Teoremi di Poisson

In alcune vecchie formulazioni *binomiali* dei Teoremi Gaussiani (vedi ad esempio il Teorema Limite Locale nella Appendice F) la convergenza verso una legge normale veniva discussa tentando di approssimare i valori di una distribuzione binomiale mediante valori di funzioni Gaussiane, ma – a causa delle differenze strutturali che intercorrono fra le leggi binomiali e quelle Gaussiane – la validità di tale approssimazione si rivela sempre più incerta man mano che ci si allontana dal *centro* verso

¹⁵A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

le *code* delle distribuzioni. Queste difficoltà, peraltro, divengono particolarmente evidenti se p è molto prossimo a 0 oppure a 1. Infatti una curva gaussiana è perfettamente simmetrica attorno al suo massimo, mentre una distribuzione binomiale lo è solo quando $p = \frac{1}{2}$: se p si discosta da $\frac{1}{2}$ avvicinandosi a 0 oppure a 1 tale simmetria si perde. Per questi valori di p dunque non è ragionevole pretendere che una curva normale approssimi bene la distribuzione binomiale, se non nelle immediate vicinanze del suo massimo. Queste osservazioni ci suggeriscono allora di cercare altre approssimazioni asintotiche (per $n \rightarrow \infty$) della distribuzione binomiale quando p si avvicina a 0 oppure a 1, anche perché in alcuni particolari tipi di problemi saremo obbligati a costruire modelli probabilistici piuttosto diversi da quello originario di Bernoulli. Più precisamente può capitare che la probabilità p non sia la stessa al variare del numero di tentativi n , ma ne dipenda esplicitamente, e in particolare noi ci interesseremo al caso in cui $p(n) \rightarrow 0$ per $n \rightarrow \infty$ come vedremo discutendo un semplice esempio

Esempio 4.29. Istanti aleatori: *Supponiamo che un centralino telefonico riceva, in istanti casuali, delle telefonate il cui numero medio è proporzionale alla lunghezza dell'intervallo di tempo considerato. In particolare sia noto che arrivano in media $\lambda = 1.5$ telefonate al minuto, cioè 90 telefonate all'ora. Detta allora S la v -a che indica il numero delle telefonate che arrivano in un intervallo di $T = 3$ minuti, ci chiediamo quale è la distribuzione di S . Osserveremo allora innanzitutto che S assume solo valori interi $k = 0, 1, \dots$ non limitati superiormente: l'insieme dei possibili valori di S coincide quindi con tutto $\mathbf{N} \cup \{0\}$. Inizieremo allora cercando una procedura per approssimare i risultati esatti: siccome arrivano mediamente $\lambda = 1.5$ telefonate al minuto, inizieremo col suddividere T in un numero n di sottointervalli uguali fra loro, e abbastanza piccoli da far in modo che in ciascuno di essi ci si attenda di ricevere non più di una telefonata. Ad esempio con $n = 9$ il numero medio di telefonate nei sottointervalli diventa*

$$\frac{\lambda T}{n} = 1.5 \times \frac{3}{9} = \frac{1}{2}$$

e quindi, in prima approssimazione, potremo supporre che non arrivi più di una telefonata. Il nostro modello sarà allora composto di $n = 9$ tentativi indipendenti di verificare che in ciascuno dei 9 sottointervalli cada una telefonata, e a questo scopo definiremo 9 v -a di Bernoulli $X_j^{(9)}$ ($j = 1, \dots, 9$) che assumono valore 1 se capita una telefonata nel j -mo sottointervallo, e valore 0 in caso contrario. Siccome $X_j^{(9)} \sim \mathfrak{B}(1; p)$, e siccome nel nostro modello $\mathbf{E}[X_j^{(9)}] = \frac{1}{2}$, da (3.27) avremo anche che $p = p(9) = \frac{1}{2}$. Conseguentemente la v -a $S_9 = X_1^{(9)} + \dots + X_9^{(9)}$ che rappresenta approssimativamente il numero di telefonate in T sarà binomiale $\mathfrak{B}(9; \frac{1}{2})$, cioè

$$\mathbf{P}\{S_9 = k\} = \binom{9}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{9-k} = \binom{9}{k} \left(\frac{1}{2}\right)^9 \quad k = 0, 1, \dots, 9$$

Il difetto di questa prima approssimazione, come è chiaro, sta nell'ipotesi che in ogni sottointervallo possa arrivare al più una telefonata, cioè nell'aver approssimato una

v-a S che assume infiniti valori, con una *v*-a $S_9 \sim \mathfrak{B}(9; \frac{1}{2})$ i cui possibili valori sono invece solo 10. Questa osservazione ci suggerisce, però, anche la strada per migliorare l'approssimazione: se il numero n dei sottointervalli di T aumenta, da un lato otteniamo delle *v*-a S_n con un numero sempre crescente di possibili valori, e dall'altro, rendendo sempre più piccola la durata dei sottointervalli, saremo sempre più sicuri del fatto che in ognuno di essi non arrivi più di una sola telefonata. Così, ad esempio, con $n = 18$ sottointervalli otteniamo

$$p(18) = \frac{\lambda T}{n} = 1.5 \times \frac{3}{18} = \frac{1}{4}$$

e quindi $S_{18} \sim \mathfrak{B}(18; \frac{1}{4})$, cioè

$$\mathbf{P}\{S_{18} = k\} = \binom{18}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{18-k} \quad k = 0, 1, \dots, 18$$

Potremo pertanto continuare a migliorare la nostra approssimazione aumentando il valore di n , nel qual caso risulterà

$$p(n) = \frac{\lambda T}{n} = 1.5 \times \frac{3}{n} \xrightarrow{n \rightarrow \infty} 0 \quad np(n) = \lambda T = \alpha, \quad \forall n \in \mathbf{N}$$

e dovremo chiederci a quale distribuzione limite ($n \rightarrow \infty$) tende, in queste condizioni, la successione di leggi binomiali $\mathfrak{B}(n; p(n))$

$$p_n(k) = \mathbf{P}\{S_n = k\} = \binom{n}{k} p(n)^k (1 - p(n))^{n-k} \quad (4.32)$$

La risposta è contenuta nel successivo teorema che riporteremo comunque nella sua forma classica binomiale¹⁶ prima di darne le versioni più moderne

Teorema 4.30. Teorema di Poisson per v-a binomiali: *Data la successione di v-a binomiali $S_n \sim \mathfrak{B}(n; p(n))$ come in (4.32), se esiste $\alpha > 0$ tale che*

$$p(n) \rightarrow 0 \quad q(n) = 1 - p(n) \rightarrow 1 \quad np(n) \rightarrow \alpha \quad n \rightarrow \infty$$

allora S_n converge in distribuzione verso la legge di Poisson $\mathfrak{P}(\alpha)$ nel senso che

$$S_n \xrightarrow{d} \mathfrak{P}(\alpha) \quad \text{ovvero} \quad \lim_n p_n(k) = \frac{\alpha^k e^{-\alpha}}{k!}, \quad k = 0, 1, \dots$$

Dimostrazione: Siccome, qualunque sia $\alpha > 0$, da un certo n in poi si ha $\alpha/n < 1$, a partire da quell' n le ipotesi fatte ci autorizzano a scrivere

$$p(n) = \frac{\alpha}{n} + o(n^{-1})$$

¹⁶**S.D. Poisson**, RECHERCHES SUR LA PROBABILITÉ DES JUGEMENTS EN MATIÈRE CRIMINELLE ET EN MATIÈRE CIVILE, Bachelier (Paris, 1837)

sicché per $k = 0, 1, \dots, n$ avremo

$$p_n(k) = \frac{n(n-1)\dots(n-k+1)}{k!} \left[\frac{\alpha}{n} + o(n^{-1}) \right]^k \left[1 - \frac{\alpha}{n} + o(n^{-1}) \right]^{n-k}$$

Ricordando allora un ben noto limite notevole si ha

$$\begin{aligned} n(n-1)\dots(n-k+1) \left[\frac{\alpha}{n} + o(n^{-1}) \right]^k &= \frac{n(n-1)\dots(n-k+1)}{n^k} [\alpha + o(1)]^k \\ &= \left(1 - \frac{1}{n} \right) \dots \left(1 - \frac{k-1}{n} \right) [\alpha + o(1)]^k \xrightarrow{n} \alpha^k \\ \left[1 - \frac{\alpha}{n} + o(n^{-1}) \right]^{n-k} &= \left[1 - \frac{\alpha}{n} + o(n^{-1}) \right]^n \left[1 - \frac{\alpha}{n} + o(n^{-1}) \right]^{-k} \xrightarrow{n} e^{-\alpha} \end{aligned}$$

da cui segue immediatamente la tesi. \square

Teorema 4.31. Teorema di Poisson per vett-a multinomiali: *Data una successione di vett-a multinomiali $\mathbf{S}_n = (X_1, \dots, X_r) \sim \mathfrak{B}(n; p_1, \dots, p_r)$ con*

$$\mathbf{P}\{X_1 = k_1, \dots, X_r = k_r\} = \frac{n!}{k_0! k_1! \dots k_r!} p_0^{k_0} p_1^{k_1} \dots p_r^{k_r} \quad \begin{cases} p_0 + p_1 + \dots + p_r = 1 \\ k_0 + k_1 + \dots + k_r = n \end{cases}$$

se per $j = 1, \dots, r$ e per $n \rightarrow \infty$ esistono degli $\alpha_j > 0$ tali che

$$p_j = p_j(n) \rightarrow 0 \quad p_0 = p_0(n) \rightarrow 1 \quad np_j(n) \rightarrow \alpha_j$$

allora

$$\mathbf{S}_n = (X_1, \dots, X_r) \xrightarrow{d} \mathfrak{P}(\alpha_1) \cdot \dots \cdot \mathfrak{P}(\alpha_r).$$

Dimostrazione: Omessa: analoga a quella del Teorema 4.30 \square

Il precedente Teorema 4.30 è stato dimostrato usando esplicitamente le proprietà delle leggi binomiali ottenute sommando v -a di Bernoulli *iid*. Esso può essere generalizzato anche a somme di v -a di Bernoulli indipendenti ma *non identicamente distribuite*: in questo caso però le leggi delle somme non sono più binomiali e la dimostrazione precedente non trova applicazione. Per fissare le idee consideriamo una successione di esperimenti, e per ogni n della successione siano date n v -a indipendenti $X_1^{(n)}, \dots, X_n^{(n)}$ di Bernoulli con $X_k^{(n)} \sim \mathfrak{B}\left(1; p_k^{(n)}\right)$, cioè

$$\mathbf{P}\{X_k^{(n)} = 1\} = p_k^{(n)} \quad \mathbf{P}\{X_k^{(n)} = 0\} = q_k^{(n)} \quad p_k^{(n)} + q_k^{(n)} = 1 \quad k = 1, \dots, n$$

La v -a somma $S_n = X_1^{(n)} + \dots + X_n^{(n)}$ prenderà allora certamente valori da 0 ad n , ma in generale non sarà più una v -a binomiale dato che gli addendi non sono più identicamente distribuiti. Nei casi più generali avremo infatti

- per ogni k fissato: le $p_k^{(n)}$ variano al variare di n sicché le v -a $X_k^{(n)}$ cambiano distribuzione al variare di n ; cioè nel passaggio da un n al successivo le v -a che si trovano in un determinato posto k vengono aggiornate

- per ogni n fissato le $X_k^{(n)}$ non sono identicamente distribuite, sicché le v -a S_n non sono binomiali

In sostanza avremo a che fare con schemi triangolari del tipo

$$\begin{array}{ll} X_1^{(1)} & p_1^{(1)} \\ X_1^{(2)}, X_2^{(2)} & p_1^{(2)}, p_2^{(2)} \\ \vdots & \vdots \\ X_1^{(n)}, \dots, X_n^{(n)} & p_1^{(n)}, \dots, p_n^{(n)} \\ \vdots & \vdots \end{array}$$

In ciascuna riga le $X_k^{(n)}$ sono indipendenti, ma non identicamente distribuite; lungo le colonne i valori delle $p_k^{(n)}$ (e quindi le leggi) in generale si modificano. Il successivo teorema stabilisce le condizioni sotto le quali le S_n convergono ancora in distribuzione verso $\mathfrak{P}(\alpha)$

Teorema 4.32. *Date per ogni $n \in \mathbf{N}$ e $k = 1, \dots, n$ le v -a indipendenti $X_k^{(n)}$ con*

$$\mathbf{P}\{X_k^{(n)} = 1\} = p_k^{(n)} \quad \mathbf{P}\{X_k^{(n)} = 0\} = q_k^{(n)} \quad p_k^{(n)} + q_k^{(n)} = 1$$

e posto $S_n = X_1^{(n)} + \dots + X_n^{(n)}$, se

$$\max_{1 \leq k \leq n} p_k^{(n)} \xrightarrow{n} 0 \quad \sum_{k=1}^n p_k^{(n)} \xrightarrow{n} \alpha > 0$$

allora risulta

$$S_n \xrightarrow{d} \mathfrak{P}(\alpha)$$

Dimostrazione: Data l'indipendenza delle $X_k^{(n)}$, e ricordando (4.7), si ha

$$\varphi_{S_n}(u) = \mathbf{E}[e^{iuS_n}] = \prod_{k=1}^n [p_k^{(n)} e^{iu} + q_k^{(n)}] = \prod_{k=1}^n [1 + p_k^{(n)}(e^{iu} - 1)]$$

Siccome per ipotesi $p_k^{(n)} \xrightarrow{n} 0$, dallo sviluppo in serie del logaritmo si ha

$$\ln \varphi_{S_n}(u) = \sum_{k=1}^n \ln[1 + p_k^{(n)}(e^{iu} - 1)] = \sum_{k=1}^n [p_k^{(n)}(e^{iu} - 1) + o(p_k^{(n)})] \xrightarrow{n} \alpha(e^{iu} - 1)$$

e data la continuità del logaritmo

$$\varphi_{S_n}(u) \xrightarrow{n} e^{\alpha(e^{iu}-1)}$$

Ricordando allora la (4.9), dal Teorema 4.16 si ha $S_n \xrightarrow{d} \mathfrak{P}(\alpha)$. □

Teorema 4.33. *Se S è una v -a di Poisson $\mathfrak{P}(\alpha)$ allora*

$$S^* = \frac{S - \alpha}{\sqrt{\alpha}} \xrightarrow{d} \mathfrak{N}(0, 1) \quad \alpha \rightarrow +\infty$$

Dimostrazione: Detta φ_α la *fc* di S^* , da (4.9) e dallo sviluppo in serie di un esponenziale si ha per $\alpha \rightarrow +\infty$

$$\begin{aligned} \varphi_\alpha(u) &= \mathbf{E} [e^{iuS^*}] = e^{-iu\sqrt{\alpha}} \mathbf{E} [e^{iuS/\sqrt{\alpha}}] \\ &= \exp \left[-iu\sqrt{\alpha} + \alpha \left(e^{iu/\sqrt{\alpha}} - 1 \right) \right] \\ &= \exp \left[-iu\sqrt{\alpha} - \alpha + \alpha \left(1 + \frac{iu}{\sqrt{\alpha}} - \frac{u^2}{2\alpha} + o\left(\frac{1}{\alpha}\right) \right) \right] \rightarrow e^{-u^2/2} \end{aligned}$$

Il risultato segue allora dal Teorema 4.16. □

4.6 Applicabilità dei teoremi limite

I teoremi limite sono affermazioni piuttosto generali sui comportamenti di somme di un numero molto grande v -a indipendenti, ma non bisogna pensare che essi possano essere applicati in maniera del tutto indiscriminata. In particolare va esercitata una certa cautela nel verificare che le loro ipotesi siano effettivamente rispettate, specialmente quelle che riguardano l'esistenza dei momenti $\mathbf{E}[X_n]$ e $\mathbf{E}[X_n^2]$. A questo scopo discuteremo brevemente un esempio che mette in evidenza come dei problemi rilevanti possano nascere già in situazioni molto comuni.

Esempio 4.34. *Supponiamo che, come nella Figura 4.3, un raggio di luce emesso da una sorgente luminosa collocata in A su una parete colpisca uno specchio collocato in C ad una distanza a , e libero di muoversi ruotando attorno ad un perno. La posizione dello specchio sia aleatoria nel senso che l'angolo Θ di riflessione del raggio sia una v -a uniforme $\mathfrak{U}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Sia poi $X = a \tan \Theta$ la distanza da A del punto B in cui il raggio riflesso colpisce la parete. Mostriamo allora prima di tutto che X è una v -a con legge di Cauchy $\mathfrak{C}(a, 0)$. Infatti siccome nel nostro modello*

$$f_\Theta(\theta) = \begin{cases} 1/\pi & \text{se } |\theta| \leq \pi/2 \\ 0 & \text{se } |\theta| > \pi/2 \end{cases}$$

mentre X risulta dall'applicazione ad Θ della funzione $x = g(\theta) = a \tan \theta$ che è monotona su $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, posto

$$\theta_1(x) = g^{-1}(x) = \arctan \frac{x}{a}$$

e osservato che

$$\theta_1'(x) = \frac{a}{a^2 + x^2}$$

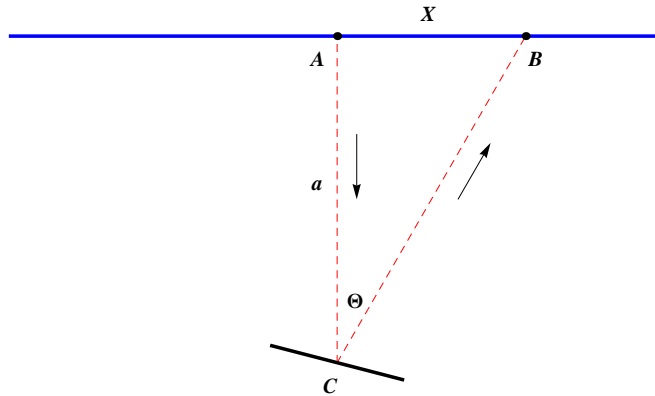


Figura 4.3: Costruzione di v - a con legge $\mathfrak{C}(a)$

dalla regola di trasformazione (3.63) si ottiene per X la ddp di Cauchy $\mathfrak{C}(a, 0)$

$$f_X(x) = f_{\Theta}(\theta_1(x))|\theta_1'(x)| = \frac{1}{\pi} \frac{a}{a^2 + x^2}$$

visto che f_{Θ} assume costantemente il valore $\frac{1}{\pi}$ in $[-\frac{\pi}{2}, \frac{\pi}{2}]$, e che $\theta_1(x) \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Resta così accertato che una legge di Cauchy può emergere in maniera naturale in casi anche molto semplici e realistici. Per provare ad applicare i teoremi limite al nostro modello, possiamo ora supporre di ripetere, in maniera indipendente, varie misure di X ottenendo una successione $(X_n)_{n \in \mathbf{N}}$ di v - a iid tutte con legge di Cauchy $\mathfrak{C}(a, 0)$. Da (4.16) sappiamo però che la fc di queste leggi è

$$\varphi(u) = \mathbf{E} [e^{iuX_n}] = e^{-a|u|}$$

e quindi che, posto $S_n = X_1 + \dots + X_n$, la fc della media aritmetica S_n/n è sempre

$$\varphi_n(u) = \mathbf{E} \left[e^{iu \frac{S_n}{n}} \right] = \mathbf{E} \left[\prod_{k=0}^n e^{iu \frac{X_k}{n}} \right] = \left[\varphi \left(\frac{u}{n} \right) \right]^n = \left(e^{-\frac{a|u|}{n}} \right)^n = e^{-a|u|} = \varphi(u)$$

Ne segue allora banalmente che

$$\varphi_n(u) \xrightarrow{n} \varphi(u) \quad \forall u \in \mathbf{R}$$

e quindi dal Teorema 4.16

$$\frac{S_n}{n} \xrightarrow{d} X$$

Pertanto non si riscontra nessuna convergenza degenera del tipo previsto dalla Legge dei Grandi Numeri (debole o forte). Inoltre, con le stesse procedure, si verifica anche facilmente che con λ_n arbitrario risulta sempre $\lambda_n S_n \sim \mathfrak{C}(n\lambda_n a, 0)$, sicché in nessun caso le somme di v - a di Cauchy indipendenti mostrano una tendenza a convergere verso leggi normali come previsto dai Teoremi Gaussiani

Il nostro esempio, quindi, si pone al di fuori della giurisdizione dei teoremi limite classici studiati finora, e una sua trattazione completa richiederebbe l'introduzione di concetti più evoluti sulle possibili *leggi limite* e sui corrispondenti teoremi (si veda anche a questo proposito la successiva Sezione 7.1.3). Noi concluderemo queste rapide osservazioni ricordando solo che il comportamento apparentemente anomalo del nostro esempio, che sembra violare teoremi enunciati in precedenza, è sostanzialmente dovuto al fatto che le ipotesi di questi teoremi non sono soddisfatte: come abbiamo già osservato nell'Esempio 3.25, una v -a X di Cauchy $\mathfrak{C}(a, 0)$ non ha un valore d'attesa definito, mentre l'esistenza di $\mathbf{E}[X]$ (nel senso di un integrale di Lebesgue) è un'ipotesi importante nelle dimostrazioni dei Teoremi Gaussiani e delle Leggi dei Grandi Numeri

Parte II
Processi Stocastici

Capitolo 5

Concetti generali

Il concetto di p -s $X(t) = X(\omega; t)$ su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ con $t > 0$ è stato già introdotto nella Sezione 3.2 dove è stato fatto osservare che esso può essere guardato da due punti di vista complementari:

- come un'applicazione che ad ogni fissato $t > 0$ associa una v -a $X(\omega; t) = X(t)$ che descrive lo *stato* del sistema all'istante t ; in questo caso il p -s è una famiglia di v -a al variare di t ;
- come un'applicazione che ad ogni fissato $\omega \in \Omega$ associa una intera *traiettoria* $X(\omega; t) = x(t)$ del processo; in questo secondo caso il p -s consiste nell'insieme (pesato) di tutte le traiettorie possibili del processo.

I due punti di vista sono sostanzialmente equivalenti e vengono adottati secondo le necessità. Sarà opportuno osservare subito, comunque, che t è qui considerato come un *tempo* solo per fissare le idee: in effetti qualunque famiglia di v -a $X(\alpha)$ descritta da un o più parametri α costituisce un p -s. L'idea che il parametro sia un tempo nasce solo dal fatto che gli esempi più noti sono di questo tipo. In generale i nostri p -s saranno definiti per $t \geq 0$, ma non sono escluse altre possibilità: ad esempio $t \in \mathbf{R}$. Come vedremo in un capitolo successivo, inoltre, un p -s può avere più di una sola componente: $\mathbf{X}(t) = (X_1(t), \dots, X_M(t))$, anche se in questa prima parte del corso ci limiteremo per semplicità solo al caso $M = 1$. Inoltre nel seguito parleremo spesso anche degli **incrementi** $\mathbf{X}(s) - \mathbf{X}(t)$ di $\mathbf{X}(t)$ sull'intervallo $[t, s]$, usando la notazione $\Delta\mathbf{X}(t) = \mathbf{X}(t + \Delta t) - \mathbf{X}(t)$ con $\Delta t = s - t > 0$, e del **processo degli incrementi** $\Delta\mathbf{X}(t)$ al variare di t con $\Delta t > 0$ fissato.

5.1 Leggi e identità dei p -s

Ad ogni p -s dato può essere associata un'intera gerarchia consistente di leggi finito-dimensionali che – come stabilito nel Teorema di Kolmogorov 2.37 – determinano univocamente la **legge globale del p -s**. Tale associazione viene effettuata nel modo seguente: fissato un arbitrario numero finito n di arbitrari istanti di tempo

t_1, \dots, t_n , si considerano tutte le leggi congiunte dei vettori $(X(t_1), \dots, X(t_n))$ così ottenuti. Tali leggi possono essere rappresentate mediante le loro *fc*

$$\varphi(u_1, t_1; \dots; u_n, t_n)$$

ma noi nel seguito useremo anche le loro distribuzioni discrete (tipicamente con valori interi $k, \ell \dots$), ovvero – se *ac*, come supporremo spesso – le loro *ddp* con le notazioni rispettive

$$p(k_1, t_1; \dots; k_n, t_n) \qquad f(x_1, t_1; \dots; x_n, t_n)$$

Allo stesso modo, introdotti altri m istanti s_1, \dots, s_m , potremo parlare di probabilità e di *ddp* condizionate secondo le notazioni della Sezione 3.4.1

$$p(k_1, t_1; \dots; k_n, t_n | \ell_1, s_1; \dots; \ell_m, s_m) = \frac{p(k_1, t_1; \dots; k_n, t_n; \ell_1, s_1; \dots; \ell_m, s_m)}{p(\ell_1, s_1; \dots; \ell_m, s_m)}$$

$$f(x_1, t_1; \dots; x_n, t_n | y_1, s_1; \dots; y_m, s_m) = \frac{f(x_1, t_1; \dots; x_n, t_n; y_1, s_1; \dots; y_m, s_m)}{f(y_1, s_1; \dots; y_m, s_m)}$$

L'ordine delle t_i, s_j è per ora irrilevante, ma di solito si scelgono in modo che

$$t_n \geq \dots \geq t_1 \geq s_m \geq \dots \geq s_1 \geq 0$$

In particolare si chiamano **ddp e probabilità di transizione** le *ddp* e probabilità condizionate a due istanti $f(x, t | y, s)$ e $p(k, t | \ell, s)$. Passiamo ora definire i diversi tipi di **uguaglianza di due p-s** $X(t)$ e $Y(t)$:

1. $X(t)$ e $Y(t)$ si dicono *indistinguibili, o identici P-qo* se tutte le traiettorie dei due processi coincidono in ogni t , con la possibile eccezione di un loro sottoinsieme di probabilità nulla, cioè se

$$\mathbf{P}\{X(t) = Y(t), \forall t > 0\} = 1$$

2. $X(t)$ e $Y(t)$ si dicono *equivalenti* (e in questo caso si dice anche che $X(t)$ è una *modificazione* di $Y(t)$ e viceversa) se invece per ogni dato t le traiettorie coincidono **P-qo**, cioè se

$$\mathbf{P}\{X(t) = Y(t)\} = 1, \quad \forall t > 0$$

3. $X(t)$ e $Y(t)$ si dicono infine *equivalenti in senso lato, ovvero uguali in distribuzione* se tutte le loro leggi congiunte finito-dimensionali (e quindi anche le loro leggi globali) coincidono

Le tre definizioni sono piuttosto diverse. Ad esempio è facile vedere che due processi indistinguibili sono equivalenti, ma non è vero il viceversa: posto infatti

$$N_t = \overline{\{X(t) = Y(t)\}} \quad t > 0$$

$$N = \overline{\{X(t) = Y(t), \forall t > 0\}} = \bigcup_{t > 0} N_t$$

l'indistinguibilità richiede $\mathbf{P}\{N\} = 0$, e quindi anche $\mathbf{P}\{N_t\} = 0, \forall t > 0$, cioè implica l'equivalenza. La semplice equivalenza, invece, richiede solo che $\mathbf{P}\{N_t\} = 0, \forall t > 0$, ma questo non basta per avere anche $\mathbf{P}\{N\} = \mathbf{P}\{\bigcup N_t\} = 0$ a causa del fatto che l'insieme degli istanti $t > 0$ non è numerabile. Allo stesso modo processi equivalenti hanno anche le stesse *ddp* finito-dimensionali, ma non si può dire in generale che sia vero il viceversa.

Come già ricordato il Teorema di Kolmogorov 2.37 garantisce che la conoscenza di una intera famiglia *consistente* di *ddp* finito-dimensionali è sufficiente per determinare una probabilità su tutto lo spazio $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$ con $T = [0, +\infty)$. È stato però fatto anche notare nell'Esempio 4 che $\mathcal{B}(\mathbf{R}^T)$ non è sufficientemente grande per i nostri propositi: affermazioni come *il processo è continuo in un dato istante*, ad esempio, corrispondono a sottoinsiemi di traiettorie che non trovano posto in tale σ -algebra. Si rende quindi necessaria un'estensione del nostro spazio di probabilità. Senza entrare in dettagli tecnici ci limiteremo qui solo a ricordare che tale estensione è sempre possibile in maniera coerente e non ambigua se il nostro processo gode di una proprietà detta *separabilità*¹, che eviteremo di definire precisamente. In questo caso le *ddp* finito-dimensionali determinano una misura di probabilità estendibile a tutti i sottoinsiemi di traiettorie di interesse pratico, e per di più questa limitazione si rivela essere più formale che sostanziale alla luce del risultato seguente

Teorema 5.1. *Comunque assegnato un processo è sempre possibile costruirne una modificazione separabile: ogni processo è quindi equivalente a un processo separabile*

Dimostrazione: Omessa² □

Conseguentemente potremo sempre supporre di avere a che fare con processi separabili per i quali la conoscenza delle *ddp* finito-dimensionali determina la probabilità su un opportuno spazio di traiettorie comprendente tutti gli eventi necessari.

Generalizzando poi le idee sulle *v-a* canoniche espone nella Sezione 3.1.3, potremo anche dire che, assegnata – tramite una famiglia consistente di leggi finito-dimensionali – una probabilità \mathbf{P} sull'insieme delle traiettorie $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$, sarà sempre possibile costruire un *p-s* che abbia proprio \mathbf{P} come legge globale

Definizione 5.2. *Data una probabilità \mathbf{P} sullo spazio delle traiettorie $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$, si chiama **processo canonico** il *p-s* $X(t)$ definito come l'applicazione identica da $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T), \mathbf{P})$ in $(\mathbf{R}^T, \mathcal{B}(\mathbf{R}^T))$, la cui distribuzione coinciderà con la \mathbf{P} assegnata*

Queste osservazioni spiegano perchè i *p-s* (come le *v-a*) sono essenzialmente classificati sulla base delle loro leggi, anche se ad una data distribuzione possono comunque corrispondere molti differenti processi che condividono solo la loro legge globale

Definizione 5.3. *Diremo che $X(t)$ è un **processo Gaussiano** quando, comunque scelti t_1, t_2, \dots, t_n per ogni $n = 0, 1, \dots$, risulta $(X(t_1), \dots, X(t_n)) \sim \mathfrak{N}(\mathbf{b}, \mathbb{A})$ dove \mathbf{b} e \mathbb{A} sono vettori delle medie e matrici delle covarianze dipendenti da t_1, t_2, \dots, t_n*

¹J.L. Doob, STOCHASTIC PROCESSES, Wiley (New York, 1953)

²J.L. Doob, STOCHASTIC PROCESSES, Wiley (New York, 1953)

5.2 Attese e correlazioni

Una buona parte delle informazioni relative a un p -s $X(t)$ (che per ora supporremo avere solo una componente) può essere ottenuta anche guardando soltanto i valori d'attesa (quando essi esistono) in uno o più istanti di tempo, e senza dover fornire tutti i dettagli sulle distribuzioni finito-dimensionali. Innanzitutto è possibile introdurre il **valore d'attesa** e la **varianza** di un p -s in ogni istante t :

$$m(t) = \mathbf{E} [X(t)] \quad \sigma^2(t) = \mathbf{V} [X(t)] \quad (5.1)$$

In secondo luogo è importante introdurre momenti del secondo ordine che descrivono la correlazione tra i valori di un p -s in diversi istanti di tempo: a questo scopo si definisce prima di tutto l'**autocorrelazione** di un p -s $X(t)$ come la funzione (simmetrica nei suoi due argomenti)

$$R(s, t) = R(t, s) = \mathbf{E} [X(s)X(t)] \quad (5.2)$$

Quando non ci sono equivoci possibili trascureremo in futuro di indicare gli indici del processo nelle attese e autocorrelazioni. Sarà bene osservare subito, però, che qui la parola *correlazione* viene usata con un significato leggermente diverso rispetto a quanto fatto per la correlazione di due v -a. Per questo motivo definiremo quindi anche un'**autocovarianza** e un **coefficiente di correlazione** di $X(t)$ rispettivamente come

$$C(s, t) = \mathbf{cov} [X(s), X(t)] = R(s, t) - m(s)m(t) \quad (5.3)$$

$$\rho(s, t) = \rho [X(s), X(t)] = \frac{C(s, t)}{\sigma(s)\sigma(t)} \quad (5.4)$$

e osserveremo che in particolare l'autocovarianza $C(s, t)$ di un processo contiene anche la sua varianza dato che evidentemente

$$\mathbf{V} [X(t)] = \sigma^2(t) = C(t, t) \quad (5.5)$$

Infine si dice **processo centrato** quello per il quale risulta

$$m(t) = 0 \quad R(s, t) = C(s, t)$$

Ci vuol poco per riconoscere che, assegnato un arbitrario p -s $X(t)$, il processo

$$\tilde{X}(t) = X(t) - m(t) = X(t) - \mathbf{E} [X(t)]$$

risulta sempre centrato. La conoscenza delle funzioni $m(t)$, $R(s, t)$, $C(s, t)$ e $\rho(s, t)$ a uno e due tempi, pur non esaurendo in generale le informazioni relative a un p -s, consente di avere un'idea abbastanza precisa della sua natura, e in alcuni casi particolari esse possono fornire addirittura una descrizione completa delle leggi del processo.

5.3 Convergenza e continuità

I tipi di convergenza introdotti nella Definizione 4.1 per le successioni di v -a $(X_n)_{n \in \mathbf{N}}$ possono essere facilmente estesi ai p -s per definire i corrispondenti tipi di convergenza di $X(t)$ verso una qualche v -a X_0 per $t \rightarrow t_0$. Così ad esempio la *convergenza in media quadratica* (mq)

$$X(t) \xrightarrow{mq} X_0 \quad t \rightarrow t_0$$

sarà definita da

$$\lim_{t \rightarrow t_0} \mathbf{E} [|X(t) - X_0|^2] = 0 \quad (5.6)$$

Allo stesso modo si possono introdurre le *convergenze \mathbf{P} -qo, in probabilità, in L^p e in distribuzione* che rispetteranno anche le mutue relazioni elencate nel Teorema 4.4 per le convergenze delle successioni. Inoltre è anche possibile dimostrare la validità dei *criteri di Cauchy* corrispondenti ad ogni tipo di convergenza. Sarà inoltre utile introdurre un nuovo concetto: quello della convergenza di una *successione di processi* verso un altro processo

Definizione 5.4. Diremo che una *successione di processi* $\{X_n(t)\}_{n \in \mathbf{N}}$ *converge in distribuzione* verso il processo $X(t)$ quando tutte le distribuzioni k -dimensionali dei vettori $(X_n(t_1), \dots, X_n(t_k))$ convergono debolmente verso le corrispondenti distribuzioni dei vettori $(X(t_1), \dots, X(t_k))$, con $k = 1, 2, \dots$

Siamo in grado a questo punto di introdurre i vari concetti di *continuità di un processo*, con la precisazione dei tipi di convergenza adottati. Bisogna però chiarire subito che in questo caso dovremo anche distinguere le continuità in ogni dato, arbitrario istante t , dalla continuità globale delle traiettorie del processo per ogni t , e adotteremo quindi lo schema seguente

Definizione 5.5. Dato un p -s $X(t)$ con $t \geq 0$ su $(\Omega, \mathcal{F}, \mathbf{P})$ diremo che esso è

- *continuo \mathbf{P} -qo, in probabilità, in L^p o in distribuzione* se in ogni arbitrario, ma fissato $t \geq 0$, e per $s \rightarrow t$ risulta $X(s) \rightarrow X(t)$ rispettivamente \mathbf{P} -qo, in probabilità, in L^p o in distribuzione; un p -s continuo in probabilità si dice anche *stocasticamente continuo*;
- *continuo globalmente* (ovvero *sample continuous*) se quasi ogni traiettoria risulta continua in ogni $t \geq 0$, ovvero se

$$\mathbf{P}\{\omega \in \Omega : X(t; \omega) \text{ è continua } \forall t \geq 0\} = 1$$

Quest'ultimo tipo di continuità globale non va confuso con la continuità \mathbf{P} -qo definita nel punto precedente: il p -s è continuo \mathbf{P} -qo se ogni istante t è quasi certamente un punto di continuità; esso invece è continuo globalmente se l'insieme delle traiettorie non continue anche in un solo t è di probabilità zero

I vari tipi di continuità non sono ovviamente equivalenti fra loro, ma soddisfano delle relazioni del tutto analoghe a quelle fra le diverse convergenze elencate nel Teorema 4.4. In particolare la continuità in L^2 (cioè la continuità in mq) è condizione sufficiente per la continuità stocastica. Sarà allora utile osservare che la continuità in mq può essere controllata esaminando le proprietà di continuità delle funzioni di autocorrelazione

Proposizione 5.6. *Un p -s $X(t)$ è continuo in mq – e quindi è stocasticamente continuo – se e solo se l'autocorrelazione $R(s, t)$ è continua in $s = t$*

Dimostrazione: Infatti siccome

$$\begin{aligned} \mathbf{E} [|X(s) - X(t)|^2] &= \mathbf{E} [X^2(s)] + \mathbf{E} [X^2(t)] - 2\mathbf{E} [X(s)X(t)] \\ &= R(t, t) + R(s, s) - 2R(s, t) \end{aligned}$$

dalla definizione e da (5.6) ricaviamo che la continuità in mq in t equivale alla continuità di $R(s, t)$ in $s = t$. \square

Delle opportune condizioni sufficienti per la continuità globale di un processo di Markov saranno invece introdotte successivamente nella Sezione 7.1.7, e sarà bene osservare infine che la continuità globale implica evidentemente tutti gli altri tipi di continuità considerati

5.4 Derivazione e integrazione in mq

Anche la derivazione e l'integrazione di un p -s richiedono delle opportune procedure di limite e quindi possono essere definite in vario modo secondo il tipo di convergenza invocato. Nei capitoli che seguono saranno affrontati alcuni punti importanti in merito a questi argomenti, mentre qui per cominciare ci limiteremo ad esaminare solo la convergenza in mq (e quindi anche in probabilità). Innanzitutto, secondo le nostre definizioni, un p -s $X(t)$ è **derivabile in mq** se esiste un altro processo $\dot{X}(t)$ tale che per ogni $t \geq 0$

$$\frac{X(t + \Delta t) - X(t)}{\Delta t} \xrightarrow{mq} \dot{X}(t) \quad \Delta t \rightarrow 0$$

ovvero se

$$\lim_{\Delta t \rightarrow 0} \mathbf{E} \left[\left| \frac{X(t + \Delta t) - X(t)}{\Delta t} - \dot{X}(t) \right|^2 \right] = 0 \quad (5.7)$$

Possiamo allora mostrare che, come per la continuità in mq , anche la derivabilità in mq può essere messa in relazione alla derivabilità della funzione di autocorrelazione:

Proposizione 5.7. *Un p -s $X(t)$ è derivabile in mq in t , se e solo se la derivata seconda mista $R_{1,1} = \partial_s \partial_t R$ dell'autocorrelazione $R(s, t)$ esiste in $s = t$. In tal caso si ha anche $\mathbf{E} [\dot{X}(t)] = \dot{m}(t)$*

Dimostrazione: L'esistenza della derivata in ogni t è garantita dal rispetto del corrispondente criterio di Cauchy per il limite (5.7), cioè da

$$\lim_{\Delta s, \Delta t \rightarrow 0} \mathbf{E} \left[\left| \frac{X(t + \Delta s) - X(t)}{\Delta s} - \frac{X(t + \Delta t) - X(t)}{\Delta t} \right|^2 \right] = 0$$

Siccome d'altra parte

$$\begin{aligned} \mathbf{E} \left[\frac{X(t + \Delta t) - X(t)}{\Delta t} \frac{X(t + \Delta s) - X(t)}{\Delta s} \right] &= \frac{R(t + \Delta s, t + \Delta t) - R(t + \Delta s, t) - R(t, t + \Delta t) + R(t, t)}{\Delta t \Delta s} \\ \mathbf{E} \left[\left| \frac{X(t + \Delta t) - X(t)}{\Delta t} \right|^2 \right] &= \frac{R(t + \Delta t, t + \Delta t) - R(t + \Delta t, t) - R(t, t + \Delta t) + R(t, t)}{\Delta t^2} \end{aligned}$$

i cui limiti per $\Delta s, \Delta t \rightarrow 0$ – se convergono – valgono $R_{1,1}(t, t)$, si ha che il criterio di Cauchy è soddisfatto e $\dot{X}(t)$ esiste se e solo se $R_{1,1}(t, t)$ esiste, dato che in questo caso

$$\begin{aligned} \mathbf{E} \left[\left| \frac{X(t + \Delta s) - X(t)}{\Delta s} - \frac{X(t + \Delta t) - X(t)}{\Delta t} \right|^2 \right] \\ \rightarrow R_{1,1}(t, t) - 2R_{1,1}(t, t) + R_{1,1}(t, t) = 0 \end{aligned}$$

Che poi risulti anche $\mathbf{E} [\dot{X}(t)] = m(t)$, si prova verificando le condizioni per lo scambio dell'attesa con i limiti □

Quanto agli **integrali stocastici**, dei quali parleremo più a lungo nella Sezione 8.2, qui ci limiteremo solo ad esaminare quelli del tipo

$$\int_a^b X(t) dt \tag{5.8}$$

che – se convergono in qualche senso – definiscono una nuova $v-a$. Noi per ora li considereremo definiti secondo la procedura di Riemann e ne esamineremo la convergenza in mq , mentre rimanderemo a un momento successivo una discussione più approfondita dell'integrazione stocastica più generale. Considerata dunque una decomposizione di $[a, b]$ in intervalli di ampiezza Δt_j , detti τ_j dei punti appartenenti a tali intervalli, e posto $\delta = \max\{\Delta t_j\}$, diremo che l'integrale (5.8) esiste in mq se esiste il limite

$$\lim_{\delta \rightarrow 0} \text{-mq} \sum_j X(\tau_j) \Delta t_j \tag{5.9}$$

indipendentemente dalla scelta dei punti τ_j negli intervalli della decomposizione.

Proposizione 5.8. *L'integrale (5.8) esiste in mq se e solo se l'autocorrelazione $R(s, t)$ di $X(t)$ è integrabile nel senso che*

$$\int_a^b \int_a^b |R(s, t)| ds dt < +\infty$$

In tal caso si ha anche

$$\int_a^b \int_a^b R(s, t) ds dt = \mathbf{E} \left[\left| \int_a^b X(t) dt \right|^2 \right] \geq 0 \quad (5.10)$$

Dimostrazione: L'esistenza del limite (5.9) è garantita dal rispetto del corrispondente criterio di Cauchy, cioè

$$\lim_{\epsilon, \delta \rightarrow 0} \mathbf{E} \left[\left| \sum_j X(\rho_j) \Delta s_j - \sum_k X(\tau_k) \Delta t_k \right|^2 \right] = 0$$

Siccome d'altra parte si ha

$$\begin{aligned} \mathbf{E} \left[\sum_j X(\rho_j) \Delta s_j \cdot \sum_k X(\tau_k) \Delta t_k \right] &= \sum_{j,k} R(\rho_j, \tau_k) \Delta s_j \Delta t_k \\ \mathbf{E} \left[\sum_j X(\tau_j) \Delta t_j \cdot \sum_k X(\tau_k) \Delta t_k \right] &= \sum_{j,k} R(\tau_j, \tau_k) \Delta t_j \Delta t_k \end{aligned}$$

i cui limiti – se esistono – coincidono con l'integrale a primo membro di (5.10), l'integrabilità in mq di $X(t)$ coinciderà con l'integrabilità di $R(s, t)$. Verificando le condizioni per lo scambio di limiti e attese si ritrova poi anche (5.10) \square

5.5 Stazionarietà ed ergodicità

Definizione 5.9. *Diremo che $X(t)$ è un **processo stazionario** quando comunque scelto $s \in \mathbf{R}$ i due processi $X(t)$ e $X(t + s)$ sono uguali in distribuzione. Diremo invece che il processo ha **incrementi stazionari** quando, fissato $\Delta t > 0$, risulta*

$$\Delta X(t) \stackrel{d}{=} \Delta X(s) \quad \forall s, t \in \mathbf{R}$$

In altri termini la legge globale di un processo stazionario deve essere invariante per un arbitrario cambiamento dell'origine dei tempi, cioè (se il processo è dotato di ddp) comunque scelti t_1, \dots, t_n e s deve risultare

$$f(x_1, t_1; \dots; x_n, t_n) = f(x_1, t_1 + s; \dots; x_n, t_n + s) \quad (5.11)$$

Si noti invece che la stazionarietà degli incrementi richiede solo che la loro legge al tempo t dipenda da Δt ma non da t : essa quindi non implica né la stazionarietà del processo degli incrementi (che richiederebbe delle condizioni anche sulle loro leggi congiunte), né tanto meno quella del processo $X(t)$ stesso. Viceversa la stazionarietà del processo $X(t)$ implica la stazionarietà degli incrementi. Infatti dalla (5.11) segue innanzitutto che la ddp a un tempo di un processo stazionario deve essere costante, e che la sua ddp congiunta a due tempi deve dipendere solo dalla differenza di questi

$$f(x, t) = f(x) \quad (5.12)$$

$$f(x_1, t_1; x_2, t_2) = f(x_1, x_2; \tau) \quad \tau = t_2 - t_1 \quad (5.13)$$

Avremo allora per la fdd degli incrementi $\Delta X(t)$

$$\begin{aligned} F_{\Delta X}(x, t) &= \mathbf{P}\{\Delta X(t) \leq x\} = \int \mathbf{P}\{X(t + \tau) - X(t) \leq x | X(t) = y\} f(y) dy \\ &= \int \mathbf{P}\{X(t + \tau) \leq x + y | X(t) = y\} f(y) dy \end{aligned}$$

da cui con una derivazione e tenendo conto di (5.13) si ricava la ddp

$$f_{\Delta X}(x) = \int f(x + y, t + \tau | y, t) f(y) dy = \int f(x + y, y; \tau) dy \quad (5.14)$$

che non dipende più da t , ma solo da τ : cioè gli incrementi sono stazionari

Ovviamente l'attesa e la varianza di un processo stazionario sono costanti, mentre l'autocorrelazione dipende solo dalla differenza dei tempi

$$\mathbf{E}[X(t)] = m \quad \mathbf{E}[X(t)X(t + \tau)] = R(\tau) \quad \forall t \geq 0 \quad (5.15)$$

Per un processo stazionario sarà dunque vero che

$$C(\tau) = R(\tau) - m^2 \quad \sigma^2(t) = \sigma^2 = C(0) = R(0) - m^2 \quad \rho(\tau) = \frac{R(\tau) - m^2}{R(0) - m^2}$$

e inoltre, data la simmetria di $R(s, t)$ nei suoi due argomenti, la $R(\tau)$ risulterà essere anche una funzione *pari*. Si noti però che, mentre le relazioni (5.15) sono sempre verificate per un processo stazionario, il viceversa non è vero: le sole condizioni (5.15) non sono sufficienti a garantire la stazionarietà nel senso della Definizione 5.9. Comunque data la loro importanza, quando di un processo sappiamo solo che soddisfa le relazioni (5.15), si usa dire che esso è **stazionario in senso lato**.

Proposizione 5.10. *Un processo stazionario in senso lato $X(t)$ con autocorrelazione $R(\tau)$ è continuo in mq se $R(\tau)$ è continua in $\tau = 0$, è derivabile in mq se $R''(\tau)$ esiste in $\tau = 0$, e infine è integrabile in mq su $[-T, T]$ se*

$$\int_{-2T}^{2T} (2T - |\tau|) R(\tau) d\tau < +\infty$$

Dimostrazione: Si tratta ovviamente di semplici corollari delle Proposizioni 5.6, 5.7 e 5.8. In particolare la condizione di integrabilità si ottiene da quella della Proposizione 5.8 con un cambiamento di variabili e una integrazione elementare. \square

Per un processo stazionario diventa ragionevole supporre – con una estensione della Legge dei Grandi Numeri – che i *valori d’attesa* possano essere sostituiti da qualche forma di limite di *medie temporali* del processo. Quando ciò accade si usa dire che il processo è **ergodico**. Esamineremo ora le condizioni sotto le quali si può dire che un p -s stazionario è anche *ergodico per l’attesa e per l’autocorrelazione*. Preliminarmente introduciamo a questo scopo, con $T > 0$ arbitrario, le v -a

$$\bar{X}_T = \frac{1}{2T} \int_{-T}^T X(t) dt \quad R_T(\tau) = \frac{1}{T} \int_0^T X(t)X(t+\tau) dt$$

che rappresentano rispettivamente la *media temporale* del p -s, e l’*autocorrelazione* nella forma di media temporale come era stata inizialmente introdotta da G.I. Taylor nel 1920, e definiamo inoltre la funzione

$$r(\tau, \sigma) = \mathbf{E} [X(t+\sigma+\tau)X(t+\sigma)X(t+\tau)X(t)] - R^2(t)$$

Teorema 5.11. *Dato un p -s stazionario $X(t)$, risulta*

$$\lim_{T \rightarrow \infty} \text{-mq} \bar{X}_T = m \quad \lim_{T \rightarrow \infty} \text{-mq} R_T(\tau) = R(\tau) \quad (5.16)$$

se sono soddisfatte rispettivamente le condizioni

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-2T}^{2T} \left(1 - \frac{|\tau|}{2T}\right) C(\tau) d\tau = 0 \quad (5.17)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-2T}^{2T} \left(1 - \frac{|\sigma|}{2T}\right) r(\tau, \sigma) d\sigma = 0 \quad (5.18)$$

Diremo allora che il processo è **ergodico per l’attesa e per l’autocorrelazione**

Dimostrazione: Le convergenze in *mq* (5.16) possono essere verificate tramite i criteri (4.1) del Teorema 4.6 che nel caso dell’ergodicità per l’attesa sono

$$\lim_{T \rightarrow \infty} \mathbf{E} [\bar{X}_T] = m \quad \lim_{T \rightarrow \infty} \mathbf{V} [\bar{X}_T] = 0 \quad (5.19)$$

Trascurando di verificare che lo scambio fra integrali e valori d’attesa è possibile, il primo criterio è banalmente soddisfatto dato che per ogni $T > 0$ si ha

$$\mathbf{E} [\bar{X}_T] = \frac{1}{2T} \int_{-T}^T \mathbf{E} [X(t)] dt = \frac{m}{2T} \int_{-T}^T dt = m$$

Quanto al secondo criterio (5.19) osserviamo che

$$\begin{aligned} \mathbf{V} [\bar{X}_T] &= \mathbf{E} [\bar{X}_T^2] - \mathbf{E} [\bar{X}_T]^2 = \frac{1}{4T^2} \mathbf{E} \left[\int_{-T}^T \int_{-T}^T X(s)X(t) dsdt \right] - m^2 \\ &= \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T [R(t-s) - m^2] dsdt = \frac{1}{4T^2} \iint_D C(t-s) dsdt \end{aligned}$$

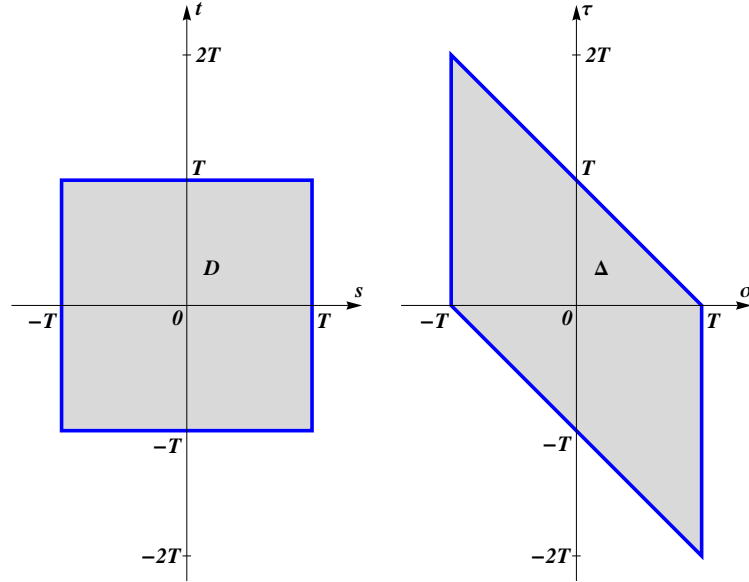


Figura 5.1: Modifica del dominio di integrazione per il cambiamento di variabili (5.20)

e che, con il seguente cambiamento di variabili di integrazione rappresentato anche nella Figura 5.1

$$\tau = t - s \quad \sigma = s \quad |J| = 1 \quad (5.20)$$

si ottiene

$$\begin{aligned} \mathbf{V} [\bar{X}_T] &= \frac{1}{4T^2} \iint_D C(t-s) ds dt = \frac{1}{4T^2} \iint_{\Delta} C(\tau) d\sigma d\tau \\ &= \frac{1}{4T^2} \left[\int_{-2T}^0 d\tau C(\tau) \int_{-T-\tau}^T d\sigma + \int_0^{2T} d\tau C(\tau) \int_{-T}^{T-\tau} d\sigma \right] \\ &= \frac{1}{4T^2} \left[\int_{-2T}^0 C(\tau)(2T + \tau) d\tau + \int_0^{2T} C(\tau)(2T - \tau) d\tau \right] \\ &= \frac{1}{2T} \int_{-2T}^{2T} C(\tau) \left(1 - \frac{|\tau|}{2T} \right) d\tau \end{aligned}$$

e quindi la seconda equazione (5.19) è verificata se l'ipotesi (5.17) è soddisfatta. Tralascieremo invece l'analoga dimostrazione esplicita dell'ergodicità per l'autocorrelazione. \square

Corollario 5.12. *Un p -s stazionario $X(t)$ è ergodico per l'attesa e l'autocorrelazione se risulta*

$$\int_0^{+\infty} |C(\tau)| d\tau < +\infty \quad (5.21)$$

Dimostrazione: Infatti, siccome

$$\left| 1 - \frac{|\tau|}{2T} \right| \leq 1 \quad -2T \leq \tau \leq 2T$$

dalla simmetria di $C(\tau)$ abbiamo facilmente

$$\left| \frac{1}{T} \int_{-2T}^{2T} \left(1 - \frac{|\tau|}{2T} \right) C(\tau) d\tau \right| \leq \frac{1}{T} \int_{-2T}^{2T} |C(\tau)| d\tau = \frac{2}{T} \int_0^{2T} |C(\tau)| d\tau$$

ed quindi la (5.17) è soddisfatta se la (5.21) è verificata. Tralascieremo invece di verificare la condizione per l'ergodicità dell'autocorrelazione \square

5.6 Spettro di potenza

Si potrebbe pensare di eseguire una analisi in frequenza di un p -s $X(t)$ (tipicamente nel caso in cui $X(t)$ è un *segnale*) semplicemente calcolandone la trasformata di Fourier, ma ci si accorge facilmente del fatto che in generale le traiettorie non sono di quadrato integrabile su $[0, +\infty)$, sicché una trasformata definita in maniera diretta semplicemente non esiste. Si ricorre allora ad un espediente che si avvale dalla definizione di una trasformata *troncata*

$$\tilde{X}_T(\varpi) = \int_0^T X(t) e^{-i\varpi t} dt \quad (5.22)$$

che per ogni $T > 0$ esiste ed è un nuovo p -s con indice ϖ . Partendo allora dall'idea secondo la quale il modulo quadro di una trasformata di Fourier rappresenta il contributo energetico delle singole componenti del segnale, lo *spettro di potenza* viene inizialmente introdotto come

$$S(\varpi) = \lim_{T \rightarrow \infty} \text{-mq} \frac{1}{T} \left| \tilde{X}_T(\varpi) \right|^2 \quad (5.23)$$

Questo $S(\varpi)$ è in linea di principio ancora un p -s con parametro ϖ , ed è quindi un risultato notevole il fatto che, per p -s stazionari ed ergodici, $S(\varpi)$ risulti invece una funzione ordinaria (non aleatoria) calcolabile anche come trasformata di Fourier della funzione di autocorrelazione $R(\tau)$

Teorema 5.13. Teorema di Wiener-Khinchin: *Dato un p -s stazionario $X(t)$, se esso è ergodico nel senso che soddisfa le ipotesi del Teorema 5.11, risulta*

$$S(\varpi) = \lim_{T \rightarrow \infty} \text{-mq} \frac{1}{T} \left| \tilde{X}_T(\varpi) \right|^2 = \int_{-\infty}^{+\infty} R(\tau) e^{-i\varpi \tau} d\tau \quad (5.24)$$

Dimostrazione: Le ipotesi formulate garantiscono la convergenza dei limiti e la legittimità dei passaggi formali che eseguiremo, ma noi trascureremo di verificare esplicitamente questi aspetti. Ci limiteremo invece a mostrare in che modo il risultato (5.24) emerge dalla definizione (5.23) e dal Teorema 5.11. Dalle definizioni (5.22) e (5.23), e con il cambiamento di variabili di integrazione (5.20) si ha innanzitutto

$$\begin{aligned}
 S(\varpi) &= \lim_{T \rightarrow \infty} \text{-mq} \frac{1}{T} \int_0^T \int_0^T X(t)X(s)e^{-i\varpi(t-s)} dt ds \\
 &= \lim_{T \rightarrow \infty} \text{-mq} \frac{1}{T} \left[\int_{-T}^0 d\tau \int_{-\tau}^T d\sigma e^{-i\varpi\tau} X(\sigma)X(\sigma + \tau) \right. \\
 &\quad \left. + \int_0^T d\tau \int_0^{T-\tau} d\sigma e^{-i\varpi\tau} X(\sigma)X(\sigma + \tau) \right] \\
 &= \lim_{T \rightarrow \infty} \text{-mq} \frac{1}{T} \int_0^T d\tau \left[e^{i\varpi\tau} \int_{\tau}^T X(\sigma)X(\sigma - \tau) d\sigma \right. \\
 &\quad \left. + e^{-i\varpi\tau} \int_0^{T-\tau} X(\sigma)X(\sigma + \tau) d\sigma \right]
 \end{aligned}$$

e siccome con un altro cambiamento di variabili si verifica che

$$\int_{\tau}^T X(\sigma)X(\sigma - \tau) d\sigma = \int_0^{T-\tau} X(\sigma)X(\sigma + \tau) d\sigma$$

in definitiva avremo

$$S(\varpi) = \lim_{T \rightarrow \infty} \text{-mq} \int_0^T d\tau \cos \varpi\tau \frac{2}{T} \int_0^{T-\tau} X(\sigma)X(\sigma + \tau) d\sigma$$

Per semplificare la dimostrazione supporremo ora che il limite in questione possa essere eseguito in due passaggi separati e successivi

$$S(\varpi) = \lim_{T \rightarrow \infty} \int_0^T d\tau \cos \varpi\tau \lim_{T' \rightarrow \infty} \text{-mq} \frac{2}{T'} \int_0^{T'-\tau} X(\sigma)X(\sigma + \tau) d\sigma$$

Eseguendo prima il limite in T' con τ arbitrario, ma finito e fissato in $[0, T]$, potremo fare uso dell'ipotizzata ergodicità per l'autocovarianza come espressa nella seconda di (5.16) nel Teorema 5.11 per ottenere

$$S(\varpi) = \lim_{T \rightarrow \infty} 2 \int_0^T R(\tau) \cos \varpi\tau d\tau = \int_{-\infty}^{+\infty} R(\tau) e^{-i\varpi\tau} d\tau$$

dove abbiamo tenuto conto del fatto che $R(\tau)$ è una funzione reale e pari. □

La rilevanza pratica di questo risultato ha fatto sì che nel corso degli anni sia prevalsa la tendenza a considerare (5.24) come la definizione stessa di spettro di potenza trascurando completamente la sua origine. Adotteremo anche noi questo punto di vista e da ora in poi ci uniformeremo alla seguente definizione che non fa più esplicito riferimento all'ergodicità del processo, ma solo alla sua stazionarietà

Definizione 5.14. *Dato un processo stazionario $X(t)$ chiameremo **spettro di potenza** la trasformata di Fourier della sua autocorrelazione definita dalle relazioni di reciprocità*

$$S(\varpi) = \int_{-\infty}^{+\infty} R(\tau) e^{-i\varpi\tau} d\tau \qquad R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S(\varpi) e^{i\varpi\tau} d\varpi \qquad (5.25)$$

Si noti che la tipica condizione di ergodicità (5.21) richiede che $C(\tau)$ sia infinitesima per $\tau \rightarrow \pm\infty$, ovvero che $X(t)$ e $X(t + \tau)$ siano non correlate per grandi separazioni τ . In questo caso dalle definizioni si ha ovviamente $R(\tau) \rightarrow m^2$ per $\tau \rightarrow \pm\infty$, e quindi la trasformata di Fourier (5.25) non esiste se m non si annulla. Per evitare questi problemi si usa allora introdurre anche una seconda definizione che fa appello all'autocovarianza $C(\tau)$ invece che all'autocorrelazione $R(\tau)$

Definizione 5.15. *Dato un processo stazionario $X(t)$ chiameremo **spettro di covarianza** la trasformata di Fourier della sua autocovarianza definita dalle relazioni di reciprocità*

$$S_c(\varpi) = \int_{-\infty}^{+\infty} C(\tau) e^{-i\varpi\tau} d\tau \qquad C(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_c(\varpi) e^{i\varpi\tau} d\varpi \qquad (5.26)$$

Capitolo 6

Costruzioni euristiche

6.1 Processo di Poisson

In questa prima discussione introdurremo il processo di Poisson con la esplicita costruzione delle sue traiettorie dalle quali ricaveremo poi anche tutte le principali proprietà statistiche. Come vedremo questa procedura, che è ovviamente suggestiva e istruttiva, non è facilmente replicabile nel caso di altri processi per i quali la definizione dovrà essere data come limite o seguendo metodi più generali

6.1.1 Processi di punto e rinnovi

Supponiamo che su un asse di tempi siano lanciati *a caso* dei punti, e poniamoci prima di tutto il problema di studiare la *v-a* che conta il numero di punti che cadono in un intervallo $[s, t]$ di ampiezza $\Delta t = t - s > 0$. Così formulata, però, la domanda è piuttosto vaga e imprecisa: innanzitutto bisogna chiarire cosa si intende dicendo che i punti sono lanciati *a caso*; inoltre, se un punto è lanciato a caso (qualunque cosa ciò voglia dire, tranne eventualità molto particolari) su un *asse infinito* la probabilità di cadere in un intervallo finito $[s, t]$ sarà semplicemente zero; infine si deve precisare quale è il *numero dei punti lanciati*. Per trovare una risposta coerente a queste domande seguiremo allora una procedura di approssimazioni successive

Proposizione 6.1. *Se su un asse illimitato viene lanciato a caso (nel senso definito precisamente più oltre) un numero infinito di punti indipendenti, e se il numero medio di punti per unità di tempo (*intensità*) è λ , la *v-a**

$$N = \text{numero dei punti che cadono in un intervallo di ampiezza } \Delta t$$

segue una distribuzione di Poisson di parametro $\lambda\Delta t$, cioè $N \sim \mathfrak{P}(\lambda\Delta t)$ con

$$P\{N = k\} = e^{-\lambda\Delta t} \frac{(\lambda\Delta t)^k}{k!}$$

*Inoltre le analoghe *v-a* $N_1 \sim \mathfrak{P}(\lambda\Delta t_1)$ e $N_2 \sim \mathfrak{P}(\lambda\Delta t_2)$ relative a intervalli non sovrapposti sono indipendenti*

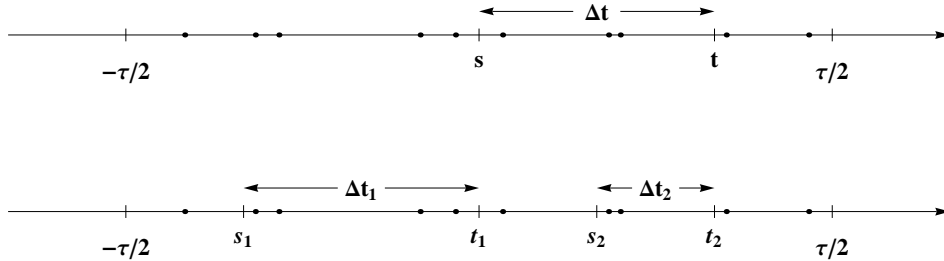


Figura 6.1: Istanti aleatori lanciati su intervalli finiti $[-\frac{\tau}{2}, \frac{\tau}{2}]$.

Dimostrazione: Per affrontare gradualmente il problema partiremo con un intervallo finito $[-\tau/2, \tau/2]$ con $\tau > 0$ e contenente $[s, t]$ (vedi Fig. 6.1), e lanceremo n punti *a caso* nel senso che:

- la posizione di ogni punto in $[-\tau/2, \tau/2]$ è aleatoria e indipendente da quella degli altri,
- la distribuzione di questa posizione aleatoria è uniforme in $[-\tau/2, \tau/2]$.

Abbiamo così precisato che ciascuno degli n punti cadrà in $[s, t]$ con probabilità

$$p = \frac{\Delta t}{\tau}$$

e che, essendo gli n lanci indipendenti, la *v-a* $X = \text{numero dei punti che cadono in } [s, t]$ seguirà una distribuzione binomiale $\mathfrak{B}(n; p)$. Ovviamente X dipende da n e τ arbitrariamente scelti, e noi supporremo di far crescere contemporaneamente n e τ all'infinito, mantenendo Δt costante e supponendo che il rapporto n/τ (numero di punti per unità di tempo) si mantenga limitato e converga verso un numero finito e positivo λ ; cioè supporremo che

$$n \rightarrow \infty \quad \tau \rightarrow +\infty \quad \frac{n}{\tau} \rightarrow \lambda > 0 \tag{6.1}$$

$$p = \frac{\Delta t}{\tau} \rightarrow 0 \quad np = \frac{n}{\tau} \Delta t \rightarrow \lambda \Delta t \tag{6.2}$$

In questo modo, al variare di n e τ , abbiamo costruito una famiglia di *v-a* binomiali $X \sim \mathfrak{B}(n; p)$ che nel limite (6.1) soddisfano le ipotesi del Teorema binomiale di Poisson 4.30, per cui potremo dire che in distribuzione avremo

$$X \xrightarrow{d} N \sim \mathfrak{P}(\lambda \Delta t)$$

e potremo quindi trarre una prima conclusione: se viene lanciato su un asse illimitato un numero infinito di punti aleatori e indipendenti, e se il numero di punti per unità

di tempo è *mediamente* λ (un numero che ha quindi le dimensioni di una *frequenza*), la *v-a* limite $N = \text{numero dei punti che cadono in un intervallo di ampiezza } \Delta t$ segue una distribuzione di Poisson $\mathfrak{P}(\lambda\Delta t)$, cioè

$$\mathbf{P}\{N = k\} = e^{-\lambda\Delta t} \frac{(\lambda\Delta t)^k}{k!}$$

Ripetiamo ora la procedura prendendo, come in Figura 6.1, *due* intervalli disgiunti $[s_1, t_1]$ e $[s_2, t_2]$ in $[-\tau/2, \tau/2]$ con $\Delta t_1 = t_1 - s_1$ e $\Delta t_2 = t_2 - s_2$: lanciando n punti con le medesime caratteristiche del caso precedente, e ponendo $X_1 = \text{numero dei punti che cadono in } [s_1, t_1]$, $X_2 = \text{numero dei punti che cadono in } [s_2, t_2]$, e $X_0 = \text{numero dei punti che cadono altrove in } [-\tau/2, \tau/2]$ avremo che il *vett-a* $\mathbf{X} = (X_1, X_2)$ seguirà una distribuzione multinomiale $\mathfrak{B}(n; p_1, p_2)$ con $r = 2$ e con

$$p_1 = \frac{\Delta t_1}{\tau}, \quad p_2 = \frac{\Delta t_2}{\tau}$$

Se allora supponiamo come prima di far crescere contemporaneamente n e τ verso l'infinito, sempre in modo che le (6.1) siano rispettate e mantenendo Δt_1 e Δt_2 costanti, avremo questa volta

$$\begin{aligned} p_1 &= \frac{\Delta t_1}{\tau} \rightarrow 0, & np_1 &= \frac{n}{\tau} \Delta t_1 \rightarrow \lambda \Delta t_1 \\ p_2 &= \frac{\Delta t_2}{\tau} \rightarrow 0, & np_2 &= \frac{n}{\tau} \Delta t_2 \rightarrow \lambda \Delta t_2 \end{aligned}$$

e quindi in base al Teorema multinomiale di Poisson 4.31

$$\mathbf{X} = (X_1, X_2) \xrightarrow{d} (N_1, N_2) \sim \mathfrak{P}(\lambda\Delta t_1) \cdot \mathfrak{P}(\lambda\Delta t_2)$$

Pertanto le due *v-a* limite N_1 e N_2 , si comporteranno come due *v-a* di Poisson *indipendenti*. Si noti che invece, lungo il procedimento di limite, per ogni n finito le *v-a* X_1 e X_2 *non* sono indipendenti. \square

Nell'Appendice G verrà spiegato come si modificano questi risultati se l'intensità λ dei punti non è più supposta uniforme. Introduremo poi la successione di *v-a* T_n che rappresentano le posizioni temporali dei nostri punti aleatori. Per definire tale successione, però, dovremo assegnare un ordine e degli indici ai punti fissando prima di tutto un'arbitraria origine non aleatoria $T_0 = 0$, e stabilendo poi che T_1 è l'istante in cui cade il primo punto *a destra* dell'origine, T_2 l'istante del secondo, mentre T_{-1} è quello del primo *a sinistra*, e così via. Ne risulta una successione bilaterale T_n , con $n = \pm 1, \pm 2, \dots$, di *v-a* che, però, *non sono più indipendenti* dato che ora ogni T_n non può precedere T_{n-1} . Si noti che viceversa il tempo d'attesa $\Delta T_n = T_{n+1} - T_n$ dell' $(n+1)$ -*mo* punto resta indipendente da T_n . La successione dei T_n così introdotta è un tipico esempio particolare di **processo di punto**, e noi ne studieremo ora la distribuzione di probabilità

Proposizione 6.2. *Le v -a T_n con $n \geq 1$, quelle cioè che cadono dopo $T_0 = 0$, seguono le leggi di Erlang $\mathfrak{E}_n(\lambda)$; quelle che cadono prima di $T_0 = 0$, per simmetria, verificano la relazione $-T_{-n} \stackrel{d}{=} T_n$; infine i tempi d'attesa fra due punti successivi $\Delta T_n = T_{n+1} - T_n$ sono v -a iid positive che seguono tutte la legge esponenziale $\mathfrak{E}(\lambda)$*

Dimostrazione: Consideriamo innanzitutto il caso $n \geq 1$ dei punti che cadono a destra di (cioè dopo) $T_0 = 0$: se N è il numero di punti in $[0, t]$ con legge $\mathfrak{P}(\lambda t)$, e $\vartheta(t)$ è la funzione di Heaviside (2.13), la fdd di $T_n > 0$ sarà

$$F_n(t) = \mathbf{P}\{T_n \leq t\} = \mathbf{P}\{N \geq n\} = 1 - \mathbf{P}\{N < n\} = \left[1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} \right] \vartheta(t)$$

e quindi la ddp

$$f_n(t) = F'_n(t) = \left[\sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} - \sum_{k=1}^{n-1} \frac{(\lambda t)^{k-1}}{(k-1)!} \right] \lambda e^{-\lambda t} \vartheta(t) = \frac{(\lambda t)^{n-1}}{(n-1)!} \lambda e^{-\lambda t} \vartheta(t) \quad (6.3)$$

coincide con la ddp (4.27) di una *distribuzione di Erlang* $\mathfrak{E}_n(\lambda)$. In particolare la legge di T_1 è la $\mathfrak{E}_1(\lambda)$, cioè un'esponenziale $\mathfrak{E}(\lambda)$ con ddp $f_1(t) = \lambda e^{-\lambda t} \vartheta(t)$. Analogamente si dimostra che $-T_{-n} \stackrel{d}{=} T_n$, ma per brevità trascureremo di farlo. Per studiare la legge dei tempi d'attesa ΔT_n , invece, cominciamo con il ricordare che da (4.26) le fc delle T_n di Erlang sono

$$\varphi_n(u) = \mathbf{E} [e^{iuT_n}] = \int_0^{+\infty} \frac{(\lambda t)^{n-1}}{(n-1)!} \lambda e^{-\lambda t} e^{iut} dt = \left(\frac{\lambda}{\lambda - iu} \right)^n$$

e inoltre che T_n e ΔT_n sono v -a indipendenti. Pertanto, detta $\varphi(u)$ la fc di ΔT_n , la v -a $T_{n+1} = T_n + \Delta T_n$ avrà fc $\varphi_{n+1}(u) = \varphi_n(u)\varphi(u)$ e quindi

$$\varphi(u) = \frac{\varphi_{n+1}(u)}{\varphi_n(u)} = \varphi_1(u) = \frac{\lambda}{\lambda - iu}$$

cioè le ΔT_n sono tutte v -a iid con legge esponenziale $\mathfrak{E}(\lambda)$. □

Si noti che le v -a ΔT_n sono un particolare esempio di **successione di rinnovi**, cioè di v -a iid $Z_n > 0$, che possono essere usate a loro volta come punto di partenza per costruire un processo di punto secondo le relazioni

$$T_n = \sum_{k=0}^{n-1} Z_k \quad Z_n = \Delta T_n \quad (6.4)$$

In pratica ogni successione di rinnovi genera un processo di punto e viceversa, ma va detto che in generale i rinnovi possono seguire una legge arbitraria diversa da $\mathfrak{E}(\lambda)$. È importante allora notare che una successione di rinnovi *esponenziali* genera sempre un processo di punto T_n con leggi di Erlang (vedi Esempio 4.22), e numeri di punti arrivati in intervalli finiti N distribuiti secondo leggi di Poisson, come mostreremo nella proposizione seguente

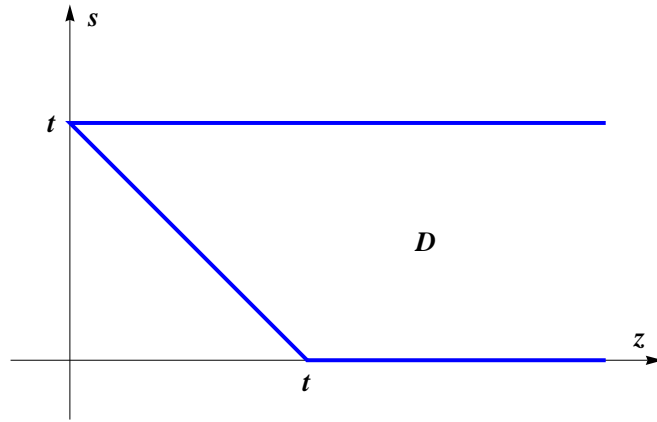


Figura 6.2: Dominio di integrazione per l'integrale (6.5)

Proposizione 6.3. Se Z_n è una successione di rinnovi con legge $\mathfrak{E}(\lambda)$, e T_n è il corrispondente processo di punto (6.4) con leggi di Erlang $\mathfrak{E}_n(\lambda)$, il numero N dei punti T_n che cadono in $[0, t]$ segue la legge di Poisson $\mathfrak{P}(\lambda t)$

Dimostrazione: Infatti, con uno scambio nell'ordine delle integrazioni sul dominio D rappresentato nella Figura 6.2

$$\begin{aligned}
 \mathbf{P}\{N = n\} &= \mathbf{P}\{T_n \leq t, T_{n+1} > t\} = \mathbf{P}\{T_n \leq t, T_n + Z_n > t\} \\
 &= \mathbf{E} [\mathbf{P}\{T_n \leq t, T_n + Z_n > t \mid Z_n\}] \\
 &= \int_0^{+\infty} \mathbf{P}\{T_n \leq t, T_n + Z_n > t \mid Z_n = z\} \lambda e^{-\lambda z} dz \\
 &= \int_0^{+\infty} \mathbf{P}\{t - z < T_n \leq t\} \lambda e^{-\lambda z} dz \\
 &= \int_0^{+\infty} dz \lambda e^{-\lambda z} \int_{t-z}^t \frac{(\lambda s)^{n-1}}{(n-1)!} \lambda e^{-\lambda s} \vartheta(s) ds \\
 &= \iint_D \lambda e^{-\lambda z} \frac{(\lambda s)^{n-1}}{(n-1)!} \lambda e^{-\lambda s} \vartheta(s) dz ds \\
 &= \int_0^t ds \frac{(\lambda s)^{n-1}}{(n-1)!} \lambda e^{-\lambda s} \int_{t-s}^{+\infty} \lambda e^{-\lambda z} dz = \int_0^t \frac{(\lambda s)^{n-1}}{(n-1)!} \lambda e^{-\lambda s} e^{-\lambda(t-s)} ds \\
 &= e^{-\lambda t} \frac{\lambda^n}{n!} \int_0^t n s^{n-1} ds = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \tag{6.5}
 \end{aligned}$$

Ne consegue che N è distribuita secondo la legge di Poisson $\mathfrak{P}(\lambda t)$ come nel processo di punto introdotto all'inizio di questa sezione \square

6.1.2 Processo di Poisson

Definizione 6.4. *Assegnato un processo di punto T_n di intensità λ , il **processo di Poisson standard di intensità** λ è il p -s $N(t)$ con $t > 0$ che conta il numero aleatorio di punti T_n che cadono in $[0, t]$, con la condizione iniziale $N(0) = 0$, P -qo. Sulla base del processo di punto T_n il p -s $N(t)$ potrà quindi essere anche rappresentato come*

$$N(t) = \sum_{k=1}^{\infty} \vartheta(t - T_k) \quad (6.6)$$

dove ϑ è la funzione di Heaviside (2.13). Inoltre, con $\Delta t > 0$ fissato, si definisce il corrispondente **processo degli incrementi di Poisson** $\Delta N(t) = N(t + \Delta t) - N(t)$ che conta il numero di punti T_n che cadono in un intervallo di ampiezza Δt

Proposizione 6.5. *Il processo di Poisson $N(t)$ ha incrementi indipendenti e stazionari; inoltre, posto per comodità*

$$q_\alpha(k) = e^{-\alpha} \frac{\alpha^k}{k!} \quad (6.7)$$

le distribuzioni e le fc di $N(t)$ e $\Delta N(t)$ rispettivamente sono

$$p_N(k, t) = \mathbf{P}\{N(t) = k\} = q_{\lambda t}(k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad (6.8)$$

$$p_{\Delta N}(k) = \mathbf{P}\{\Delta N(t) = k\} = q_{\lambda \Delta t}(k) = e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^k}{k!} \quad (6.9)$$

$$\varphi_N(u, t) = e^{\lambda t(e^{iu} - 1)} \quad \varphi_{\Delta N}(u) = e^{\lambda \Delta t(e^{iu} - 1)} \quad (6.10)$$

infine le **probabilità di transizione** (cioè le probabilità condizionate a due tempi) di $N(t)$, con $\Delta t > 0$ e $k \geq \ell$, sono

$$p_N(k, t + \Delta t | \ell, t) = q_{\lambda \Delta t}(k - \ell) \quad (6.11)$$

Dimostrazione: Gli incrementi relativi a intervalli non sovrapposti (con al più un estremo in comune) sono v -a indipendenti per costruzione, e inoltre l'incremento $\Delta N(t)$ risulta anche indipendente da $N(t)$: pertanto il processo di Poisson costituisce il primo esempio di *processo ad incrementi indipendenti*, una categoria di p -s della quale ripareremo nella Sezione 7.1.3. Dalla discussione svolta nelle sezioni precedenti sappiamo inoltre che, per ogni $t > 0$, $N(t) \sim \mathfrak{P}(\lambda t)$, mentre $\Delta N(t) \sim \mathfrak{P}(\lambda \Delta t)$, e quindi che valgono le (6.8), (6.9) e (6.10). Ne segue che le leggi degli incrementi $\Delta N(t)$ – diversamente da quanto avviene per quelle di $N(t)$ – non cambiano al variare di t ma dipendono solo da Δt , per cui potremo anche dire che $N(t)$ ha *incrementi stazionari* (vedi Sezione 5.5). Infine, per l'indipendenza degli incrementi, si ha

$$\begin{aligned} p_N(k, t + \Delta t | \ell, t) &= \mathbf{P}\{N(t + \Delta t) = k | N(t) = \ell\} \\ &= \mathbf{P}\{N(t + \Delta t) - N(t) + N(t) = k | N(t) = \ell\} \\ &= \mathbf{P}\{\Delta N(t) = k - \ell\} \\ &= q_{\lambda \Delta t}(k - \ell) = e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^{k - \ell}}{(k - \ell)!} \end{aligned}$$

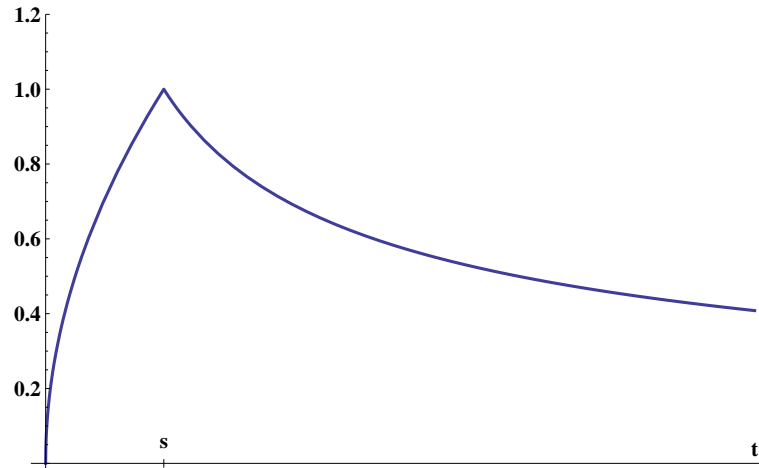


Figura 6.3: Coefficiente di correlazione $\rho_N(s, t)$ (6.15) del processo di Poisson standard $N(t)$.

cioè la (6.11). Si noti come questa distribuzione dipenda solo da Δt , e non da t , a causa della stazionarietà degli incrementi (Sezione 5.5); e da $k - \ell$, ma non da k e ℓ separatamente, a causa dell'indipendenza degli incrementi (Sezione 7.1.3) \square

Proposizione 6.6. *Le proprietà statistiche principali del processo di Poisson standard $N(t)$ sono*

$$m_N(t) = \sigma_N^2(t) = \lambda t \quad (6.12)$$

$$R_N(s, t) = \lambda \min\{s, t\} + \lambda^2 st \quad (6.13)$$

$$C_N(s, t) = \lambda \min\{s, t\} \quad (6.14)$$

$$\rho_N(s, t) = \frac{\min\{s, t\}}{\sqrt{st}} = \begin{cases} \sqrt{s/t} & \text{se } s < t \\ \sqrt{t/s} & \text{se } t < s \end{cases} \quad (6.15)$$

Dimostrazione: Le (6.12) discendono immediatamente da (6.8). Per dimostrare la (6.13) si comincia con l'osservare che dai risultati precedenti per $s = t$ si ha

$$R_N(t, t) = \mathbf{E} [N^2(t)] = \mathbf{V} [N(t)] + \mathbf{E} [N(t)]^2 = \lambda t + \lambda^2 t^2$$

e poi che da questa, e dall'indipendenza degli incrementi, per $s < t$ avremo

$$\begin{aligned} R_N(s, t) &= \mathbf{E} [N(s)N(t)] = \mathbf{E} [N(s)(N(t) - N(s) + N(s))] \\ &= \mathbf{E} [N(s)] \mathbf{E} [N(t) - N(s)] + R_N(s, s) \\ &= \lambda s \cdot \lambda(t - s) + \lambda s + \lambda^2 s^2 = \lambda s + \lambda^2 st \end{aligned}$$

cioè in definitiva la (6.13) per scelte arbitrarie di s e t ; (6.14) e (6.15) seguono poi dalla definizione (5.3) e da (6.12). Da (6.14) ritroviamo coerentemente anche la varianza (6.12) $\sigma_N^2(t) = C(t, t) = \lambda t$ \square

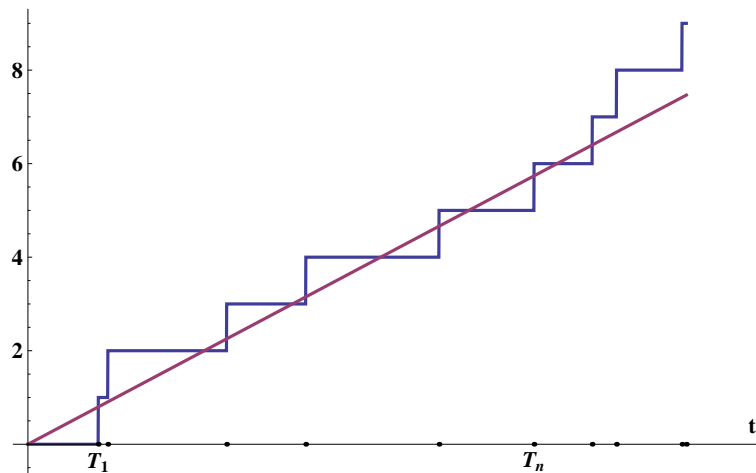


Figura 6.4: Esempio di traiettoria del processo di Poisson $N(t)$ con 10 punti. Nel grafico sono mostrati anche il processo di punto T_n e la funzione $m_N(t) = \lambda t$

Si noti che la varianza (6.12) aumenta linearmente con il tempo: un comportamento, detto talora *diffusivo*, che ritroveremo anche in altri processi. L'andamento del coefficiente di correlazione (6.15), invece, può essere visto nella Figura 6.3 nella quale, per fissare le idee, abbiamo considerato s come costante e t come variabile. Come si osserva chiaramente la correlazione diminuisce – benché piuttosto lentamente – quando i due istanti di tempo si allontanano: il processo all'istante t progressivamente *dimentica* il suo stato all'istante s man mano che il tempo trascorre. Il processo $N(t)$ gode anche di altre proprietà che incontreremo nel seguito. Ad esempio si verifica facilmente per sostituzione che le distribuzioni (6.8) sono soluzioni dell'equazione

$$\partial_t p_N(n, t) = -\lambda [p_N(n, t) - p_N(n - 1, t)] \quad p_N(n, 0) = \delta_{n0} \quad (6.16)$$

che costituisce un primo esempio di *master equation*: un tipo di equazione che esamineremo con maggior dettaglio più innanzi (Sezione 7.2.3). Peraltro anche le probabilità di transizione $p_N(k, t|\ell, s)$ in (6.11) sono soluzioni della medesima *master equation*, ma con condizioni iniziali $p_N(k, s^+) = \delta_{k\ell}$. Dallo sviluppo in serie di potenze dell'esponenziale per $t \rightarrow 0^+$ si ricavano poi i seguenti comportamenti

$$p_N(n, t) = [1 - \lambda t + o(t)] \frac{(\lambda t)^n}{n!} = \begin{cases} 1 - \lambda t + o(t) & n = 0 \\ \lambda t + o(t) & n = 1 \\ o(t) & n \geq 2 \end{cases} \quad (6.17)$$

che sono caratteristici delle distribuzioni del processo di Poisson: si potrebbe dimostrare infatti (ma noi trascureremo di farlo) che se la $p_N(n, t)$ di un *processo di nascita* – come è il processo di Poisson – soddisfa le (6.17), allora essa è anche soluzione di (6.16) e quindi deve avere la forma (6.8).

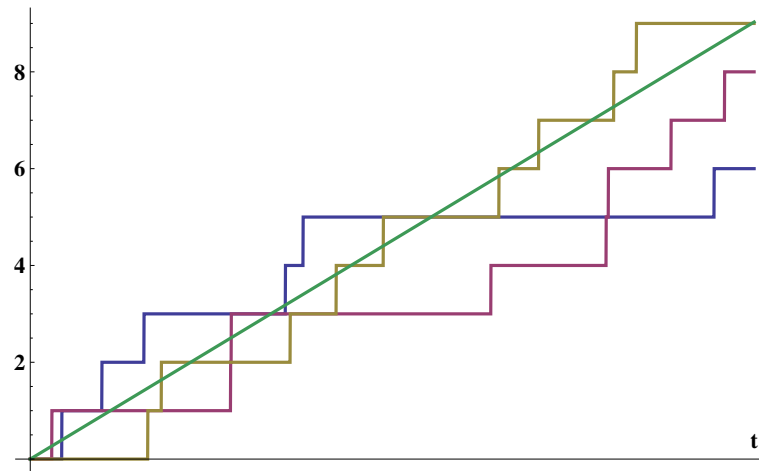


Figura 6.5: Esempio di alcune traiettorie del processo di Poisson $N(t)$ che si distribuiscono attorno all'attesa $m_N(t) = \lambda t$

Le tipiche **traiettorie** del processo di Poisson $N(t)$ per $t > 0$ sono mostrate nella Figura 6.4: esse consistono in *scalinate* crescenti e infinite con gradini di lunghezza aleatoria (gli istanti del processo di punto), ma di altezza deterministicamente uguale a 1, e rappresentano il *conteggio* del numero di punti aleatori che cadono nell'intervallo $[0, t]$. La relazione fra le traiettorie e la funzione lineare $m_N(t)$ è mostrata nelle figure: osservate su un piccolo intervallo temporale le singole traiettorie in Figura 6.5 si discostano poco dalla loro *tendenza media* e nel complesso esse si distribuiscono equamente attorno a $m_N(t) = \lambda t$. Se poi consideriamo tempi più lunghi come in Figura 6.6 le traiettorie, continuano a distribuirsi equamente attorno alla retta λt , ma se ne allontanano progressivamente per effetto dell'aumento della varianza. Il fatto che nella Figura 6.6 le traiettorie sembrano allontanarsi poco da $m_N(t) = \lambda t$ dipende principalmente da un effetto di scala: la deviazione standard è $\sqrt{\lambda t}$, mentre la scala dell'asse verticale cresce come λt .

Proposizione 6.7. *Per ogni $t > 0$ il processo di Poisson $N(t)$ è continuo ma non derivabile in m ; invece esso è continuo e derivabile \mathbf{P} -qo, e in questo senso risulta $\dot{N}(t) = 0$. Inoltre $N(t)$ non è stazionario.*

Dimostrazione: In base alla Proposizione 5.6 la continuità in m deriva dal fatto che l'autocorrelazione (6.13) è una funzione continua. Non esiste invece in $s = t$ la derivata mista $\partial_s \partial_t R_N$: infatti la derivata prima è

$$\partial_t R_N(s, t) = \lambda \vartheta(s - t) + \lambda^2 s$$

dove ϑ è la funzione di Heaviside (2.13) con una discontinuità in $s = t$, e quindi la derivata seconda $\partial_s \partial_t R_N$ non esiste in $s = t$. Dalla Proposizione 5.7 segue allora che il processo non è derivabile in m in nessun $t > 0$. Invece per dimostrare che la

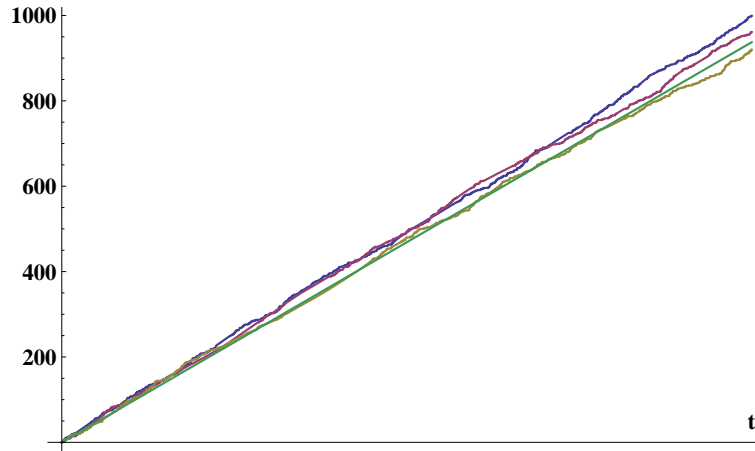


Figura 6.6: Traiettorie del processo di Poisson $N(t)$ con 1000 punti. La deviazione standard aumenta come $\sqrt{\lambda t}$, ma le traiettorie sembrano discostarsi poco dall'attesa $m_N(t) = \lambda t$ per un effetto di scala: $\sqrt{1000} \simeq 32$

derivata \mathbf{P} -qo esiste e si annulla dovremmo provare che

$$\mathbf{P} \left\{ \lim_{\Delta t \rightarrow 0} \frac{\Delta N(t)}{\Delta t} = 0 \right\} = 1 \quad \forall t > 0$$

ma noi ci accontenteremo di mostrare solo che per ogni $t > 0$

$$\lim_{\Delta t \rightarrow 0} \mathbf{P} \left\{ \left| \frac{\Delta N(t)}{\Delta t} \right| < \epsilon \right\} = 1 \quad \forall \epsilon > 0 \quad (6.18)$$

Infatti, fissato un arbitrario $\epsilon > 0$ e scelto $|\Delta t| < 1/\epsilon$, da (6.9) su intervalli di ampiezza $|\Delta t|$ si ha al limite per $\Delta t \rightarrow 0$

$$\mathbf{P} \left\{ \left| \frac{\Delta N(t)}{\Delta t} \right| < \epsilon \right\} = \mathbf{P} \{ \Delta N(t) = 0 \} = e^{-\lambda |\Delta t|} \longrightarrow 1$$

Per la continuità \mathbf{P} -qo poi si adoperano degli argomenti analoghi a quelli usati per la derivabilità \mathbf{P} -qo. Infine la non stazionarietà (anche solo in senso lato) di $N(t)$ si verifica osservando che l'attesa (6.12) non è costante, e che l'autocorrelazione (6.13) non dipende solo da $t - s$, ma da s e t separatamente \square

Data la conformazione *a salti* delle traiettorie del processo di Poisson, le continuità in m_q e \mathbf{P} -qo della Proposizione 6.7 possono risultare sorprendenti. Bisogna però ricordare a questo proposito che tali continuità (come pure la continuità stocastica) affermano solo che ogni t è un punto di continuità in m_q e \mathbf{P} -qo. Non viene invece affatto garantita la continuità *globale* delle traiettorie, e infatti, come vedremo più oltre, il processo di Poisson non soddisfa i criteri di questo secondo – e più forte –

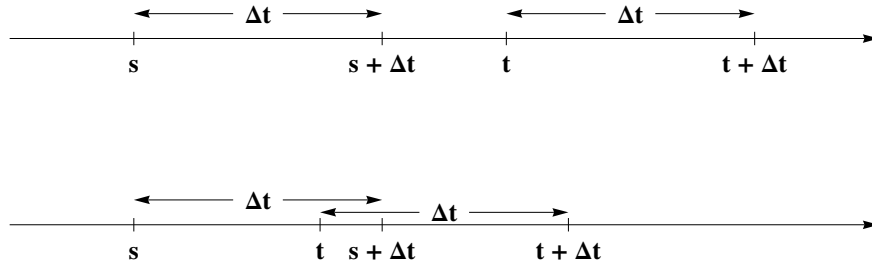


Figura 6.7: Possibili disposizioni degli intervalli per il calcolo dell'autocorrelazione (6.20) degli incrementi di Poisson con $\Delta t > 0$

tipo di continuità. Allo stesso modo, intuitivamente la derivabilità \mathbf{P} -qo di $N(t)$ dipende dal fatto che le traiettorie del processo $N(t)$ sono tutte costanti a tratti tranne che nell'insieme di punti (di misura di Lebesgue nulla) in cui si presentano le discontinuità. Quanto all'apparente incongruità fra la non derivabilità in mq e la derivabilità \mathbf{P} -qo osserveremo solo che essa è una tipica manifestazione del diverso significato delle due convergenze: questo può essere anche sottolineato ricordando ad esempio che – in analogia con (6.18) – l'esistenza della derivata $\dot{N}(t) = 0$ in mq richiederebbe che la quantità

$$\mathbf{E} \left[\left| \frac{\Delta N(t)}{\Delta t} \right|^2 \right] = \frac{\mathbf{E} [(\Delta N(t))^2]}{\Delta t^2} = \frac{\lambda \Delta t + \lambda^2 \Delta t^2}{\Delta t^2} = \frac{\lambda}{\Delta t} + \lambda^2$$

sia infinitesima per $\Delta t \rightarrow 0$, il che evidentemente non avviene. Questa discussione sulla derivabilità sarà ripresa nella Sezione 6.3 sul rumore bianco

Proposizione 6.8. *Il processo degli incrementi di Poisson $\Delta N(t)$ con $\Delta t > 0$ fissato è stazionario in senso lato, infatti si ha*

$$m_{\Delta N} = \sigma_{\Delta N}^2 = \lambda \Delta t \quad (6.19)$$

$$R_{\Delta N}(\tau) = \begin{cases} \lambda^2 \Delta t^2 & \text{se } |\tau| \geq \Delta t \\ \lambda^2 \Delta t^2 + \lambda(\Delta t - |\tau|) & \text{se } |\tau| < \Delta t \end{cases} \quad (6.20)$$

$$C_{\Delta N}(\tau) = \begin{cases} 0 & \text{se } |\tau| \geq \Delta t \\ \lambda(\Delta t - |\tau|) & \text{se } |\tau| < \Delta t \end{cases} \quad (6.21)$$

$$\rho_{\Delta N}(\tau) = \begin{cases} 0 & \text{se } |\tau| \geq \Delta t \\ 1 - \frac{|\tau|}{\Delta t} & \text{se } |\tau| < \Delta t \end{cases} \quad (6.22)$$

$$S_{\Delta N}(\varpi) = 2\lambda(\Delta t)^2 \frac{1 - \cos \varpi \Delta t}{(\varpi \Delta t)^2} \quad (6.23)$$

dove $S_{\Delta N}$ è lo spettro di covarianza (5.26).

Dimostrazione: Avevamo già osservato che $N(t)$ ha *incrementi indipendenti e stazionari*: qui si mostra che il p -s $\Delta N(t)$ è anche *stazionario in senso lato*. Il risultato (6.19) dipende dal fatto che gli incrementi $\Delta N(t)$ sono distribuiti secondo la legge $\mathfrak{P}(\lambda\Delta t)$. Per calcolare la funzione di autocorrelazione

$$R_{\Delta N}(s, t) = \mathbf{E} [\Delta N(s)\Delta N(t)]$$

bisogna poi ricordare che gli incrementi su intervalli di tempo non sovrapposti (vedi Figura 6.7) sono indipendenti, e che quindi se $|t - s| \geq \Delta t$

$$R_{\Delta N}(s, t) = \mathbf{E} [\Delta N(s)]\mathbf{E} [\Delta N(t)] = \lambda^2\Delta t^2 \quad |t - s| \geq \Delta t$$

Se invece $|t - s| < \Delta t$, supponiamo dapprima $t > s$ e osserviamo che (vedi Figura 6.7)

$$\begin{aligned} \Delta N(s)\Delta N(t) &= [N(s + \Delta t) - N(s)][N(t + \Delta t) - N(t)] \\ &= [N(s + \Delta t) - N(t) + N(t) - N(s)][N(t + \Delta t) - N(t)] \\ &= [N(t) - N(s)][N(t + \Delta t) - N(t)] + [N(s + \Delta t) - N(t)]^2 \\ &\quad + [N(s + \Delta t) - N(t)][N(t + \Delta t) - N(s + \Delta t)] \end{aligned}$$

per cui, data la disposizione degli intervalli, si avrà

$$\begin{aligned} R_{\Delta N}(s, t) &= \lambda(t - s) \cdot \lambda\Delta t + \lambda(\Delta t - t + s) + \lambda^2(\Delta t - t + s)^2 \\ &\quad + \lambda(\Delta t - t + s) \cdot \lambda(t - s) \\ &= \lambda^2\Delta t^2 + \lambda[\Delta t - (t - s)] \end{aligned}$$

Se invece $t < s$ basterà eseguire uno scambio di s con t : mettendo insieme tutti i casi si ottiene così la (6.20) con $\tau = t - s$. Da questi risultati discendono immediatamente anche l'autocovarianza (6.21) e il coefficiente di correlazione (6.22), mentre lo spettro di covarianza (6.23) si calcola con un'elementare trasformata di Fourier. \square

6.1.3 Processo di Poisson compensato

Il processo di Poisson e il corrispondente processo di punto sono anche il primo passo per definire altri importanti processi: si chiama processo di Poisson *compensato* il processo

$$\tilde{N}(t) = N(t) - \lambda t \tag{6.24}$$

Siccome λt è l'attesa di $N(t)$, si vede subito che il processo $\tilde{N}(t)$ non è altro che il processo di Poisson *centrato* con traiettorie come quelle della Figura 6.8. Tenendo conto di (6.13) si ha dunque

$$\begin{aligned} m_{\tilde{N}}(t) &= 0 & \sigma_{\tilde{N}}^2(t) &= \lambda t \\ R_{\tilde{N}}(s, t) &= C_{\tilde{N}}(s, t) = \lambda \min\{s, t\} & \rho_{\tilde{N}}(s, t) &= \frac{\min\{s, t\}}{\sqrt{st}} \end{aligned}$$

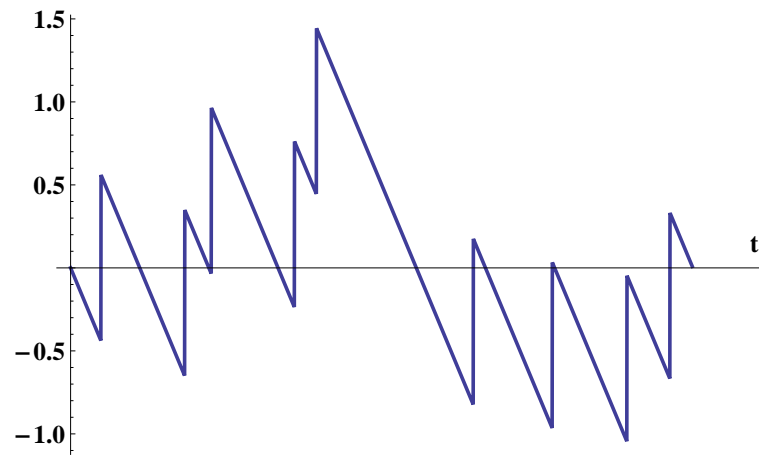


Figura 6.8: Campione di traiettoria del processo di Poisson compensato (6.24).

per cui la varianza ancora una volta aumenterà linearmente nel tempo e le traiettorie si disporranno equamente attorno all'asse orizzontale allontanandosi gradatamente come in Figura 6.9. Il processo $\tilde{N}(t)$ avrà quindi la tendenza a diffondere attorno al valore 0. Si noti però che ora rispetto alla Figura 6.6 la *diffusione* è divenuta più visibile: l'effetto di scala infatti è stato eliminato con la centratura. Inoltre, siccome l'autocorrelazione di $\tilde{N}(t)$ coincide essenzialmente con quella di $N(t)$, i risultati della Proposizione 6.7 valgono anche per il processo di Poisson compensato. Infine da (6.10) si ricava per la *fc* che

$$\varphi_{\tilde{N}}(u, t) = \varphi_N(u, t)e^{-iu\lambda t} = e^{\lambda t(e^{iu} - iu - 1)}$$

Il processo di Poisson compensato gioca un ruolo particolarmente importante nella teoria delle equazioni differenziali stocastiche per processi a salti, un aspetto che non potremo affrontare in queste lezioni.

6.1.4 Processo di Poisson composto

Definizione 6.9. Dato un processo di punto T_n , generalizzando l'espressione (6.6) chiameremo **processo di Poisson composto** il processo

$$X(t) = \sum_{k=1}^{\infty} X_k \vartheta(t - T_k) \quad (6.25)$$

dove le X_k sono una successione di v -a indipendenti dalle T_k . Se poi $N(t)$ è il processo di Poisson standard associato alle T_n , il processo (6.25) potrà essere rappresentato anche come la somma del numero aleatorio $N(t)$ di v -a X_k arrivate entro il tempo t

$$X(t) = \sum_{k=1}^{N(t)} X_k \quad (6.26)$$

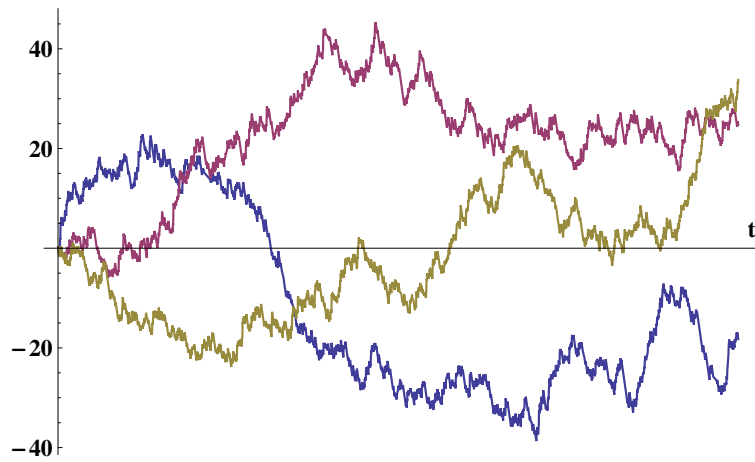


Figura 6.9: Traiettorie del processo di Poisson compensato (6.8) di 1000 passi

In pratica il processo di Poisson composto ha delle **traiettorie** analoghe a quelle del processo di Poisson standard, ma in corrispondenza di ogni istante T_k , invece di compiere un salto di lunghezza deterministicamente uguale a 1, esso compie ora un salto di lunghezza aleatoria X_k . Nella Figura 6.10 se ne può vedere un esempio per il quale abbiamo scelto delle X_k tutte distribuite come normali standard indipendenti. Naturalmente il processo di Poisson standard è a sua volta un caso particolare di Processo di Poisson composto: quello che si ottiene quando $X_k = 1$, **P**-qo. Nella Figura 6.11 sono rappresentati alcuni esempi di traiettorie per tempi lunghi che, ad una osservazione puramente qualitativa, non appaiono molto diverse da quelle del processo di Poisson compensato. Come vedremo nella successiva proposizione, la seconda rappresentazione (6.26) del processo $X(t)$ risulta particolarmente utile nel calcolo delle sue caratteristiche statistiche principali

Proposizione 6.10. *Dato un processo di Poisson composto $X(t)$ (6.25), se le X_k sono v-a iid con $\mathbf{E}[X_k] = \mu$, e $\mathbf{V}[X_k] = \sigma^2$, allora risulta*

$$m_X(t) = \lambda\mu t \tag{6.27}$$

$$\sigma_X^2(t) = \lambda(\mu^2 + \sigma^2)t \tag{6.28}$$

$$R_X(s, t) = \lambda(\mu^2 + \sigma^2) \min\{s, t\} + \lambda^2\mu^2 st \tag{6.29}$$

$$C_X(s, t) = \lambda(\mu^2 + \sigma^2) \min\{s, t\} \tag{6.30}$$

$$\rho_X(s, t) = \frac{\min\{s, t\}}{\sqrt{st}} \tag{6.31}$$

Dimostrazione: Si ottiene innanzitutto la (6.27) utilizzando la rappresentazio-

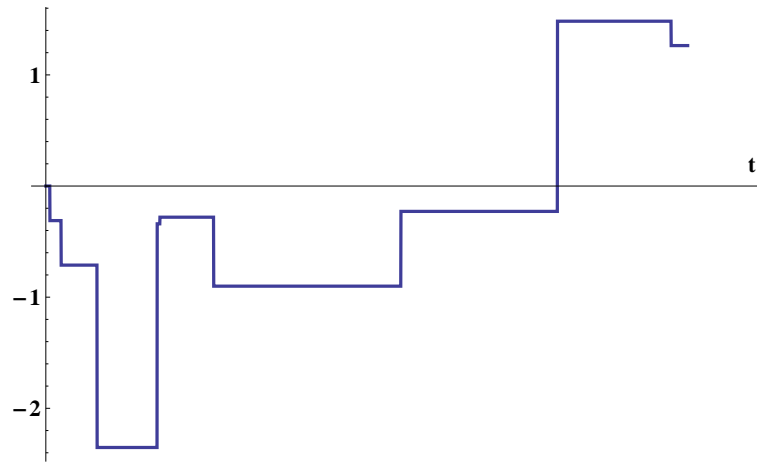


Figura 6.10: Campione di traiettoria di un processo di Poisson composto (6.25) con componenti X_k iid e $\mathfrak{N}(0, 1)$

ne (6.26) del processo $X(t)$ e le usuali proprietà delle attese condizionate:

$$\begin{aligned}
 m_X(t) &= \mathbf{E} \left[\sum_{k=1}^{N(t)} X_k \right] = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \mathbf{E} \left[\sum_{k=1}^{N(t)} X_k \mid N(t) = n \right] \\
 &= \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \mathbf{E} \left[\sum_{k=1}^n X_k \right] = \sum_{n=1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{(n-1)!} \mu \\
 &= \lambda \mu t \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} = \lambda \mu t
 \end{aligned}$$

Tenendo inoltre conto della relazione generale $\mathbf{E}[X_k X_\ell] = \mu^2 + \sigma^2 \delta_{k\ell}$ facilmente verificabile sulla base delle ipotesi, è poi possibile calcolare prima

$$\begin{aligned}
 R_X(t, t) &= \mathbf{E}[X(t)^2] = \mathbf{E} \left[\sum_{k,\ell=1}^{N(t)} X_k X_\ell \right] = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \mathbf{E} \left[\sum_{k,\ell=1}^n X_k X_\ell \right] \\
 &= \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \sum_{k,\ell=1}^n (\mu^2 + \sigma^2 \delta_{k\ell}) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} (n^2 \mu^2 + n \sigma^2) \\
 &= \mu^2 \sum_{n=1}^{\infty} n e^{-\lambda t} \frac{(\lambda t)^n}{(n-1)!} + \sigma^2 \sum_{n=1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{(n-1)!} \\
 &= \mu^2 \lambda t \sum_{n=0}^{\infty} (n+1) e^{-\lambda t} \frac{(\lambda t)^n}{n!} + \sigma^2 \lambda t \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\
 &= (\lambda t)^2 \mu^2 + \lambda t (\mu^2 + \sigma^2)
 \end{aligned}$$

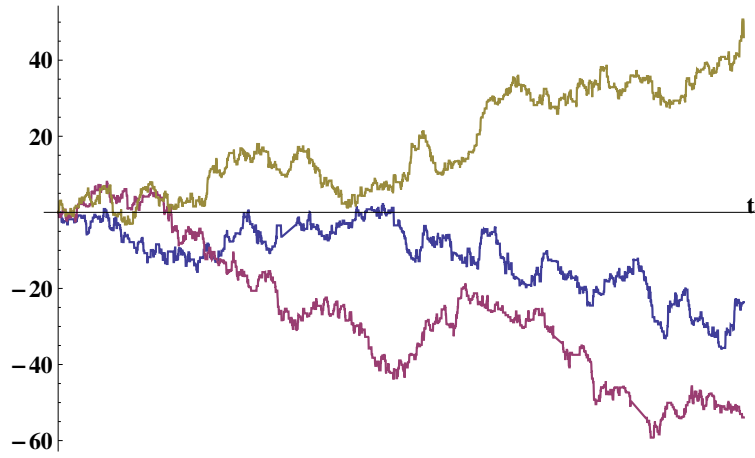


Figura 6.11: Esempi di traiettorie di 1000 passi di un processo di Poisson composto (6.25) con componenti X_k *iid* e $\mathfrak{N}(0, 1)$

e poi l'autocorrelazione (6.29) tenendo conto dell'indipendenza degli incrementi: scegliendo infatti, per fissare le idee, $s < t$ si ha

$$\begin{aligned} R_X(s, t) &= \mathbf{E}[X(s)X(t)] = \mathbf{E}[X(s)(X(t) - X(s) + X(s))] \\ &= \mathbf{E}[X(s)] \mathbf{E}[X(t) - X(s)] + R(s, s) \\ &= \lambda^2 \mu^2 s(t - s) + (\lambda s)^2 \mu^2 + \lambda s(\mu^2 + \sigma^2) = \lambda^2 \mu^2 st + \lambda s(\mu^2 + \sigma^2) \end{aligned}$$

e quindi in generale la (6.29) con s, t arbitrari. L'autocovarianza (6.30), la varianza (6.28) e il coefficiente di correlazione (6.31) si ottengono poi banalmente dalle definizioni (5.3) e (5.5). \square

La varianza avrà quindi ancora un carattere diffusivo aumentando linearmente con il tempo, come si vede anche dalla Figura 6.11. Si noti che il coefficiente di correlazione (6.29) coincide con quello dei processi di Poisson standard e compensato (6.15), mentre autocorrelazione e autocovarianza si riducono alle corrispondenti formule (6.13) per $\mu = 1, \sigma = 0$. Pertanto anche le proprietà di stazionarietà, continuità e derivabilità del processo di Poisson composto coincidono con quelle del processo di Poisson standard riassunte nella Proposizione 6.7. Infine si noti che, se $\varphi(u)$ è la *fc* comune a tutte le X_k , la *fc* del processo di Poisson composto è

$$\begin{aligned} \varphi_X(u, t) &= \mathbf{E}[e^{iuX(t)}] = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \mathbf{E}[e^{iu \sum_{k=1}^n X_k}] \\ &= \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \varphi(u)^n = e^{\lambda t[\varphi(u)-1]} \end{aligned} \tag{6.32}$$

Anche questa *fc* si riduce a quella del processo di Poisson standard (6.10) se $X_k = 1$, \mathbf{P} -qo per ogni k dato che in questo caso $\varphi(u) = e^{iu}$

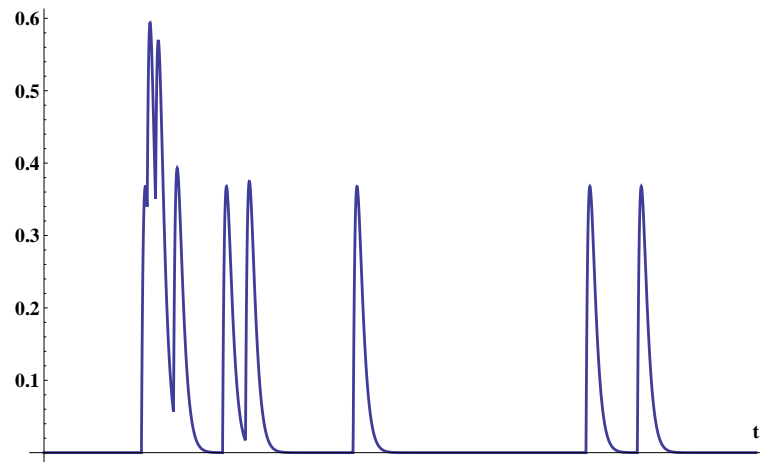


Figura 6.12: Campione di traiettoria del processo di rumore granulare (6.33) con funzione $h(t)$ della forma (6.34).

6.1.5 Rumore granulare

Definizione 6.11. Dato un processo di punto di intensità λ , chiameremo **rumore granulare (shot noise)** il p-s

$$X(t) = \sum_{k=1}^{\infty} h(t - T_k) \quad (6.33)$$

dove $h(t)$ è un'arbitraria funzione che in genere (ma non necessariamente) è diversa da 0 solo per $t > 0$

Esempio tipico di funzione $h(t)$ è

$$h(t) = q a t e^{-at} \vartheta(t) \quad a > 0, q > 0 \quad (6.34)$$

che produce campioni di traiettorie come quelle della Figura 6.12. Si può immaginare per fissare le idee che questo processo descriva le correnti prodotte dall'arrivo aleatorio di elettroni isolati al catodo di una valvola per effetto termoelettrico: in questo caso i tempi di arrivo costituiscono ovviamente un processo di punto, e ogni elettrone produce nel circuito un impulso di corrente $h(t)$ che tipicamente si spegne con andamento esponenziale. Inoltre arrivi ravvicinati generano sovrapposizioni di impulsi che producono effetti come quelli della Figura 6.12.

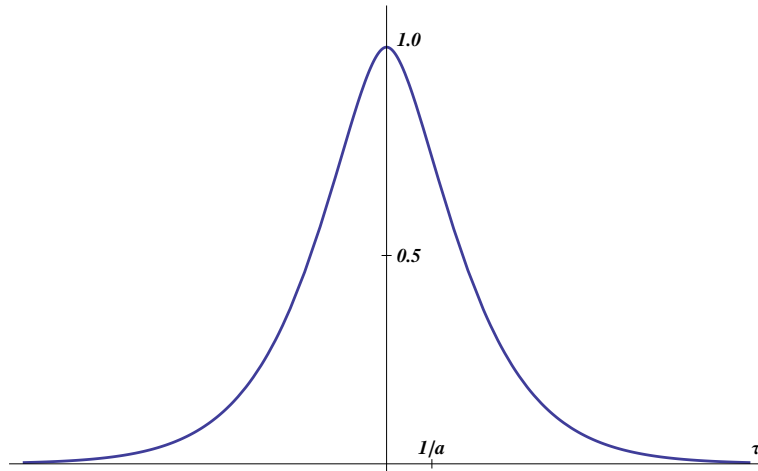


Figura 6.13: Coefficiente di correlazione $\rho_X(\tau)$ (6.38) di un rumore granulare con funzione $h(t)$ della forma (6.34).

Proposizione 6.12. *Un rumore granulare $X(t)$ (6.33) è stazionario in senso lato, e posto $\tau = t - s$ risulta*

$$m_X(t) = \lambda H \qquad \sigma_X^2(t) = \lambda g(0) \qquad (6.35)$$

$$R_X(\tau) = \lambda g(|\tau|) + \lambda^2 H^2 \qquad C_X(\tau) = \lambda g(|\tau|) \qquad \rho_X(\tau) = \frac{g(|\tau|)}{g(0)} \qquad (6.36)$$

$$H = \int_{-\infty}^{+\infty} h(t) dt \qquad g(t) = \int_{-\infty}^{+\infty} h(t+s)h(s) ds \qquad (6.37)$$

Dimostrazione: Omessa¹. Osserveremo soltanto che la stazionarietà è coerente con la natura delle traiettorie che, consistendo in treni di impulsi di forma h e intensità λ , non hanno più la tendenza a diffondere allontanandosi dall'asse orizzontale \square

Esempio 6.13. *I risultati della Proposizione (6.12) sono del tutto generali, ma le proprietà analitiche di un rumore granulare $X(t)$ dipenderanno esplicitamente dalla scelta della funzione h : ad esempio con una $h(t)$ della forma (6.34) si hanno i seguenti risultati (vedi anche Figura 6.13)*

$$\begin{aligned} H &= \frac{q}{a} & g(t) &= \frac{q^2}{4a}(1 + a|t|)e^{-a|t|} \\ m_X(t) &= \frac{\lambda q}{a} & \sigma_X^2(t) &= \frac{\lambda q^2}{4a} & R_X(\tau) &= \frac{\lambda q^2}{4a}(1 + a|\tau|)e^{-a|\tau|} + \frac{\lambda^2 q^2}{a^2} \\ C_X(\tau) &= \frac{\lambda q^2}{4a}(1 + a|\tau|)e^{-a|\tau|} & \rho_X(\tau) &= (1 + a|\tau|)e^{-a|\tau|} \end{aligned} \qquad (6.38)$$

¹A. Papoulis, **PROBABILITY, RANDOM VARIABLES AND STOCHASTIC PROCESSES**, McGraw Hill (Boston, 2002)

e quindi, tenendo conto dei risultati presentati nella Sezione 5, potremo dire che

- $X(t)$ è continuo in mq perchè la sua autocorrelazione $R_X(\tau)$ è una funzione continua in $\tau = 0$ (Proposizione 5.6)
- $X(t)$ è derivabile in mq perchè si potrebbe mostrare esplicitamente che la derivata seconda $R_X''(\tau)$ esiste in $\tau = 0$ (Proposizione 5.10)
- $X(t)$ è ergodico per l'attesa e per l'autocorrelazione perchè la condizione (5.21) è ovviamente soddisfatta dalla nostra $C_X(\tau)$ (Corollario 5.12)
- Lo spettro di covarianza del rumore granulare del nostro esempio si calcola infine dalla (6.38) e dalla (5.26) e si ha

$$S_X(\varpi) = \frac{\lambda a^2 q^2}{2\pi (a^2 + \varpi^2)^2} \quad (6.39)$$

6.2 Processo di Wiener

Il processo di Wiener (detto anche *moto Browniano*, nome che però noi riserveremo per il fenomeno fisico che sarà discusso nella Sezione 6.4 e nel Capitolo 9) può essere introdotto, come faremo più oltre, definendo in maniera formale le sue proprietà. In questa introduzione, invece, seguiremo una strada più intuitiva che passa attraverso la costruzione esplicita delle traiettorie, analogamente a quanto fatto per il processo di Poisson. A differenza di quest'ultimo, però, il processo di Wiener risulterà dal limite di una *successione di processi* elementari noti come *random walk*, e per questo motivo le sue traiettorie potranno solo essere *approssimate* da quelle di un *random walk* con un gran numero di passi.

6.2.1 Random walk

Definizione 6.14. Dati $s > 0$, $\tau > 0$, e la successione $(X_j)_{j \geq 0}$ di v -a iid

$$X_0 = 0, \mathbf{P}\text{-}qo \quad X_j = \begin{cases} +s, & \text{con probabilità } p \\ -s, & \text{con probabilità } q = 1 - p \end{cases} \quad j = 1, 2, \dots$$

chiameremo **random walk** il p -s

$$X(t) = \sum_{j=0}^{\infty} X_j \vartheta(t - j\tau) \quad (6.40)$$

per il quale ovviamente risulta

$$X(t) = \begin{cases} X_0 = 0 & 0 \leq t < \tau \\ X_1 + \dots + X_n & n\tau \leq t < (n+1)\tau, \quad n = 1, 2, \dots \end{cases}$$

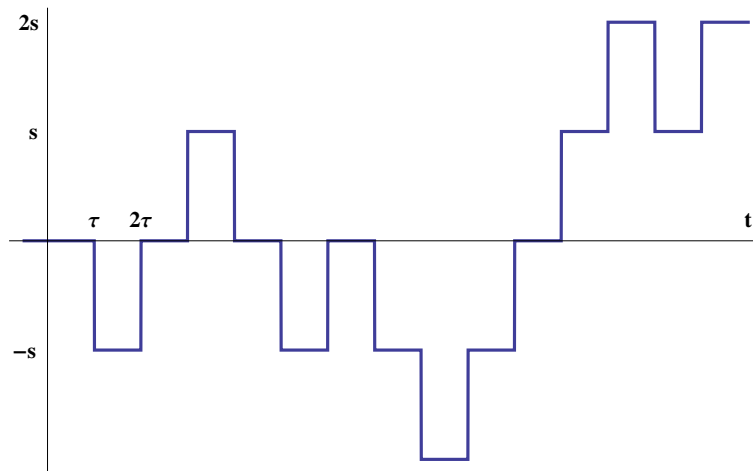


Figura 6.14: Tipica traiettoria di un *random walk* simmetrico con 15 passi.

Le possibili traiettorie di un *random walk* sono pertanto delle gradinate ascendenti e discendenti come quelle mostrate in Figura 6.14 che a prima vista somigliano a quelle di un processo di Poisson composto (6.25). Diversamente da queste, invece, la lunghezza dei gradini ora non è più aleatoria ma costante e uguale a τ , mentre la loro altezza prende solo i due possibili valori $\pm s$. Siccome inoltre risulta

$$\mathbf{E}[X_j] = (p - q)s \quad \mathbf{E}[X_j^2] = s^2 \quad \mathbf{V}[X_j] = 4pqs^2$$

è anche facile rendersi conto del fatto che, con $n = 0, 1, 2, \dots$ si ha

$$\mathbf{E}[X(t)] = (p - q)ns \quad \mathbf{V}[X(t)] = 4pqns^2 \quad n\tau \leq t < (n + 1)\tau \quad (6.41)$$

per cui, se ad esempio $p \neq q$, il valore assoluto dell'attesa crescerà con n , cioè con t . Quando invece $p = q = \frac{1}{2}$ si parla di **random walk simmetrico**, e in questo caso $\mathbf{E}[X(t)] = 0$ per ogni t . La varianza invece cresce con n , e quindi con t , in ogni caso. Naturalmente, per costruzione, anche in questo caso gli incrementi $\Delta X(t)$ saranno indipendenti se i corrispondenti intervalli sono disgiunti

6.2.2 Processo di Wiener

Definizione 6.15. Considerata al variare di $s > 0$ e $\tau > 0$ la famiglia di tutti i *random walk* simmetrici $X(t)$ con $p = q = \frac{1}{2}$, chiameremo **processo di Wiener standard** $W(t)$, con $W(0) = 0$, \mathbf{P} -go, il p -s limite in distribuzione (nel senso della Definizione 5.4) dei *random walk* $X(t)$ quando

$$\tau \rightarrow 0 \quad s \rightarrow 0 \quad \frac{s^2}{\tau} \rightarrow D > 0 \quad (6.42)$$

Inoltre, assegnato un $\Delta t > 0$, si definisce anche il **processo degli incrementi di Wiener** $\Delta W(t) = W(t + \Delta t) - W(t)$

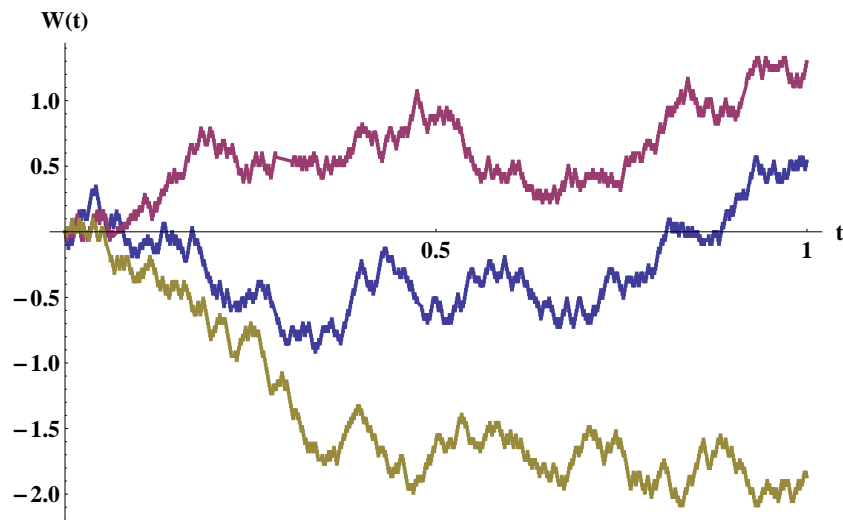


Figura 6.15: Tipiche traiettorie con 1000 passi di un *random walk* simmetrico che può essere considerato come un'approssimazione di un processo di Wiener $W(t)$ con $D = 1$.

Noi supporremo per semplicità che il limite su cui si basa la precedente definizione esista, cioè che – nel limite (6.42) – tutte le distribuzioni finito-dimensionali di $X(t)$ convergano verso delle distribuzioni finito-dimensionali consistenti che definiscono la legge di $W(t)$. Definito dunque $W(t)$ secondo la procedura euristica qui delineata, osserviamo innanzitutto che le sue traiettorie (diversamente da quelle del processo di Poisson o di un *random walk*) non potranno più essere riprodotte in maniera esatta perché si tratta di un *processo limite*. Possiamo però assumere intuitivamente – sulla base della discussione fin qui svolta – che le traiettorie di un *random walk* con un gran numero di passi appaiano come una buona approssimazione delle traiettorie di $W(t)$. Tali approssimazioni assumono un aspetto simile a quello proposto nella Figura 6.15 dove i campioni sono quelli di un *random walk* di 1000 passi. Si noti come su tempi lunghi queste curve somiglino qualitativamente a quelle di un processo di Poisson compensato della Figura 6.9, o composto della Figura 6.11. La differenza essenziale fra le vere traiettorie di $W(t)$ e le sue approssimazioni sta nel fatto che se osservassimo più da vicino la traiettoria approssimate di Figura 6.15 ritroveremmo quella del *random walk* di Figura 6.14, mentre se ingrandissimo a qualunque livello una vera traiettoria di $W(t)$ ritroveremmo sempre lo stesso tipo di traiettorie irregolari. In altri termini i campioni di un processo di Wiener appaiono sempre dello stesso tipo quale che sia la scala spazio-temporale alla quale le si osserva. Nella discussione seguente converrà per brevità indicare con la notazione

$$\phi_{a^2}(x) = \frac{e^{-x^2/2a^2}}{\sqrt{2\pi a^2}} \quad \Phi_{a^2}(x) = \int_{-\infty}^x \phi_{a^2}(y) dy$$

rispettivamente la *ddp* e la *fdd* di una legge $\mathfrak{N}(0, a^2)$

Proposizione 6.16. *Un processo di Wiener standard $W(t)$ ha incrementi indipendenti e stazionari; inoltre risulta*

$$W(t) \sim \mathfrak{N}(0, Dt) \quad \Delta W(t) \sim \mathfrak{N}(0, D\Delta t) \quad (6.43)$$

e quindi le rispettive *ddp* e *fc* sono

$$f_W(x, t) = \phi_{Dt}(x) = \frac{e^{-x^2/2Dt}}{\sqrt{2\pi Dt}} \quad f_{\Delta W}(x) = \phi_{D\Delta t}(x) = \frac{e^{-x^2/2D\Delta t}}{\sqrt{2\pi D\Delta t}} \quad (6.44)$$

$$\varphi_W(u, t) = e^{-Dtu^2/2} \quad \varphi_{\Delta W}(u) = e^{-D\Delta t u^2/2} \quad (6.45)$$

Infine le **ddp di transizione** (*ddp condizionate*) con $\Delta t > 0$ sono $\mathfrak{N}(y, D\Delta t)$ cioè

$$f_W(x, t + \Delta t | y, t) = \phi_{D\Delta t}(x - y) = f_{\Delta W}(x - y) = \frac{e^{-(x-y)^2/2D\Delta t}}{\sqrt{2\pi D\Delta t}} \quad (6.46)$$

Dimostrazione: Siccome il processo $W(t)$ è stato definito come limite di processi di *random walk* $X(t)$ che hanno per definizione incrementi indipendenti, è molto ragionevole affermare che anche tutti gli incrementi $\Delta W(t)$ corrispondenti a intervalli non sovrapposti siano indipendenti: il processo di Wiener costituisce quindi il secondo esempio di *processo a incrementi indipendenti* che incontriamo. Il risultato (6.43) discende invece dal Teorema Limite Centrale 4.27: infatti, scelto un t arbitrario ma fissato, si consideri la successione di *random walk* simmetrici $X(t)$ con

$$\tau = \frac{t}{n} \quad s^2 = \frac{Dt}{n} \quad n = 1, 2, \dots$$

in modo che le (6.42) siano soddisfatte per $n \rightarrow \infty$. In tal caso

$$X(t) = X(n\tau) = X_0 + X_1 + \dots + X_n = S_n \quad n = 0, 1, 2, \dots$$

con $\mathbf{E}[S_n] = 0$ e $\mathbf{V}[S_n] = Dt$, come si ricava da (6.41) per $p = q = \frac{1}{2}$. Dal Teorema Limite Centrale 4.27 abbiamo allora che per $n \rightarrow \infty$

$$S_n^* = \frac{X(t)}{\sqrt{Dt}} \xrightarrow{d} \mathfrak{N}(0, 1)$$

Se allora $W(t)$ è il limite in distribuzione di $X(t)$ potremo dire che

$$\frac{W(t)}{\sqrt{Dt}} \sim \mathfrak{N}(0, 1)$$

ovvero $W(t) \sim \mathfrak{N}(0, Dt)$ cioè (6.43), e la *ddp* del processo di Wiener sarà quindi quella riportata in (6.44). La *fc* (6.45) si calcola poi facilmente da (4.13) tenendo conto di (6.43). Anche la legge (6.43) degli incrementi $\Delta W(t)$ e le sue *ddp* (6.44)

e *fc* (6.45) si ricavano sulla base di analoghe considerazioni. Infine per la *ddp* di transizione, ricordando (6.43) e l'indipendenza degli incrementi, si ha per $\Delta t > 0$

$$\begin{aligned} F_W(x, t + \Delta t | y, t) &= \mathbf{P}\{W(t + \Delta t) \leq x | W(t) = y\} \\ &= \mathbf{P}\{W(t + \Delta t) - W(t) + W(t) \leq x | W(t) = y\} \\ &= \mathbf{P}\{\Delta W(t) \leq x - y | W(t) = y\} \\ &= \mathbf{P}\{\Delta W(t) \leq x - y\} = \Phi_{D\Delta t}(x - y) \end{aligned}$$

e quindi (6.46) si ottiene per derivazione rispetto a x □

Proposizione 6.17. *Le caratteristiche statistiche del processo di Wiener standard $W(t)$ sono*

$$m_W(t) = 0 \quad \sigma_W^2(t) = Dt \quad (6.47)$$

$$R_W(s, t) = C_W(s, t) = D \min\{s, t\} \quad (6.48)$$

$$\rho_W(s, t) = \frac{\min\{s, t\}}{\sqrt{st}} = \begin{cases} \sqrt{s/t} & \text{se } s < t \\ \sqrt{t/s} & \text{se } t < s \end{cases} \quad (6.49)$$

Dimostrazione: Le (6.47) discendono direttamente da (6.43). Per calcolare l'autocorrelazione (6.48) (che coincide con l'autocovarianza dato che l'attesa è nulla) consideriamo dapprima il caso $s = t$ per il quale da (6.43) abbiamo

$$R_W(t, t) = \mathbf{E} [W^2(t)] = \mathbf{V} [W(t)] = Dt \quad (6.50)$$

e poi, con $s < t$, osserviamo che gli incrementi $W(s) = W(s) - W(0)$ e $W(t) - W(s)$ sono *v-a* indipendenti con leggi $\mathfrak{N}(0, Ds)$ e $\mathfrak{N}(0, D(t - s))$ rispettivamente, sicché

$$\begin{aligned} R_W(s, t) &= \mathbf{E} [W(s)W(t)] = \mathbf{E} [W(s)(W(t) - W(s) + W(s))] \\ &= \mathbf{E} [W^2(s)] = R_W(s, s) = Ds \end{aligned}$$

In definitiva, abbandonando anche la limitazione $s < t$, si ricava facilmente (6.48) e conseguentemente anche (6.49). □

Nonostante le evidenti differenze fra i suoi processi, i risultati di questa proposizione sul processo di Wiener sono molto simili a quelli della corrispondente Proposizione 6.6 per il processo di Poisson standard, con il coefficiente D che gioca un ruolo analogo a quello dell'intensità λ . Il coefficiente $D > 0$ (la cui esistenza è stata ipotizzata all'interno del processo di limite che definisce $W(t)$, e che per costruzione ha dimensioni m^2/sec) prende il nome di **coefficiente di diffusione** ed è un parametro caratteristico del processo. Si noti, infine, che anche in questo caso, come per il processo di Poisson, la varianza aumenta linearmente con il tempo t , sicché anche il processo di Wiener può essere considerato (in questo senso) una *diffusione*.

Proposizione 6.18. *Per ogni $t > 0$ un processo di Wiener standard $W(t)$ è continuo in m_q e \mathbf{P} -qo, ma non è derivabile né in m_q né \mathbf{P} -qo. Inoltre $W(t)$ non è stazionario, ma è gaussiano e, comunque scelti $t_1 < t_2 < \dots < t_n$ (l'ordine è scelto per comodità), risulta $(W(t_1), \dots, W(t_n)) \sim \mathfrak{N}(0, \mathbb{A})$ con matrice delle covarianze*

$$\mathbb{A} = D \begin{pmatrix} t_1 & t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & t_2 & & t_2 \\ t_1 & t_2 & t_3 & & t_3 \\ \vdots & & & \ddots & \vdots \\ t_1 & t_2 & t_3 & \dots & t_n \end{pmatrix} \quad (6.51)$$

Dimostrazione: La continuità, la non derivabilità in m_q e la non stazionarietà di $W(t)$ si dimostrano con argomenti simili a quelli usati nella dimostrazione della Proposizione 6.7 per il processo di Poisson $N(t)$: infatti le funzioni di autocorrelazione dei due processi sono essenzialmente identiche. Quanto alla non derivabilità \mathbf{P} -qo questa può essere dedotta dalla *negazione del punto 1.* nel Teorema 4.4 dimostrando che $W(t)$ non è derivabile neanche in probabilità, visto che risulta

$$\left| \frac{\Delta W(t)}{\Delta t} \right| \xrightarrow{\mathbf{P}} +\infty \quad \text{per } \Delta t \rightarrow 0, \quad \forall t > 0$$

ovvero

$$\lim_{\Delta t \rightarrow 0} \mathbf{P} \left\{ \left| \frac{\Delta W(t)}{\Delta t} \right| > M \right\} = 1 \quad \forall M > 0, \quad \forall t > 0 \quad (6.52)$$

Infatti, comunque scelto $M > 0$, da (6.44) si ha

$$\mathbf{P} \left\{ \left| \frac{\Delta W(t)}{\Delta t} \right| > M \right\} = 1 - \mathbf{P} \left\{ \left| \frac{\Delta W(t)}{\Delta t} \right| \leq M \right\} = 1 - \int_{-M|\Delta t|}^{M|\Delta t|} \frac{e^{-x^2/2D|\Delta t|}}{\sqrt{2\pi D|\Delta t|}} dx$$

e quindi la (6.52) discende dall'osservazione che posto $y = x/\sqrt{|\Delta t|}$ risulta

$$\lim_{\Delta t \rightarrow 0} \int_{-M|\Delta t|}^{M|\Delta t|} \frac{e^{-x^2/2D|\Delta t|}}{\sqrt{2\pi D|\Delta t|}} dx = \lim_{\Delta t \rightarrow 0} \int_{-M\sqrt{|\Delta t|}}^{M\sqrt{|\Delta t|}} \frac{e^{-y^2/2D}}{\sqrt{2\pi D}} dy = 0$$

Per la continuità \mathbf{P} -qo in ogni $t > 0$ ci limiteremo invece ad osservare – anticipando i risultati della Proposizione 7.23 – che in realtà il processo di Wiener $W(t)$ non è solo continuo in m_q e \mathbf{P} -qo, ma lo è addirittura nel senso globale (*sample continuous*) della Definizione 5.5: quasi ogni traiettoria di $W(t)$ risulta continua in ogni t . La semplice continuità \mathbf{P} -qo in ogni $t > 0$ discende pertanto dal fatto (vedi la fine della Sezione 5.3) che la continuità globale implica anche tutti gli altri tipi di continuità. Infine, dalla Proposizione 6.16 già sappiamo che la *ddp* di $W(t)$ e la *ddp* di transizione sono Gaussiane. Questo ci permette di ottenere facilmente anche le *ddp* congiunte²

²Si noti che dalle *ddp* congiunte (6.53) si possono ritrovare le marginali (6.44)

$$f_W(x, t) = \int_{-\infty}^{+\infty} f_W(x, t; y, s) dy = [\phi_{D(t-s)} * \phi_{Ds}](x) = \phi_{Dt}(x)$$

dei vettori $(W(s), W(t))$: preso per comodità $s < t$, avremo la *ddp*

$$f_W(x, t; y, s) = f_W(x, t|y, s)f_W(y, s) = \phi_{D(t-s)}(x - y)\phi_{Ds}(y) \quad (6.53)$$

che – scritte esplicitamente e confrontate con la *ddp* (2.24) – risultano essere normali bivariate $\mathfrak{N}(0, \mathbb{A})$ con matrice delle covarianze

$$\mathbb{A} = D \begin{pmatrix} s & s \\ s & t \end{pmatrix}$$

Una volta determinate le leggi a due istanti, si passa a quelle con $n = 3, 4, \dots$ istanti iterando il procedimento: scelti ad esempio tre istanti $t_1 < t_2 < t_3$ (l'ordine è fissato per comodità), e sfruttando l'indipendenza degli incrementi, si calcola prima – con un procedimento analogo a quello che conduce alla (6.46) – la *ddp* condizionata

$$f_W(x_3, t_3 | x_2, t_2; x_1, t_1) = \phi_{D(t_3-t_2)}(x_3 - x_2)$$

e poi, tenendo conto di (6.53), la *ddp* congiunta

$$f_W(x_3, t_3; x_2, t_2; x_1, t_1) = f_W(x_3, t_3 | x_2, t_2; x_1, t_1)f_W(x_2, t_2; x_1, t_1)$$

che risulta Gaussiana con matrice delle covarianze della forma (6.51). Ripetendo la procedura, comunque presi $t_1 < \dots < t_n$, si mostra che le *ddp* congiunte dei vettori $(W(t_1), \dots, W(t_n))$ sono tutte $\mathfrak{N}(0, \mathbb{A})$ con matrici delle covarianze (6.51), e quindi che il processo di Wiener $W(t)$ è un *processo Gaussiano*. \square

Si verifica infine con un calcolo diretto che le *ddp* $f_W(x, t|y, s)$ sono soluzioni dell'equazione

$$\partial_t f(x, t) = \frac{D}{2} \partial_x^2 f(x, t), \quad f(x, s^+) = \delta(x - y) \quad (6.54)$$

che costituisce un primo esempio di *equazione di Fokker-Planck* (Sezione 7.2.3)

Proposizione 6.19. *Il processo degli incrementi di Wiener $\Delta W(t)$ con $\Delta t > 0$ è stazionario in senso lato, infatti risulta*

$$m_{\Delta W} = 0 \quad \sigma_{\Delta W}^2 = D\Delta t \quad (6.55)$$

$$R_{\Delta W}(\tau) = C_{\Delta W}(\tau) = \begin{cases} 0 & \text{se } |\tau| \geq \Delta t \\ D(\Delta t - |\tau|) & \text{se } |\tau| < \Delta t \end{cases} \quad (6.56)$$

$$\rho_{\Delta W}(\tau) = \begin{cases} 0 & \text{se } |\tau| \geq \Delta t \\ 1 - \frac{|\tau|}{\Delta t} & \text{se } |\tau| < \Delta t \end{cases} \quad (6.57)$$

$$S_{\Delta W}(\varpi) = 2D(\Delta t)^2 \frac{1 - \cos \varpi \Delta t}{(\varpi \Delta t)^2} \quad (6.58)$$

e che il calcolo si esegue facilmente utilizzando la proprietà riproduttiva (3.70) delle leggi normali

$$\mathfrak{N}(0, D(t-s)) * \mathfrak{N}(0, Ds) = \mathfrak{N}(0, Dt)$$

Dimostrazione: Avevamo già osservato che gli incrementi $\Delta W(t)$ sono indipendenti e stazionari: qui si mostra inoltre che il processo degli incrementi è stazionario in senso lato. I risultati (6.55) derivano direttamente da (6.43), cioè dal fatto che $\Delta W(t) \sim \mathfrak{N}(0, D\Delta t)$. Per calcolare invece le funzioni di autocorrelazione e autocovarianza si ripercorre la procedura utilizzata nel caso degli incrementi di Poisson: ricordando che gli incrementi su intervalli di tempo non sovrapposti sono indipendenti, se $|t - s| \geq \Delta t$ si ha

$$R_{\Delta W}(s, t) = \mathbf{E} [\Delta W(s)\Delta W(t)] = \mathbf{E} [\Delta W(s)] \mathbf{E} [\Delta W(t)] = 0$$

Se invece $|t - s| < \Delta t$, supponendo prima $t > s$ risulta (vedi Figura 6.7)

$$\begin{aligned} \Delta W(s)\Delta W(t) &= [W(s + \Delta t) - W(s)] [W(t + \Delta t) - W(t)] \\ &= [W(s + \Delta t) - W(t) + W(t) - W(s)] [W(t + \Delta t) - W(t)] \\ &= [W(t) - W(s)][W(t + \Delta t) - W(t)] + [W(s + \Delta t) - W(t)]^2 \\ &\quad + [W(s + \Delta t) - W(t)][W(t + \Delta t) - W(s + \Delta t)] \end{aligned}$$

per cui – per l’indipendenza degli incrementi coinvolti e per l’annullarsi delle loro attese – con $\tau = t - s > 0$ si avrà

$$R_{\Delta W}(\tau) = \mathbf{E} [[W(s + \Delta t) - W(t)]^2] = D(\Delta t - \tau)$$

Nel caso $t < s$ basta poi uno scambio di s con t : mettendo insieme tutti i casi si può scrivere quindi la (6.56). Con questi risultati discendono immediatamente dalle loro definizioni anche (6.57) e (6.58). \square

6.2.3 Processo di Wiener geometrico

Come nel caso del processo di Poisson, anche ora potremo ricavare dal processo di Wiener molti altri processi, ma noi ci soffermeremo brevemente solo sul cosiddetto **processo di Wiener (o moto Browniano) geometrico** definito come

$$X(t) = e^{W(t)} \quad X(0) = 1 \tag{6.59}$$

dove $W(t)$ è un processo di Wiener. Questo tipo di processi è particolarmente importante nel campo della finanza matematica e abbiamo già osservato nella Sezione 3.47 che la sua *ddp* è *log-normale*

$$f_X(x, t) = \frac{e^{-\ln^2 x / 2Dt}}{x\sqrt{2\pi Dt}} \quad x > 0 \tag{6.60}$$

mentre per l’attesa e la varianza si ha

$$m_X(t) = e^{Dt/2} \quad \sigma_X^2(t) = (e^{Dt} - 1) e^{Dt}$$

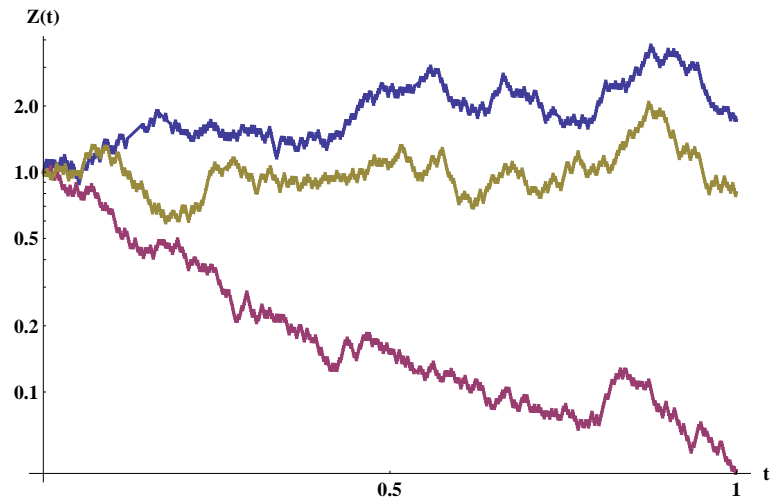


Figura 6.16: Traiettorie (scala logaritmica e $D = 1$) di un moto Browniano geometrico (6.61) approssimate con esponenziali di random walks di 1 000 passi

Ovviamente le traiettorie di $X(t)$ – diversamente da quelle del processo di Wiener $W(t)$ – non prenderanno mai valori negativi (Figura 6.16 in scala logaritmica) e questo è uno dei motivi principali per i quali il moto Browniano geometrico è considerato un buon modello per descrivere l'andamento dei prezzi sul mercato. Nelle applicazioni viene però piuttosto usata una variante di $X(t)$ centrata attorno al valore 1 definita da

$$Z(t) = \frac{X(t)}{m_X(t)} = e^{W(t) - Dt/2} \quad (6.61)$$

per la quale facilmente si verifica che

$$m_Z(t) = 1 \quad \sigma_Z^2(t) = e^{Dt} - 1$$

6.3 Rumore bianco

Da un punto di vista strettamente formale il rumore bianco (*white noise*) non esiste come processo stocastico, nello stesso senso in cui la *delta di Dirac* $\delta(x)$ non esiste come funzione. Trascurando per brevità le procedure rigorose con le quali si può dare un significato preciso a questo concetto, noi ci limiteremo qui ad alcune osservazioni euristiche per metterne in luce le opportunità e i rischi

Definizione 6.20. Chiameremo **rumore bianco** (*white noise*) un processo $B(t)$ per il quale l'autocovarianza sia

$$C_B(s, t) = q(t)\delta(t - s) \quad (6.62)$$

dove $q(s) > 0$ è detta **intensità** del rumore bianco. Il rumore bianco è **stazionario** quando l'intensità q è costante e in questo caso con $\tau = t - s$ avremo

$$C_B(\tau) = q\delta(\tau) \quad S_B(\varpi) = q \quad (6.63)$$

per cui lo spettro di covarianza risulta piatto e giustifica il nome adottato.

I rumori bianchi possono quindi essere considerati dei *processi singolari*, come peraltro rivelato dalla presenza della δ di Dirac nella loro definizione. La loro irregolarità è inoltre confermata dal fatto che la (6.62) implica in particolare che i valori del processo $B(t)$ in istanti differenti sono sempre non correlati, e quindi che in un certo senso essi assumono valori non prevedibili sulla base di osservazioni in altri istanti. In particolare noi esamineremo alcuni esempi e mostreremo come la loro singolarità sia spesso legata all'introduzione di processi ottenuti come derivate di processi non derivabili, nello stesso senso in cui la δ di Dirac può essere considerata come la derivata della funzione di Heaviside che notoriamente non è derivabile.

Esempio 6.21. Rumore bianco di Poisson: *Un primo esempio è il caso particolare di rumore granulare (6.33) che si ottiene prendendo $h(t) = \delta(t)$: con questa scelta il processo può essere immaginato come costituito da una successione di impulsi δ ai tempi aleatori T_k , e può essere formalmente visto come la derivata, traiettoria per traiettoria, del processo di Poisson nella forma (6.6)*

$$\dot{N}(t) = \sum_{k=1}^{\infty} \delta(t - T_k) \quad (6.64)$$

che prende anche il nome di **processo degli impulsi di Poisson**. Per mostrare che (6.64) è proprio un rumore bianco stazionario con intensità λ basterà poi osservare che da (6.35) con $h(t) = \delta(t)$ si ha $H = 1$ e $g(t) = \delta(t)$, e quindi

$$m_{\dot{N}} = \lambda \quad R_{\dot{N}}(\tau) = \lambda^2 + \lambda\delta(\tau) \quad C_{\dot{N}}(\tau) = \lambda\delta(\tau) \quad (6.65)$$

Un altro rumore bianco stazionario, ma con attesa nulla, si ricava poi derivando un processo di Poisson compensato (6.24)

$$\tilde{\dot{N}}(t) = \dot{N}(t) - \lambda \quad (6.66)$$

Si tratta quindi del processo degli impulsi centrato dato che da (6.65) è facile ricavare

$$m_{\tilde{\dot{N}}} = m_{\dot{N}} - \lambda = 0$$

Che anche in questo caso si tratti di un rumore bianco stazionario si deduce osservando che da (6.65) e (6.66) si ha

$$R_{\tilde{\dot{N}}}(\tau) = C_{\tilde{\dot{N}}}(\tau) = R_{\dot{N}}(\tau) - \lambda^2 = \lambda\delta(\tau)$$

È utile infine notare che tutti gli altri rumori granulari (6.33) con $h(t)$ diverse da $\delta(t)$ si possono ricavare per convoluzione dal processo degli impulsi e dalla funzione $h(t)$ secondo la relazione

$$X(t) = [\dot{N} * h](t) \quad (6.67)$$

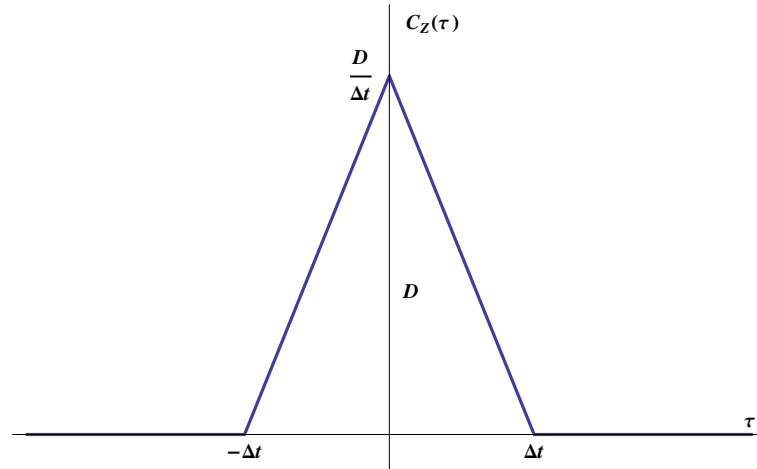


Figura 6.17: Autocovarianza $C_Z(\tau)$ (6.69) del rapporto incrementale (6.68) di un processo di Wiener: l'area del triangolo è sempre D per qualunque valore di Δt , ma la sua forma diventa sempre più stretta e alta per $\Delta t \rightarrow 0$.

Esempio 6.22. Rumore bianco di Wiener: *Un altro tipo di rumore bianco è infine associato al processo di Wiener $W(t)$: come abbiamo già osservato $W(t)$ non è derivabile (in nessun senso) e quindi possiamo immaginare che la sua derivata formale $\dot{W}(t)$ possa avere le proprietà singolari di un rumore bianco. Infatti se, con Δt fissato, consideriamo il processo dei rapporti incrementali*

$$Z(t) = \frac{\Delta W(t)}{\Delta t} \quad (6.68)$$

vediamo facilmente dai risultati della Proposizione 6.19 relativi al processo degli incrementi $\Delta W(t)$ su intervalli di ampiezza $|\Delta t|$ che

$$m_Z = 0 \quad R_Z(\tau) = C_Z(\tau) = \begin{cases} 0 & \text{se } |\tau| \geq |\Delta t| \\ \frac{D}{|\Delta t|} \left(1 - \frac{|\tau|}{|\Delta t|}\right) & \text{se } |\tau| < |\Delta t| \end{cases} \quad (6.69)$$

La funzione $R_Z(\tau) = C_Z(\tau)$ è mostrata in Figura 6.17 e da essa si vede che

$$C_Z(\tau) \rightarrow D \delta(\tau) \quad \Delta t \rightarrow 0$$

Pertanto, se ammettiamo in qualche senso che esista $\dot{W}(t)$ come il limite

$$Z(t) = \frac{\Delta W(t)}{\Delta t} \rightarrow \dot{W}(t) \quad \Delta t \rightarrow 0$$

ci attendiamo anche che risulti

$$m_{\dot{W}} = 0 \quad R_{\dot{W}}(\tau) = C_{\dot{W}}(\tau) = D \delta(\tau) \quad (6.70)$$

cioè che $\dot{W}(t)$ sia proprio un rumore bianco stazionario con intensità D .

Queste considerazioni ci permettono di anticipare intuitivamente alcune proprietà degli incrementi che saranno riprese in seguito in maniera più sistematica. Consideriamo i processi degli incrementi finora introdotti $\Delta N(t)$, $\Delta \tilde{N}(t)$ e $\Delta W(t)$ con Δt arbitrario ma fissato, e osserviamo che, essendo $m_{\Delta \tilde{N}}(t) = m_{\Delta W}(t) = 0$ da (6.19) e (6.55), su intervalli di ampiezza $|\Delta t|$ abbiamo

$$\mathbf{E} \left[\Delta \tilde{N}^2(t) \right] = \sigma_{\Delta \tilde{N}}^2 = \lambda |\Delta t| \quad \mathbf{E} \left[\Delta W^2(t) \right] = \sigma_{\Delta W}^2 = D |\Delta t| \quad (6.71)$$

$$\mathbf{E} \left[\Delta N^2(t) \right] = \sigma_{\Delta N}^2 + m_{\Delta N}^2 = \lambda |\Delta t| + \lambda^2 \Delta t^2 \quad (6.72)$$

e in definitiva – ma in un senso per ora solo simbolico – al limite per $\Delta t \rightarrow 0$

$$\mathbf{E} \left[d\tilde{N}^2(t) \right] = \mathbf{E} \left[dN^2(t) \right] = \lambda |dt| \quad \mathbf{E} \left[dW^2(t) \right] = D |dt| \quad (6.73)$$

Queste relazioni suggeriscono l'idea che in realtà gli incrementi infinitesimi dei nostri processi non siano dell'ordine di dt , ma siano piuttosto dell'ordine di \sqrt{dt} , ovvero, ancora con una notazione simbolica,

$$dN(t) = O\left(\sqrt{dt}\right) \quad d\tilde{N}(t) = O\left(\sqrt{dt}\right) \quad dW(t) = O\left(\sqrt{dt}\right)$$

Pur nella sua attuale imprecisione, questo risultato spiega intuitivamente perché i limiti di rapporti incrementali del tipo (6.68) non esistano, e giustifica quindi la non derivabilità dei processi considerati. Comunque, se strettamente parlando i rumori bianchi $\dot{N}(t)$, $\dot{\tilde{N}}(t)$ e $\dot{W}(t)$ non esistono, gli incrementi $\Delta N(t)$, $\Delta \tilde{N}(t)$ e $\Delta W(t)$ invece esistono, e in un certo senso (come vedremo più precisamente quando introdurremo il calcolo stocastico) esistono e giocano un ruolo importante anche i loro *differenziali stocastici* $dN(t)$, $d\tilde{N}(t)$ e $dW(t)$ che per ora possono essere solo intuitivamente pensati come incrementi infinitesimi in dt , ma dell'ordine di \sqrt{dt} . Quello che invece non è possibile scrivere senza correre il rischio di commettere errori seri è una diretta generalizzazione delle solite formule del calcolo infinitesimale che legano differenziali e derivate: insomma per i processi stocastici i simboli

$$dN(t) \quad d\tilde{N}(t) \quad dW(t)$$

possono essere correttamente definiti (come vedremo nel Capitolo 8); essi però non possono essere identificati in maniera indiscriminata con le rispettive espressioni

$$\dot{N}(t) dt \quad \dot{\tilde{N}}(t) dt \quad \dot{W}(t) dt$$

sia perché le derivate coinvolte non sono ben definite, sia perché gli incrementi infinitesimi sono dell'ordine di \sqrt{dt} , e non di dt . Vedremo nell'Appendice H a quali rischi esporrebbe un uso ingenuo delle solite regole del calcolo differenziale, e spiegheremo nel Capitolo 8 come questi problemi possano essere adeguatamente affrontati

6.4 Moto Browniano

Dopo le osservazioni di Robert Brown³ nel 1827 sul moto delle particelle di polline immerse in un fluido (da allora noto come *moto Browniano*) si aprì un lungo dibattito sulla natura di questo fenomeno e si avanzò anche l'ipotesi che le particelle di polline fossero *vive*. Esperimenti successivi mostrarono che questo non era il caso, ma l'origine del movimento restava misteriosa. Bisognò attendere i lavori di Einstein⁴ (1905) e Smoluchowski⁵ (1906) per avere una teoria che desse una risposta soddisfacente. In particolare Einstein costruì un modello fisico del fenomeno basato sull'interazione delle particelle con le molecole del fluido circostante, stabilì che il movimento è caratterizzato da un coefficiente di diffusione D legato alla temperatura, e prevede che in un tempo t lo spostamento quadratico medio in ogni direzione è proporzionale a \sqrt{Dt} . Tutte queste affermazioni – che oggi sembrano banali – erano all'epoca piuttosto controverse, ed è importante osservare che il successo del modello di Einstein contribuì in maniera significativa a stabilire l'idea che la materia fosse composta di atomi e molecole (come poi mostrato definitivamente da Jean Perrin⁶ nel 1909): un'idea tutt'altro che generalmente condivisa in quei tempi

Bisogna inoltre ricordare che una teoria coerente e rigorosa dei processi stocastici è un risultato piuttosto recente. Le prime idee pionieristiche su modelli che possono essere ricondotti al processo di Wiener sono contenute in un lavoro di Thorvald Thiele (1880) sul metodo dei minimi quadrati, e nella tesi di dottorato di Louis Bachelier (1900). Quest'ultimo lavoro in particolare è rimasto a lungo ignorato perché il suo argomento era una descrizione dei prezzi sui mercati finanziari, un problema che solo di recente è diventato popolare in ambiente fisico e matematico. Per questo motivo i primi lavori che hanno aperto effettivamente la strada allo studio moderno dei processi sono quelli di Einstein nel 1905, di Smoluchowski nel 1906 e di Langevin⁷ nel 1908. In particolare questo gruppo di articoli individua fin dall'inizio le due strade che si possono percorrere per esaminare l'evoluzione di un fenomeno aleatorio:

1. Si può studiare l'evoluzione delle *leggi*, cioè – nella nostra notazione – delle $ddp f_X(x, t)$ e delle ddp di transizione $f_X(x, t|y, s)$, determinando le opportune equazioni differenziali alle derivate parziali che queste funzioni devono soddisfare; in questo caso, quindi, l'attenzione è rivolta alle distribuzioni del processo, e non al processo stesso e alle sue traiettorie.

³R. Brown, *A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies*, Phil. Mag. 4 (1828) 161-173

⁴A. Einstein, *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, Ann. Phys. 17 (1905) 549-560

⁵M. von Smoluchowski, *Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen*, Ann. Phys. 21 (1906) 757-779

⁶J. Perrin, *Mouvement brownien et réalité moléculaire*, Ann. Chim. Phys. 8-ième série 18 (1909) 5-114

⁷P. Langevin, *On the theory of Brownian motion*, C. R. Acad. Sci. (Paris) 146 (1908) 530-533

2. Alternativamente si possono esaminare le traiettorie $x(t)$ del processo considerandole come generalizzazioni di funzioni tradizionali, e in questo caso avremo a che fare con equazioni differenziali sul processo $X(t)$ come nella tradizionale meccanica Newtoniana, ma bisognerà esercitare particolare cura per definire correttamente che cosa queste equazioni possono significare. Infatti, mentre le ddp del processo sono funzioni ordinarie, i processi $X(t)$ di cui ci occupiamo non sono in generale derivabili.

Storicamente gli articoli di Einstein e Smoluchowski si pongono nel primo filone, mentre quello di Langevin ha aperto la seconda strada. Noi esamineremo ora brevemente i problemi posti da questi articoli per introdurre l'argomento in maniera intuitiva, rinviando al Capitolo 9 una discussione più approfondita del moto Browniano.

6.4.1 Einstein (1905)

Proviamo a ripercorrere – con una notazione e un linguaggio adattati ai nostri – gli argomenti del lavoro di Einstein: prendiamo in considerazione innanzitutto un intervallo di tempo τ che sia contemporaneamente *abbastanza piccolo* rispetto ai tempi macroscopici delle osservazioni, e *abbastanza grande* rispetto ai tempi microscopici del movimento delle molecole del fluido. Questa scelta, come vedremo, è importante: essa da un lato riflette l'osservazione secondo la quale le particelle di polline sono *piccole* su scale macroscopiche, ma sono anche *grandi* su scale molecolari; dall'altro consente di avanzare realisticamente l'ipotesi che due spostamenti in intervalli τ successivi siano indipendenti. Infatti, se τ è grande rispetto ai tempi caratteristici dell'agitazione termica molecolare, potremo pensare che gli spostamenti corrispondenti risultino dalla somma di molti urti individuali e quindi siano complessivamente indipendenti fra di loro. La piccolezza di τ su scale macroscopiche, invece, permette di introdurre degli opportuni sviluppi in serie. La scala di un parametro come τ che è contemporaneamente piccolo su scale macroscopiche e grande su scale microscopiche si dice anche *mesoscopica*. Va detto comunque che una descrizione del moto Browniano a scale più piccole (che noi discuteremo nel Capitolo 9) è stata successivamente proposta da Ornstein e Uhlenbeck in un altro celebre lavoro⁸ (1930) nel quale viene definito un nuovo processo che prende il nome dai suoi due proponenti e che noi discuteremo in dettaglio nei prossimi capitoli

Sia allora Z la v -a che rappresenta lo spostamento della particella di polline in τ e $g(z)$ la sua ddp che supporremo simmetrica e concentrata attorno a valori prossimi a $z = 0$

$$\int_{-\infty}^{+\infty} g(z) dz = 1, \quad g(-z) = g(z)$$

$$g(z) \neq 0 \quad \text{solo per piccoli valori di } z$$

⁸L.S. Ornstein, G.E. Uhlenbeck, *On the theory of Brownian Motion*, Phys. Rev. 36 (1930) 823-841

Einstein dimostra innanzitutto che, se $X(t)$ è la posizione della particella di polline all'istante t , la sua ddp $f(x, t)$ soddisfa l'equazione

$$f(x, t + \tau) = \int_{-\infty}^{+\infty} f(x + z, t)g(z) dz \quad (6.74)$$

Il suo ragionamento è prevalentemente fisico, ma noi preferiamo darne una giustificazione in un linguaggio più aderente alla nostra notazione. Sappiamo infatti dalle regole sul condizionamento che per le due $v-a$ $X(t)$ e $X(t + \tau)$ si ha in ogni caso

$$\begin{aligned} f(x, t + \tau) dx &= \mathbf{P}\{x \leq X(t + \tau) < x + dx\} \\ &= \int_{-\infty}^{+\infty} \mathbf{P}\{x \leq X(t + \tau) < x + dx \mid X(t) = y\} f(y, t) dy \end{aligned}$$

D'altra parte sulla base delle nostre ipotesi potremo dire che $Z = X(t + \tau) - X(t)$ è indipendente da $X(t)$ per cui

$$\begin{aligned} \mathbf{P}\{x \leq X(t + \tau) < x + dx \mid X(t) = y\} &= \mathbf{P}\{x \leq X(t) + Z < x + dx \mid X(t) = y\} \\ &= \mathbf{P}\{x \leq y + Z < x + dx \mid X(t) = y\} \\ &= \mathbf{P}\{x - y \leq Z < x - y + dx\} \\ &= g(x - y) dx \end{aligned}$$

Ponendo allora $z = y - x$ e usando le proprietà di simmetria di $g(z)$ ne segue che

$$\begin{aligned} f(x, t + \tau) dx &= dx \int_{-\infty}^{+\infty} g(x - y) f(y, t) dy \\ &= dx \int_{-\infty}^{+\infty} f(x + z, t) g(z) dz \end{aligned}$$

ossia la richiesta equazione (6.74). Siccome ora τ è piccolo e $g(z)$ è non nulla solo per piccoli valori di z , potremo adottare in (6.74) i seguenti sviluppi di Taylor

$$\begin{aligned} f(x, t + \tau) &= f(x, t) + \tau \partial_t f(x, t) + o(\tau) \\ f(x + z, t) &= f(x, t) + z \partial_x f(x, t) + \frac{z^2}{2} \partial_x^2 f(x, t) + o(z^2) \end{aligned}$$

ottenendo quindi (in notazione sintetica)

$$f + \tau \partial_t f = f \int g(z) dz + \partial_x f \int z g(z) dz + \partial_x^2 f \int \frac{z^2}{2} g(z) dz$$

e siccome per le nostre ipotesi

$$\int_{-\infty}^{+\infty} g(z) dz = 1 \quad \int_{-\infty}^{+\infty} z g(z) dz = 0$$

in definitiva si ha

$$\partial_t f(x, t) = \partial_x^2 f(x, t) \frac{1}{2\tau} \int_{-\infty}^{+\infty} z^2 g(z) dz$$

Ora è evidente che $g(z)$ (la *ddp* dell'incremento Z in τ) dipende da τ , e che per simmetria $\mathbf{E}[Z] = 0$, sicché

$$\int_{-\infty}^{+\infty} z^2 g(z) dz$$

è la varianza di Z e potremo ragionevolmente supporre che essa sia infinitesima per $\tau \rightarrow 0$ (nel senso che l'incremento Z tende ad essere invariabilmente nullo per $\tau \rightarrow 0$). Se allora introduciamo l'ipotesi che

$$\lim_{\tau \rightarrow 0} \frac{1}{\tau} \int_{-\infty}^{+\infty} z^2 g(z) dz = D$$

la nostra equazione si riduce a

$$\partial_t f(x, t) = \frac{D}{2} \partial_x^2 f(x, t)$$

cioè coincide con l'equazione (6.54) che abbiamo trovato essere soddisfatta dalle *ddp* del processo di Wiener, per cui le sue soluzioni, con la condizione iniziale $f(x, 0^+) = \delta(x)$ (ossia $X(0) = 0$, **P**-qo), saranno ovviamente le *ddp* normali $\mathfrak{N}(0, Dt)$ in modo tale che

$$\mathbf{E}[X(t)] = 0 \quad \mathbf{V}[X(t)] = Dt \quad (6.75)$$

La prima importante conclusione è quindi che nel modello di Einstein la **posizione** della particella Browniana può essere descritta da un **processo di Wiener** con coefficiente di diffusione D . Sulla base di considerazioni termodinamiche, infine, Einstein fu anche in grado di legare il valore della costante di diffusione alla temperatura del fluido mediante la relazione

$$D = \frac{kT}{3\pi\eta a} \quad (6.76)$$

dove k è la costante di Boltzmann, T la temperatura, η la viscosità e a il diametro della particella supposta sferica. La (6.76) e gli altri risultati qui riportati saranno ricavati nuovamente, e in maniera più semplice, nella successiva trattazione di Langevin

6.4.2 Langevin (1908)

Nel 1908 Langevin ottenne sostanzialmente gli stessi risultati di Einstein trattando direttamente le traiettorie delle particelle con un'accorta (anche se poco rigorosa) generalizzazione delle equazioni del moto di Newton. In questo modello $X(t)$ indica la posizione e $V(t) = \dot{X}(t)$ la velocità della particella, e si suppone che quest'ultima sia soggetta a due tipi di forze dovute al mezzo in cui essa è immersa:

- una forza di attrito viscoso che, con le notazioni introdotte poco prima, è proporzionale alla velocità e vale $-6\pi\eta aV(t)$;
- una forza aleatoria prodotta dagli urti delle molecole rappresentata da un processo $B(t)$ con media nulla $\mathbf{E}[B(t)] = 0$, e non correlato con $X(t)$.

Conseguentemente l'equazione di Newton del moto è

$$m\ddot{X}(t) = -6\pi\eta a\dot{X}(t) + B(t) \quad (6.77)$$

e questa, con $V(t) = \dot{X}(t)$, può essere anche scritta sotto forma di una equazione del primo ordine per la velocità

$$m\dot{V}(t) = -6\pi\eta aV(t) + B(t) \quad (6.78)$$

che prende il nome di **equazione di Langevin**

Mostreremo nella Sezione 8.1 che a tutti gli effetti il processo $B(t)$ si comporta come il rumore bianco di Wiener discusso nell'Esempio 6.22, e questo costituirà il punto di partenza per la formulazione del calcolo stocastico di Itô. Per il momento però ci limiteremo a ricavare alcune informazioni sulla varianza del processo calcolando $\mathbf{E}[X^2(t)]$. Si moltiplica allora (6.77) per $X(t)$

$$mX(t)\ddot{X}(t) = -6\pi\eta aX(t)\dot{X}(t) + X(t)B(t)$$

e si osserva che usando le solite regole del calcolo differenziale

$$\frac{d}{dt} [X^2(t)] = 2X(t)\dot{X}(t) \quad (6.79)$$

$$\frac{d^2}{dt^2} [X^2(t)] = 2\dot{X}^2(t) + 2X(t)\ddot{X}(t) = 2V^2(t) + 2X(t)\ddot{X}(t) \quad (6.80)$$

potremo anche scrivere

$$\frac{m}{2} \frac{d^2}{dt^2} [X^2(t)] - mV^2(t) = -3\pi\eta a \frac{d}{dt} [X^2(t)] + X(t)B(t)$$

Prendendo ora le attese di ambedue i membri, e ricordando che per le nostre ipotesi $\mathbf{E}[X(t)B(t)] = \mathbf{E}[X(t)]\mathbf{E}[B(t)] = 0$, otteniamo l'equazione

$$\frac{m}{2} \frac{d^2}{dt^2} \mathbf{E}[X^2(t)] - \mathbf{E}[mV^2(t)] = -3\pi\eta a \frac{d}{dt} \mathbf{E}[X^2(t)]$$

Ma dalla meccanica statistica sappiamo che, all'equilibrio, l'energia cinetica media di una particella deve soddisfare la relazione

$$\mathbf{E}\left[\frac{m}{2}V^2(t)\right] = \frac{kT}{2} \quad (6.81)$$

dove k è la costante di Boltzmann e T la temperatura. Conseguentemente la nostra equazione diviene

$$\frac{m}{2} \frac{d^2}{dt^2} \mathbf{E} [X^2(t)] + 3\pi\eta a \frac{d}{dt} \mathbf{E} [X^2(t)] = kT$$

Una prima integrazione conduce allora a

$$\frac{d}{dt} \mathbf{E} [X^2(t)] = \frac{kT}{3\pi\eta a} + C e^{-6\pi\eta a t/m}$$

dove C è una costante di integrazione, e dopo un breve transitorio (l'esponenziale è infinitesimo con un tempo caratteristico dell'ordine di 10^{-8} sec) si ha

$$\frac{d}{dt} \mathbf{E} [X^2(t)] = \frac{kT}{3\pi\eta a} = D$$

cioè la (6.76), e con una seconda integrazione si ritrova lo stesso risultato dell'analisi di Einstein

$$\mathbf{V} [X(t)] = \mathbf{E} [X^2(t)] = Dt \tag{6.82}$$

dove abbiamo anche supposto che $X(0) = 0$, in modo che $\mathbf{E} [X^2(0)] = 0$. I vantaggi della trattazione di Langevin sono nel fatto che essa si presenta come piuttosto intuitiva essendo basata su un modello fisico con una semplice equazione del moto; e anche i calcoli sono piuttosto elementari. Essa presenta, però, anche dei rischi a causa delle sue basi matematiche poco solide come viene esposto nell'Appendice H. Le basi rigorose per una riformulazione convincente di questo modello di Langevin sono rinviate al Capitolo 8

Capitolo 7

Processi di Markov

7.1 Processi di Markov

7.1.1 Proprietà di Markov

Sebbene la proprietà di Markov sia in genere data con una direzione temporale preferenziale, quella dal *passato* verso il *futuro*, essa è in realtà simmetrica nelle due direzioni, e potrebbe essere intuitivamente espressa dicendo che *gli eventi del futuro e quelli del passato risultano reciprocamente indipendenti condizionatamente alla conoscenza dell'informazione disponibile al presente*. Per sottolineare questa simmetria partiremo dando la seguente definizione di **proprietà di Markov**, riservandoci di mostrare poi che essa è equivalente ad altre, più familiari formulazioni

Definizione 7.1. Diremo che un processo vettoriale con M componenti $\mathbf{X}(t) = (X_1(t), \dots, X_M(t))$ è un **processo di Markov** se, comunque scelti gli istanti di tempo $s_1 \leq \dots \leq s_m \leq s \leq t_1 \leq \dots \leq t_n$, e i vettori $\mathbf{y}_1, \dots, \mathbf{y}_m, \mathbf{y}, \mathbf{x}_1, \dots, \mathbf{x}_n$, risulta

$$\begin{aligned} F(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1 | \mathbf{y}, s) \\ = F(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1 | \mathbf{y}, s) F(\mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1 | \mathbf{y}, s) \end{aligned} \quad (7.1)$$

dove le F sono le fdd congiunte e condizionate del processo

In pratica quindi, più che l'*ordine* degli istanti di tempo, nella Markovianità è importante la loro *separazione*, tramite un *presente* s , in due gruppi (il *passato* s_1, \dots, s_m , e il *futuro* t_1, \dots, t_n) che però giocano ruoli simmetrici. Inoltre, proprio per questo motivo, anche l'ordine degli istanti all'interno dei due gruppi è irrilevante. Ovviamente, se $\mathbf{X}(t)$ è un processo *ac* dotato di *ddp* la (7.1) si scriverà come

$$\begin{aligned} f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1 | \mathbf{y}, s) \\ = f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1 | \mathbf{y}, s) f(\mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1 | \mathbf{y}, s) \end{aligned} \quad (7.2)$$

mentre se invece è discreto con valori interi \mathbf{k}, ℓ , e distribuzione p , si ha

$$\begin{aligned} & p(\mathbf{k}_n, t_n; \dots; \mathbf{k}_1, t_1; \ell_m, s_m; \dots; \ell_1, s_1 | \ell, s) \\ &= p(\mathbf{k}_n, t_n; \dots; \mathbf{k}_1, t_1 | \ell, s) p(\ell_m, s_m; \dots; \ell_1, s_1 | \ell, s) \end{aligned} \quad (7.3)$$

Proposizione 7.2. $\mathbf{X}(t)$ è un processo di Markov se e solo se

$$F(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1 | \mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1) = F(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1 | \mathbf{y}, s) \quad (7.4)$$

comunque scelti $s_1 \leq \dots \leq s_m \leq s \leq t_1 \leq \dots \leq t_n$, e $\mathbf{y}_1, \dots, \mathbf{y}_m, \mathbf{y}, \mathbf{x}_1, \dots, \mathbf{x}_n$; inoltre in (7.4) i ruoli del passato s_1, \dots, s_m e del futuro t_1, \dots, t_n possono essere scambiati

Dimostrazione: Sarà utile osservare che nei casi in cui $\mathbf{X}(t)$ è o *ac*, o discreto con valori interi \mathbf{k}, ℓ , la condizione (7.4) si scrive rispettivamente come

$$f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1 | \mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1) = f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1 | \mathbf{y}, s) \quad (7.5)$$

$$p(\mathbf{k}_n, t_n; \dots; \mathbf{k}_1, t_1 | \ell, s; \ell_m, s_m; \dots; \ell_1, s_1) = p(\mathbf{k}_n, t_n; \dots; \mathbf{k}_1, t_1 | \ell, s) \quad (7.6)$$

Per comodità di scrittura dimostreremo allora la nostra proposizione solo nella forma (7.5): innanzitutto, se è vera la (7.5) si ha

$$\begin{aligned} & f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1 | \mathbf{y}, s) \\ &= \frac{f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1; \mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1)}{f(\mathbf{y}, s)} \\ &= \frac{f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1; \mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1)}{f(\mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1)} \frac{f(\mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1)}{f(\mathbf{y}, s)} \\ &= f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1 | \mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1) f(\mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1 | \mathbf{y}, s) \\ &= f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1 | \mathbf{y}, s) f(\mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1 | \mathbf{y}, s) \end{aligned}$$

cioè si ritrova (7.2) e il processo è Markoviano. Viceversa, se il processo è Markoviano, cioè se è vera (7.2), si ha

$$\begin{aligned} & f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1 | \mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1) \\ &= \frac{f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1; \mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1)}{f(\mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1)} \\ &= \frac{f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1; \mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1)}{f(\mathbf{y}, s)} \frac{f(\mathbf{y}, s)}{f(\mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1)} \\ &= \frac{f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1 | \mathbf{y}, s)}{f(\mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1 | \mathbf{y}, s)} = f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1 | \mathbf{y}, s) \end{aligned}$$

cioè si ritrova (7.4). Inoltre, data la simmetria passato-futuro delle (7.1), l'ordine degli istanti di tempo non è rilevante per la dimostrazione e può essere modificato scambiando passato e futuro, ma sempre nel rispetto della loro separazione \square

In questa seconda formulazione – nella versione dal *passato* al *futuro* – la proprietà di Markov esprime l'idea che l'informazione contenuta nell'ultima osservazione disponibile (il *presente*, qui rappresentato dall'istante s) riassume tutto il *passato* (gli istanti s_1, \dots, s_m) ai fini della previsione del *futuro* t : a questo scopo infatti conta conoscere *dove* si trova in processo nell'istante presente s , ma non serve sapere *come* (cioè lungo quale percorso) ci si è arrivati. Ovviamente il futuro non è in assoluto indipendente dal passato, ma quest'ultimo risulta ridondante in quanto la storia precedente è riassunta nel presente s . Si tratta della maniera più semplice per introdurre una dipendenza non triviale fra i valori del processo in istanti differenti. Si noti che se il *futuro* si riduce a un solo istante t le (7.5) e (7.6) assumono una forma semplificata che mette in evidenza il ruolo delle *ddp* e delle *probabilità di transizione*

$$f(\mathbf{x}, t | \mathbf{y}, s; \mathbf{y}_m, s_m; \dots; \mathbf{y}_1, s_1) = f(\mathbf{x}, t | \mathbf{y}, s) \quad (7.7)$$

$$p(\mathbf{k}, t | \boldsymbol{\ell}, s; \boldsymbol{\ell}_m, s_m; \dots; \boldsymbol{\ell}_1, s_1) = p(\mathbf{k}, t | \boldsymbol{\ell}, s) \quad (7.8)$$

Corollario 7.3. $\mathbf{X}(t)$ è un **processo di Markov** se e solo se, comunque scelti gli istanti di tempo $s_1 \leq \dots \leq s_m \leq s \leq t_1 \leq \dots \leq t_n$, e assegnata un'arbitraria funzione di Borel limitata $g(\mathbf{x}_1, \dots, \mathbf{x}_n)$, risulta **P**-qo

$$\begin{aligned} \mathbf{E} [g(\mathbf{X}(t_1), \dots, \mathbf{X}(t_n)) | \mathbf{X}(s), \mathbf{X}(s_m), \dots, \mathbf{X}(s_1))] \\ = \mathbf{E} [g(\mathbf{X}(t_1), \dots, \mathbf{X}(t_n)) | \mathbf{X}(s)] \end{aligned} \quad (7.9)$$

ovvero equivalentemente se, per quasi ogni (in $\mathbf{P}_{\mathbf{X}}$) scelta di $\mathbf{y}_1, \dots, \mathbf{y}_m, \mathbf{y}$, risulta

$$\begin{aligned} \mathbf{E} [g(\mathbf{X}(t_1), \dots, \mathbf{X}(t_n)) | \mathbf{X}(s) = \mathbf{y}, \mathbf{X}(s_m) = \mathbf{y}_m, \dots, \mathbf{X}(s_1) = \mathbf{y}_1] \\ = \mathbf{E} [g(\mathbf{X}(t_1), \dots, \mathbf{X}(t_n)) | \mathbf{X}(s) = \mathbf{y}] \end{aligned} \quad (7.10)$$

Per la Definizione 3.41, la condizione (7.9) è anche equivalente a richiedere che per ogni fissata $g(\mathbf{x}_1, \dots, \mathbf{x}_n)$ esista un'altra funzione di Borel $h(\mathbf{x})$ tale che **P**-qo

$$\mathbf{E} [g(\mathbf{X}(t_1), \dots, \mathbf{X}(t_n)) | \mathbf{X}(s), \mathbf{X}(s_m), \dots, \mathbf{X}(s_1)] = h(\mathbf{X}(s)) \quad (7.11)$$

Anche in questo caso i ruoli del passato e del futuro possono essere scambiati

Dimostrazione: Se è vera la (7.10), basterà considerare il caso in cui $g = \chi_A$ è la funzione indicatrice dell'evento $A = (-\infty, \mathbf{x}_1] \times \dots \times (-\infty, \mathbf{x}_n]$, per avere

$$\begin{aligned} F(\mathbf{x}_n, t_n, \dots, \mathbf{x}_1, t_1 | \mathbf{y}, s; \mathbf{y}_m, s_m \dots; \mathbf{y}_1, s_1) \\ = \mathbf{P}\{(\mathbf{X}(t_1), \dots, \mathbf{X}(t_n)) \in A | \mathbf{X}(s) = \mathbf{y}; \dots; \mathbf{X}(s_1) = \mathbf{y}_1\} \\ = \mathbf{E} [\chi_A(\mathbf{X}(t_1), \dots, \mathbf{X}(t_n)) | \mathbf{X}(s) = \mathbf{y}; \dots; \mathbf{X}(s_1) = \mathbf{y}_1] \\ = \mathbf{E} [\chi_A(\mathbf{X}(t_1), \dots, \mathbf{X}(t_n)) | \mathbf{X}(s) = \mathbf{y}] \\ = \mathbf{P}\{(\mathbf{X}(t_1), \dots, \mathbf{X}(t_n)) \in A | \mathbf{X}(s) = \mathbf{y}\} = F(\mathbf{x}_n, t_n, \dots, \mathbf{x}_1, t_1 | \mathbf{y}, s) \end{aligned}$$

cioè la (7.4). D'altra parte, se viceversa è vera (7.4), la (7.10) ne discende facilmente perché le attese condizionate si calcolano proprio a partire dalla conoscenza delle *fdd* condizionate \square

Bisogna precisare che la Markovianità di un processo vettoriale $\mathbf{X}(t)$ non implica affatto la Markovianità delle singole componenti $X_j(t)$ o di un loro sottoinsieme: le singole componenti, infatti, contengono meno informazione dell'intero vettore, e quindi spesso risultano non Markoviane. Questa osservazione, però, può essere guardata anche da una prospettiva rovesciata: ad esempio, se un dato processo con una sola componente $X_1(t)$ non è Markoviano, è possibile in generale aggiungere altre componenti per costruire un vettore che sia Markoviano. In un certo senso questo allarga molto il dominio di applicabilità delle proprietà di Markov, perché in molti casi di interesse pratico potremo sempre considerare i nostri processi come componenti di qualche opportuno processo vettoriale Markoviano (vedi in particolare la discussione sul moto Browniano nella Sezione 9.3)

Proposizione 7.4. Regola di moltiplicazione a catena: *La legge di un processo di Markov $\mathbf{X}(t)$ è completamente determinata dalla distribuzione a un istante e dalla distribuzione di transizione, ovvero – rispettivamente nei casi ac e discreto – da*

$$f(\mathbf{x}, t) \quad e \quad f(\mathbf{x}, t | \mathbf{y}, s) \quad (7.12)$$

$$p(\mathbf{k}, t) \quad e \quad p(\mathbf{k}, t | \mathbf{l}, s) \quad (7.13)$$

secondo le regole di moltiplicazione a catena

$$f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1) = f(\mathbf{x}_n, t_n | \mathbf{x}_{n-1}, t_{n-1}) \dots f(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1) f(\mathbf{x}_1, t_1) \quad (7.14)$$

$$p(\mathbf{k}_n, t_n; \dots; \mathbf{k}_1, t_1) = p(\mathbf{k}_n, t_n | \mathbf{k}_{n-1}, t_{n-1}) \dots p(\mathbf{k}_2, t_2 | \mathbf{k}_1, t_1) p(\mathbf{k}_1, t_1) \quad (7.15)$$

con i tempi in ordine crescente $t_1 \leq \dots \leq t_n$, ovvero nell'ordine inverso

Dimostrazione: Come già ricordato nella Sezione 5.1, la legge globale di un processo è determinata dalla conoscenza di tutte le leggi congiunte finito-dimensionali in un numero arbitrario di istanti arbitrari; ma è facile vedere – ad esempio nel caso ac – che se $\mathbf{X}(t)$ è Markoviano tali leggi congiunte si ricavano da (7.12). Presi infatti $t_1 \leq \dots \leq t_n$ arbitrari (eventualmente anche nell'ordine inverso), dalla definizione di *ddp* condizionata (3.53) e da (7.7) si ottiene

$$\begin{aligned} f(\mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1) &= f(\mathbf{x}_n, t_n | \mathbf{x}_{n-1}, t_{n-1}; \dots; \mathbf{x}_1, t_1) \cdot \\ &\quad f(\mathbf{x}_{n-1}, t_{n-1} | \mathbf{x}_{n-2}, t_{n-2}; \dots; \mathbf{x}_1, t_1) \cdot \dots \\ &\quad \cdot f(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1) f(\mathbf{x}_1, t_1) \\ &= f(\mathbf{x}_n, t_n | \mathbf{x}_{n-1}, t_{n-1}) f(\mathbf{x}_{n-1}, t_{n-1} | \mathbf{x}_{n-2}, t_{n-2}) \cdot \dots \\ &\quad \cdot f(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1) f(\mathbf{x}_1, t_1) \end{aligned}$$

cioè la regola (7.14) per cui le *ddp* congiunte si ricavano tutte dalla conoscenza delle (7.12). Nel caso discreto la dimostrazione è del tutto analoga. Si noti che per poter ottenere la (7.14) dalla Markovianità (7.7) i tempi devono essere ordinati in maniera crescente o decrescente: ogni altra possibilità è esclusa \square

7.1.2 Equazioni di Chapman-Kolmogorov

L'importanza della Markovianità si apprezza immediatamente se si riflette al fatto che in base alla Proposizione 7.4 tale proprietà consente di ricostruire completamente tutta la gerarchia delle *ddp* congiunte di $\mathbf{X}(t)$ – cioè la sua legge globale – a partire dalla sola conoscenza delle leggi a un istante e delle leggi di transizione (7.12) o (7.13): una considerevole semplificazione sicuramente non disponibile per tutti gli altri tipi di processi. A questo proposito è bene osservare subito, però, che le funzioni (7.12) e (7.13) sono sempre definite per qualunque tipo di processo, anche non Markoviano; in quest'ultima eventualità però esse non sono più sufficienti per determinare completamente la legge del processo. Inoltre bisogna ricordare che in generale possono esistere processi differenti (non tutti di Markov, ovviamente) che condividano fra loro le medesime (7.12) o (7.13) come legge a un istante e legge di transizione. In pratica, una volta data (ad esempio nel caso *ac*) la coppia di funzioni (7.12), ci saranno in generale differenti processi che le ammettono come *ddp* a un istante e *ddp* di transizione: se fra tali processi ve ne è uno Markoviano (ma potrebbe anche non esserci) esso non può che essere uno solo e ricostruibile, tramite la regola di moltiplicazione a catena (7.14), a partire dalle sole (7.12). Queste considerazioni suggeriscono allora che la semplice assegnazione a priori di una coppia di funzioni di tipo (7.12) non garantisce affatto che sia possibile costruire un processo di Markov: tali funzioni potrebbero infatti essere associate a un processo non Markoviano, oppure a una pluralità di processi dei quali uno solo è Markoviano. Sarà quindi importante innanzitutto essere in grado di capire se delle *ddp* del tipo (7.12) assegnate possono o meno generare un processo di Markov: in effetti vedremo subito che, mentre la forma di $f(\mathbf{x}, t)$ è arbitraria, non ogni possibile *ddp* di transizione $f(\mathbf{x}, t | \mathbf{y}, s)$ è Markoviana. Nel seguito gli integrali e le somme senza ulteriori indicazioni sono eseguiti su tutto il dominio disponibile, ad esempio \mathbf{R}^M , \mathbf{N} , mentre $d\mathbf{x}$ è una semplificazione di $d^M\mathbf{x}$

Proposizione 7.5. Equazioni di Chapman-Kolmogorov: *Se $\mathbf{X}(t)$ è un processo di Markov, e $s \leq r \leq t$, nel caso *ac* saranno soddisfatte le equazioni*

$$f(\mathbf{x}, t) = \int f(\mathbf{x}, t | \mathbf{y}, s) f(\mathbf{y}, s) d\mathbf{y} \quad (7.16)$$

$$f(\mathbf{x}, t | \mathbf{y}, s) = \int f(\mathbf{x}, t | \mathbf{z}, r) f(\mathbf{z}, r | \mathbf{y}, s) d\mathbf{z} \quad (7.17)$$

mentre nel caso discreto avremo

$$p(\mathbf{k}, t) = \sum_{\ell} p(\mathbf{k}, t | \ell, s) p(\ell, s) \quad (7.18)$$

$$p(\mathbf{k}, t | \ell, s) = \sum_j p(\mathbf{k}, t | \mathbf{j}, r) p(\mathbf{j}, r | \ell, s) \quad (7.19)$$

L'ordine degli istanti s, r, t può anche essere invertito, ma r deve sempre essere intermedio fra s e t

Dimostrazione: Prima di tutto osserviamo che le equazioni (7.16) e (7.18) sono soddisfatte da qualunque processo (anche non Markoviano): infatti, se ad esempio $\mathbf{X}(t)$ è dotato di *ddp*, la (7.16) deriva direttamente dalle definizioni

$$f(\mathbf{x}, t) = \int f(\mathbf{x}, t; \mathbf{y}, s) d\mathbf{y} = \int f(\mathbf{x}, t | \mathbf{y}, s) f(\mathbf{y}, s) d\mathbf{y}$$

Nel caso discreto la (7.18) si ricava nello stesso modo. Invece solo i processi di Markov soddisfano le equazioni (7.17) e (7.19): così, ad esempio, per ricavare (7.17) basterà osservare che

$$\begin{aligned} f(\mathbf{x}, t | \mathbf{y}, s) &= \int f(\mathbf{x}, t; \mathbf{z}, r | \mathbf{y}, s) dz = \int f(\mathbf{x}, t | \mathbf{z}, r; \mathbf{y}, s) f(\mathbf{z}, r | \mathbf{y}, s) dz \\ &= \int f(\mathbf{x}, t | \mathbf{z}, r) f(\mathbf{z}, r | \mathbf{y}, s) dz \end{aligned}$$

se si usano non solo le definizioni, ma anche la proprietà di Markov (7.7) valida sia nell'ordine $s \leq r \leq t$, sia nell'ordine inverso $t \leq r \leq s$ □

Si noti pertanto che, se da un lato le (7.16) e (7.18) sono solo relazioni che permettono di calcolare le leggi a tempi successivi t partendo da tempi precedenti s (le *ddp* e le probabilità di transizione in questo caso fungono da *propagatori*), dall'altro le equazioni (7.17) e (7.19) sono delle vere e proprie *condizioni di compatibilità* per le *ddp* e per le probabilità di transizione: se (7.17) e (7.19) non sono rispettate, le $f(\mathbf{x}, t | \mathbf{y}, s)$ e $p(\mathbf{k}, t | \ell, s)$ non possono essere usate per costruire processi Markoviani; se invece esse sono soddisfatte, esse potranno facilmente generare processi Markov

Definizione 7.6. *Chiameremo leggi di transizione Markoviane quelle per le quali $f(\mathbf{x}, t | \mathbf{y}, s)$ oppure $p(\mathbf{k}, t | \ell, s)$ soddisfano rispettivamente (7.17) oppure (7.19); inoltre, fissato l'istante del condizionamento $s \geq 0$, distingueremo in ciascuna di esse una regione avanzata $t \in [0, s]$ e una regione ritardata $t \in [s, +\infty)$*

Proposizione 7.7. *Ad ogni legge di transizione Markoviana assegnata nella regione ritardata*

$$f(\mathbf{x}, t | \mathbf{y}, s) \quad \text{ovvero} \quad p(\mathbf{k}, t | \ell, s) \quad 0 \leq s \leq t$$

è associata un'intera famiglia di processi di Markov: uno per ogni legge iniziale data tramite una $f_0(\mathbf{x})$ oppure una $p_0(\ell)$

Dimostrazione: Limitandoci per brevità al caso *ac*, data un'arbitraria *ddp* iniziale $f(\mathbf{x}, 0) = f_0(\mathbf{x})$, la *ddp* Markoviana $f(\mathbf{x}, t | \mathbf{y}, s)$ nella regione ritardata permetterà di determinare le *ddp* in ogni istante

$$f(\mathbf{x}, t) = \int f(\mathbf{x}, t | \mathbf{y}, 0) f_0(\mathbf{y}) d\mathbf{y}$$

completando così la coppia di funzioni (7.12) necessaria per ricavare la legge del processo di Markov dalla regola di moltiplicazione a catena (7.14). Si noti infatti

$\mathbf{X}(t_n) - \mathbf{X}(t_{n-1})$ risultano indipendenti. In particolare ogni incremento $\Delta\mathbf{X}(t) = \mathbf{X}(t + \Delta t) - \mathbf{X}(t)$ con $\Delta t > 0$ sarà indipendente¹ da ogni $\mathbf{X}(s)$ con $s \leq t$

Abbiamo già incontrato diversi esempi di processi che per costruzione risultavano avere incrementi indipendenti (ad esempio i processi di Poisson e di Wiener), e ora vogliamo mostrare che i processi che godono di questa proprietà sono tutti processi di Markov, anzi ne costituiscono una delle categorie più importanti

Proposizione 7.9. *Ogni processo $\mathbf{X}(t)$ ad incrementi indipendenti è un processo di Markov la cui legge è completamente determinata (a meno di una condizione iniziale) dalla legge degli incrementi. In particolare, con $\Delta t > 0$, si ha*

$$\mathbf{E}[g(\mathbf{X}(t + \Delta t)) | \mathbf{X}(t) = \mathbf{x}] = \mathbf{E}[g(\Delta\mathbf{X}(t) + \mathbf{x})] \quad (7.20)$$

Dimostrazione: Dati i tempi $t_1 \leq \dots \leq t_n \leq t \leq t + \Delta t$ e un'arbitraria $g(\mathbf{x})$, dalla proprietà 3 della Proposizione 3.42 – opportunamente adeguata – e dall'indipendenza dell'incremento $\Delta\mathbf{X}(t) = \mathbf{X}(t + \Delta t) - \mathbf{X}(t)$ dai precedenti valori del processo si ricava innanzitutto la proprietà di Markov nella forma (7.10):

$$\begin{aligned} \mathbf{E}[g(\mathbf{X}(t + \Delta t)) | \mathbf{X}(t) = \mathbf{x}, \mathbf{X}(t_n) = \mathbf{x}_n, \dots, \mathbf{X}(t_1) = \mathbf{x}_1] \\ &= \mathbf{E}[g(\Delta\mathbf{X}(t) + \mathbf{X}(t)) | \mathbf{X}(t) = \mathbf{x}, \dots, \mathbf{X}(t_1) = \mathbf{x}_1] \\ &= \mathbf{E}[g(\Delta\mathbf{X}(t) + \mathbf{x}) | \mathbf{X}(t) = \mathbf{x}, \dots, \mathbf{X}(t_1) = \mathbf{x}_1] \\ &= \mathbf{E}[g(\Delta\mathbf{X}(t) + \mathbf{x}) | \mathbf{X}(t) = \mathbf{x}] \\ &= \mathbf{E}[g(\Delta\mathbf{X}(t) + \mathbf{X}(t)) | \mathbf{X}(t) = \mathbf{x}] = \mathbf{E}[g(\mathbf{X}(t + \Delta t)) | \mathbf{X}(t) = \mathbf{x}] \end{aligned}$$

D'altra parte dalla precedente catena di uguaglianze discende anche (7.20)

$$\mathbf{E}[g(\mathbf{X}(t + \Delta t)) | \mathbf{X}(t) = \mathbf{x}] = \mathbf{E}[g(\Delta\mathbf{X}(t) + \mathbf{x}) | \mathbf{X}(t) = \mathbf{x}] = \mathbf{E}[g(\Delta\mathbf{X}(t) + \mathbf{x})]$$

dato che $\Delta\mathbf{X}(t)$ è indipendente da $\mathbf{X}(t)$. Per mostrare infine che – a meno di una condizione iniziale – la legge del nostro processo è completamente determinata dalla legge dei suoi incrementi $\Delta\mathbf{X}(t)$ osserveremo (limitandoci per brevità al caso *ac*) che la conoscenza della *ddp* degli incrementi $f_{\Delta\mathbf{X}}(\mathbf{x}; \Delta t, t)$ equivale alla conoscenza della *ddp* di transizione nella regione ritardata: infatti

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}, t + \Delta t | \mathbf{y}, t) &= \mathbf{P}\{\mathbf{X}(t + \Delta t) \leq \mathbf{x} | \mathbf{X}(t) = \mathbf{y}\} \\ &= \mathbf{P}\{\mathbf{X}(t + \Delta t) - \mathbf{X}(t) \leq \mathbf{x} - \mathbf{y} | \mathbf{X}(t) = \mathbf{y}\} \\ &= \mathbf{P}\{\Delta\mathbf{X}(t) \leq \mathbf{x} - \mathbf{y}\} = F_{\Delta\mathbf{X}}(\mathbf{x} - \mathbf{y}; \Delta t, t) \end{aligned}$$

e quindi per derivazione

$$f_{\mathbf{X}}(\mathbf{x}, t + \Delta t | \mathbf{y}, t) = f_{\Delta\mathbf{X}}(\mathbf{x} - \mathbf{y}; \Delta t, t) \quad \Delta t > 0 \quad (7.21)$$

¹Infatti, con $0 \leq s \leq t \leq t + \Delta t$, i vett-*a*

$$\mathbf{X}(0) \quad \mathbf{X}(s) - \mathbf{X}(0) \quad \mathbf{X}(t) - \mathbf{X}(s) \quad \Delta\mathbf{X}(t) = \mathbf{X}(t + \Delta t) - \mathbf{X}(t)$$

sono indipendenti per definizione, e quindi lo sono anche $\Delta\mathbf{X}(t)$ e $\mathbf{X}(s) = [\mathbf{X}(s) - \mathbf{X}(0)] + \mathbf{X}(0)$

D'altra parte, se il processo è ad incrementi indipendenti esso è Markoviano e dalla Proposizione 7.7 sappiamo che – a meno di una condizione iniziale – la sua legge è determinata dalla *ddp* di transizione ritardata, cioè dalla *ddp* degli incrementi \square

Questo risultato è cruciale per capire la relazione che intercorre fra la teoria dei processi ad incrementi indipendenti e i *teoremi limite* studiati nel Capitolo 4. Infatti, se $\mathbf{X}(t)$ è un processo ad incrementi indipendenti, per studiare la legge degli incrementi $\Delta\mathbf{X}(t) = \mathbf{X}(t) - \mathbf{X}(s)$ possiamo suddividere l'intervallo $[s, t]$ in n sotto-intervalli mediante i punti

$$s = t_0 < t_1 < \dots < t_{n-1} < t_n = t$$

e osservare che gli n incrementi $\mathbf{X}(t_k) - \mathbf{X}(t_{k-1})$ con $k = 1, \dots, n$ sono tutti indipendenti. Conseguentemente l'incremento $\Delta\mathbf{X}(t)$ risulterà essere somma di n *v-a* indipendenti, e data l'arbitrarietà di n e della scelta dei punti si intuisce facilmente che in generale la legge degli incrementi $\Delta\mathbf{X}(t)$ coinciderà la legge limite di qualche successione di somme di *v-a* indipendenti. Risulta fondamentale allora nello studio dei processi ad incrementi indipendenti (cioè di una larga classe di processi di Markov) essere capaci di caratterizzare tutte le possibili leggi limite di somme di *v-a* indipendenti. Noi già ne conosciamo alcune: la legge normale (Teorema Limite Centrale), la legge degenere (Legge dei Grandi Numeri) e infine la legge di Poisson (Teorema di Poisson). È però possibile dimostrare che questi sono solo gli esempi più importanti di una classe molto più vasta di leggi limite che prendono il nome di *leggi infinitamente divisibili*, introdotte nel 1929 da B. de Finetti e poi completamente classificate e caratterizzate negli anni '30 con i lavori di P. Lévy, A. Khintchin e di altri. Noi non potremo, per ragioni di spazio, entrare in una discussione dettagliata di queste distribuzioni che sono alla base dei *processi di Lévy*² (vedi Sezione 7.1.6) e ci limiteremo solo a ricordarne la definizione

Definizione 7.10. *Si dice che una legge con fc $\varphi(\mathbf{u})$ è **infinitamente divisibile** quando per ogni $n = 1, 2, \dots$ esiste un'altra fc $\varphi_n(\mathbf{u})$ tale che $\varphi(\mathbf{u}) = [\varphi_n(\mathbf{u})]^n$.*

Come si può capire facilmente da (4.5) ciò vuol dire che una *v-a* con legge φ si può sempre decomporre nella somma di un numero arbitrario di altre *v-a iid* con legge φ_n , e questo ovviamente giustifica il nome dato a queste leggi. Oltre le tre leggi limite già ricordate (Gauss, degenere e Poisson), sono infinitamente divisibili anche le leggi di Cauchy, quelle di Student, le esponenziali e molte altre famiglie di leggi continue e discrete. Viceversa ci sono importanti famiglie di leggi che non sono infinitamente divisibili: in particolare ricorderemo che nessuna distribuzione concentrata su intervalli limitati (come le uniformi o le beta) può essere infinitamente divisibile. Ovviamente le leggi che non sono infinitamente divisibili non possono essere leggi di incrementi di processi di Markov con incrementi indipendenti.

²**K.I. Sato**, LÉVY PROCESSES AND INFINITELY DIVISIBLE DISTRIBUTIONS, Cambridge UP (Cambridge, 1999)

7.1.4 Stazionarietà e omogeneità

Sappiamo che in generale la stazionarietà degli incrementi di un p -s non garantisce né la sua stazionarietà globale (Definizione 5.9), né la sua stazionarietà in senso lato. Capita infatti (e i processi di Poisson e di Wiener sono casi tipici) che le leggi degli incrementi siano stazionarie, ma che attese e varianze non siano costanti. Peraltro anche le condizioni (5.12) e (5.13) sulle leggi a uno e due tempi (che pure garantiscono almeno la stazionarietà in senso lato) non sono sufficienti per la stazionarietà globale del processo. Per i processi di Markov, invece, la situazione è diversa: per brevità nel seguito limiteremo l'esposizione a p -s ac dotati di ddp

Proposizione 7.11. *Un processo di Markov $\mathbf{X}(t)$ risulta stazionario se e solo se*

$$f(\mathbf{x}, t) = f(\mathbf{x}) \quad (7.22)$$

$$f(\mathbf{x}, t; \mathbf{y}, s) = f(\mathbf{x}, \mathbf{y}; \tau) \quad \tau = t - s \quad (7.23)$$

e inoltre, se $\mathbf{X}(t)$ ha anche incrementi indipendenti, con $\tau > 0$ risulterà

$$f(\mathbf{x}, \mathbf{y}; \tau) = f_{\Delta\mathbf{X}}(\mathbf{x} - \mathbf{y}; \tau)f(\mathbf{y}) \quad (7.24)$$

Dimostrazione: La condizione di stazionarietà

$$f(\mathbf{x}_1, t_1; \dots; \mathbf{x}_n, t_n) = f(\mathbf{x}_1, t_1 + \tau; \dots; \mathbf{x}_n, t_n + \tau) \quad (7.25)$$

per ogni t_1, \dots, t_n e τ deriva da (7.22) e (7.23) quando le ddp congiunte del processo sono calcolate con la regola della moltiplicazione a catena della Proposizione 7.4. Il viceversa è banale. Siccome $\mathbf{X}(t)$ è stazionario i suoi incrementi $\Delta\mathbf{X}(t)$ sono stazionari (vedi discussione nella Sezione 5.5): se essi sono anche indipendenti, da (7.21), (7.22) e (7.23) si ha

$$f(\mathbf{x}, \mathbf{y}; \tau) = f(\mathbf{x}, t + \tau | \mathbf{y}, t)f(\mathbf{y}, t) = f_{\Delta\mathbf{X}}(\mathbf{x} - \mathbf{y}; \tau)f(\mathbf{y})$$

cioè la (7.24) □

Naturalmente le ddp di transizione (ritardate) di un processo di Markov stazionario $\mathbf{X}(t)$ dipenderanno solo da $\tau = t - s > 0$, per cui useremo la notazione

$$f(\mathbf{x}, t | \mathbf{y}, s) = f(\mathbf{x}, \tau | \mathbf{y}) \quad \tau = t - s > 0 \quad (7.26)$$

Se poi gli incrementi sono anche indipendenti, le ddp di transizione saranno determinate dalle leggi (stazionarie) degli incrementi

$$f(\mathbf{x}, t | \mathbf{y}, s) = f(\mathbf{x}, \tau | \mathbf{y}) = f_{\Delta\mathbf{X}}(\mathbf{x} - \mathbf{y}, \tau) \quad \tau = t - s > 0 \quad (7.27)$$

Corrispondentemente le equazioni di Chapman-Kolmogorov per processi di Markov stazionari si riscriveranno nella forma ridotta (indicando con $s > 0, t > 0$ le ampiezze)

$$f(\mathbf{x}) = \int f(\mathbf{x}, t | \mathbf{y})f(\mathbf{y}) d\mathbf{y} \quad (7.28)$$

$$f(\mathbf{x}, t + s | \mathbf{y}) = \int f(\mathbf{x}, t | \mathbf{z})f(\mathbf{z}, s | \mathbf{y}) d\mathbf{z} \quad (7.29)$$

e se gli incrementi sono anche indipendenti

$$f(\mathbf{x}) = \int f_{\Delta\mathbf{X}}(\mathbf{x} - \mathbf{y}, t) f(\mathbf{y}) d\mathbf{y} \quad (7.30)$$

$$f_{\Delta\mathbf{X}}(\mathbf{x} - \mathbf{y}, t + s) = \int f_{\Delta\mathbf{X}}(\mathbf{x} - \mathbf{z}, t) f_{\Delta\mathbf{X}}(\mathbf{z} - \mathbf{y}, s) dz \quad (7.31)$$

Si noti che le due prime equazioni (7.28) e (7.30) esprimono il fatto che $f(\mathbf{x})$ è una *ddp* invariante che resta fissata dalla condizione iniziale; le altre due equazioni invece sono le condizioni di Markovianità rispettivamente per $f(\mathbf{x}, t | \mathbf{y})$ e $f_{\Delta\mathbf{X}}(\mathbf{x}, t)$: quando la (7.31) è soddisfatta parleremo anche di **incrementi Markoviani**

Corollario 7.12. *La condizione di Markovianità per incrementi indipendenti e stazionari (7.31) con ampiezze $s > 0, t > 0$ si può porre sotto forma di convoluzione*

$$f_{\Delta\mathbf{X}}(\mathbf{x}, t + s) = [f_{\Delta\mathbf{X}}(t) * f_{\Delta\mathbf{X}}(s)](\mathbf{x})$$

o di semplice prodotto delle corrispondenti *fc*

$$\varphi_{\Delta\mathbf{X}}(\mathbf{u}, t + s) = \varphi_{\Delta\mathbf{X}}(\mathbf{u}, t) \varphi_{\Delta\mathbf{X}}(\mathbf{u}, s)$$

e coincide con la richiesta che ogni incremento su un intervallo di ampiezza $s + t$ sia somma di due incrementi indipendenti su intervalli di ampiezza s e t

Dimostrazione: Per mostrare che la (7.31) assume la forma di una convoluzione basta eseguire il cambiamento di variabili $\mathbf{x} - \mathbf{y} \rightarrow \mathbf{x}$, e $\mathbf{z} - \mathbf{y} \rightarrow \mathbf{z}$ con determinante Jacobiano uguale a 1, ottenendo

$$f_{\Delta\mathbf{X}}(\mathbf{x}, t + s) = \int f_{\Delta\mathbf{X}}(\mathbf{x} - \mathbf{z}, t) f_{\Delta\mathbf{X}}(\mathbf{z}, s) dz$$

La forma per le *fc* segue poi banalmente dal teorema di convoluzione. \square

Definizione 7.13. *Diremo che un processo di Markov è **omogeneo (nel tempo)** quando la sua *ddp* di transizione (7.26) dipende solo dalla differenza $\tau = t - s$; in tal caso la seconda equazione di Chapman-Kolmogorov assume la forma (7.29).*

Corollario 7.14. *Ogni processo con incrementi indipendenti e stazionari è un processo di Markov omogeneo e – se esiste una distribuzione invariante che sia presa come distribuzione iniziale – esso è anche stazionario.*

Dimostrazione: Si vede subito da (7.21) che se gli incrementi sono indipendenti e stazionari la *ddp* di transizione soddisfa anche la (7.26) e quindi il processo è Markoviano e omogeneo. Se poi anche la distribuzione a un tempo è invariante, cioè soddisfa (7.22), è facile vedere che vale anche (7.23) e quindi il processo è stazionario in base alla Proposizione 7.11. Si noti però che in generale non è affatto detto che la legge a un tempo sia invariante, per cui un processo di Markov omogeneo può

benissimo non essere stazionario (neanche in senso lato). Questo può accadere o perchè la distribuzione invariante non esiste affatto (come nei casi dei p -s di Wiener e Poisson), o perchè comunque l'eventuale ddp invariante non è stata presa come distribuzione iniziale lungo l'evoluzione descritta dall'equazione di Chapman-Kolmogorov (7.30). In questi casi quindi il processo non è stazionario (neanche in senso lato) secondo la Definizione 5.9, ma solo omogeneo secondo la Definizione 7.13 \square

7.1.5 Ergodicità

Supponiamo di voler costruire (la legge di) un processo di Markov *stazionario* partendo dalla assegnazione (almeno nella regione ritardata $t > 0$) di una ddp di transizione *omogenea* $f(\mathbf{x}, t | \mathbf{y})$ del tipo (7.26). Dovremo innanzitutto verificare che sia soddisfatta la condizione di Markovianità di Chapman-Kolmogorov nella forma (7.29): se questo accade, per la Proposizione 7.7 saremo in grado di costruire un'intera famiglia di processi scegliendo arbitrariamente la ddp iniziale $f_0(\mathbf{x})$. I processi di Markov così ottenuti saranno sicuramente omogenei per definizione, ma – riproducendo la discussione del Corollario 7.14 – essi potrebbero essere tutti non stazionari. La ddp iniziale $f_0(\mathbf{x})$ iniziale, infatti, può non essere invariante; anzi una ddp invariante potrebbe anche non esistere affatto per la legge $f(\mathbf{x}, t | \mathbf{y})$ assegnata. Se questo avviene (o perchè la legge invariante non esiste, o perchè se ne è scelta un'altra come legge iniziale) il processo sarà omogeneo, ma non stazionario e si avrà l'evoluzione

$$f(\mathbf{x}, t) = \int f(\mathbf{x}, t | \mathbf{y}) f_0(\mathbf{y}) d\mathbf{y} \quad (7.32)$$

Se invece $f_0(\mathbf{x})$ è proprio la legge stazionaria $f(\mathbf{x})$, allora l'equazione (7.32) si riduce alla (7.28) con $f(\mathbf{x}) = f_0(\mathbf{x})$ e il processo è anche globalmente stazionario. Acquista quindi particolare interesse studiare le procedure tramite le quali si possa determinare – se esiste – una distribuzione invariante per una assegnata ddp di transizione omogenea

Riprendendo allora la discussione sulla ergodicità svolta nella Sezione 5.5, consideriamo un processo di Markov *ac* e stazionario $\mathbf{X}(t)$ con ddp invariante $f(\mathbf{x})$ e poniamoci il problema di stimarne la distribuzione

$$\mathbf{P}\{\mathbf{X}(t) \in B\} = \mathbf{E}[\chi_B(\mathbf{X}(t))] = \int_B f(\mathbf{x}) d\mathbf{x} \quad B \in \mathcal{B}(\mathbf{R}^M) \quad (7.33)$$

(ovviamente $\chi_B(\mathbf{x})$ è l'indicatore dell'insieme B) tramite una media temporale su un lungo intervallo $[-T, T]$. Per fare questo si definisce innanzitutto il processo $Y(t) = \chi_B(\mathbf{X}(t))$, per il quale da (7.33) si ha

$$\mathbf{E}[Y(t)] = \int_B f(\mathbf{x}) d\mathbf{x} \quad (7.34)$$

e poi la v -a

$$\bar{Y}_T = \frac{1}{2T} \int_{-T}^T Y(t) dt = \frac{1}{2T} \int_{-T}^T \chi_B(\mathbf{X}(t)) dt$$

che rappresenta la frazione di $[-T, T]$ durante la quale $\mathbf{X}(t)$ si trova in B (cioè una misura della *frequenza relativa* dei ritrovamenti del processo in B)

Teorema 7.15. *Sia data la famiglia di processi di Markov ac e omogenei associati a una ddp di transizione Markoviana e omogenea $f(\mathbf{x}, t | \mathbf{y})$: se esiste una ddp asintotica $\bar{f}(\mathbf{x})$, cioè tale che*

$$\lim_{t \rightarrow +\infty} f(\mathbf{x}, t | \mathbf{y}) = \bar{f}(\mathbf{x}) \quad \forall \mathbf{y} \in \mathbf{R}^M \quad (7.35)$$

allora $\bar{f}(\mathbf{x})$ è anche una ddp invariante, e per ogni altro processo omogeneo ma non stazionario, con ddp iniziale $f_0(\mathbf{x})$ non invariante, risulterà

$$\lim_{t \rightarrow +\infty} f(\mathbf{x}, t) = \bar{f}(\mathbf{x}) \quad (7.36)$$

Inoltre, se $\mathbf{X}(t)$ è il processo stazionario con ddp invariante $\bar{f}(\mathbf{x})$, e se la convergenza (7.35) è abbastanza rapida da garantire anche che

$$\int_0^{+\infty} |C_Y(\tau)| d\tau < +\infty \quad (7.37)$$

dove $C_Y(\tau)$ è l'autocovarianza di $Y(t)$, allora risulterà

$$\lim_{T \rightarrow \infty} \text{-mq} \bar{Y}_T = \int_B \bar{f}(\mathbf{x}) d\mathbf{x} \quad (7.38)$$

e diremo che il processo $\mathbf{X}(t)$ è **ergodico per la distribuzione**

Dimostrazione: Supponendo di poter scambiare limiti e integrali, da (7.35) e dall'equazione di Chapman-Kolmogorov (7.29) avremo innanzitutto

$$\begin{aligned} \int f(\mathbf{x}, t | \mathbf{y}) \bar{f}(\mathbf{y}) d\mathbf{y} &= \int f(\mathbf{x}, t | \mathbf{y}) \lim_{s \rightarrow +\infty} f(\mathbf{y}, s | \mathbf{z}) d\mathbf{y} \\ &= \lim_{s \rightarrow +\infty} \int f(\mathbf{x}, t | \mathbf{y}) f(\mathbf{y}, s | \mathbf{z}) d\mathbf{y} = \lim_{s \rightarrow +\infty} f(\mathbf{x}, t + s | \mathbf{z}) = \bar{f}(\mathbf{x}) \end{aligned}$$

per cui la ddp limite è anche una ddp invariante. Inoltre, comunque scelta una condizione iniziale $f_0(\mathbf{x})$ non invariante, dall'equazione di Chapman-Kolmogorov (7.32) e da (7.35) si ha anche

$$\lim_{t \rightarrow +\infty} f(\mathbf{x}, t) = \lim_{t \rightarrow +\infty} \int f(\mathbf{x}, t | \mathbf{y}) f_0(\mathbf{y}) d\mathbf{y} = \bar{f}(\mathbf{x}) \int f_0(\mathbf{y}) d\mathbf{y} = \bar{f}(\mathbf{x})$$

Infine, nella nostra notazione, la convergenza in mq (7.38) equivale a richiedere che il processo $Y(t) = \chi_B(\mathbf{X}(t))$ sia ergodico per l'attesa nel senso del Teorema 5.11, e

questo, in base al Corollario 5.12, è garantito se risulta verificata la condizione (7.37). D'altra parte, dalla stazionarietà di $\mathbf{X}(t)$, per l'autocovarianza di $Y(t)$ si ha

$$\begin{aligned}
 C_Y(\tau) &= \mathbf{E}[Y(t+\tau)Y(t)] - \mathbf{E}[Y(t+\tau)]\mathbf{E}[Y(t)] \\
 &= \mathbf{E}[Y(t+\tau)Y(t)] - \mathbf{E}[Y(t)]^2 \\
 &= \int \int \chi_B(\mathbf{x})\chi_B(\mathbf{y})f(\mathbf{x}, t+\tau; \mathbf{y}, t) d\mathbf{x}d\mathbf{y} - \left(\int_B \bar{f}(\mathbf{x}) d\mathbf{x} \right)^2 \\
 &= \int_B \int_B f(\mathbf{x}, \tau | \mathbf{y})\bar{f}(\mathbf{y}) d\mathbf{x}d\mathbf{y} - \int_B \int_B \bar{f}(\mathbf{x})\bar{f}(\mathbf{y}) d\mathbf{x}d\mathbf{y} \\
 &= \int_B \int_B [f(\mathbf{x}, \tau | \mathbf{y}) - \bar{f}(\mathbf{x})] \bar{f}(\mathbf{y}) d\mathbf{x}d\mathbf{y}
 \end{aligned}$$

per cui la condizione (7.35) garantisce innanzitutto che $C_Y(\tau)$ sia infinitesima per $\tau \rightarrow +\infty$. Se poi la convergenza di (7.35) è anche abbastanza rapida, potremo ritenere che la condizione (7.37) sia soddisfatta e quindi, a norma del Corollario 5.12, che $Y(t)$ sia ergodico per l'attesa, e $\mathbf{X}(t)$ ergodico per la distribuzione \square

In conclusione, per un processo di Markov omogeneo ed ergodico che ammetta una *ddp* limite nel senso della (7.35), tutte le *ddp* iniziali $f_0(\mathbf{x})$ seguono nel tempo delle evoluzioni $f(\mathbf{x}, t)$ che tendono verso la medesima *ddp* stazionaria $\bar{f}(\mathbf{x})$. In pratica il processo perde progressivamente memoria della legge di partenza e tende verso una legge limite che coincide anche con la *ddp* stazionaria $\bar{f}(\mathbf{x})$. Processi con questa proprietà, dunque, pur non essendo stazionari propriamente parlando, tendono asintoticamente verso un processo stazionario

Si noti infine che in condizioni particolari (ad esempio quando lo spazio disponibile risulta suddiviso in regioni segregate e non comunicanti) potrebbero esistere più di una *ddp* invariante o asintotica. In questo caso si deve esercitare una certa cura per individuare le *ddp* stazionarie di interesse e le loro relazioni con le *ddp* asintotiche: noi però trascureremo di elaborare ulteriormente su questo punto

7.1.6 Processi di Lévy

Definizione 7.16. Diremo che un p -s $\mathbf{X}(t)$ è un **processo di Lévy** se

1. $\mathbf{X}(0) = 0$ \mathbf{P} -qo
2. ha incrementi indipendenti e stazionari
3. è stocasticamente continuo nel senso della Definizione 5.5, cioè (tenendo conto anche di 1. e 2.) se per ogni $\epsilon > 0$

$$\lim_{t \rightarrow 0^+} \mathbf{P}\{|\mathbf{X}(t)| > \epsilon\} = 0 \tag{7.39}$$

Da quanto detto nelle sezioni precedenti si ricava dunque facilmente che ogni processo di Lévy è un processo di Markov omogeneo, e che le leggi dei suoi incrementi devono essere infinitamente divisibili. Pertanto la legge globale di un processo di Lévy sarà completamente determinata dalla legge stazionaria dei suoi incrementi per cui assume notevole importanza pratica il seguente risultato

Proposizione 7.17. *Se $\mathbf{X}(t)$ è un processo di Lévy, allora esiste una fc infinitamente divisibile $\varphi(\mathbf{u})$ e una scala di tempi $T > 0$ tali che la fc $\varphi_{\Delta\mathbf{X}}(\mathbf{u}, t)$ degli incrementi di ampiezza t sia*

$$\varphi_{\Delta\mathbf{X}}(\mathbf{u}, t) = [\varphi(\mathbf{u})]^{t/T} \quad t \geq 0 \quad (7.40)$$

Viceversa, comunque data una fc infinitamente divisibile $\varphi(\mathbf{u})$ e una scala di tempi $T > 0$, la (7.40) rappresenterà la fc degli incrementi di un processo di Lévy

Dimostrazione: Senza entrare nei dettagli di una discussione che eccederebbe i limiti di queste lezioni³, osserveremo solo in merito al *viceversa* che l'infinita divisibilità della fc assegnata garantisce che la (7.40) continui a restare una fc infinitamente divisibile comunque scelto $t \geq 0$, e che la forma (7.40) della fc degli incrementi è la maniera più semplice per assicurare anche il rispetto delle condizioni di Markovianità di Chapman-Kolmogorov del Corollario 7.12. \square

In pratica la costruzione di un generico processo di Lévy parte dalla scelta di una legge infinitamente divisibile con fc $\varphi(\mathbf{u})$, e definisce poi la legge del processo assegnando la distribuzione degli incrementi indipendenti e stazionari mediante la loro fc (7.40). La forma esplicita della ddp dell'incremento sarà poi ottenuta – ove possibile – invertendo la fc .

Esempio 7.18. *Abbiamo già incontrato due esempi di processi di Lévy: il processo di Poisson standard $N(t)$ è un processo stocasticamente continuo, e la fc dei suoi incrementi indipendenti (6.10) si ricava da (7.40) partendo da una legge infinitamente divisibile $\mathfrak{P}(\alpha)$ con fc $\varphi(u) = e^{\alpha(e^{iu}-1)}$ e ponendo $\lambda = \alpha/T$. Allo stesso modo il processo di Wiener standard $W(t)$ è un processo stocasticamente continuo, e la fc dei suoi incrementi indipendenti (6.45) si ricava da (7.40) partendo da una legge infinitamente divisibile $\mathfrak{N}(0, a^2)$ con fc $\varphi(u) = e^{-a^2 u^2/2}$ e ponendo $D = a^2/T$. Un terzo esempio, il processo di Cauchy standard, sarà introdotto con la Definizione 7.25, e la fc dei suoi incrementi indipendenti (7.50) si ricava da (7.40) partendo da una legge infinitamente divisibile $\mathfrak{C}(\alpha)$ con fc $\varphi(u) = e^{-\alpha|u|}$ e ponendo $a = \alpha/T$.*

7.1.7 Continuità e salti

I diversi tipi di continuità di un p -s sono già stati introdotti nella Definizione 5.5, e le condizioni per la continuità in mq (e quindi anche per la continuità stocastica) sono

³**K.I. Sato**, LÉVY PROCESSES AND INFINITELY DIVISIBLE DISTRIBUTIONS, Cambridge UP (Cambridge, 1999). **D. Applebaum**, LÉVY PROCESSES AND STOCHASTIC CALCULUS, Cambridge UP (Cambridge, 2009)

state discusse nella Proposizione 5.6. Ora vogliamo invece discutere della **continuità globale** di un processo di Markov $\mathbf{X}(t)$, intendendo con questo che tutte le sue traiettorie sono continue per ogni t , fatta eccezione al più per un sottoinsieme di traiettorie di misura nulla.

Teorema 7.19. *Un processo di Markov $\mathbf{X}(t)$ è continuo (globalmente) se e solo se sono soddisfatte le **condizioni di Lindeberg***

$$\sup_{\mathbf{y}, t} \mathbf{P}\{|\Delta \mathbf{X}(t)| > \epsilon \mid \mathbf{X}(t) = \mathbf{y}\} = o(\Delta t) \quad \forall \epsilon > 0 \quad \Delta t \rightarrow 0 \quad (7.41)$$

Se il processo ha anche incrementi indipendenti e stazionari tali condizioni si riducono a

$$\mathbf{P}\{|\Delta \mathbf{X}(t)| > \epsilon\} = o(\Delta t) \quad \forall \epsilon > 0 \quad \Delta t \rightarrow 0 \quad (7.42)$$

Dimostrazione: Omessa⁴ □

In base a questo risultato la continuità del processo di Markov è garantita se (uniformemente rispetto alle condizioni \mathbf{y} e t) la probabilità che in Δt l'incremento superi un'arbitraria soglia $\epsilon > 0$ è un infinitesimo di ordine superiore a Δt . Si vede in questo modo che la condizione di Lindeberg è una richiesta del tutto ragionevole sulla rapidità con la quale si deve annullare $|\Delta \mathbf{X}(t)|$ quando $\Delta t \rightarrow 0$. Si noti infine che se il processo è *ac* la condizione (7.41) si scrive anche

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \sup_{\mathbf{y}, t} \int_{|\mathbf{x}-\mathbf{y}| > \epsilon} f(\mathbf{x}, t + \Delta t \mid \mathbf{y}, t) d\mathbf{x} = 0 \quad \forall \epsilon > 0 \quad (7.43)$$

ovvero

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{|\mathbf{x}-\mathbf{y}| > \epsilon} f(\mathbf{x}, t + \Delta t \mid \mathbf{y}, t) d\mathbf{x} = 0 \quad \forall \epsilon > 0 \quad (7.44)$$

uniformemente in \mathbf{y} e t . Se poi il processo ha anche incrementi indipendenti e stazionari, con la sostituzione $\mathbf{x} - \mathbf{y} \rightarrow \mathbf{x}$ le condizioni di Lindeberg si riducono a

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{|\mathbf{x}| > \epsilon} f_{\Delta \mathbf{X}}(\mathbf{x}, \Delta t) d\mathbf{x} = 0 \quad \forall \epsilon > 0 \quad (7.45)$$

Va osservato che c'è una sorta di competitività fra la Markovianità e la continuità di un processo in base alla **scala dei tempi** sui quali si eseguono le osservazioni. Potremmo dire che più sono brevi i tempi di osservazione, più un processo risulta continuo, ma contemporaneamente perde in Markovianità. In un certo senso questo dipende dal fatto che una descrizione continua richiede maggiori informazioni sul passato della traiettoria e quindi entra in conflitto con la Markovianità. Ad esempio la descrizione del moto di una molecola in un gas con un modello di sfere dure e urti

⁴**W. Feller**, AN INTRODUCTION TO PROBABILITY THEORY AND ITS APPLICATIONS - II, Wiley (New York, 1971). **C.W. Gardiner**, HANDBOOK OF STOCHASTIC METHODS, Springer (Berlin, 1997)

istantanei produce traiettorie rettilinee a tratti con improvvisi cambiamenti nella velocità: in questo caso il processo della velocità è discontinuo, mentre quello della posizione è continuo e Markoviano (la posizione dopo ogni urto dipende dal punto di partenza, ma non da come ci si è arrivati). Se però si passa a scale di tempi più brevi e si introducono dettagli più accurati nella descrizione del fenomeno (elasticità e deformazioni dei corpi) anche la velocità può diventare un processo continuo, ma la posizione perde la Markovianità perché una parte della storia del sistema sarà richiesta per prevedere, anche solo probabilisticamente, il futuro.

Naturalmente, processi che non sono continui in senso globale possono comunque esserlo stocasticamente, ma in questo caso bisognerà tenere in conto la possibilità che le traiettorie presentino delle **discontinuità (salti)**: si tratta, come vedremo, di una eventualità tutt'altro che rara, ad esempio, fra i processi di Lévy. Tipicamente i salti dei nostri processi Markoviani capiteranno in istanti aleatori e saranno di ampiezza altrettanto aleatoria; in generale, però, si tratterà di *discontinuità di prima specie* nel senso che le traiettorie saranno comunque di tipo *cadlag*, cioè continue da destra con limite finito da sinistra in ogni istante. L'analisi delle discontinuità e delle loro distribuzioni costituisce una parte rilevante dello studio dei processi di Lévy che noi, però, non potremo sviluppare.

7.1.8 Processi di Poisson, Wiener e Cauchy

In questa e nella prossima sezione esamineremo le leggi di univariate di alcuni fra i processi di Markov più conosciuti ricavandole da leggi di transizione Markoviane, e le utilizzeremo per studiare le proprietà del processo. In alcuni casi (processi di Poisson e Wiener) si tratterà di processi già introdotti euristicamente costruendo direttamente le loro traiettorie e ricavando da queste le informazioni essenziali sulle loro distribuzioni. Negli altri casi il processo sarà introdotto per la prima volta proprio partendo dalla legge di transizione che consentirà di ricavare tutte le altre proprietà del processo

Processi di Poisson

Definizione 7.20. Diremo che $N(t)$ è un **processo di Poisson** di intensità λ se esso è un p -s di Markov, stocasticamente continuo, a valori interi con incrementi indipendenti, e con probabilità di transizione Markoviana e omogenea (nella regione ritardata $\Delta t > 0$)

$$p_N(n, t + \Delta t | m, t) = e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^{n-m}}{(n-m)!} \quad n = 0, 1, 2, \dots; \quad 0 \leq m \leq n \quad (7.46)$$

che si ricava dalla distribuzione (6.9) degli incrementi di ampiezza $\Delta t > 0$. Il **processo di Poisson standard** con $N(0) = 0$, \mathbf{P} -qo è un processo di Lévy

Ricordiamo inoltre da (6.10) che la *fc* degli incrementi è

$$\varphi_{\Delta N}(u, t) = e^{\lambda t(e^{iu} - 1)} \quad (7.47)$$

Pertanto, siccome abbiamo anche mostrato nella Proposizione 6.7 che il processo di Poisson è continuo in mq , e quindi anche stocasticamente, il processo di Poisson standard è un processo di Lévy costruito a partire dalla fc infinitamente divisibile

$$\varphi(u) = e^{\lambda T(e^{iu} - 1)}$$

dove $T > 0$ è una scala di tempi. Data però la natura delle sue traiettorie discussa nella Sezione 6.1.2 è intuitivo che $N(t)$ non sarà continuo in senso globale, come confermato dal seguente risultato ottenuto sulla base delle condizioni di Lindeberg

Proposizione 7.21. *Un processo di Poisson $N(t)$ non soddisfa le condizioni di Lindeberg (7.42) e pertanto non è globalmente continuo*

Dimostrazione: Si ha infatti che per $0 < \epsilon < 1$, e per $t \rightarrow 0$

$$\begin{aligned} \mathbf{P}\{|\Delta N| > \epsilon\} &= 1 - \mathbf{P}\{|\Delta N| \leq \epsilon\} = 1 - \mathbf{P}\{|\Delta N| = 0\} = 1 - e^{-\lambda t} \\ &= 1 - \left(1 - \lambda t + \dots + (-1)^n \frac{\lambda^n t^n}{n!} + \dots\right) = \lambda t + o(t) \end{aligned}$$

è del primo ordine in t , e quindi la condizione di Lindeberg non è soddisfatta \square

Il processo $N(t)$ è omogeneo, ma non stazionario, e le sue probabilità di transizione non convergono verso nessuna distribuzione di probabilità per $t \rightarrow +\infty$: pertanto le distribuzioni di transizione non ammettono leggi invarianti, e $N(t)$ non è ergodico

Processi di Wiener

Definizione 7.22. *Diremo che $W(t)$ è un **processo di Wiener** con coefficiente di diffusione D se esso è un p -s di Markov, stocasticamente continuo, con incrementi indipendenti, e con ddp di transizione Markoviana e omogenea*

$$f_W(x, t + \Delta t | y, t) = \frac{e^{-(x-y)^2/2D\Delta t}}{\sqrt{2\pi D\Delta t}} \quad \Delta t > 0 \quad (7.48)$$

che si ricava dalla distribuzione (6.44) degli incrementi di ampiezza $\Delta t > 0$. Il **processo di Wiener standard** con $W(0) = 0$, \mathbf{P} -qo è un processo di Lévy

Ricordiamo inoltre da (6.45) la fc degli incrementi risulta

$$\varphi_{\Delta W}(u, t) = e^{-Dt u^2/2} \quad (7.49)$$

Pertanto, siccome abbiamo mostrato nella Proposizione 6.7 che il processo di Wiener è continuo in mq , e quindi anche stocasticamente, il processo di Wiener standard è un processo di Lévy costruito a partire dalla fc infinitamente divisibile

$$\varphi(u) = e^{-DT u^2/2}$$

dove $T > 0$ è una scala di tempi. A differenza dal processo di Poisson, però, $W(t)$ risulta anche continuo in senso globale, come mostra il seguente risultato

Proposizione 7.23. *Un processo di Wiener $W(t)$ soddisfa sempre le condizioni di Lindeberg (7.42) e quindi è globalmente continuo*

Dimostrazione: Si ha infatti da (7.48) che per $t \rightarrow 0^+$

$$P\{|\Delta W| > \epsilon\} = \int_{|x|>\epsilon} f_{\Delta W}(x, t) dx = 2 \left[1 - \Phi \left(\frac{\epsilon}{\sqrt{Dt}} \right) \right] \rightarrow 0$$

dove abbiamo posto

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy \quad \Phi'(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

Con $\alpha = \epsilon/\sqrt{Dt}$, avremo allora dal Teorema di l'Hôpital

$$\lim_{t \rightarrow 0} \frac{1}{t} \left[1 - \Phi \left(\frac{\epsilon}{\sqrt{Dt}} \right) \right] = \lim_{t \rightarrow 0} \frac{\epsilon}{2t} \frac{e^{-\epsilon^2/2Dt}}{\sqrt{2\pi Dt}} = \lim_{\alpha \rightarrow +\infty} \frac{D\alpha^3}{\epsilon^2} \frac{e^{-\alpha^2/2}}{\sqrt{2\pi}} = 0$$

Pertanto la condizione di Lindeberg è rispettata e $W(t)$ è continuo □

Infine, come già sappiamo, $W(t)$ è omogeneo ma non stazionario e le sue ddp di transizione non convergono verso nessuna altra ddp per $t \rightarrow +\infty$ (con l'eccezione della densità di Lebesgue uniforme su tutto \mathbf{R} che però, come è noto, non è una vera ddp): pertanto le ddp di transizione non ammettono leggi invarianti e $W(t)$ non è ergodico

Processi di Cauchy

Proposizione 7.24. *Gli incrementi stazionari $\Delta X(t) \sim \mathfrak{C}(a\Delta t)$ per $a > 0, \Delta t > 0$, con una fc*

$$\varphi_{\Delta X}(u, \Delta t) = e^{-a\Delta t|u|} \tag{7.50}$$

e una ddp

$$f_{\Delta X}(x, \Delta t) = \frac{1}{\pi} \frac{a\Delta t}{x^2 + a^2\Delta t^2} \tag{7.51}$$

sono Markoviani e permettono quindi di definire una famiglia di processi di Markov ad incrementi indipendenti

Dimostrazione: Infatti le fc (7.50) soddisfano banalmente la condizione richiesta dal Corollario 7.12 dato che per incrementi $s > 0$ e $t > 0$ si ha

$$\varphi_{\Delta X}(u, t+s) = e^{-a(t+s)|u|} = e^{-at|u|} e^{-as|u|} = \varphi_{\Delta X}(u, t) \varphi_{\Delta X}(u, s)$$

Pertanto, in base alla Proposizione 7.7 potremo coerentemente costruire (le leggi di) una intera famiglia di processi $X(t)$ con incrementi indipendenti e stazionari con legge (7.51) □

Definizione 7.25. Diremo che $X(t)$ è un **processo di Cauchy** se è un p -s di Markov con incrementi indipendenti, e con ddp di transizione Markoviana

$$f_X(x, t + \Delta t | y, t) = \frac{1}{\pi} \frac{a\Delta t}{(x - y)^2 + (a\Delta t)^2} \quad \Delta t > 0 \quad (7.52)$$

che si ricava dalla distribuzione (7.51) degli incrementi di ampiezza $\Delta t > 0$. Lo chiameremo inoltre **processo di Cauchy standard** se $X(0) = 0$, P -qo

Date le proprietà delle leggi di Cauchy bisogna osservare che per il processo $X(t)$ non sarà possibile parlare di attese, varianze, autocorrelazioni, o di altri momenti di ordine maggiore o uguale a 1. Le uniche quantità che potremo sempre calcolare sono le probabilità, e conseguentemente mediane e quantili, ma dovremo fare a meno degli altri concetti più usuali

Proposizione 7.26. Un processo di Cauchy $X(t)$ è sempre stocasticamente continuo ma non è continuo in senso globale: pertanto il processo di Cauchy standard è un processo di Lévy

Dimostrazione: Infatti per $\epsilon > 0$ e $\Delta t \rightarrow 0$ da (7.51) si ha

$$\mathbf{P}\{|\Delta X| > \epsilon\} = \int_{|x|>\epsilon} f_{\Delta X}(x, \Delta t) dx = 1 - \frac{2}{\pi} \arctan \frac{\epsilon}{a\Delta t} \rightarrow 0$$

per cui dal Teorema di l'Hôpital risulta

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left(1 - \frac{2}{\pi} \arctan \frac{\epsilon}{a\Delta t} \right) = \frac{2\epsilon}{a\pi} \lim_{\Delta t \rightarrow 0} \frac{a^2}{\epsilon^2 + a^2\Delta t^2} = \frac{2a}{\epsilon\pi} > 0$$

Pertanto la condizione di Lindeberg non è rispettata e il processo di Cauchy non è globalmente continuo. Ciononostante $X(t)$ risulta stocasticamente continuo, e quindi il processo standard è un processo di Lévy: ricordando infatti che la ddp e la fc di un processo di Cauchy standard sono

$$f_X(x, t) = \frac{1}{\pi} \frac{at}{x^2 + a^2t^2} \quad \varphi_X(u, t) = e^{-at|u|} \quad (7.53)$$

da (7.53) si ha

$$\mathbf{P}\{|X(t)| > \epsilon\} = \int_{|x|>\epsilon} f_X(x, t) dx = 1 - \frac{2}{\pi} \arctan \frac{\epsilon}{at} \rightarrow 0$$

e quindi la condizione (7.39) è soddisfatta □

Si noti che – diversamente da quanto accade per il processo di Poisson – il processo di Cauchy è un processo a salti pur prendendo valori continui in \mathbf{R} . Naturalmente si potrebbe dimostrare che i salti del processo di Cauchy hanno ampiezze comunque piccole che si addensano attorno a valori infinitesimi, ma questo non contrasta con l'esistenza delle discontinuità. Anche il processo di Cauchy infine è omogeneo, ma non è né stazionario né ergodico perché le sue ddp di transizione non convergono verso nessuna ddp limite

7.1.9 Processi di Ornstein-Uhlenbeck

Proposizione 7.27. *La ddp di transizione omogenea ($\alpha > 0$, $\beta > 0$, $\Delta t > 0$)*

$$f(x, t + \Delta t | y, t) = f(x, \Delta t | y) = \frac{e^{-(x - ye^{-\alpha\Delta t})^2 / 2\beta^2(1 - e^{-2\alpha\Delta t})}}{\sqrt{2\pi\beta^2(1 - e^{-2\alpha\Delta t})}} \quad (7.54)$$

corrispondente alla legge $\mathfrak{N}(ye^{-\alpha\Delta t}, \beta^2(1 - e^{-2\alpha\Delta t}))$, è Markoviana, ergodica e ammette come ddp invariante la legge normale $\mathfrak{N}(0, \beta^2)$ con ddp

$$f(x) = \frac{e^{-x^2/2\beta^2}}{\sqrt{2\pi\beta^2}} \quad (7.55)$$

Dimostrazione: Siccome la (7.54) è omogenea nel tempo, per dimostrare la Markovianità dovremo verificare la seconda equazione di Chapman-Kolmogorov nella forma (7.29), e a questo scopo osserveremo che ponendo $u = ze^{-\alpha t}$, e tenendo conto delle proprietà riproduttive (3.70) delle leggi normali, si ottiene

$$\begin{aligned} \int f(x, t | z) f(z, s | y) dz &= \int \frac{e^{-(x - ze^{-\alpha t})^2 / 2\beta^2(1 - e^{-2\alpha t})}}{\sqrt{2\pi\beta^2(1 - e^{-2\alpha t})}} \frac{e^{-(z - ye^{-\alpha s})^2 / 2\beta^2(1 - e^{-2\alpha s})}}{\sqrt{2\pi\beta^2(1 - e^{-2\alpha s})}} dz \\ &= \int \frac{e^{-(x - u)^2 / 2\beta^2(1 - e^{-2\alpha t})}}{\sqrt{2\pi\beta^2(1 - e^{-2\alpha t})}} \frac{e^{-(u - ye^{-\alpha(t+s)})^2 / 2\beta^2 e^{-2\alpha t}(1 - e^{-2\alpha s})}}{\sqrt{2\pi\beta^2 e^{-2\alpha t}(1 - e^{-2\alpha s})}} du \\ &= \mathfrak{N}(0, \beta^2(1 - e^{-2\alpha t})) * \mathfrak{N}(ye^{-\alpha(t+s)}, \beta^2 e^{-2\alpha t}(1 - e^{-2\alpha s})) \\ &= \mathfrak{N}(ye^{-\alpha(t+s)}, \beta^2(1 - e^{-2\alpha(t+s)})) = f(x, t + s | y) \end{aligned}$$

come richiesto per la Markovianità. L'ergodicità è poi assicurata dal fatto che

$$\mathfrak{N}(ye^{-\alpha\tau}, \beta^2(1 - e^{-2\alpha\tau})) \longrightarrow \mathfrak{N}(0, \beta^2) \quad \tau \rightarrow +\infty$$

con ddp limite data da (7.55), e si verifica che tale legge limite è anche invariante: infatti, adottando ancora una volta il cambiamento di variabile $u = ye^{-\alpha t}$, si ha

$$\begin{aligned} \int f(x, t | y) f(y) dy &= \int \frac{e^{-(x - ye^{-\alpha t})^2 / 2\beta^2(1 - e^{-2\alpha t})}}{\sqrt{2\pi\beta^2(1 - e^{-2\alpha t})}} \frac{e^{-y^2/2\beta^2}}{\sqrt{2\pi\beta^2}} dy \\ &= \int \frac{e^{-(x - u)^2 / 2\beta^2(1 - e^{-2\alpha t})}}{\sqrt{2\pi\beta^2(1 - e^{-2\alpha t})}} \frac{e^{-u^2/2\beta^2 e^{-2\alpha t}}}{\sqrt{2\pi\beta^2 e^{-2\alpha t}}} du \\ &= \mathfrak{N}(0, \beta^2(1 - e^{-2\alpha t})) * \mathfrak{N}(0, \beta^2 e^{-2\alpha t}) = \mathfrak{N}(0, \beta^2) = f(x) \end{aligned}$$

dove abbiamo tenuto conto delle proprietà riproduttive (3.70), □

Definizione 7.28. *Diremo che $X(t)$ è un **processo di Ornstein-Uhlenbeck** se esso è un p -s di Markov omogeneo ed ergodico con ddp di transizione (7.54); quest'ultima ammette la ddp invariante (7.55) che individua il **processo di Ornstein-Uhlenbeck stazionario**. Diremo infine che $X(t)$ è un **processo di Ornstein-Uhlenbeck standard** se $X(0) = 0$, \mathbf{P} -qo*

Proposizione 7.29. *Un processo di Ornstein-Uhlenbeck $X(t)$ è continuo (globalmente, e quindi anche stocasticamente), ed è anche un processo Gaussiano se $X(0)$ è Gaussiana. In particolare il processo stazionario è Gaussiano e comunque scelti t_1, \dots, t_n si ha $(X(t_1), \dots, X(t_n)) \sim \mathfrak{N}(0, \mathbb{A})$ con matrice delle covarianze ($\tau_{jk} = t_j - t_k$)*

$$\mathbb{A} = \beta^2 \begin{pmatrix} 1 & e^{-\alpha|\tau_{12}|} & \dots & e^{-\alpha|\tau_{1n}|} \\ e^{-\alpha|\tau_{21}|} & 1 & & e^{-\alpha|\tau_{2n}|} \\ \vdots & & \ddots & \vdots \\ e^{-\alpha|\tau_{n1}|} & e^{-\alpha|\tau_{n2}|} & \dots & 1 \end{pmatrix} \quad (7.56)$$

e inoltre risulta $\mathbf{E}[X(t)] = 0$, $\mathbf{V}[X(t)] = \beta^2$, e

$$R(\tau) = C(\tau) = \beta^2 e^{-\alpha|\tau|} \quad \rho(\tau) = e^{-\alpha|\tau|} \quad S(\varpi) = \frac{\beta^2}{\pi} \frac{\alpha}{\alpha^2 + \varpi^2} \quad (7.57)$$

Infine gli incrementi $\Delta X = X(t) - X(s)$ di un processo di Ornstein-Uhlenbeck non sono indipendenti, e pertanto il processo di Ornstein-Uhlenbeck standard non è un processo di Lévy

Dimostrazione: Rinviando alla Sezione 7.2.4 sulle equazioni di evoluzione una discussione della continuità del processo, passiamo a verificare la Gaussianità del processo stazionario studiandone le distribuzioni congiunte: dalle proprietà di $X(t)$ si ha infatti con $\tau = t - s > 0$

$$\begin{aligned} f(x, t; y, s) &= f(x, t | y, s) f(y, s) = f(x, \tau | y) f(y) \\ &= \frac{e^{-(x - ye^{-\alpha\tau})^2 / 2\beta^2(1 - e^{-2\alpha\tau})}}{\sqrt{2\pi\beta^2(1 - e^{-2\alpha\tau})}} \frac{e^{-y^2/2\beta^2}}{\sqrt{2\pi\beta^2}} = \frac{e^{-(x^2 + y^2 - 2xye^{-\alpha\tau}) / 2\beta^2(1 - e^{-2\alpha\tau})}}{\sqrt{2\pi\beta^2} \sqrt{2\pi\beta^2(1 - e^{-2\alpha\tau})}} \end{aligned}$$

sicché da un rapido paragone con la forma generale di una *ddp* normale bivariata (2.24) si ricava che il vett-*a* $(X(s), X(t))$ segue una legge Gaussiana $\mathfrak{N}(0, \mathbb{A})$ con

$$\mathbb{A} = \beta^2 \begin{pmatrix} 1 & e^{-\alpha\tau} \\ e^{-\alpha\tau} & 1 \end{pmatrix}$$

La generalizzazione al caso di n istanti di tempo si ottiene iterando con un po' di lavoro la procedura seguita per il caso di 2 istanti. I risultati (7.57) seguono poi da una semplice applicazione della Gaussianità delle leggi multivariate alle rispettive definizioni. Rinviando invece alla Sezione 8.5.4 la verifica esplicita della Gaussianità dei processi con generiche condizioni iniziali $X(0)$ Gaussiane. Per semplicità, infine, discuteremo la non indipendenza degli incrementi solo per il processo stazionario: innanzitutto, per determinare la legge degli incrementi, osserviamo che con $\tau = t - s > 0$ dalla stazionarietà del processo, tenendo conto di (5.14), con in cambiamento di variabile $z = -y(1 - e^{-\alpha\tau})$ e l'uso delle solite proprietà riproduttive delle leggi

normali, si ha

$$\begin{aligned}
 f_{\Delta X}(x) &= \int f(x+y, \tau | y) f(y) dy = \int \frac{e^{-(x+y-ye^{-\alpha\tau})^2/2\beta^2(1-e^{-2\alpha\tau})}}{\sqrt{2\pi\beta^2(1-e^{-2\alpha\tau})}} \frac{e^{-y^2/2\beta^2}}{\sqrt{2\pi\beta^2}} dy \\
 &= \int \frac{e^{-(x-z)^2/2\beta^2(1-e^{-2\alpha\tau})}}{\sqrt{2\pi\beta^2(1-e^{-2\alpha\tau})}} \frac{e^{-z^2/2\beta^2(1-e^{-\alpha\tau})^2}}{\sqrt{2\pi\beta^2(1-e^{-\alpha\tau})^2}} dz \\
 &= \mathfrak{N}(0, \beta^2(1-e^{-2\alpha\tau})) * \mathfrak{N}(0, \beta^2(1-e^{-\alpha\tau})^2) = \mathfrak{N}(0, 2\beta^2(1-e^{-\alpha\tau}))
 \end{aligned}$$

Pertanto la legge degli incrementi è $\Delta X = X(t) - X(s) \sim \mathfrak{N}(0, 2\beta^2(1-e^{-\alpha\tau}))$ e quindi in particolare si ha costantemente $\mathbf{E}[\Delta X] = 0$. Per dimostrare allora che incrementi non sovrapposti di un processo di Ornstein-Uhlenbeck stazionario non sono indipendenti, useremo questa legge per provare che due incrementi relativi ad intervalli disgiunti hanno covarianza non nulla, e quindi sono correlati. Infatti, con $s_1 < s_2 < t_1 < t_2$, dai risultati precedenti si ha

$$\begin{aligned}
 \mathbf{E}[(X(t_2) - X(t_1))(X(s_2) - X(s_1))] &= \mathbf{E}[X(t_2)X(s_2)] + \mathbf{E}[X(t_1)X(s_1)] \\
 &\quad - \mathbf{E}[X(t_2)X(s_1)] - \mathbf{E}[X(t_1)X(s_2)] \\
 &= \beta^2 [e^{-\alpha(t_2-s_2)} + e^{-\alpha(t_1-s_1)} \\
 &\quad - e^{-\alpha(t_2-s_1)} - e^{-\alpha(t_1-s_2)}] \\
 &= \beta^2 (e^{-\alpha t_2} - e^{-\alpha t_1}) (e^{\alpha s_2} - e^{\alpha s_1})
 \end{aligned}$$

un'espressione che in generale non si annulla ed esclude quindi l'indipendenza \square

7.1.10 Processi Gaussiani non Markoviani

Esistono processi stocastici non Markoviani (un importante esempio è riportato nell'Appendice J) che ovviamente non godono delle proprietà fin qui esaminate. In particolare la semplice conoscenza delle ddp di transizione non permette più di ricostruire la legge globale del processo tramite la regola di moltiplicazione a catena: queste semplificazioni si perdono e in generale la gerarchia di tutte le leggi finito-dimensionali necessarie a definire la loro distribuzione deve essere assegnata con altri mezzi

Esempio 7.30. Leggi di transizione non Markoviane: *Supponiamo di considerare la ddp condizionata **uniforme** in $[y - \alpha(t-s), y + \alpha(t-s)]$*

$$f(x, t | y, s) = \begin{cases} \frac{1}{2\alpha(t-s)} & \text{se } |x - y| \leq \alpha(t-s) \\ 0 & \text{se } |x - y| > \alpha(t-s) \end{cases} \quad (7.58)$$

Si può allora vedere facilmente che usando la notazione $|[a, b]| = |b - a|$ si ha

$$\begin{aligned} & \int f(x, t | z, r) f(z, r | y, s) dz \\ &= \frac{|[x - \alpha(t - r), x + \alpha(t - r)] \cap [y - \alpha(r - s), y + \alpha(r - s)]|}{4\alpha^2(t - r)(r - s)} \\ &\neq f(x, t | y, s) \end{aligned}$$

e quindi l'equazione di Chapman-Kolmogorov (7.17) risulta non soddisfatta pur essendo (7.58) una legittima ddp condizionata: essa non potrà dunque essere il punto di partenza per la costruzione delle leggi di un processo di Markov.

Un altro esempio di questo genere è offerto dalla seguente famiglia di ddp condizionate di **Student** ($a > 0, \nu > 0$)

$$f(x, t + \Delta t | y, t) = \frac{1}{a\Delta t B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left(\frac{a^2 \Delta t^2}{(x - y)^2 + a^2 \Delta t^2} \right)^{\frac{\nu+1}{2}} \quad (7.59)$$

dove B è la funzione Beta di Riemann definita dalla relazione

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)} \quad (7.60)$$

tramite la funzione Gamma introdotta in (3.71). Si vede facilmente che le (7.59) non sono altro che una semplice generalizzazione delle ddp di transizione di Cauchy (7.52) che si ottengono per $\nu = 1$. Anche in questo caso si tratta dunque di ddp assolutamente legittime, ma un calcolo non elementare che non riportiamo consente di verificare che ancora una volta l'equazione di Chapman-Kolmogorov (7.17) non è soddisfatta con l'eccezione di un solo caso: quello di Cauchy $\nu = 1$ che come già sappiamo è Markoviano. In realtà un processo di Lévy (e quindi di Markov) detto processo di Student esiste, ma le sue ddp di transizione non sono della forma (7.59), e non sono neanche conosciute in forma esplicita con l'eccezione del caso di Cauchy

Sarà importante allora notare che c'è un'altra famiglia di processi (in generale non Markoviani) le cui leggi possono essere comunque assegnate con metodi piuttosto elementari: i *processi Gaussiani* della Definizione 5.3. La semplificazione principale discende dalle proprietà delle leggi Gaussiane multivariate che risultano sempre definite a partire dalle matrici di covarianza e dai vettori delle medie. Applicando queste idee a un processo Gaussiano $X(t)$ (per semplicità con una sola componente) potremo dire in base alla Definizione 4.17 che la sua legge sarà completamente definita tramite l'assegnazione delle funzioni (5.1) e (5.3)

$$m(t) = \mathbf{E} [X(t)] \quad C(t, s) = \mathbf{E} [X(t)X(s)] - m(t)m(s)$$

Infatti se sono note una generica $m(t)$ e una $C(t, s)$ definita non negativa, comunque assegnati t_1, \dots, t_n il vett- a $(X(t_1), \dots, X(t_n))$ avrà legge $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ dove

$$b_j = m(t_j) \quad a_{jk} = C(t_j, t_k)$$

e conseguentemente la *fc* di un'arbitraria distribuzione finito-dimensionale del nostro processo $X(t)$ avrà la forma

$$\varphi(u_1, t_1; \dots; u_n, t_n) = e^{i \sum_j m(t_j) u_j - \frac{1}{2} \sum_{jk} C(t_j, t_k) u_j u_k}$$

La legge di un processo Gaussiano $X(t)$ è quindi completamente conosciuta una volta che siano assegnate $m(t)$ e $C(t, s)$ (definita non negativa). Noi abbiamo già incontrato alcuni casi di processi Gaussiani che sono contemporaneamente anche processi di Markov: ad esempio il processo di Wiener standard $W(t)$, per il quale dalle Proposizioni 6.17 e 6.18 sappiamo che

$$m_W(t) = 0 \quad C_W(t, s) = D \min\{s, t\}$$

o il processo di Ornstein-Uhlenbeck stazionario $X(t)$ per il quale dalla Proposizione 7.29 abbiamo

$$m_X(t) = 0 \quad C_X(t, s) = \beta^2 e^{-\alpha|t-s|}$$

ma i casi rilevanti non si limitano a questi due e vari altri tipi di processi possono essere introdotti in questo modo. Un importante esempio di processo Gaussiano, ma non Markoviano, il **moto Browniano frazionario**, è presentato nell'Appendice J

7.2 Equazioni di evoluzione

L'equazione di Chapman-Kolmogorov (7.17) è una importante condizione di compatibilità per *ddp* di transizione Markoviane, ma – trattandosi di un'equazione integrale e non lineare – è difficile usarla come vera e propria equazione per determinare le leggi di un processo. Sarà importante allora mostrare che, per una vasta classe di processi di Markov, essa assume una forma più maneggevole. Bisogna notare subito, inoltre, che le equazioni che troveremo in questo modo non sono altro che le generalizzazioni delle equazioni di diffusione trovate da Einstein nel 1905, e quindi si pongono all'interno del primo dei due filoni di ricerca che abbiamo menzionato nella Sezione 6.4: si tratta cioè di equazioni per le leggi e non per le traiettorie del processo. I processi di Markov $\mathbf{X}(t) = (X_1(t), \dots, X_M(t))$ ai quali faremo riferimento in questa sezione sono dotati di *ddp* di transizione $f(\mathbf{x}, t | \mathbf{y}, s)$ nella regione ritardata $t > s$, e soddisfano le seguenti **condizioni**⁵:

1. esiste $\ell(\mathbf{x}|\mathbf{z}, t) \geq 0$ tale che, per ogni $\epsilon > 0$, e uniformemente in $\mathbf{x}, \mathbf{z}, t$, risulta

$$\lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} f(\mathbf{x}, t + \Delta t | \mathbf{z}, t) = \ell(\mathbf{x}|\mathbf{z}, t) \quad \text{con } |\mathbf{x} - \mathbf{z}| > \epsilon \quad (7.61)$$

$\ell(\mathbf{x}|\mathbf{z}, t)$ si chiamerà anche **densità di Lévy**

⁵W. Feller, AN INTRODUCTION TO PROBABILITY THEORY AND ITS APPLICATIONS - II, Wiley (New York, 1971). C.W. Gardiner, HANDBOOK OF STOCHASTIC METHODS, Springer (Berlin, 1997)

2. esiste $\mathbf{A}(\mathbf{z}, t)$ tale che, per ogni $\epsilon > 0$, e uniformemente in \mathbf{z}, t , risulta

$$\lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \int_{|\mathbf{x}-\mathbf{z}| < \epsilon} (x_i - z_i) f(\mathbf{x}, t + \Delta t | \mathbf{z}, t) d\mathbf{x} = A_i(\mathbf{z}, t) + O(\epsilon) \quad (7.62)$$

Ovviamente questo vuol dire che, con $\chi_B(\mathbf{x})$ indicatore di B ,

$$A_i(\mathbf{z}, t) = \lim_{\epsilon \rightarrow 0^+} \lim_{\Delta t \downarrow 0} \mathbf{E} \left[\frac{\Delta X_i(t)}{\Delta t} \chi_{[0, \epsilon]}(|\Delta \mathbf{X}(t)|) \mid \mathbf{X}(t) = \mathbf{z} \right] \quad (7.63)$$

il vettore $\mathbf{A}(\mathbf{z}, t)$ si chiamerà anche **vettore di trascinamento o di deriva**

3. esiste $\mathbb{B}(\mathbf{z}, t)$ tale che, per ogni $\epsilon > 0$, e uniformemente in \mathbf{z}, t , risulta

$$\lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \int_{|\mathbf{x}-\mathbf{z}| < \epsilon} (x_i - z_i)(x_j - z_j) f(\mathbf{x}, t + \Delta t | \mathbf{z}, t) d\mathbf{x} = B_{ij}(\mathbf{z}, t) + O(\epsilon) \quad (7.64)$$

Ovviamente questo vuol dire che

$$B_{ij}(\mathbf{z}, t) = \lim_{\epsilon \rightarrow 0^+} \lim_{\Delta t \downarrow 0} \mathbf{E} \left[\frac{\Delta X_i(t) \Delta X_j(t)}{\Delta t} \chi_{[0, \epsilon]}(|\Delta \mathbf{X}(t)|) \mid \mathbf{X}(t) = \mathbf{z} \right] \quad (7.65)$$

la matrice $\mathbb{B}(\mathbf{z}, t)$ si chiamerà anche **matrice di diffusione**

Si potrebbe dimostrare che, nelle condizioni date, potenze di ordine superiore al secondo in formule del tipo (7.62) e (7.64) condurrebbero a valori nulli: esse non vengono quindi prese in considerazione per questo motivo. Si noti inoltre che se $\ell(\mathbf{x}|\mathbf{z}, t) = 0$ per $\mathbf{x} \neq \mathbf{z}$ la condizione di Lindeberg (7.44) è soddisfatta e il processo sarà continuo. Una $\ell(\mathbf{x}|\mathbf{z}, t)$ non nulla è legata quindi alla presenza di discontinuità nelle traiettorie, cioè alla presenza di *termini di salto* nelle equazioni

7.2.1 Equazioni in avanti

Teorema 7.31. *Se $\mathbf{X}(t)$ è un processo di Markov ac con $\mathbf{X}(0) = \mathbf{X}_0$, \mathbf{P} -qo, e se esso soddisfa le ipotesi 1, 2 e 3, le sue ddp $f(\mathbf{x}, t)$ saranno soluzioni dell'equazione*

$$\begin{aligned} \partial_t f(\mathbf{x}, t) = & - \sum_i \partial_i [A_i(\mathbf{x}, t) f(\mathbf{x}, t)] + \frac{1}{2} \sum_{i,j} \partial_i \partial_j [B_{ij}(\mathbf{x}, t) f(\mathbf{x}, t)] \\ & + \int_{\mathbf{z} \neq \mathbf{x}} [\ell(\mathbf{x}|\mathbf{z}, t) f(\mathbf{z}, t) - \ell(\mathbf{z}|\mathbf{x}, t) f(\mathbf{x}, t)] d\mathbf{z} \end{aligned} \quad (7.66)$$

con condizione iniziale

$$f(\mathbf{x}, 0^+) = f_0(\mathbf{x}) \quad (7.67)$$

dove $f_0(\mathbf{x})$ è la ddp di \mathbf{X}_0 . In particolare le ddp di transizione $f(\mathbf{x}, t | \mathbf{y}, s)$ nella regione ritardata $t > s$ saranno le soluzioni di (7.66) corrispondenti alle condizioni

$$f(\mathbf{x}, s^+) = \delta(\mathbf{x} - \mathbf{y}) \quad (7.68)$$

L'equazione (7.66) prende anche il nome di **equazione in avanti**

Dimostrazione: Tenendo conto del fatto che l'uniformità delle convergenze richiesta nelle nostre ipotesi ci permetterà di eseguire i limiti necessari anche sotto segno di integrale, cominceremo con il dimostrare che una *ddp* di transizione Markoviana $f(\mathbf{x}, t | \mathbf{y}, s)$ nella regione ritardata $t \geq s$ è soluzione di (7.66). Presa a questo scopo una funzione $h(\mathbf{x})$ derivabile almeno due volte (per poter usare la formula di Taylor fino al secondo ordine), ricordando che le derivate – se esistono – coincidono con le derivate *destre*, avremo

$$\begin{aligned} \partial_t \mathbf{E}[h(\mathbf{X}(t)) | \mathbf{X}(s) = \mathbf{y}] &= \partial_t \int h(\mathbf{x}) f(\mathbf{x}, t | \mathbf{y}, s) d\mathbf{x} = \int h(\mathbf{x}) \partial_t f(\mathbf{x}, t | \mathbf{y}, s) d\mathbf{x} \\ &= \lim_{\Delta t \downarrow 0} \int h(\mathbf{x}) \frac{f(\mathbf{x}, t + \Delta t | \mathbf{y}, s) - f(\mathbf{x}, t | \mathbf{y}, s)}{\Delta t} d\mathbf{x} \end{aligned}$$

Ma, con $\Delta t > 0$ e ridenominando alcune variabili, dall'equazione di Chapman-Kolmogorov (7.17) e dalle relazioni di normalizzazione potremo scrivere che

$$\begin{aligned} \int h(\mathbf{x}) f(\mathbf{x}, t + \Delta t | \mathbf{y}, s) d\mathbf{x} &= \iint h(\mathbf{x}) f(\mathbf{x}, t + \Delta t | \mathbf{z}, t) f(\mathbf{z}, t | \mathbf{y}, s) d\mathbf{x} d\mathbf{z} \\ \int h(\mathbf{x}) f(\mathbf{x}, t | \mathbf{y}, s) d\mathbf{x} &= \int h(\mathbf{z}) f(\mathbf{z}, t | \mathbf{y}, s) d\mathbf{z} \\ &= \iint h(\mathbf{z}) f(\mathbf{x}, t + \Delta t | \mathbf{z}, t) f(\mathbf{z}, t | \mathbf{y}, s) d\mathbf{x} d\mathbf{z} \end{aligned}$$

per cui, dividendo il dominio di integrazione in $|\mathbf{x} - \mathbf{z}| < \epsilon$ e $|\mathbf{x} - \mathbf{z}| \geq \epsilon$ con $\epsilon > 0$ arbitrario, avremo

$$\begin{aligned} &\int h(\mathbf{x}) \partial_t f(\mathbf{x}, t | \mathbf{y}, s) d\mathbf{x} \\ &= \lim_{\Delta t \downarrow 0} \iint [h(\mathbf{x}) - h(\mathbf{z})] \frac{f(\mathbf{x}, t + \Delta t | \mathbf{z}, t)}{\Delta t} f(\mathbf{z}, t | \mathbf{y}, s) d\mathbf{x} d\mathbf{z} \\ &= \lim_{\epsilon \downarrow 0} \lim_{\Delta t \downarrow 0} \left[\iint_{|\mathbf{x} - \mathbf{z}| < \epsilon} [h(\mathbf{x}) - h(\mathbf{z})] \frac{f(\mathbf{x}, t + \Delta t | \mathbf{z}, t)}{\Delta t} f(\mathbf{z}, t | \mathbf{y}, s) d\mathbf{x} d\mathbf{z} \right. \\ &\quad \left. + \iint_{|\mathbf{x} - \mathbf{z}| \geq \epsilon} [h(\mathbf{x}) - h(\mathbf{z})] \frac{f(\mathbf{x}, t + \Delta t | \mathbf{z}, t)}{\Delta t} f(\mathbf{z}, t | \mathbf{y}, s) d\mathbf{x} d\mathbf{z} \right] \end{aligned}$$

Si consideri innanzitutto l'integrazione nel dominio $|\mathbf{x} - \mathbf{z}| < \epsilon$: con $\epsilon \rightarrow 0$ si potrà usare la formula di Taylor di $h(\mathbf{x})$ attorno al punto \mathbf{z} arrestata al secondo ordine

$$\begin{aligned} h(\mathbf{x}) &= h(\mathbf{z}) + \sum_i (x_i - z_i) \partial_i h(\mathbf{z}) \\ &\quad + \frac{1}{2} \sum_{i,j} (x_i - z_i)(x_j - z_j) \partial_i \partial_j h(\mathbf{z}) + |\mathbf{x} - \mathbf{z}|^2 R(\mathbf{x}, \mathbf{z}) \end{aligned}$$

nella quale per il resto dovrà intendersi che $R(\mathbf{x}, \mathbf{z}) \rightarrow 0$ quando $|\mathbf{x} - \mathbf{z}| \rightarrow 0$. Si ha allora per l'integrale su $|\mathbf{x} - \mathbf{z}| < \epsilon$

$$\begin{aligned} & \iint_{|\mathbf{x}-\mathbf{z}|<\epsilon} [h(\mathbf{x}) - h(\mathbf{z})] \frac{f(\mathbf{x}, t + \Delta t | \mathbf{z}, t)}{\Delta t} f(\mathbf{z}, t | \mathbf{y}, s) d\mathbf{x}d\mathbf{z} \\ &= \iint_{|\mathbf{x}-\mathbf{z}|<\epsilon} \left[\sum_i (x_i - z_i) \partial_i h(\mathbf{z}) \right. \\ & \quad \left. + \frac{1}{2} \sum_{i,j} (x_i - z_i)(x_j - z_j) \partial_i \partial_j h(\mathbf{z}) \right] \frac{f(\mathbf{x}, t + \Delta t | \mathbf{z}, t)}{\Delta t} f(\mathbf{z}, t | \mathbf{y}, s) d\mathbf{x}d\mathbf{z} \\ & \quad + \iint_{|\mathbf{x}-\mathbf{z}|<\epsilon} |\mathbf{x} - \mathbf{z}|^2 R(\mathbf{x}, \mathbf{z}) \frac{f(\mathbf{x}, t + \Delta t | \mathbf{z}, t)}{\Delta t} f(\mathbf{z}, t | \mathbf{y}, s) d\mathbf{x}d\mathbf{z} \end{aligned}$$

e quindi, al limite per $\Delta t \rightarrow 0$ e $\epsilon \rightarrow 0$, da (7.62) e (7.64) e con delle integrazioni per parti otteniamo

$$\begin{aligned} & \int \left[\sum_i A_i(\mathbf{z}, t) \partial_i h(\mathbf{z}) + \sum_{i,j} \frac{1}{2} B_{ij}(\mathbf{z}, t) \partial_i \partial_j h(\mathbf{z}) \right] f(\mathbf{z}, t | \mathbf{y}, s) d\mathbf{z} \\ &= \int h(\mathbf{z}) \left\{ - \sum_i \partial_{z_i} [A_i(\mathbf{z}, t) f(\mathbf{z}, t | \mathbf{y}, s)] + \frac{1}{2} \sum_{i,j} \partial_{z_i} \partial_{z_j} [B_{ij}(\mathbf{z}, t) f(\mathbf{z}, t | \mathbf{y}, s)] \right\} d\mathbf{z} \end{aligned}$$

Nel dominio $|\mathbf{x} - \mathbf{z}| \geq \epsilon$ invece – con uno scambio dei nomi delle variabili di integrazione \mathbf{x} e \mathbf{z} nel primo dei due addendi – si ha

$$\begin{aligned} & \iint_{|\mathbf{x}-\mathbf{z}|\geq\epsilon} [h(\mathbf{x}) - h(\mathbf{z})] \frac{f(\mathbf{x}, t + \Delta t | \mathbf{z}, t)}{\Delta t} f(\mathbf{z}, t | \mathbf{y}, s) d\mathbf{x}d\mathbf{z} \\ &= \iint_{|\mathbf{x}-\mathbf{z}|\geq\epsilon} h(\mathbf{z}) \left[\frac{f(\mathbf{z}, t + \Delta t | \mathbf{x}, t)}{\Delta t} f(\mathbf{x}, t | \mathbf{y}, s) \right. \\ & \quad \left. - \frac{f(\mathbf{x}, t + \Delta t | \mathbf{z}, t)}{\Delta t} f(\mathbf{z}, t | \mathbf{y}, s) \right] d\mathbf{x}d\mathbf{z} \end{aligned}$$

e passando poi al limite per $\Delta t \rightarrow 0$ da (7.61) si ottiene prima

$$\iint_{|\mathbf{x}-\mathbf{z}|\geq\epsilon} h(\mathbf{z}) [\ell(\mathbf{z}|\mathbf{x}, t) f(\mathbf{x}, t | \mathbf{y}, s) - \ell(\mathbf{x}|\mathbf{z}, t) f(\mathbf{z}, t | \mathbf{y}, s)] d\mathbf{x}d\mathbf{z}$$

e successivamente per $\epsilon \rightarrow 0$

$$\int h(\mathbf{z}) \left\{ \int_{\mathbf{x} \neq \mathbf{z}} [\ell(\mathbf{z}|\mathbf{x}, t) f(\mathbf{x}, t | \mathbf{y}, s) - \ell(\mathbf{x}|\mathbf{z}, t) f(\mathbf{z}, t | \mathbf{y}, s)] d\mathbf{x} \right\} d\mathbf{z}$$

dove abbiamo introdotto la notazione abbreviata

$$\int_{\mathbf{x} \neq \mathbf{z}} \dots d\mathbf{x} = \lim_{\epsilon \rightarrow 0} \int_{|\mathbf{x}-\mathbf{z}|\geq\epsilon} \dots d\mathbf{x}$$

Raccogliendo allora tutti i termini avremo

$$\begin{aligned} & \int h(\mathbf{z}) \partial_t f(\mathbf{z}, t | \mathbf{y}, s) d\mathbf{z} \\ &= \int h(\mathbf{z}) \left\{ - \sum_i \partial_{z_i} [A_i(\mathbf{z}, t) f(\mathbf{z}, t | \mathbf{y}, s)] + \frac{1}{2} \sum_{i,j} \partial_{z_i} \partial_{z_j} [B_{ij}(\mathbf{z}, t) f(\mathbf{z}, t | \mathbf{y}, s)] \right. \\ & \quad \left. + \int_{\mathbf{x} \neq \mathbf{z}} [\ell(\mathbf{z} | \mathbf{x}, t) f(\mathbf{x}, t | \mathbf{y}, s) - \ell(\mathbf{x} | \mathbf{z}, t) f(\mathbf{z}, t | \mathbf{y}, s)] d\mathbf{x} \right\} d\mathbf{z} \end{aligned}$$

e data l'arbitrarietà di $h(\mathbf{x})$ potremo in definitiva scrivere (scambiando per comodità la \mathbf{z} con la \mathbf{x})

$$\begin{aligned} \partial_t f(\mathbf{x}, t | \mathbf{y}, s) &= - \sum_i \partial_{x_i} [A_i(\mathbf{x}, t) f(\mathbf{x}, t | \mathbf{y}, s)] + \frac{1}{2} \sum_{i,j} \partial_{x_i} \partial_{x_j} [B_{ij}(\mathbf{x}, t) f(\mathbf{x}, t | \mathbf{y}, s)] \\ & \quad + \int_{\mathbf{z} \neq \mathbf{x}} [\ell(\mathbf{x} | \mathbf{z}, t) f(\mathbf{z}, t | \mathbf{y}, s) - \ell(\mathbf{z} | \mathbf{x}, t) f(\mathbf{x}, t | \mathbf{y}, s)] d\mathbf{z} \quad (7.69) \end{aligned}$$

Questo mostra che le *ddp* di transizione soddisfano l'equazione integro-differenziale *in avanti* (7.66). Per ottenere poi l'equazione per le *ddp* $f(\mathbf{x}, t)$ basterà moltiplicare ambedue i membri della (7.69) per $f(\mathbf{y}, s)$ e integrare in $d\mathbf{y}$: la prima equazione di Chapman-Kolmogorov (7.16) si incaricherà allora di mostrare che anche le $f(\mathbf{x}, t)$ sono soluzioni di (7.66), in particolare anche per $s = 0$ \square

Teorema 7.32. *Se sono assegnati*

1. *un vettore di trascinamento* $\mathbf{A}(\mathbf{x}, t)$
2. *una matrice di covarianza definita non negativa* $\mathbb{B}(\mathbf{x}, t)$
3. *una densità di Lévy non negativa* $\ell(\mathbf{x} | \mathbf{y}, t)$

*sotto adeguate ipotesi esiste sempre un'unica soluzione non negativa e normalizzata dell'equazione in avanti (7.66) con condizione iniziale degenere (7.67) e opportune condizioni ai limiti; tale soluzione $f(\mathbf{x}, t | \mathbf{y}, s)$ con $t > s$ è una *ddp* di transizione nella regione ritardata che soddisfa l'equazione di Chapman-Kolmogorov (7.17) e quindi è Markoviana: pertanto in base alla Proposizione 7.7 essa individua una intera famiglia di processi di Markov, uno per ogni possibile condizione iniziale*

Dimostrazione: Omessa⁶ \square

Si noti che per ottenere questi risultati abbiamo supposto che il processo fosse dotato di *ddp* f : è evidente quindi che gli enunciati devono essere parzialmente modificati

⁶**I.F. Gihman, A.V. Skorohod**, THE THEORY OF STOCHASTIC PROCESSES - II, Springer (Berlin, 1975). **C.W. Gardiner**, HANDBOOK OF STOCHASTIC METHODS, Springer (Berlin, 1997)

per trattare il caso in cui il processo non sia dotato di *ddp*. Per limitarci ai casi più semplici consideriamo allora un processo che assuma solo valori interi, come ad esempio il processo di Poisson $N(t)$, e riformuliamo innanzitutto le ipotesi (7.61), (7.62) e (7.64). A questo scopo si noti che per un processo a valori interi le condizioni $|\mathbf{x} - \mathbf{z}| > \epsilon$ e $|\mathbf{x} - \mathbf{z}| \leq \epsilon$ con $\epsilon > 0$ arbitrario si traducono semplicemente in $\mathbf{n} \neq \mathbf{m}$ e $\mathbf{n} = \mathbf{m}$. Ne consegue immediatamente che la seconda e terza ipotesi (7.62) e (7.64) banalmente si riducono a chiedere che $\mathbf{A} = 0$ e $\mathbb{B} = 0$. Quanto alla prima ipotesi (7.61), invece, essa viene sostituita dalla sua versione discreta

$$\lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} p(\mathbf{n}, t + \Delta t | \mathbf{m}, t) = \ell(\mathbf{n} | \mathbf{m}, t) \quad \text{con } \mathbf{n} \neq \mathbf{m} \quad (7.70)$$

e uniformemente in \mathbf{n}, \mathbf{m} e t . Sotto queste condizioni si può dimostrare allora che l'equazione (7.69) viene sostituita da una **master equation** della forma (vedi Sezione 7.2.3 per ulteriori dettagli)

$$\partial_t p(\mathbf{n}, t) = \sum_{\mathbf{k}} [\ell(\mathbf{n} | \mathbf{k}, t) p(\mathbf{k}, t) - \ell(\mathbf{k} | \mathbf{n}, t) p(\mathbf{n}, t)] \quad (7.71)$$

la cui soluzione è la probabilità di transizione $p(\mathbf{n}, t | \mathbf{m}, s)$ se $p(\mathbf{n}, s^+) = \delta_{\mathbf{n}\mathbf{m}}$. Naturalmente per processi di questo tipo, data la natura intrinsecamente discontinua dei valori del processo, è piuttosto scontato che il ruolo principale sia giocato dalla $\ell(\mathbf{n} | \mathbf{m}, t)$. Non si deve però pensare che la funzione $\ell(\mathbf{x} | \mathbf{y}, t)$ sia sempre nulla nei casi di processi con valori continui. È già stato osservato alla fine della Sezione 7.1.1, infatti, che anche questi possono presentare delle discontinuità nelle loro traiettorie come ad esempio avviene per il processo di Cauchy che riprenderemo nel seguito

7.2.2 Equazioni *all'indietro*

Le equazioni di evoluzione che abbiamo così determinato si chiamano anche equazioni *in avanti* (*forward*) perché esse – intese come equazioni per le *ddp* di transizione nella regione ritardata – sono equazioni sulle variabili *finali* \mathbf{x}, t delle $f(\mathbf{x}, t | \mathbf{y}, s)$ con *condizioni iniziali* all'istante $s < t$. Esiste però anche una formulazione di queste equazioni detta *all'indietro* (*backward*): in questo caso le variabili coinvolte nelle operazioni integro-differenziali sono quelle iniziali \mathbf{y}, s , mentre le soluzioni devono soddisfare delle *condizioni finali* al tempo t . Si potrebbe infatti mostrare che le *ddp* di transizione ritardate del nostro processo di Markov soddisfano anche la seguente **equazione all'indietro**

$$\begin{aligned} \partial_s f(\mathbf{x}, t | \mathbf{y}, s) = & - \sum_i A_i(\mathbf{y}, s) \partial_{y_i} f(\mathbf{x}, t | \mathbf{y}, s) \\ & - \frac{1}{2} \sum_{i,j} B_{ij}(\mathbf{y}, s) \partial_{y_i} \partial_{y_j} f(\mathbf{x}, t | \mathbf{y}, s) \\ & + \int_{\mathbf{z} \neq \mathbf{y}} \ell(\mathbf{z} | \mathbf{y}, s) [f(\mathbf{x}, t | \mathbf{y}, s) - f(\mathbf{x}, t | \mathbf{z}, s)] d\mathbf{z} \end{aligned} \quad (7.72)$$

con **condizione finale** $f(\mathbf{x}, t | \mathbf{y}, t^-) = \delta(\mathbf{x} - \mathbf{y})$. Le due equazioni in avanti (7.69) e all'indietro (7.72) sono equivalenti: le *ddp* ritardate $f(\mathbf{x}, t | \mathbf{y}, s)$ soluzioni dell'equazione in avanti in \mathbf{x}, t con condizioni iniziali \mathbf{y}, s , sono anche soluzioni dell'equazione all'indietro in \mathbf{y}, s con i medesimi coefficienti e con condizioni finali \mathbf{x}, t . Le equazioni possono quindi essere ambedue usate secondo le esigenze: quelle in avanti sono più comuni nelle ordinarie applicazioni, ma anche quelle all'indietro trovano impiego in diversi problemi come quelli dei tempi di primo passaggio.

Da un punto di vista matematico, comunque, le equazioni all'indietro (7.72) sono considerate preferibili proprio per il fatto che esse operano sulle variabili del condizionamento \mathbf{y}, s . Infatti per poter scrivere l'equazione in avanti nella forma (7.69) bisogna supporre che il processo sia dotato di *ddp* $f(\mathbf{x}, t | \mathbf{y}, s)$ visto che le operazioni integro-differenziali coinvolte lavorano sulle variabili finali \mathbf{x}, t . D'altra parte è chiaro che tale ipotesi non è sempre soddisfatta, e quindi (come nel caso dei processi a valori interi) bisogna stabilire una formulazione distinta per i casi senza *ddp*. Viceversa le variabili del condizionamento \mathbf{y}, s sono sempre esplicitamente disponibili in qualunque tipo di legge o attesa condizionata dall'evento $\{\mathbf{X}(s) = \mathbf{y}\}$. Questo permette di dare una versione del tutto generale delle equazioni di evoluzione senza operare nessuna distinzione sulla base dell'esistenza o meno delle *ddp*.

Così ad esempio, trascurando per semplicità i termini di salto le equazioni all'indietro di tipo (7.72) si riducono alla forma

$$\begin{aligned} \partial_s f(\mathbf{x}, t | \mathbf{y}, s) = & - \sum_i A_i(\mathbf{y}, s) \partial_{y_i} f(\mathbf{x}, t | \mathbf{y}, s) \\ & - \frac{1}{2} \sum_{i,j} B_{ij}(\mathbf{y}, s) \partial_{y_i} \partial_{y_j} f(\mathbf{x}, t | \mathbf{y}, s) \end{aligned} \quad (7.73)$$

e in questo caso, posto per un'arbitraria funzione $h(\mathbf{x})$

$$g(\mathbf{y}, s) = \mathbf{E} [h(\mathbf{X}(t)) | \mathbf{X}(s) = \mathbf{y}] = \int h(\mathbf{x}) f(\mathbf{x}, t | \mathbf{y}, s) d\mathbf{x}$$

con una moltiplicazione per $h(\mathbf{x})$ e una integrazione la (7.73) diviene

$$\partial_s g(\mathbf{y}, s) = - \sum_i A_i(\mathbf{y}, s) \partial_i g(\mathbf{y}, s) - \frac{1}{2} \sum_{i,j} B_{ij}(\mathbf{y}, s) \partial_i \partial_j g(\mathbf{y}, s) \quad (7.74)$$

con condizione finale $g(\mathbf{y}, t^-) = h(\mathbf{y})$. La (7.74) è anche nota in letteratura come **equazione di Kolmogorov** e può essere enunciata senza fare riferimento alcuno all'esistenza delle *ddp* del processo

7.2.3 Classi di equazioni *in avanti*

Esamineremo ora alcune particolari categorie di equazioni di evoluzione *in avanti* scegliendo opportunamente i coefficienti \mathbf{A}, \mathbb{B} e ℓ e svolgendo alcune considerazioni euristiche per studiare la natura delle soluzioni

Processo di salto: Master equation

Innanzitutto consideriamo il caso in cui $\mathbf{A} = \mathbb{B} = 0$, mentre solo $\ell \neq 0$: l'equazione in avanti (7.66) si riduce allora a

$$\partial_t f(\mathbf{x}, t) = \int_{\mathbf{z} \neq \mathbf{x}} [\ell(\mathbf{x}|\mathbf{z}, t) f(\mathbf{z}, t) - \ell(\mathbf{z}|\mathbf{x}, t) f(\mathbf{x}, t)] d\mathbf{z} \quad (7.75)$$

e in particolare, nel caso in cui il processo è a valori interi (come il processo di Poisson), essa assume la forma discreta (7.71). Equazioni come queste prendono il nome di **master equation**, e i processi $\mathbf{X}(t)$ (anche quelli con valori continui) le cui leggi soddisfano un'equazione del tipo (7.75) si chiamano **processi di puro salto**

Proposizione 7.33. *Se $\mathbf{X}(t)$ è un processo di puro salto con valori continui, la probabilità che esso compia un salto di lunghezza finita in un intervallo $[t, t + dt]$ è*

$$\mathbf{P}\{\mathbf{X}(t + dt) \neq \mathbf{X}(t) \mid \mathbf{X}(t) = \mathbf{y}\} = dt \int_{\mathbf{x} \neq \mathbf{y}} \ell(\mathbf{x}|\mathbf{y}, t) d\mathbf{x} \quad (7.76)$$

Dimostrazione: La *ddp* di transizione ritardata $f(\mathbf{x}, t|\mathbf{y}, s)$, con $t > s$ e condizione iniziale $f(\mathbf{x}, s^+|\mathbf{y}, s) = \delta(\mathbf{x} - \mathbf{y})$, è soluzione della master equation (7.75) nella forma

$$\partial_t f(\mathbf{x}, t|\mathbf{y}, s) = \int_{\mathbf{z} \neq \mathbf{x}} [\ell(\mathbf{x}|\mathbf{z}, t) f(\mathbf{z}, t|\mathbf{y}, s) - \ell(\mathbf{z}|\mathbf{x}, t) f(\mathbf{x}, t|\mathbf{y}, s)] d\mathbf{z}$$

che, prendendo in particolare $s = t$, può anche essere simbolicamente scritta come

$$\frac{f(\mathbf{x}, t + dt|\mathbf{y}, t) - \delta(\mathbf{x} - \mathbf{y})}{dt} = \int_{\mathbf{z} \neq \mathbf{x}} [\ell(\mathbf{x}|\mathbf{z}, t)\delta(\mathbf{z} - \mathbf{y}) - \ell(\mathbf{z}|\mathbf{x}, t)\delta(\mathbf{x} - \mathbf{y})] d\mathbf{z}$$

ovvero anche come

$$f(\mathbf{x}, t + dt|\mathbf{y}, t) = \left[1 - dt \int_{\mathbf{z} \neq \mathbf{x}} \ell(\mathbf{z}|\mathbf{x}, t) d\mathbf{z} \right] \delta(\mathbf{x} - \mathbf{y}) + dt \int_{\mathbf{z} \neq \mathbf{x}} \ell(\mathbf{x}|\mathbf{z}, t)\delta(\mathbf{z} - \mathbf{y}) d\mathbf{z}$$

Avremo pertanto

$$\begin{aligned} \mathbf{P}\{\mathbf{X}(t + dt) \neq \mathbf{X}(t) \mid \mathbf{X}(t) = \mathbf{y}\} &= \int_{\mathbf{x} \neq \mathbf{y}} f(\mathbf{x}, t + dt|\mathbf{y}, t) d\mathbf{x} \\ &= \int_{\mathbf{x} \neq \mathbf{y}} d\mathbf{x} \left[1 - dt \int_{\mathbf{z} \neq \mathbf{x}} \ell(\mathbf{z}|\mathbf{x}, t) d\mathbf{z} \right] \delta(\mathbf{x} - \mathbf{y}) \\ &\quad + dt \int_{\mathbf{x} \neq \mathbf{y}} d\mathbf{x} \int_{\mathbf{z} \neq \mathbf{x}} \ell(\mathbf{x}|\mathbf{z}, t)\delta(\mathbf{z} - \mathbf{y}) d\mathbf{z} \end{aligned}$$

e quindi la (7.76) discende dal fatto che per le proprietà della δ di Dirac

$$\begin{aligned} \int_{\mathbf{x} \neq \mathbf{y}} d\mathbf{x} \left[1 - dt \int_{\mathbf{z} \neq \mathbf{x}} \ell(\mathbf{z}|\mathbf{x}, t) d\mathbf{z} \right] \delta(\mathbf{x} - \mathbf{y}) &= 0 \\ \int_{\mathbf{x} \neq \mathbf{y}} d\mathbf{x} \int_{\mathbf{z} \neq \mathbf{x}} \ell(\mathbf{x}|\mathbf{z}, t)\delta(\mathbf{z} - \mathbf{y}) d\mathbf{z} &= \int_{\mathbf{x} \neq \mathbf{y}} \ell(\mathbf{x}|\mathbf{y}, t) d\mathbf{x} \end{aligned}$$

Questo risultato permette di riconoscere la natura discontinua del processo e il significato della funzione $\ell(\mathbf{x}|\mathbf{y}, t)$: per intervalli di tempo infinitesimi il termine $\ell(\mathbf{x}|\mathbf{y}, t) dt$ gioca infatti il ruolo di una densità di probabilità di *non* restare nella posizione iniziale \mathbf{y} , compiendo un salto di lunghezza finita $\mathbf{x} - \mathbf{y}$ come indicato in (7.76). Si tenga presente però che questa interpretazione non può essere spinta oltre un certo limite dato che $\ell(\mathbf{x}|\mathbf{y}, t)$ non è normalizzata (e potrebbe anche non essere normalizzabile) e quindi non può essere considerata come una vera *ddp*: l'integrale di $\ell(\mathbf{x}|\mathbf{y}, t)$ può infatti approssimare una probabilità solo come un infinitesimo, cioè se moltiplicato per dt . \square

Processo di diffusione: equazione di Fokker-Planck

Consideriamo ora il caso in cui $\ell = 0$, ma $\mathbb{B} \neq 0$: il criterio di Lindeberg garantisce allora che quasi ogni traiettoria del processo $\mathbf{X}(t)$ è continua, e l'equazione in avanti si riduce a

$$\partial_t f(\mathbf{x}, t) = - \sum_i \partial_{x_i} [A_i(\mathbf{x}, t) f(\mathbf{x}, t)] + \frac{1}{2} \sum_{i,j} \partial_{x_i} \partial_{x_j} [B_{ij}(\mathbf{x}, t) f(\mathbf{x}, t)] \quad (7.77)$$

che prende il nome di **equazione di Fokker-Planck** e, a differenza di (7.66), è un'equazione puramente differenziale. In particolare l'equazione (6.54) per il processo di Wiener con una sola componente è proprio di questo tipo con $\mathbf{A} = 0$ e $\mathbb{B} = D$. Quando le leggi di un processo di Markov $\mathbf{X}(t)$ soddisfano un'equazione di Fokker-Planck si usa anche dire che $\mathbf{X}(t)$ è un **processo di diffusione**

Proposizione 7.34. *Se la legge di un processo di Markov è soluzione dell'equazione di Fokker-Planck (7.77) con condizione $\mathbf{X}(t) = \mathbf{y}$, \mathbf{P} -qo (cioè con $\mathbf{X}(t) \sim \mathfrak{N}(\mathbf{y}, 0) = \delta_{\mathbf{y}}$), allora dopo un tempo infinitesimo $dt > 0$ la legge del processo evolve in*

$$\mathbf{X}(t + dt) \sim \mathfrak{N}(\mathbf{y} + \mathbf{A}(\mathbf{y}, t)dt, \mathbb{B}(\mathbf{y}, t)dt) \quad (7.78)$$

Dimostrazione: La *ddp* di transizione ritardata del nostro processo $f(\mathbf{x}, t|\mathbf{y}, s)$, con $t > s$ e condizione iniziale $f(\mathbf{x}, s^+|\mathbf{y}, s) = \delta(\mathbf{x} - \mathbf{y})$, è soluzione dell'equazione di Fokker-Planck (7.77) nella forma

$$\partial_t f(\mathbf{x}, t|\mathbf{y}, s) = - \sum_i \partial_{x_i} [A_i(\mathbf{x}, t) f(\mathbf{x}, t|\mathbf{y}, s)] + \frac{1}{2} \sum_{i,j} \partial_{x_i} \partial_{x_j} [B_{ij}(\mathbf{x}, t) f(\mathbf{x}, t|\mathbf{y}, s)]$$

Se l'intervallo di tempo $[s, t]$ è infinitesimo la *ddp* di transizione f al tempo t vicino ad s sarà ancora molto concentrata attorno alla posizione iniziale \mathbf{y} , e quindi – attorno a \mathbf{y} dove è non nulla – presenterà derivate spaziali molto grandi. Potremo allora supporre che le derivate (spaziali) di \mathbf{A} e \mathbb{B} siano trascurabili rispetto a quelle di f , e che in prima approssimazione \mathbf{A} e \mathbb{B} possano essere considerate praticamente

costanti e uguali ai loro valori in \mathbf{y} al tempo s . L'equazione in avanti (7.77) si ridurrà così alla forma semplificata

$$\partial_t f(\mathbf{x}, t | \mathbf{y}, s) = - \sum_i A_i(\mathbf{y}, s) \partial_{x_i} f(\mathbf{x}, t | \mathbf{y}, s) + \frac{1}{2} \sum_{i,j} B_{ij}(\mathbf{y}, s) \partial_{x_i} \partial_{x_j} f(\mathbf{x}, t | \mathbf{y}, s)$$

nella quale \mathbf{A} e \mathbb{B} dipendono solo dalle variabili \mathbf{y}, s non coinvolte nelle operazioni differenziali. Si noti che, nonostante la somiglianza formale, questa equazione differisce dall'equazione all'indietro senza termini di salto (7.73): non solo il segno del termine di diffusione \mathbb{B} è invertito, ma in (7.73) le derivate riguardano \mathbf{y}, s e non \mathbf{x}, t sicché i termini \mathbf{A} e \mathbb{B} non possono essere considerati come termini costanti. La nostra equazione di Fokker-Planck approssimata con coefficienti costanti può essere ora facilmente risolta: ponendo infatti $t - s = dt > 0$ si verifica con un calcolo diretto che la soluzione è

$$f(\mathbf{x}, t | \mathbf{y}, s) = \sqrt{\frac{|\mathbb{B}^{-1}(\mathbf{y}, s)|}{(2\pi)^M dt}} e^{-[\mathbf{x} - \mathbf{y} - \mathbf{A}(\mathbf{y}, s)dt] \cdot \mathbb{B}^{-1}(\mathbf{y}, s) [\mathbf{x} - \mathbf{y} - \mathbf{A}(\mathbf{y}, s)dt] / 2dt}$$

dove $|\mathbb{B}^{-1}|$ indica il determinante della matrice \mathbb{B}^{-1} . Se ora in questa espressione, per adeguare le notazioni all'enunciato della proposizione, eseguiamo le sostituzioni $s \rightarrow t$, e $t \rightarrow t + dt$, potremo affermare che, partendo da $\mathbf{X}(t) = \mathbf{y}$, la soluzione di (7.77) in un intervallo infinitesimo $[t, t + dt]$ evolve proprio verso la legge annunciata nella relazione (7.78). In particolare ciò mostra esplicitamente sia il ruolo di velocità di trascinamento di \mathbf{A} , sia quello di coefficiente di diffusione di \mathbb{B} e ne giustifica pertanto i nomi. \square

Processo degenero: equazione di Liouville

Consideriamo infine $\ell = \mathbb{B} = 0$ e solo $\mathbf{A} \neq 0$, caso in cui l'equazione in avanti diviene

$$\partial_t f(\mathbf{x}, t) = - \sum_i \partial_i [A_i(\mathbf{x}, t) f(\mathbf{x}, t)] \tag{7.79}$$

e prende il nome di **equazione di Liouville**. Si tratta di un'equazione differenziale che governa l'evoluzione di una ddp senza né salti né diffusioni, e quindi è l'equazione di un caso **degenero**: mostreremo infatti che la soluzione con condizione iniziale $f(\mathbf{x}, s^+) = \delta(\mathbf{x} - \mathbf{y})$ evolve mantenendo sempre inalterata la sua concentrazione in un punto che segue una traiettoria deterministica.

Proposizione 7.35. *Dato un sistema dinamico $\mathbf{x}(t)$ governato dall'equazione*

$$\dot{\mathbf{x}}(t) = \mathbf{A}[\mathbf{x}(t), t] \quad t \geq s \tag{7.80}$$

e detta $\mathbf{x}(t|\mathbf{y}, s)$ la soluzione che corrisponde alla condizione iniziale $\mathbf{x}(s) = \mathbf{y}$, la soluzione dell'equazione di Liouville (7.79) corrispondente alla condizione iniziale $f(\mathbf{x}, s^+ | \mathbf{y}, s) = \delta(\mathbf{x} - \mathbf{y})$ è

$$f(\mathbf{x}, t | \mathbf{y}, s) = \delta[\mathbf{x} - \mathbf{x}(t|\mathbf{y}, s)] \tag{7.81}$$

cioè essa resta sempre concentrata attorno al punto $\mathbf{x}(t|\mathbf{y}, s)$ seguendone fedelmente la traiettoria senza diffondere

Dimostrazione: Per ottenere questo risultato richiameremo la seguente proprietà delle distribuzioni δ

$$\partial_t \delta[\mathbf{x} - \mathbf{g}(t)] = - \sum_i \dot{g}_i(t) \partial_i \delta[\mathbf{x} - \mathbf{g}(t)]$$

Infatti, limitandoci per semplicità al caso unidimensionale, per una generica funzione di prova $\varphi(x)$ si ha

$$\begin{aligned} \int \varphi(x) \partial_t \delta[x - g(t)] dx &= \partial_t \int \varphi(x) \delta[x - g(t)] dx \\ &= \partial_t [\varphi(g(t))] = \dot{g}(t) \varphi'(g(t)) \\ - \int \varphi(x) \dot{g}(t) \partial_x \delta[x - g(t)] dx &= \dot{g}(t) \int \varphi'(x) \delta[x - g(t)] dx = \dot{g}(t) \varphi'(g(t)) \end{aligned}$$

Tenendo allora conto di (7.80) si ha

$$\begin{aligned} - \sum_i \partial_i [A_i(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}(t|\mathbf{y}, s))] &= - \sum_i \partial_i [A_i(\mathbf{x}(t|\mathbf{y}, s), t) \delta(\mathbf{x} - \mathbf{x}(t|\mathbf{y}, s))] \\ &= - \sum_i A_i(\mathbf{x}(t|\mathbf{y}, s), t) \partial_i \delta(\mathbf{x} - \mathbf{x}(t|\mathbf{y}, s)) \\ &= - \sum_i \dot{x}_i(t|\mathbf{y}, s) \partial_i \delta(\mathbf{x} - \mathbf{x}(t|\mathbf{y}, s)) \\ &= \partial_t \delta(\mathbf{x} - \mathbf{x}(t|\mathbf{y}, s)) \end{aligned}$$

e questo prova che la (7.81) è soluzione dell'equazione di Liouville (7.79) □

7.2.4 Equazioni di processi particolari

Esamineremo ora alcuni esempi classici di equazioni *in avanti*: per semplicità e per riconnettersi alla discussione dei capitoli precedenti, in questa sezione ci limiteremo a prendere in considerazione solo processi con una componente per i quali l'equazione (7.66) del caso *ac* con *ddp* f si riduce a

$$\begin{aligned} \partial_t f(x, t) &= -\partial_x [A(x, t) f(x, t)] + \frac{1}{2} \partial_x^2 [B(x, t) f(x, t)] \\ &\quad + \int_{z \neq x} [\ell(x|z, t) f(z, t) - \ell(z|x, t) f(x, t)] dz \end{aligned} \quad (7.82)$$

mentre la master equation (7.71) di un processo a valori interi diventa

$$\partial_t p(n, t) = \sum_k [\ell(n|k, t) p(k, t) - \ell(k|n, t) p(n, t)] \quad (7.83)$$

Nel seguito determineremo esplicitamente i coefficienti A , B e ℓ delle equazioni *in avanti* a partire dalle leggi di transizione Markoviane che individuano le diverse famiglie di processi introdotti nelle sezioni precedenti, e daremo anche qualche indicazione sui metodi di risoluzione

Processo di Poisson

Proposizione 7.36. *Le leggi dei processi di Poisson con probabilità di transizione (7.46) soddisfano la master equation*

$$\partial_t p(n, t) = -\lambda [p(n, t) - p(n-1, t)] \quad (7.84)$$

Dimostrazione: Trattandosi di processi di conteggio con valori interi, le traiettorie saranno tipicamente discontinue per cui – in base alle osservazioni conclusive della Sezione 7.2.1 – avremo innanzitutto $A = B = 0$, sicché l'equazione di evoluzione in avanti sarà una master equation di tipo (7.83). Per calcolare $\ell(n|m, t)$ da (7.70) consideriamo allora la probabilità di transizione (7.46) con $n \neq m$: al limite per $\Delta t \rightarrow 0$ avremo

$$\frac{1}{\Delta t} p_N(n, t + \Delta t | m, t) = \begin{cases} \lambda e^{-\lambda \Delta t} \rightarrow \lambda & \text{se } n = m + 1 \\ O(\Delta t^{n-m}) \rightarrow 0 & \text{se } n \geq m + 2 \end{cases}$$

Complessivamente potremo quindi dire che per un processo di Poisson

$$\ell(n|m, t) = \lambda \delta_{n, m+1} \quad (7.85)$$

e quindi che in generale le distribuzioni soddisferanno la master equation (7.84) e al variare delle condizioni iniziali esauriranno tutta la famiglia dei processi di Poisson della Definizione 7.20. In particolare la probabilità di transizione (7.46) corrisponderà alla condizione iniziale degenere $p(n, t) = \delta_{nm}$ \square

Se viceversa una master equation viene assegnata, il nostro problema sarà quello di risolverla con una condizione iniziale per determinare la legge del corrispondente processo di Markov discreto $N(t)$. Il metodo più noto per farlo è quello della **funzione generatrice**: ad esempio, nel caso più semplice della master equation (7.84) con condizione iniziale degenere $p(n, 0) = \delta_{n0}$, il metodo consiste nel verificare, con un uso diretto della (7.84), che la funzione generatrice della legge di $N(t)$, definita come

$$\gamma(u, t) = \mathbf{E} [u^{N(t)}] = \sum_n u^n p(n, t) \quad (7.86)$$

soddisfa a sua volta l'equazione

$$\partial_t \gamma(u, t) = \lambda(u-1)\gamma(u, t) \quad \gamma(u, 0) = 1$$

Si vede allora facilmente che la soluzione è

$$\gamma(u, t) = e^{\lambda(u-1)t}$$

e, confrontando lo sviluppo in serie di Taylor di quest'ultima attorno a $u = 0$

$$\gamma(u, t) = e^{-\lambda t} \sum_n u^n \frac{(\lambda t)^n}{n!}$$

con la definizione (7.86), si ottiene in definitiva

$$p(n, t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

Questo risultato ovviamente conferma che la soluzione della master equation (7.84) con condizione iniziale degenere è proprio la legge di un processo di Poisson standard

Processo di Wiener

Proposizione 7.37. *Le leggi dei processi di Wiener con ddp di transizione (7.48) soddisfano l'equazione di Fokker-Planck*

$$\partial_t f(x, t) = \frac{D}{2} \partial_x^2 f(x, t) \quad (7.87)$$

Dimostrazione: Innanzitutto, usando la regola di l'Hôpital, si ottiene facilmente da (7.48) che per $x \neq y$ e $\Delta t \rightarrow 0$

$$\frac{1}{\Delta t} f_W(x, t + \Delta t | y, t) = \frac{e^{-(x-y)^2/2D\Delta t}}{\Delta t \sqrt{2\pi D\Delta t}} \rightarrow 0$$

cioè $\ell(x|y, t) = 0$, coerentemente con il risultato della Proposizione 7.23 secondo il quale un processo di Wiener è sempre (globalmente) continuo. Inoltre sarà $A = 0$ per simmetria, mentre per il coefficiente di diffusione B , ponendo $y = (x-z)/\sqrt{D\Delta t}$, avremo

$$\frac{1}{\Delta t} \int_{z-\epsilon}^{z+\epsilon} (x-z)^2 f_W(x, t + \Delta t | z, t) dx = D \int_{-\epsilon/\sqrt{D\Delta t}}^{+\epsilon/\sqrt{D\Delta t}} \frac{y^2 e^{-y^2/2}}{\sqrt{2\pi}} dy \xrightarrow{\Delta t \rightarrow 0} D$$

Raccogliendo tutte queste osservazioni avremo in definitiva che la ddp di un processo di Wiener soddisferà l'equazione di Fokker-Planck (7.87) – della forma già incontrata in (6.54) – e che la ddp di transizione (7.48) sarà la particolare soluzione con condizione iniziale $f(x, t^+) = \delta(x - y)$ \square

Se invece viene assegnata un'equazione di Fokker-Planck con condizione iniziale degenere

$$\partial_t f(x, t) = \frac{D}{2} \partial_x^2 f(x, t) \quad f(x, s^+) = \delta(x - y) \quad (7.88)$$

potremo determinare la ddp di transizione del processo di Wiener associato. Il metodo di risoluzione più noto è quello della **trasformata di Fourier**, ovvero la fc dell'incognita

$$\varphi(u, t) = \int_{-\infty}^{+\infty} e^{iux} f(x, t) dx$$

per la quale si verifica che essa soddisfa l'equazione

$$\partial_t \varphi(u, t) = -\frac{Du^2}{2} \varphi(u, t) \quad \varphi(u, s) = e^{iuy}$$

la cui soluzione banalmente è

$$\varphi(u, t) = e^{iuy} e^{-Du^2(t-s)/2}$$

L'inversione di questa fc è immediata e fornisce un risultato coerente con la ddp di Wiener (7.48)

$$f(x, t | y, s) = \frac{e^{-(x-y)^2/2D(t-s)}}{\sqrt{2\pi D(t-s)}}$$

Processo di Cauchy

Proposizione 7.38. *Le leggi dei processi di Cauchy con ddp di transizione (7.52) soddisfano la master equation*

$$\partial_t f(x, t) = \frac{a}{\pi} \int_{z \neq x} \frac{f(z, t) - f(x, t)}{(x-z)^2} dz \quad (7.89)$$

Dimostrazione: Infatti innanzitutto abbiamo

$$\frac{1}{\Delta t} f_X(x, t + \Delta t | y, t) = \frac{a}{\pi} \frac{1}{(x-y)^2 + (a\Delta t)^2} \xrightarrow{\Delta t \rightarrow 0} \ell(x|y, t) = \frac{a}{\pi(x-y)^2}$$

per cui le traiettorie del processo presenteranno sicuramente delle discontinuità. Inoltre risulta $A = 0$ per simmetria, mentre per il coefficiente di diffusione si ha con $y = (x-z)/a\Delta t$

$$\begin{aligned} \frac{1}{\Delta t} \int_{z-\epsilon}^{z+\epsilon} (x-z)^2 f_X(x, t + \Delta t | z, t) dx &= \frac{a^2 \Delta t}{\pi} \int_{-\epsilon/a\Delta t}^{+\epsilon/a\Delta t} \frac{y^2}{1+y^2} dy \\ &= \frac{2a^2 \Delta t}{\pi} \left(\frac{\epsilon}{a\Delta t} - \arctan \frac{\epsilon}{a\Delta t} \right) \xrightarrow{\Delta t \rightarrow 0} \frac{2a\epsilon}{\pi} \end{aligned}$$

e quindi per $\epsilon \rightarrow 0$ si ottiene in definitiva $B = 0$. Il processo di Cauchy è quindi un processo di puro salto e l'equazione soddisfatta dalla sua ddp è la master equation (7.89). Le ddp di transizione (7.52) sono le soluzioni individuate dalla solita condizione iniziale degenera $f(x, t) = \delta(x-y)$ \square

Diversamente dai due casi precedenti questa equazione non era già stata introdotta sulla base di considerazioni euristiche: essa infatti è derivabile solo nel quadro della discussione sui processi di Markov. Naturalmente anche in questo caso è possibile assumere un punto di vista opposto, cioè quello di ricostruire la legge del processo risolvendo l'equazione in avanti (7.89) con condizione iniziale $f(x, s^+) = \delta(x-y)$ per determinare prima la ddp di transizione e poi tutte le altre leggi; questa volta però, trattandosi di un'equazione integro-differenziale, non ci sono procedure elementari per determinare la soluzione.

Processo di Ornstein-Uhlenbeck

Proposizione 7.39. *Le leggi di un processo di Ornstein-Uhlenbeck con ddp di transizione (7.54) soddisfano l'equazione di Fokker-Planck*

$$\partial_t f(x, t) = \alpha \partial_x [x f(x, t)] + \frac{D}{2} \partial_x^2 f(x, t) \quad (7.90)$$

dove abbiamo posto $D = 2\alpha\beta^2$, e e pertanto il processo è globalmente continuo

Dimostrazione: Vedi Appendice K. Noteremo soltanto che la continuità globale del processo – lasciata in sospeso nella discussione della Proposizione 7.29 – deriva dall'osservazione che per un processo di Ornstein-Uhlenbeck il coefficiente di salto ℓ si annulla, per cui le condizioni di continuità di Lindeberg sono soddisfatte \square

Anche in questo caso i metodi per risolvere l'equazione (7.90) sono meno elementari di quelli usati in queste lezioni e quindi trascureremo di esporli, limitandoci invece ad osservare che con poco sforzo si può verificare che la ddp di transizione (7.54) è soluzione della nostra equazione per la condizione iniziale degenera $f(x, t) = \delta(x - y)$

Capitolo 8

Elementi di calcolo stocastico

8.1 Rumore bianco Wieneriano

Anche in questo capitolo, per ragioni di semplicità, prenderemo prevalentemente in considerazione processi con una sola componente. Abbiamo già osservato nella Sezione 6.3 che il *rumore bianco* è un processo singolare le cui caratteristiche dipendono dalla non derivabilità di alcuni processi. Come primo esempio abbiamo infatti mostrato esplicitamente che il processo degli impulsi di Poisson (6.64) e il relativo processo compensato (6.66) sono rumori bianchi prodotti dalla derivazione formale rispettivamente di un processo di Poisson $N(t)$ e della sua versione compensata $\tilde{N}(t)$. Allo stesso modo abbiamo fatto vedere che anche la derivata formale del processo di Wiener $W(t)$ – non derivabile in base alla Proposizione 6.18 – gode delle proprietà (6.70) di un rumore bianco, e nell'Appendice H abbiamo suggerito che tale rumore bianco $\dot{W}(t)$ giochi il ruolo di forza fluttuante $B(t)$ nell'equazione di Langevin (6.78) del moto Browniano. Dopo l'introduzione dei processi di Markov possiamo ora fornire una argomentazione matematicamente più convincente per questa ultima identificazione.

L'equazione di Langevin (6.78) si presenta come un caso particolare dell'equazione

$$\dot{X}(t) = a(X(t), t) + b(X(t), t) Z(t) \quad (8.1)$$

dove $a(x, t)$ e $b(x, t)$ sono funzioni assegnate e $Z(t)$ è un processo con $\mathbf{E}[Z(t)] = 0$ non correlato con $X(t)$. Da una integrazione formale di (8.1)

$$X(t) = X(t_0) + \int_{t_0}^t a(X(s), s) ds + \int_{t_0}^t b(X(s), s) Z(s) ds$$

si vede però che $X(t)$ risulterà costruito a partire dai valori di $Z(s)$ con $t_0 < s < t$, e pertanto è intuitivo che per garantire la non correlazione fra $X(t)$ e $Z(t)$ bisognerà richiedere anche la non correlazione di $Z(s)$ e $Z(t)$ per $s \neq t$. Inoltre la supposta irregolarità di $Z(t)$ suggerisce che la sua varianza (che nel nostro caso è semplicemente $\mathbf{E}[Z^2(t)]$) presenti un comportamento singolare, e quindi in definitiva, per

qualche costante $D > 0$, sarà piuttosto naturale supporre che

$$\mathbf{E}[Z(t)Z(s)] = D\delta(t - s)$$

cioè che $Z(t)$ sia un rumore bianco stazionario con media nulla e intensità D . Nel seguito supporremo inoltre che $Z(s)$ e $Z(t)$ siano addirittura indipendenti¹ per $s \neq t$, e che – se l’equazione (8.1) è destinata a descrivere fenomeni fisici analoghi al moto Browniano – tutti i processi coinvolti siano continui in senso globale. Mostriamo ora che sotto queste ipotesi il rumore bianco $Z(t)$ non può che essere un rumore bianco Wieneriano $\dot{W}(t)$

Proposizione 8.1. *Se $Z(t)$ è un rumore bianco stazionario di intensità $D > 0$ con $Z(s)$ indipendente da $Z(t)$ per $s \neq t$, allora il processo*

$$W(t) = \int_{t_0}^t Z(s) ds \tag{8.2}$$

se è globalmente continuo, è un processo di Wiener con coefficiente di diffusione D

Dimostrazione: Per dimostrare la tesi faremo vedere che le leggi del processo $W(t)$ definito in (8.2) soddisfano l’equazione di Fokker-Planck (7.87) di un processo di Wiener. Gli incrementi di $W(t)$ in intervalli non sovrapposti $t_1 < t_2 \leq t_3 < t_4$ hanno la forma

$$W(t_2) - W(t_1) = \int_{t_1}^{t_2} Z(s) ds \qquad W(t_4) - W(t_3) = \int_{t_3}^{t_4} Z(s) ds$$

cioè sono somme di v -a $Z(s)$ indipendenti per ipotesi, e sono quindi a loro volta indipendenti. Segue allora dalla Proposizione 7.9 che $W(t)$ è innanzitutto un processo di Markov, ed essendo continuo per ipotesi le sue leggi soddisferanno un’equazione di Fokker-Planck (con $\ell = 0$) del genere discusso nella Sezione 7.2.3. Per scoprire allora la natura del processo $W(t)$ basterà calcolare i coefficienti (7.63) e (7.65) dell’equazione i quali per il nostro caso con una sola componente si riducono a

$$\begin{aligned} A(x, t) &= \lim_{\epsilon \rightarrow 0^+} \lim_{\Delta t \rightarrow 0} \int_{|y-x| < \epsilon} \frac{y-x}{\Delta t} f(y, t + \Delta t | x, t) dy \\ B(x, t) &= \lim_{\epsilon \rightarrow 0^+} \lim_{\Delta t \rightarrow 0} \int_{|y-x| < \epsilon} \frac{(y-x)^2}{\Delta t} f(y, t + \Delta t | x, t) dy \end{aligned}$$

D’altra parte a causa della continuità sappiamo dalle condizioni di Lindeberg (7.44) che comunque scelto $\epsilon > 0$ risulterà

$$\lim_{\Delta t \rightarrow 0} \int_{|y-x| > \epsilon} \frac{1}{\Delta t} f(y, t + \Delta t | x, t) dy = 0$$

¹Si noti che questa ipotesi non è molto restrittiva: in vista del fatto che i processi in questione si riveleranno essere Gaussiani, indipendenza e non correlazione sono proprietà equivalenti

sicch , per $\Delta t \rightarrow 0$, il supporto di $f(y, t + \Delta t | x, t)$ tender  a stare tutto all'interno dell'intervallo $[x - \epsilon, x + \epsilon]$. Conseguentemente le definizioni di A e B possono essere semplificate senza modificarne il valore estendendo gli integrali a tutto $(-\infty, +\infty)$: si ottiene cos 

$$\begin{aligned} A(x, t) &= \lim_{\Delta t \rightarrow 0} \int_{-\infty}^{+\infty} \frac{y-x}{\Delta t} f(y, t + \Delta t | x, t) dy \\ &= \lim_{\Delta t \rightarrow 0} \mathbf{E} \left[\frac{\Delta W(t)}{\Delta t} \mid W(t) = x \right] \\ B(x, t) &= \lim_{\Delta t \rightarrow 0} \int_{-\infty}^{+\infty} \frac{(y-x)^2}{\Delta t} f(y, t + \Delta t | x, t) dy \\ &= \lim_{\Delta t \rightarrow 0} \mathbf{E} \left[\frac{[\Delta W(t)]^2}{\Delta t} \mid W(t) = x \right] \end{aligned}$$

Ma nel nostro caso sappiamo dalle propriet  di $Z(t)$ che

$$\begin{aligned} \mathbf{E} [\Delta W(t) | W(t) = x] &= \mathbf{E} \left[\int_t^{t+\Delta t} Z(s) ds \mid W(t) = x \right] \\ &= \int_t^{t+\Delta t} \mathbf{E} [Z(s)] ds = 0 \\ \mathbf{E} [[\Delta W(t)]^2 | W(t) = x] &= \mathbf{E} \left[\int_t^{t+\Delta t} Z(s) ds \int_t^{t+\Delta t} Z(s') ds' \mid W(t) = x \right] \\ &= \int_t^{t+\Delta t} ds \int_t^{t+\Delta t} ds' \mathbf{E} [Z(s)Z(s')] \\ &= D \int_t^{t+\Delta t} ds \int_t^{t+\Delta t} ds' \delta(s-s') \\ &= D \int_t^{t+\Delta t} ds = D\Delta t \end{aligned}$$

e quindi in definitiva si ottengono i coefficienti

$$A(x, t) = 0 \quad B(x, t) = D$$

che sono quelli dell'equazione di Fokker-Planck (7.87) per il processo di Wiener \square

Dalla proposizione precedente si ricava dunque che $W(t)$ definito da (8.2)   un processo di Wiener, e che la sua derivata formale $Z(t) = \dot{W}(t)$   un rumore bianco Wieneriano. Tale rumore bianco gioca il ruolo di forza aleatoria nell'equazione di Langevin, ma strettamente parlando non esiste per cui l'equazione (8.1) non pu  ancora essere considerata come ben definita nella forma fin qui presentata. Ricorderemo per  di aver osservato alla fine della Sezione 6.3 che, bench  la derivata di un processo di Wiener $W(t)$ non esista, sembra invece avere un senso pi  preciso

il concetto del suo differenziale $dW(t)$ inteso come limite infinitesimo per $\Delta t \rightarrow 0$ dell'incremento $\Delta W(t) = W(t + \Delta t) - W(t)$, per il quale vale ovviamente la relazione

$$\int_{t_0}^t dW(s) = W(t) - W(t_0)$$

Possiamo quindi pensare di riscrivere l'equazione (8.1) in termini di differenziali in modo che – avendo cura di sostituire sostituire $dW(t)$ alle espressioni singolari $Z(t)dt = \dot{W}(t)dt$ – essa assuma la forma

$$dX(t) = a(X(t), t)dt + b(X(t), t)dW(t)$$

il cui significato più preciso è

$$X(t) = X(t_0) + \int_{t_0}^t a(X(s), s) ds + \int_{t_0}^t b(X(s), s) dW(s)$$

Bisogna però notare subito che, mentre il primo di questi due integrali

$$\int_{t_0}^t a(X(s), s) ds$$

può essere adeguatamente definito sulla base delle considerazioni già svolte nella Sezione 5.4, resta invece aperto il problema del significato da attribuire al secondo integrale

$$\int_{t_0}^t b(X(s), s) dW(s) \tag{8.3}$$

nel quale la misura è definita a partire da un processo di Wiener: un caso non contemplato nelle nostre precedenti discussioni. La definizione di questo nuovo tipo di integrali costituirà l'argomento della prossima sezione e caratterizzerà la struttura del calcolo stocastico.

8.2 Integrali stocastici

Ci sono diverse forme di integrali stocastici che risultano ben definiti sotto varie condizioni: in ogni caso, quando esistono, essi sono sempre delle *v-a*. Noi abbiamo già discusso nella Proposizione 5.8 la convergenza in *mq* del caso più semplice di integrale stocastico (5.8) che può ora essere agevolmente generalizzato nella forma

$$\int_a^b Y(t) dx(t)$$

dove $Y(t)$ è un processo mentre $x(t)$ è una funzione che in generale viene supposta *a variazione limitata*². Quest'ultima ipotesi, sotto condizioni molto generali per

²Una funzione $\Phi(x)$ definita in $[a, b]$ si dice *a variazione limitata* se esiste $C > 0$ tale che

$$\sum_{k=1}^n |\Phi(x_k) - \Phi(x_{k-1})| < C$$

il processo $Y(t)$, consente di dimostrare³ che questo integrale, non solo esiste in mq , ma converge anche nel senso di Lebesgue–Stieltjes per quasi ogni traiettoria di $Y(t)$. In pratica l'integrale è ben definito, in un senso tradizionale, traiettoria per traiettoria. Il problema però si complica quando è l'integratore ad essere un processo stocastico $X(t)$ perché in questo caso non si può più supporre che le sue traiettorie siano a variazione limitata e quindi i procedimenti tradizionali non consentono più di garantire coerentemente la convergenza dell'integrale. Il caso tipico con il quale avremo a che fare nel resto di queste lezioni è quello in cui l'integratore è proprio il processo di Wiener $W(t)$ le cui traiettorie – non essendo derivabili in nessun punto – non sono a variazione limitata

8.2.1 Integrale di Wiener

Prendiamo prima di tutto in considerazione un integrale del tipo

$$\int_a^b y(t) dX(t) \quad (8.4)$$

dove ora $y(t)$ è una funzione non aleatoria mentre $X(t)$ è un processo: come già osservato, la definizione di (8.4) traiettoria per traiettoria con la procedura di Lebesgue–Stieltjes non può più essere adottata perché in questo caso le traiettorie del processo $X(t)$ non sono in generale funzioni a variazione limitata: come nel caso del processo di Wiener, infatti, dovremo supporre ora che in generale le traiettorie non siano derivabili, e quindi non siano a variazione limitata (vedi nota 2 in questa sezione). La forma più comune di integrali di questo genere è quella in cui l'integratore è un processo di Wiener

$$\int_a^b y(t) dW(t) \quad (8.5)$$

e in questo caso si parla di **integrale di Wiener**. Restando comunque nel caso più generale (8.4), questo integrale può essere definito rigorosamente per

per ogni decomposizione finita $a = x_0 < x_1 < \dots < x_n = b$, e in tal caso la quantità

$$V[\Phi] = \sup_{\mathcal{D}} \sum_{k=1}^n |\Phi(x_k) - \Phi(x_{k-1})|$$

dove \mathcal{D} è l'insieme delle decomposizioni finite di $[a, b]$, si chiama variazione totale di Φ . È noto che l'integrale di Lebesgue–Stieltjes

$$\int_a^b f(x) d\Phi(x)$$

può essere correttamente definito quando $\Phi(x)$ è una funzione a variazione limitata. Si noti infine che ogni funzione a variazione limitata è quasi ovunque derivabile. Per maggiori dettagli si veda **A.N. Kolmogorov, S.V. Fomin**, ELEMENTI DI TEORIA DELLE FUNZIONI E DI ANALISI FUNZIONALE, Mir (Mosca, 1980)

³**J.L. Doob**, STOCHASTIC PROCESSES, Wiley (New York, 1953)

- processi $X(t)$ con incrementi non correlati (ad esempio il processo di Wiener ha incrementi indipendenti e quindi anche non correlati)
- funzioni $y(t)$ di quadrato integrabile secondo Lebesgue

adottando una procedura che qui riportiamo in maniera semplificata⁴:

1. si definisce prima in maniera elementare un integrale per funzioni $\varphi(t)$ a gradino

$$\int_a^b \varphi(t) dX(t)$$

2. si considera poi una successione di funzioni a gradino $\varphi_n(t)$ che approssimi $y(t)$ in media quadratica (si dimostra che tale successione esiste e che la particolare scelta della successione è irrilevante)
3. si definisce infine l'integrale (8.4) come limite in media quadratica della successione di $v-a$

$$\int_a^b \varphi_n(t) dX(t)$$

Si dimostra che questa procedura conduce a una definizione perfettamente coerente, e inoltre si fa vedere che, se la funzione $y(t)$ è continua, l'integrale (8.4) può anche essere calcolato con un'altra procedura del tutto analoga a quella usata per definire il tradizionale integrale di Riemann:

1. si considera un'arbitraria partizione $a = t_0 < t_1 < \dots < t_n = b$ dell'intervallo di integrazione
2. si scelgono arbitrariamente dei punti τ_j in $[t_j, t_{j+1}]$ e si pone

$$\delta = \max_j \{t_{j+1} - t_j\}$$

3. si calcola l'integrale come il limite in media quadratica

$$\lim_{n, \delta \rightarrow 0} \text{-mq} \sum_{j=0}^{n-1} y(\tau_j) [X(t_{j+1}) - X(t_j)]$$

Quando l'integrale esiste, la seconda procedura conduce a un risultato identico a quello definito nella prima, e questo accade indipendentemente sia dalla successione di partizioni considerata, sia dalla scelta dei punti τ_j in $[t_j, t_{j+1}]$. In particolare in questo modo viene dato un significato preciso agli integrali di Wiener (8.5)

⁴**J.L. Doob**, STOCHASTIC PROCESSES, Wiley (New York, 1953)

8.2.2 Integrale di Itô

L'integrale (8.4) è un caso particolare dell'integrale di tipo più generale

$$\int_a^b Y(t) dX(t) \quad (8.6)$$

nel quale $X(t)$ e $Y(t)$ sono ambedue processi: l'integrale (8.3) nel quale ci siamo imbattuti alla fine della sezione precedente è proprio un esempio di questo genere. Una definizione rigorosa di (8.6) non è però semplice⁵ e richiede una nuova procedura introdotta da K. Itô (1944) per il caso in cui l'integratore è il processo di Wiener

$$\int_a^b Y(t) dW(t) \quad (8.7)$$

e poi estesa ad una classe di integratori $X(t)$ un po' più ristretta di quella degli integrali della Sezione 8.2.1. La nuova definizione richiede inoltre che il processo integrando $Y(t)$ soddisfi delle condizioni piuttosto generali, la più importante delle quali per il caso degli integrali di Itô di tipo (8.7) è la *non-anticipatività* rispetto al processo $W(t)$

Definizione 8.2. Dato un processo di Wiener $W(t)$ e detta $\mathcal{F}_t = \sigma\{W(s), s \leq t\}$ la famiglia crescente delle σ -algebre da esso generate (filtrazione naturale), diremo che un processo $Y(t)$ è **non-anticipativo** rispetto a $W(t)$ se

- $Y(t)$ è \mathcal{F}_t -misurabile,
- $Y(t)$ non dipende da $W(s) - W(t)$ con $s > t$.

cioè se $Y(t)$ dipende dal passato (e dal presente) di $W(t)$, ma non dal suo futuro.

Questo concetto, che esprime un requisito di causalità piuttosto naturale, è essenziale per una definizione rigorosa di integrale – e quindi di equazione differenziale stocastica – di Itô, nel senso che molti risultati possono essere dedotti solo con questa ipotesi. In particolare si verifica che tutti i seguenti processi sono non-anticipativi:

$$W(t) \quad \int_{t_0}^t h[W(s)] ds \quad \int_{t_0}^t h[W(s)] dW(s)$$

e che anche i processi

$$\int_{t_0}^t Y(s) ds \quad \int_{t_0}^t Y(s) dW(s)$$

sono non-anticipativi se $Y(t)$ lo è.

Nel seguito quindi supporremo sempre, tra le altre ipotesi, che i nostri integrandi $Y(t)$ siano non-anticipativi rispetto a $W(t)$, e all'interno di queste ipotesi adotteremo una procedura per la definizione rigorosa dell'**integrale di Itô** che per molti versi è analoga a quella dell'integrale di Wiener discussa nella Sezione 8.2.1:

⁵I. Karatzas, S.E. Shreve, BROWNIAN MOTION AND STOCHASTIC CALCULUS, Springer (Berlin, 1991). B. Øksendal, STOCHASTIC DIFFERENTIAL EQUATIONS, Springer (Berlin, 2005)

1. si definisce prima in maniera elementare un integrale di Itô per *funzioni a gradino* $\Phi(\omega; t)$ (cioè per dei particolari processi stocastici non-anticipativi):

$$\int_a^b \Phi(\omega; t) dW(\omega; t)$$

2. si considera poi una successione di funzioni a gradino $\Phi_n(t)$ che approssimi un assegnato processo non-anticipativo $Y(t)$ in media quadratica (si dimostra che tale successione esiste e che la particolare scelta della successione è irrilevante)
3. si definisce infine l'integrale di Itô (8.7) come limite in media quadratica della successione di v -a

$$\int_a^b \Phi_n(t) dW(t)$$

Naturalmente anche in questo caso si dimostra che il limite non dipende dalla particolare successione $\Phi_n(t)$ scelta in modo che la definizione risulti perfettamente coerente, ma la nuova procedura, nonostante le evidenti analogie con quella che definisce l'integrale di Wiener, prevede l'introduzione due importanti novità:

- come nel caso dell'integrale di Wiener (8.5), l'integrale di Itô può essere calcolato anche adottando una procedura alla Riemann del tipo

$$\lim_{n, \delta \rightarrow 0} \text{-mq} \sum_{j=0}^{n-1} Y(t_j) [W(t_{j+1}) - W(t_j)] \quad (8.8)$$

ma in questo caso i valori $Y(t_j)$ dell'integrando *devono* sempre essere calcolati nell'estremo sinistro dell'intervallo $[t_j, t_{j+1}]$, e non in un generico τ_j ad esso interno: infatti si potrebbe mostrare che ora il valore del limite (8.8) *dipende* da questa scelta, e si prova che solo l'uso di $Y(t_j)$ fornisce come risultato il corretto valore dell'integrale di Itô definito precedentemente con la procedura di Itô; illustreremo più oltre questo comportamento elaborando un esempio esplicito

- la definizione di integrale di Itô non arriva senza un costo aggiuntivo: essa ha come conseguenza anche l'introduzione di un nuovo *calcolo stocastico* con regole diverse da quelle dell'ordinario calcolo integrale e differenziale alle quali siamo abituati; un anticipo di questa importante novità – alla quale bisognerà imparare ad adattarsi per poter fare tesoro delle proprietà del nuovo integrale – si trova già nella discussione sui possibili errori dovuti a un uso improprio del calcolo differenziale nella Appendice H: noi comunque dedicheremo buona parte delle prossime sezioni proprio a un'analisi dettagliata di queste nuove regole

Qualche osservazione relativa a possibili definizioni alternative di integrali stocastici, come l'*integrale di Stratonovich* che permetterebbe di conservare le regole tradizionali del calcolo differenziale, è infine riportata nell'Appendice L

8.3 Calcolo stocastico di Itô

Nel seguito calcoleremo tutti gli integrali di Itô (8.7) con la procedura di Riemann (8.8) per un processo di Wiener $W(t)$ con coefficiente di diffusione D , ma con condizioni iniziali non standard $W(t_0) = w_0$, \mathbf{P} -qo. Per questa ragione i risultati delle Proposizioni 6.16, 6.17 e 6.18 vanno adeguati osservando che, se ora indichiamo con $W_0(t)$ il processo standard con condizioni $W_0(0) = 0$, \mathbf{P} -qo, ovviamente avremo $W(t) = W_0(t - t_0) + w_0$ per il processo non standard definito per $t \geq t_0$. In particolare si verifica subito allora che $W(t) \sim \mathfrak{N}(w_0, D(t - t_0))$, e inoltre

$$\mathbf{E}[W(t)] = w_0 \quad \mathbf{V}[W(t)] = D(t - t_0) \quad (8.9)$$

$$\mathbf{E}[W(s)W(t)] = D \min\{s - t_0, t - t_0\} + w_0^2 \quad (8.10)$$

Per semplicità, per ogni partizione $t_1 < \dots < t_n$ di una procedura di Riemann, adotteremo anche le notazioni sintetiche con $j = 1, \dots, n$

$$W_j = W(t_j) \quad \Delta W_j = W_j - W_{j-1} \quad \Delta t_j = t_j - t_{j-1}$$

8.3.1 Regole elementari di integrazione

Lemma 8.3. *Se $W(t)$ è un processo di Wiener con $W(t_0) = w_0$, \mathbf{P} -qo, allora*

$$W_j \sim \mathfrak{N}(w_0, D(t_j - t_0)) \quad \Delta W_j \sim \mathfrak{N}(0, D\Delta t_j) \quad (8.11)$$

$$\mathbf{E}[(\Delta W_j)^4] = 3D^2(\Delta t_j)^2 \quad (8.12)$$

Dimostrazione: La prima relazione (8.11) deriva immediatamente dal fatto già osservato che $W(t) \sim \mathfrak{N}(w_0, D(t - t_0))$ per ogni t . Quanto alla seconda di (8.11), essendo gli incrementi $\Delta W(t)$ comunque Gaussiani secondo la Proposizione 6.16, e tenendo conto di (8.9) e (8.10), basterà osservare che

$$\begin{aligned} \mathbf{E}[\Delta W_j] &= \mathbf{E}[W_j - W_{j-1}] = w_0 - w_0 = 0 \\ \mathbf{V}[\Delta W_j] &= \mathbf{E}[(\Delta W_j)^2] = \mathbf{E}[W_j^2 + W_{j-1}^2 - 2W_j W_{j-1}] \\ &= w_0^2 + D(t_j - t_0) + w_0^2 + D(t_{j-1} - t_0) - 2[w_0^2 + D(t_{j-1} - t_0)] \\ &= D(t_j - t_{j-1}) = D\Delta t_j \end{aligned}$$

Infine la (8.12) discende dal fatto del tutto generale che, se $X \sim \mathfrak{N}(0, \sigma^2)$, con una integrazione per parti si ha

$$\begin{aligned} \mathbf{E}[X^4] &= \int_{-\infty}^{+\infty} x^4 \frac{e^{-x^2/2\sigma^2}}{\sigma\sqrt{2\pi}} dx = -\sigma^2 \int_{-\infty}^{+\infty} x^3 \frac{d}{dx} \left(\frac{e^{-x^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \right) dx \\ &= 3\sigma^2 \int_{-\infty}^{+\infty} x^2 \frac{e^{-x^2/2\sigma^2}}{\sigma\sqrt{2\pi}} dx = 3\sigma^4 = 3\mathbf{E}[X^2]^2 \end{aligned}$$

e dalla (8.11) in base alla quale $\Delta W_j \sim \mathfrak{N}(0, D\Delta t_j)$ □

Proposizione 8.4. *Se $W(t)$ è un processo di Wiener con $W(t_0) = w_0$, \mathbf{P} -qo, allora*

$$\int_{t_0}^t W(s) dW(s) = \frac{1}{2} [W^2(t) - w_0^2 - D(t - t_0)] \quad (8.13)$$

$$\mathbf{E} \left[\int_{t_0}^t W(s) dW(s) \right] = 0 \quad (8.14)$$

Dimostrazione: Si noti innanzitutto che in (8.13) il termine $\frac{1}{2}D(t-t_0)$ è totalmente estraneo all'usuale formula del calcolo integrale che prevederebbe solo la presenza dei primi due termini: questo è il primo esempio di come il calcolo stocastico Itô conduca a risultati quantitativamente diversi da quelli consueti. Per dimostrare il risultato osserviamo allora che in questo caso le somme di Riemann (8.8) assumono la forma particolare

$$\begin{aligned} S_n &= \sum_{j=1}^n W_{j-1}(W_j - W_{j-1}) = \sum_{j=1}^n W_{j-1}\Delta W_j \\ &= \frac{1}{2} \sum_{j=1}^n [(W_{j-1} + \Delta W_j)^2 - W_{j-1}^2 - (\Delta W_j)^2] \\ &= \frac{1}{2} \sum_{j=1}^n [W_j^2 - W_{j-1}^2 - (\Delta W_j)^2] = \frac{1}{2} [W^2(t) - W^2(t_0)] - \frac{1}{2} \sum_{j=1}^n (\Delta W_j)^2 \\ &= \frac{1}{2} [W^2(t) - w_0^2] - \frac{1}{2} \sum_{j=1}^n (\Delta W_j)^2 \end{aligned}$$

per cui il risultato sarà dimostrato se riusciremo a provare che

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n (\Delta W_j)^2 = D(t - t_0) \quad (8.15)$$

ovvero equivalentemente secondo il Teorema 4.6

$$\lim_n \mathbf{E} \left[\sum_{j=1}^n (\Delta W_j)^2 \right] = D(t - t_0) \quad \lim_n \mathbf{V} \left[\sum_{j=1}^n (\Delta W_j)^2 \right] = 0 \quad (8.16)$$

Ricordando allora i risultati del Lemma 8.3, la prima delle (8.16) si ricava facilmente osservando che comunque scelto n si ha

$$\mathbf{E} \left[\sum_{j=1}^n (\Delta W_j)^2 \right] = \sum_{j=1}^n \mathbf{E} [(\Delta W_j)^2] = \sum_{j=1}^n D(t_j - t_{j-1}) = D(t - t_0)$$

per cui il valore rimane sempre lo stesso anche al limite per $n \rightarrow \infty$. Quanto alla seconda delle (8.16) cominciamo con l'osservare che dai risultati appena ottenuti

potremo sempre scrivere

$$\begin{aligned} \mathbf{V} \left[\sum_{j=1}^n (\Delta W_j)^2 \right] &= \mathbf{E} \left[\left(\sum_{j=1}^n (\Delta W_j)^2 - D(t - t_0) \right)^2 \right] \\ &= \mathbf{E} \left[\sum_{j=1}^n (\Delta W_j)^4 + 2 \sum_{j < k} (\Delta W_j)^2 (\Delta W_k)^2 \right. \\ &\quad \left. - 2D(t - t_0) \sum_{j=1}^n (\Delta W_j)^2 + D^2(t - t_0)^2 \right] \end{aligned}$$

e che queste attese potranno poi essere calcolate tenendo ancora conto del Lemma 8.3 – in particolare della formula (8.12) – e dell'indipendenza degli incrementi del processo di Wiener in base alla quale per $j < k$

$$\mathbf{E} [(\Delta W_j)^2 (\Delta W_k)^2] = \mathbf{E} [(\Delta W_j)^2] \mathbf{E} [(\Delta W_k)^2] = D^2(t_j - t_{j-1})(t_k - t_{k-1})$$

Pertanto, riordinando tutti i termini dell'espressione precedente, e ricordando che nella nostra procedura di Riemann

$$\sum_{j=1}^n (t_j - t_{j-1}) = t - t_0 \quad \delta = \max_j \{t_j - t_{j-1}\} \xrightarrow{n \rightarrow \infty} 0$$

complessivamente avremo

$$\begin{aligned} \mathbf{V} \left[\sum_{j=1}^n (\Delta W_j)^2 \right] &= 3D^2 \sum_{j=1}^n (t_j - t_{j-1})^2 + 2D^2 \sum_{j < k} (t_j - t_{j-1})(t_k - t_{k-1}) \\ &\quad - 2D^2(t - t_0) \sum_{j=1}^n (t_j - t_{j-1}) + D^2(t - t_0)^2 \\ &= 2D^2 \sum_{j=1}^n (t_j - t_{j-1})^2 + D^2 \sum_{j,k=1}^n (t_j - t_{j-1})(t_k - t_{k-1}) - D^2(t - t_0)^2 \\ &= 2D^2 \sum_{j=1}^n (t_j - t_{j-1})^2 \leq 2D^2 \max_j \{t_j - t_{j-1}\} \sum_{j=1}^n (t_j - t_{j-1}) \\ &= 2\delta D^2(t - t_0) \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

In conclusione la convergenza (8.15) è confermata, e conseguentemente la (8.13) è in definitiva dimostrata. Infine per provare la (8.14) basterà osservare che da (8.13) si ha

$$\mathbf{E} \left[\int_{t_0}^t W(s) dW(s) \right] = \frac{1}{2} \mathbf{E} [W^2(t) - w_0^2 - D(t - t_0)] = 0$$

dove abbiamo nuovamente fatto uso del Lemma 8.3 □

Esempio 8.5. Nella Sezione 8.2.2 abbiamo accettato senza dimostrazione l'affermazione secondo la quale il valore corretto di un integrale di $It\bar{o}$ come (8.13) è quello che si calcola con la procedura di Riemann secondo la formula (8.8) nella quale l'integrando deve essere sempre valutato negli estremi inferiori degli intervalli della partizione. Possiamo ora mostrare con un esempio esplicito che in effetti il risultato della procedura di Riemann per il calcolo di (8.13) sarebbe stato differente se avessimo valutato l'integrando in punti diversi dall'estremo inferiore usando somme del tipo

$$S_n = \sum_{j=1}^n W(\tau_j) [W(t_j) - W(t_{j-1})]$$

con i τ_j scelti in maniera arbitraria negli intervalli $[t_{j-1}, t_j]$. Per provare senza complicazioni inessenziali che il limite in mq della successione S_n dipende dalla scelta delle τ_j sarà allora sufficiente mostrare tale dipendenza anche solo per il limite delle attese $\mathbf{E}[S_n]$. Da (8.10) si ha infatti

$$\begin{aligned} \mathbf{E}[S_n] &= \mathbf{E} \left[\sum_{j=1}^n W(\tau_j) [W(t_j) - W(t_{j-1})] \right] \\ &= \sum_{j=1}^n \left(\mathbf{E}[W(\tau_j)W(t_j)] - \mathbf{E}[W(\tau_j)W(t_{j-1})] \right) \\ &= \sum_{j=1}^n [w_0^2 + D(\tau_j - t_0) - w_0^2 - D(t_{j-1} - t_0)] \\ &= \sum_{j=1}^n [D(\tau_j - t_0) - D(t_{j-1} - t_0)] = D \sum_{j=1}^n (\tau_j - t_{j-1}) \end{aligned}$$

Se allora introduciamo un parametro $\alpha \in [0, 1]$ per individuare la posizione di τ_j con

$$\tau_j = \alpha t_j + (1 - \alpha)t_{j-1}$$

potremo anche scrivere per ogni n

$$\mathbf{E}[S_n] = \alpha D \sum_{j=1}^n (t_j - t_{j-1}) = \alpha D(t - t_0)$$

per cui, scambiando il limite per $n \rightarrow \infty$ di Riemann con l'attesa, si ottiene

$$\mathbf{E} \left[\int_{t_0}^t W(s) dW(s) \right] = \alpha D(t - t_0)$$

un risultato che manifestamente dipende da α , cioè dalla scelta di τ_j all'interno dell'intervallo $[t_{j-1}, t_j]$, e che coincide con la formula corretta (8.14) per l'integrale di $It\bar{o}$ solo se $\alpha = 0$, cioè quando τ_j è l'estremo sinistro dell'intervallo $[t_{j-1}, t_j]$

8.3.2 Attese e covarianze

Proposizione 8.6. *Se $G(t)$ e $H(t)$ sono processi non anticipativi*

$$\mathbf{E} \left[\int_{t_0}^t G(s) dW(s) \right] = 0 \quad (8.17)$$

$$\mathbf{E} \left[\int_{t_0}^t G(s) dW(s) \int_{t_0}^t H(s') dW(s') \right] = D \int_{t_0}^t \mathbf{E} [G(s)H(s)] ds \quad (8.18)$$

Dimostrazione: La formula (8.17) generalizza la (8.14): per le solite somme di Riemann, dalla non anticipatività di $G(t)$ e dal Lemma 8.3 si ha infatti

$$\mathbf{E} \left[\sum_{j=1}^n G_{j-1} \Delta W_j \right] = \sum_{j=1}^n \mathbf{E} [G_{j-1}] \mathbf{E} [\Delta W_j] = 0$$

e se accettiamo (come al solito senza verifica) che le operazioni $\mathbf{E}[\cdot]$ e \lim_n -mq possano essere scambiate, il risultato segue facilmente. Quanto alla formula di covarianza (8.18), dalla non anticipatività dei processi, dal Lemma 8.3 e dall'indipendenza degli incrementi ΔW_j si ha

$$\begin{aligned} & \mathbf{E} \left[\sum_{j=1}^n G_{j-1} \Delta W_j \sum_{k=1}^n H_{k-1} \Delta W_k \right] \\ &= \mathbf{E} \left[\sum_{j=1}^n G_{j-1} H_{j-1} (\Delta W_j)^2 \right] + \mathbf{E} \left[\sum_{k>j} (G_{j-1} H_{k-1} + G_{k-1} H_{j-1}) \Delta W_j \Delta W_k \right] \\ &= \sum_{j=1}^n \mathbf{E} [G_{j-1} H_{j-1}] \mathbf{E} [(\Delta W_j)^2] \\ & \quad + \sum_{k>j} \mathbf{E} [(G_{j-1} H_{k-1} + G_{k-1} H_{j-1}) \Delta W_j] \mathbf{E} [\Delta W_k] \\ &= \sum_{j=1}^n \mathbf{E} [G_{j-1} H_{j-1}] D \Delta t_j \end{aligned}$$

e il risultato segue poi come per la prima formula passando al limite in mq sotto il segno di valore d'attesa \square

Questi risultati ci permettono ora di riformulare in una prospettiva rovesciata le considerazioni sul rumore bianco Wieneriano condensate nella Proposizione 8.1

Corollario 8.7. *Se $W(t)$ con $W(t_0) = w_0$ è un processo di Wiener con coefficiente di diffusione D , un processo $Z(t)$ tale che $\mathbf{E} [Z(t)] = 0$, e che $dW(t) = Z(t) dt$, non può che essere un rumore bianco (Wieneriano) stazionario e di intensità D*

Dimostrazione: Si tratta di una conseguenza immediata della (8.18): presi infatti due generici processi $G(t)$ e $H(t)$ non anticipativi e indipendenti da $Z(t)$, e supponendo di poter scambiare attese e integrali dalla (8.18), dalle ipotesi si ha

$$\begin{aligned}
 D \int_{t_0}^t \mathbf{E} [G(s)H(s)] ds &= \mathbf{E} \left[\int_{t_0}^t G(s) dW(s) \int_{t_0}^t H(s') dW(s') \right] \\
 &= \mathbf{E} \left[\int_{t_0}^t ds \int_{t_0}^t ds' G(s)H(s')Z(s)Z(s') \right] \\
 &= \int_{t_0}^t ds \int_{t_0}^t ds' \mathbf{E} [G(s)H(s')Z(s)Z(s')] \\
 &= \int_{t_0}^t ds \int_{t_0}^t ds' \mathbf{E} [G(s)H(s')] \mathbf{E} [Z(s)Z(s')]
 \end{aligned}$$

che può essere soddisfatta solo se

$$\mathbf{E} [Z(s)Z(s')] = D\delta(s - s')$$

cioè se $Z(t)$ è un rumore bianco stazionario di intensità D . Naturalmente, siccome richiediamo $dW(t) = Z(t)dt$, il rumore bianco in questione non può che essere Wieneriano \square

8.3.3 Infinitesimi stocastici

Proposizione 8.8. *Se $G(t)$ è un processo non anticipativo si ha con $k = 0, 1, \dots$*

$$\begin{aligned}
 \int_{t_0}^t G(s) [dW(s)]^{2+k} &= \lim\text{-mq}_n \sum_{j=1}^n G_{j-1} (\Delta W_j)^{2+k} = \delta_{k0} D \int_{t_0}^t G(s) ds \\
 \int_{t_0}^t G(s) [dW(s)]^{1+k} ds &= \lim\text{-mq}_n \sum_{j=1}^n G_{j-1} (\Delta W_j)^{1+k} \Delta t_j = 0
 \end{aligned}$$

dove δ_{kl} è la delta di Kronecker; in maniera simbolica scriveremo anche

$$[dW(t)]^{2+k} = \delta_{k0} D dt \quad [dW(t)]^{1+k} dt = 0 \quad (8.19)$$

Dimostrazione: Questa proposizione permette di dare un significato preciso all'affermazione della Sezione 6.3 secondo la quale $dW(t)$ è un infinitesimo dell'ordine di \sqrt{dt} . Più precisamente essa suggerisce di trascurare nei calcoli tutti i termini del tipo $dW(t) dt$, $[dW(t)]^3, \dots$ in quanto infinitesimi di ordine superiore a dt , ma di conservare quelli del tipo $[dW(t)]^2 = D dt$ che – contrariamente alle apparenze – sono invece dell'ordine di dt . Per evitare ripetizioni qui ci limiteremo a dimostrare solo che $[dW(t)]^2 = D dt$, ovvero

$$\int_{t_0}^t G(s) [dW(s)]^2 = D \int_{t_0}^t G(s) ds$$

tralasciando di verificare – in maniera del tutto analoga – che le altre espressioni si annullano. Dovremo dunque provare con la solita procedura di Riemann che

$$\lim\text{-mq}_n \sum_{j=1}^n G_{j-1} (\Delta W_j)^2 = D \lim\text{-mq}_n \sum_{j=1}^n G_{j-1} \Delta t_j$$

ovvero equivalentemente che

$$\lim\text{-mq}_n \sum_{j=1}^n [G_{j-1} (\Delta W_j)^2 - G_{j-1} D \Delta t_j] = \lim_n \mathcal{E}_n = 0 \quad (8.20)$$

avendo posto sinteticamente

$$\begin{aligned} \mathcal{E}_n &= \mathbf{E} \left[\left| \sum_{j=1}^n G_{j-1} [(\Delta W_j)^2 - D \Delta t_j] \right|^2 \right] \\ &= \mathbf{E} \left[\sum_{j=1}^n G_{j-1}^2 [(\Delta W_j)^2 - D \Delta t_j]^2 \right. \\ &\quad \left. + 2 \sum_{j < k} G_{j-1} G_{k-1} [(\Delta W_j)^2 - D \Delta t_j] [(\Delta W_k)^2 - D \Delta t_k] \right] \end{aligned}$$

A causa della non anticipatività di $G(t)$, il termine G_{j-1}^2 risulta indipendente da $(\Delta W_j)^2 - D \Delta t_j$, e $G_{j-1} G_{k-1} [(\Delta W_j)^2 - D \Delta t_j]$ risulta indipendente da $(\Delta W_k)^2 - D \Delta t_k$; pertanto – tenendo anche conto del fatto che in base al Lemma 8.3 si ha $\mathbf{E} [(\Delta W_k)^2 - D \Delta t_k] = 0$ – per il secondo termine dell'espressione di \mathcal{E}_n si ha

$$\begin{aligned} &\mathbf{E} [G_{j-1} G_{k-1} [(\Delta W_j)^2 - D \Delta t_j] [(\Delta W_k)^2 - D \Delta t_k]] \\ &= \mathbf{E} [G_{j-1} G_{k-1} [(\Delta W_j)^2 - D \Delta t_j]] \mathbf{E} [(\Delta W_k)^2 - D \Delta t_k] = 0 \end{aligned}$$

mentre per il primo termine, sempre tenendo conto del Lemma 8.3, si trova

$$\begin{aligned} \mathbf{E} [[(\Delta W_j)^2 - D \Delta t_j]^2] &= \mathbf{E} [(\Delta W_j)^4] + (D \Delta t_j)^2 - 2 D \Delta t_j \mathbf{E} [(\Delta W_j)^2] \\ &= 3(D \Delta t_j)^2 + (D \Delta t_j)^2 - 2(D \Delta t_j)^2 = 2(D \Delta t_j)^2 \end{aligned}$$

e quindi complessivamente

$$\mathcal{E}_n = 2D^2 \sum_{j=1}^n (\Delta t_j)^2 \mathbf{E} [G_{j-1}^2] \leq 2D^2 \max_j \{\Delta t_j\} \sum_{j=1}^n \Delta t_j \mathbf{E} [G_{j-1}^2]$$

Il risultato (8.20) resta quindi acquisito dato che nella procedura di Riemann $\delta = \max_j \{\Delta t_j\} \rightarrow 0$ mentre, con ragionevoli condizioni su $G(t)$, il secondo fattore resta limitato. \square

Proposizione 8.9. *Dato un processo di Wiener $W(t)$ con $W(t_0) = w_0$, per $n = 1, 2, \dots$ risulta*

$$\int_{t_0}^t W^n(s) dW(s) = \frac{W^{n+1}(t) - w_0^{n+1}}{n+1} - \frac{nD}{2} \int_{t_0}^t W^{n-1}(s) ds \quad (8.21)$$

Dimostrazione: La (8.21) generalizza il risultato (8.13) e può essere ricavata facilmente utilizzando i risultati simbolici della Proposizione 8.8 sugli ordini degli infinitesimi. Si ha infatti

$$\begin{aligned} dW^{n+1}(t) &= W^{n+1}(t+dt) - W^{n+1}(t) = [W(t) + dW(t)]^{n+1} - W^{n+1}(t) \\ &= \sum_{k=0}^{n+1} \binom{n+1}{k} W^{n+1-k}(t) [dW(t)]^k - W^{n+1}(t) \\ &= \sum_{k=1}^{n+1} \binom{n+1}{k} W^{n+1-k}(t) [dW(t)]^k \\ &= \binom{n+1}{1} W^n(t) dW(t) + \binom{n+1}{2} W^{n-1}(t) [dW(t)]^2 \\ &= (n+1)W^n(t) dW(t) + \frac{(n+1)n}{2} W^{n-1}(t) Ddt \end{aligned}$$

e pertanto

$$\begin{aligned} W^{n+1}(t) - W^{n+1}(t_0) &= \int_{t_0}^t dW^{n+1}(s) \\ &= (n+1) \int_{t_0}^t W^n(s) dW(s) + \frac{(n+1)n}{2} D \int_{t_0}^t W^{n-1}(s) ds \end{aligned}$$

da cui si ottiene immediatamente il risultato tenendo anche conto della condizione $W(t_0) = w_0$ □

8.3.4 Regole di differenziazione e di integrazione per parti

Proposizione 8.10. *Se $g(x, t)$ è derivabile almeno due volte in x e una in t , e se $W(t)$ è un processo di Wiener, posto*

$$g_x = \partial_x g \quad g_{xx} = \partial_x^2 g \quad g_t = \partial_t g$$

si ha la seguente regola di differenziazione

$$dg(W(t), t) = \left[g_t(W(t), t) + \frac{D}{2} g_{xx}(W(t), t) \right] dt + g_x(W(t), t) dW(t) \quad (8.22)$$

Dimostrazione: Tenendo conto della Proposizione 8.8 si ha

$$\begin{aligned}
dg(W(t), t) &= g(W(t+dt), t+dt) - g(W(t), t) \\
&= \left[g(W(t+dt), t+dt) - g(W(t), t+dt) \right] \\
&\quad + \left[g(W(t), t+dt) - g(W(t), t) \right] \\
&= \left[g(W(t), t+dt) + g_x(W(t), t+dt)dW(t) \right. \\
&\quad \left. + \frac{1}{2}g_{xx}(W(t), t+dt)[dW(t)]^2 + \dots - g(W(t), t+dt) \right] \\
&\quad + \left[g(W(t), t) + g_t(W(t), t)dt \right. \\
&\quad \left. + \frac{1}{2}g_{tt}(W(t), t)(dt)^2 + \dots - g(W(t), t) \right] \\
&= \left[g_x(W(t), t) + g_{xt}(W(t), t)dt + \dots \right] dW(t) \\
&\quad + \frac{1}{2} \left[g_{xx}(W(t), t) + g_{xxt}(W(t), t)dt + \dots \right] [dW(t)]^2 + \dots \\
&\quad + g_t(W(t), t)dt + \frac{1}{2}g_{tt}(W(t), t)(dt)^2 + \dots \\
&= g_x(W(t), t)dW(t) + \frac{D}{2}g_{xx}(W(t), t)dt + g_t(W(t), t)dt
\end{aligned}$$

cioè il risultato richiesto (8.22) □

Esempio 8.11. *In pratica nelle differenziazioni stocastiche bisogna tenere conto del fatto che $[dW(t)]^2$ è un infinitesimo dell'ordine di dt e non di ordine superiore: in particolare questo produce la presenza di termini che sarebbero altrimenti non spiegabili. Così ad esempio per un processo di Wiener geometrico (6.59) $X(t) = e^{W(t)}$ si ha $g(x, t) = e^x$ e quindi*

$$dX(t) = d(e^{W(t)}) = e^{W(t)} dW(t) + \frac{D}{2} e^{W(t)} dt$$

Allo stesso modo, se $X(t) = W^2(t)$, cioè se $g(x, t) = x^2$, si ottiene

$$dX(t) = d(W^2(t)) = 2W(t) dW(t) + D dt$$

Si noti che i termini aggiuntivi rispetto alle regole tradizionali sono contrassegnati dalla presenza del coefficiente di diffusione D : se questo si annulla i processi degenerano in traiettorie deterministiche, e le regole tornano ad essere quelle ordinarie. Naturalmente questa osservazione – adeguatamente riformulata – può essere estesa anche a tutte le altre formule del calcolo stocastico incontrate finora, ad esempio (8.13), (8.18), (8.19), (8.21) e (8.22)

Le nuove regole di differenziazione permettono anche di generalizzare le formule dell'*integrazione per parti*: è noto ad esempio dal calcolo ordinario che

$$d[x(t)h(x(t), t)] = x(t) dh(x(t), t) + h(x(t), t) dx(t)$$

da cui si ricava la formula

$$\int_a^b h(x(t), t) dx(t) = [x(t)h(x(t), t)]_a^b - \int_a^b x(t) dh(x(t), t)$$

Questa espressione si riduce a quella più familiare quando $h(x, t)$ non dipende da x : in questo caso infatti si ha

$$d[x(t)h(t)] = x(t) dh(t) + h(t) dx(t) = [x(t)\dot{h}(t) + \dot{x}(t)h(t)] dt$$

e quindi con un'integrazione

$$\int_a^b h(t)\dot{x}(t) dt = [h(t)x(t)]_a^b - \int_a^b \dot{h}(t)x(t) dt$$

Il calcolo stocastico richiede una modifica di questi risultati, ma le differenze sono visibili solo nel caso in cui $h(x, t)$ dipende anche da x

Proposizione 8.12. Integrazione per parti: *Se $h(x, t)$ è derivabile almeno due volte in x e una in t , e $W(t)$ è un processo di Wiener con $W(t_0) = w_0$, la regola di integrazione per parti è*

$$\begin{aligned} \int_{t_0}^t h(W(s), s) dW(s) &= [W(s)h(W(s), s)]_{t_0}^t - \int_{t_0}^t W(s) dh(W(s), s) \\ &\quad - D \int_{t_0}^t h_x(W(s), s) ds \end{aligned} \quad (8.23)$$

Dimostrazione: Applicando le regole di differenziazione (8.22) con $g(x, t) = xh(x, t)$ si ha

$$g_t = xh_t \quad g_x = h + xh_x \quad g_{xx} = 2h_x + xh_{xx}$$

e quindi

$$\begin{aligned} d[W(t)h(W(t), t)] &= dg(W(t), t) \\ &= \left[W(t)h_t(W(t), t) + \frac{D}{2} \left(2h_x(W(t), t) + W(t)h_{xx}(W(t), t) \right) \right] dt \\ &\quad + [h(W(t), t) + W(t)h_x(W(t), t)] dW(t) \\ &= W(t) \left[\left(h_t(W(t), t) + \frac{D}{2} h_{xx}(W(t), t) \right) dt + h_x(W(t), t) dW(t) \right] \\ &\quad + h(W(t), t) dW(t) + Dh_x(W(t), t) dt \\ &= W(t)dh(W(t), t) + h(W(t), t)dW(t) + Dh_x(W(t), t)dt \end{aligned}$$

da cui segue la formula per semplice integrazione. □

8.4 Equazioni differenziali stocastiche (*EDS*)

8.4.1 Differenziali stocastici e formula di Itô

Definizione 8.13. Diremo che un processo $X(t)$ ammette in $[0, T]$ il **differenziale stocastico**

$$dX(t) = A(t) dt + B(t) dW(t) \quad (8.24)$$

quando per ogni t_0, t con $0 \leq t_0 < t \leq T$ esso può essere rappresentato come

$$X(t) = X(t_0) + \int_{t_0}^t A(s) ds + \int_{t_0}^t B(s) dW(s)$$

dove $W(t)$ è un processo di Wiener con $W(t_0) = w_0$, e $A(t)$ e $B(t)$ sono due processi tali che

$$\mathbf{P} \left\{ \int_0^T |A(t)| dt < +\infty \right\} = 1 \quad \mathbf{P} \left\{ \int_0^T |B(t)|^2 dt < +\infty \right\} = 1$$

Proposizione 8.14. Formula di Itô: Se $X(t)$ ammette il differenziale stocastico (8.24), e se $g(x, t)$ è una funzione derivabile almeno due volte in x e una volta in t , allora anche $g(X(t), t)$ ammette il differenziale stocastico

$$\begin{aligned} dg(X(t), t) &= \left[g_t(X(t), t) + \frac{D}{2} B^2(t) g_{xx}(X(t), t) \right] dt + g_x(X(t), t) dX(t) \\ &= \left[g_t(X(t), t) + A(t) g_x(X(t), t) + \frac{D}{2} B^2(t) g_{xx}(X(t), t) \right] dt + B(t) g_x(X(t), t) dW(t) \end{aligned} \quad (8.25)$$

Dimostrazione: La (8.25) generalizza la (8.22) che si ottiene per $A(t) = 0$ e $B(t) = 1$, cioè quando da (8.24) si ha che $X(t) = W(t)$. Per dimostrare (8.25) osserviamo che da (8.24) e dalle regole (8.19) si ha simbolicamente

$$[dX(t)]^2 = [A(t)dt]^2 + [B(t)dW(t)]^2 + 2A(t)B(t)dW(t)dt = B^2(t)Ddt$$

e quindi, riproducendo la dimostrazione di (8.22) con $X(t)$ al posto di $W(t)$, avremo

$$\begin{aligned} dg(X(t), t) &= [g_x(X(t), t) + g_{xt}(X(t), t)dt + \dots] dX(t) \\ &\quad + \frac{1}{2} [g_{xx}(X(t), t) + g_{xxt}(X(t), t)dt + \dots] [dX(t)]^2 + \dots \\ &\quad + g_t(X(t), t)dt + \frac{1}{2} g_{tt}(X(t), t)(dt)^2 + \dots \\ &= g_x(X(t), t) [A(t) dt + B(t) dW(t)] \\ &\quad + \frac{D}{2} B^2(t) g_{xx}(X(t), t) dt + g_t(X(t), t) dt \end{aligned}$$

da cui discende immediatamente la formula di Itô (8.25) □

8.4.2 EDS e loro soluzioni

Definizione 8.15. Chiameremo **equazione differenziale stocastica (EDS)** la

$$\begin{aligned} dX(t) &= a(X(t), t) dt + b(X(t), t) dW(t) & 0 \leq t_0 < t \leq T & \quad (8.26) \\ X(t_0) &= X_0 & \mathbf{P}\text{-}q_0 & \end{aligned}$$

dove $W(t)$ è un processo di Wiener con $W(t_0) = w_0$, e X_0 una v -a indipendente da $W(t)$, e diremo che un processo $X(t)$ è una **soluzione** se esso ammette (8.26) come differenziale stocastico, ovvero se

$$X(t) = X_0 + \int_{t_0}^t a(X(s), s) ds + \int_{t_0}^t b(X(s), s) dW(s) \quad (8.27)$$

Si dice che la soluzione è **unica** se, comunque scelte due soluzioni $X_1(t)$ e $X_2(t)$ risulta

$$\mathbf{P}\left\{ \sup_{t_0 \leq t \leq T} |X_1(t) - X_2(t)| > 0 \right\} = 0$$

Le soluzioni di (8.26) possono essere determinate mediante varie procedure di approssimazioni successive:

1. si costruisce una *successione di processi* approssimanti mediante la prescrizione

$$\begin{aligned} X_0(t) &= X_0 \\ X_n(t) &= X_0 + \int_{t_0}^t a(X_{n-1}(s), s) ds + \int_{t_0}^t b(X_{n-1}(s), s) dW(s) \end{aligned}$$

e se ne studia il processo limite per $n \rightarrow \infty$; questo è il tipico procedimento iterativo usato per la dimostrazione dei Teoremi di esistenza e unicità;

2. si costruiscono le *traiettorie del processo* soluzione con un metodo iterativo molto usato per generare delle simulazioni: scelti degli istanti (in genere equidistanti)

$$t_0 < t_1 < \dots < t_n = t \leq T$$

si costruiscono i campioni partendo da un valore iniziale x_0 e secondo la prescrizione

$$x_{j+1} = x_j + a(x_j, t_j)\Delta t_j + b(x_j, t_j)\Delta w_j$$

dove

$$x_j = x(t_j) \quad \Delta t_j = t_{j+1} - t_j \quad \Delta w_j = w(t_{j+1}) - w(t_j)$$

e $w(t)$ è un campione del processo di Wiener $W(t)$; i campioni Δw_j di ΔW_j sono ovviamente estratti in maniera indipendente dalle x_j . Ogni valore di x_0 e ogni campione $w(t)$ del processo di Wiener determina una possibile traiettoria discretizzata. Si esegue poi un limite per $n \rightarrow \infty$ e si dice che la soluzione *esiste* se tale limite esiste quasi per ogni campione $w(t)$ del processo di Wiener; la soluzione poi è *unica* se quasi per ogni campione $w(t)$ del processo di Wiener tale traiettoria limite è unica.

Teorema 8.16. Teorema di esistenza e unicità: *L'EDS (8.26) ammette un'unica soluzione $X(t)$ se sono soddisfatte le condizioni di Lipschitz, cioè se esistono due numeri k_1 e k_2 tali che*

$$\begin{aligned} |a(x, t) - a(y, t)| + |b(x, t) - b(y, t)| &\leq k_1|x - y| \\ |a(x, t)|^2 + |b(x, t)|^2 &\leq k_2(1 + |x|^2) \end{aligned}$$

comunque presi x, y e $t \in [0, T]$. La soluzione è un processo (globalmente) continuo e non anticipativo

Dimostrazione: Omessa⁶. La dimostrazione consiste nella verifica del fatto che la successione di processi $X_n(t)$ generata con la procedura 1 converge **P**-qo e uniformemente in $[0, T]$. Siccome non sempre le funzioni $a(x, t)$ e $b(x, t)$ soddisfano le condizioni di Lipschitz, si usa introdurre il concetto di **soluzione debole** contrapposto a quello di **soluzione forte** introdotto nella Definizione 8.15. Noi non potremo soffermarci su questa distinzione \square

Corollario 8.17. Cambiamento di variabile: *Se $X(t)$ è soluzione della EDS (8.26) e $g(x, t)$ è una funzione derivabile almeno due volte in x e una volta in t , allora la formula di Itô per il processo $Y(t) = g(X(t), t)$ diviene*

$$\begin{aligned} dg(X(t), t) = &\left[g_t(X(t), t) + a(X(t), t)g_x(X(t), t) + \frac{D}{2}b^2(X(t), t)g_{xx}(X(t), t) \right] dt \\ &+ b(X(t), t)g_x(X(t), t) dW(t) \end{aligned} \quad (8.28)$$

e può essere messa sotto forma di una nuova EDS per $Y(t)$

Dimostrazione: Discende facilmente dalla Formula di Itô (8.25) \square

8.4.3 EDS ed equazioni di Fokker-Planck

Nel seguito il calcolo dei valori d'attesa sarà sempre eseguito tenendo conto di *tutte le condizioni iniziali* richieste sui processi coinvolti ($X(t)$, $W(t)$ o anche altri), e quindi con le corrispondenti distribuzioni condizionate

Proposizione 8.18. *Ogni soluzione dell'EDS (8.26) è un processo di Markov*

Dimostrazione: Ci limiteremo a una spiegazione intuitiva costruendo i campioni delle traiettorie di $X(t)$ secondo la procedura 2. Dato $X(s) = y$ per un dato $s > t_0$, l'evoluzione di $X(t)$ per $t > s$ dipenderà dal campione $w(t)$ di $W(t)$ per $t > s$. D'altra parte $X(t)$ è non anticipativo, e quindi $X(t')$ con $t' < s$ è indipendente da $W(t)$ con $t > s$. Ne segue che, noto y , i valori di $X(t)$ con $t > s$ saranno indipendenti da quelli con $t' < s$, e quindi $X(t)$ è un processo di Markov. \square

⁶I. Karatzas, S.E. Shreve, BROWNIAN MOTION AND STOCHASTIC CALCULUS, Springer (Berlin, 1991). B. Øksendal, STOCHASTIC DIFFERENTIAL EQUATIONS, Springer (Berlin, 2005)

Proposizione 8.19. *Se $X(t)$ è una soluzione ac dell'EDS (8.26) con $X(t_0) = X_0$, \mathbf{P} -qo, allora la sua ddp sarà soluzione dell'equazione di Fokker–Planck*

$$\partial_t f(x, t) = -\partial_x [A(x, t)f(x, t)] + \frac{1}{2} \partial_x^2 [B(x, t)f(x, t)] \quad f(x, t_0) = f_0(x)$$

dove f_0 è la ddp di X_0 , e

$$A(x, t) = a(x, t) \quad B(x, t) = D b^2(x, t) \quad (8.29)$$

In particolare la ddp di transizione $f(x, t | x_0, t_0)$ corrisponderà alla condizione iniziale $f(x, t_0) = \delta(x - x_0)$, cioè $X(t_0) = x_0$, \mathbf{P} -qo

Dimostrazione: Abbiamo già osservato che una soluzione dell'EDS (8.26) è un processo di Markov continuo, e pertanto la sua ddp di transizione sarà soluzione di qualche forma di equazione di Fokker–Planck (7.77). Supponiamo allora che $X(t_0) = x_0$, \mathbf{P} -qo, e che $h(x)$ sia una funzione derivabile almeno due volte in x , ma costante in t : dalla Formula del cambiamento di variabile (8.28) abbiamo allora

$$\begin{aligned} dh(X(t)) &= \left[a(X(t), t) h'(X(t)) + \frac{D}{2} b^2(X(t), t) h''(X(t)) \right] dt \\ &\quad + b(X(t), t) h'(X(t)) dW(t) \end{aligned}$$

Siccome $X(t)$ è non anticipativo avremo

$$\mathbf{E} [b(X(t), t) h'(X(t)) dW(t)] = \mathbf{E} [b(X(t), t) h'(X(t))] \mathbf{E} [dW(t)] = 0$$

e quindi integrando per parti

$$\begin{aligned} \mathbf{E} [dh(X(t))] &= \mathbf{E} \left[a(X(t), t) h'(X(t)) + \frac{D}{2} b^2(X(t), t) h''(X(t)) \right] dt \\ &= \int_{-\infty}^{+\infty} \left[a(x, t) h'(x) + \frac{D}{2} b^2(x, t) h''(x) \right] f(x, t | x_0, t_0) dx dt \\ &= \int_{-\infty}^{+\infty} \left[-\partial_x [a(x, t) f(x, t | x_0, t_0)] \right. \\ &\quad \left. + \frac{D}{2} \partial_x^2 [b^2(x, t) f(x, t | x_0, t_0)] \right] h(x) dx dt \end{aligned}$$

D'altra parte abbiamo anche

$$\begin{aligned} \mathbf{E} [dh(X(t))] &= d\mathbf{E} [h(X(t))] = \frac{d}{dt} \mathbf{E} [h(X(t))] dt \\ &= \int_{-\infty}^{+\infty} h(x) \partial_t f(x, t | x_0, t_0) dx dt \end{aligned}$$

Confrontando allora le due espressioni, il risultato per la ddp di transizione deriva dall'arbitrarietà di $h(x)$. L'equazione per condizioni iniziali più generali si ricava poi facilmente da quella per la ddp di transizione. \square

Dato il ruolo giocato dai coefficienti A e B nelle equazioni *in avanti* (vedi Sezione 7.2.3), il risultato della Proposizione 8.19 suggerisce quindi di interpretare anche i coefficienti a e b dell'EDS (8.26) rispettivamente come *velocità di trascinamento* e *coefficiente di diffusione*.

8.5 Forme particolari di EDS

Sarà utile ricordare innanzitutto che la legge di un processo di Markov $X(t)$ è completamente specificata dalla conoscenza delle *ddp* $f(x, t)$ e $f(x, t; y, s)$. Se però $X(t)$ è un *processo Gaussiano* (vedi Sezione 7.1.10) queste due *ddp* sono completamente determinate da $\mathbf{E}[X(t)]$ e $\mathbf{cov}[X(t), X(s)]$: infatti in questo caso si ha

$$f(x, t) = \mathfrak{N}(\mathbf{E}[X(t)], \mathbf{V}[X(t)]) \quad f(x, t; y, s) = \mathfrak{N}(\mathbf{b}, \mathbb{A})$$

dove

$$\mathbf{b} = \begin{pmatrix} \mathbf{E}[X(t)] \\ \mathbf{E}[X(s)] \end{pmatrix} \quad \mathbb{A} = \begin{pmatrix} \mathbf{V}[X(t)] & \mathbf{cov}[X(s), X(t)] \\ \mathbf{cov}[X(t), X(s)] & \mathbf{V}[X(s)] \end{pmatrix}$$

Inoltre, per un processo di Wiener con condizione iniziale $W(t_0) = w_0$, sarà utile definire anche il corrispondente *processo centrato* per il quale adotteremo la notazione $\widetilde{W}(t) = W(t) - w_0$ con un riadattamento delle formule (8.9) e (8.10) del tipo

$$\mathbf{E}[\widetilde{W}(t)] = 0 \quad \mathbf{E}[\widetilde{W}(s)\widetilde{W}(t)] = D \min\{s - t_0, t - t_0\}$$

8.5.1 EDS a coefficienti costanti

Il caso più semplice di EDS si ha per $a(x, t) = a$, $b(x, t) = b$ costanti: in tal caso l'equazione si riduce a

$$dX(t) = a dt + b dW(t) \quad X(t_0) = X_0 \quad (8.30)$$

e la sua soluzione esplicita è semplicemente

$$X(t) = X_0 + a(t - t_0) + b[W(t) - w_0]$$

L'equazione di Fokker-Planck corrispondente secondo la Proposizione 8.19 è poi

$$\partial_t f(x, t) = -a \partial_x f(x, t) + \frac{Db^2}{2} \partial_x^2 f(x, t) \quad f(x, t_0) = f_0(x)$$

se f_0 è la *ddp* di X_0 . Inoltre la soluzione risulta $X(t)$ Gaussiana se la condizione iniziale X_0 è Gaussiana, e in particolare se $X_0 = x_0$, \mathbf{P} -qo. È facile quindi capire che in questi casi la soluzione di (8.30) non è nient'altro che un processo di Wiener reso un po' più generale dall'introduzione di un trascinamento a e di un riscaldamento della diffusione b , in modo che la sua *ddp* di transizione è $\mathfrak{N}(a(t - t_0), Db^2(t - t_0))$. Naturalmente se in particolare $a = 0$ e $b = 1$, $X(t)$ coinciderebbe proprio con un processo di Wiener con equazione (7.87). Si noti che invece un'arbitraria condizione iniziale X_0 produrrebbe un processo $X(t)$ con la stessa *ddp* di transizione Wieneriana, ma con leggi differenti, in generale non Gaussiane

8.5.2 *EDS* a coefficienti dipendenti solo dal tempo

Con $a(x, t) = a(t)$, $b(x, t) = b(t)$ dipendenti solo dal tempo l'*EDS* (8.26) diventa

$$dX(t) = a(t) dt + b(t) dW(t) \quad X(t_0) = X_0, \quad \mathbf{P}\text{-qo} \quad (8.31)$$

e formalmente la sua soluzione esplicita è

$$X(t) = X_0 + \int_{t_0}^t a(t') dt' + \int_{t_0}^t b(t') dW(t') \quad (8.32)$$

Anche in questo caso quindi – essendo l'integrale di Wiener ovviamente Gaussiano – la soluzione risulta Gaussiana se X_0 lo è, e in particolare se $X_0 = x_0$. Inoltre la corrispondente equazione di Fokker–Planck è ora

$$\partial_t f(x, t) = -a(t)\partial_x f(x, t) + \frac{D}{2} b(t)^2 \partial_x^2 f(x, t) \quad f(x, t_0) = f_0(x)$$

Per determinare la distribuzione del processo basterà calcolare la *ddp* di transizione associata alla condizione iniziale degenera $X(t_0) = x_0$: altre soluzioni con condizioni iniziali generiche $f_0(x)$ si ottengono dalle equazioni di Chapman-Kolmogorov (7.16)

Proposizione 8.20. *La soluzione $X(t)$ dell'*EDS**

$$dX(t) = a(t) dt + b(t) dW(t) \quad X(t_0) = x_0, \quad \mathbf{P}\text{-qo} \quad (8.33)$$

è un processo Gaussiano con $(s \wedge t = \min\{s, t\})$

$$m(t) = \mathbf{E}[X(t)] = x_0 + \int_{t_0}^t a(t') dt' \quad \mathbf{cov}[X(s), X(t)] = D \int_{t_0}^{s \wedge t} b^2(t') dt' \quad (8.34)$$

*e quindi la *ddp* di transizione è $\mathfrak{N}(m(t), \sigma^2(t))$, con $\sigma^2(t) = \mathbf{V}[X(t)]$ da (8.34)*

Dimostrazione: Trattandosi di un processo Gaussiano basterà determinare l'attesa e la covarianza. Per l'attesa da (8.32) con $X_0 = x_0$ si ha

$$m(t) = \mathbf{E}[X(t)] = \mathbf{E}[X_0] + \int_{t_0}^t a(t') dt' + \int_{t_0}^t b(t') \mathbf{E}[dW(t')] = x_0 + \int_{t_0}^t a(t') dt'$$

Per la covarianza invece, supponendo $t_0 < s < t$, dai risultati precedenti con $X_0 = x_0$ si ha

$$\begin{aligned} \mathbf{cov}[X(s), X(t)] &= \mathbf{E}[(X(t) - \mathbf{E}[X(t)])(X(s) - \mathbf{E}[X(s)])] \\ &= \mathbf{E}\left[\int_{t_0}^t b(t') dW(t') \int_{t_0}^s b(s') dW(s')\right] \end{aligned}$$

e poi dall'indipendenza dei $dW(t)$ su intervalli non sovrapposti, e da (8.18)

$$\begin{aligned} \mathbf{cov} [X(s), X(t)] &= \mathbf{E} \left[\int_{t_0}^s b(t') dW(t') \int_{t_0}^s b(s') dW(s') \right] \\ &\quad + \mathbf{E} \left[\int_s^t b(t') dW(t') \int_{t_0}^s b(s') dW(s') \right] \\ &= D \int_{t_0}^s b^2(t') dt' \end{aligned}$$

per cui in definitiva risultano verificate le (8.34). In particolare poi, ricordando che risulta anche

$$\sigma^2(t) = \mathbf{V} [X(t)] = \mathbf{cov} [X(t), X(t)] = D \int_{t_0}^t b^2(t') dt'$$

si trova che la ddp di transizione è $\mathfrak{N}(m(t), \sigma^2(t))$ □

8.5.3 EDS con diffusione lineare in x

Supponiamo ora che il coefficiente di diffusione sia $b(x, t) = cx$ con $c > 0$, e per semplicità poniamo $a(x, t) = 0$ per il trascinato: la nostra EDS assume quindi la forma

$$dX(t) = cX(t) dW(t) \quad X(t_0) = X_0 > 0, \quad \mathbf{P}\text{-qo} \quad (8.35)$$

mentre la corrispondente equazione di Fokker–Planck, con $A(x, t) = a(x, t) = 0$ e $B(x, t) = Db^2(x, t) = Dc^2x^2$, è

$$\partial_t f(x, t) = \frac{Dc^2}{2} \partial_x^2 [x^2 f(x, t)] \quad f(x, t_0) = f_0(x)$$

Per risolvere la (8.35) converrà trasformarla introducendo il processo ausiliario

$$Y(t) = \ln X(t) \quad Y(t_0) = Y_0 = \ln X_0$$

la cui equazione può essere determinata con un'applicazione della formula (8.28): osservando infatti che

$$g(x, t) = \ln x \quad g_x(x, t) = \frac{1}{x} \quad g_{xx}(x, t) = -\frac{1}{x^2} \quad g_t(x, t) = 0$$

da (8.28) si ottiene immediatamente

$$dY(t) = -\frac{Dc^2}{2} dt + c dW(t) \quad Y(t_0) = Y_0 \quad (8.36)$$

Si noti che il primo termine del secondo membro non ci sarebbe se adottassimo le regole del calcolo ordinario: in pratica la formula di Itô richiede la presenza di un

ulteriore termine di trascinamento del tutto estraneo ad una teoria non stocastica. L'EDS (8.36) ha ora dei coefficienti costanti come l'equazione (8.30) trattata nella Sezione 8.5.1, e quindi la sua soluzione è semplicemente

$$Y(t) = Y_0 - \frac{Dc^2}{2} (t - t_0) + c[W(t) - w_0] \quad (8.37)$$

Possiamo a questo punto ritornare al processo iniziale con $X_0 = e^{Y_0}$ per il quale troviamo

$$X(t) = e^{Y(t)} = X_0 e^{-Dc^2(t-t_0)/2} e^{c[W(t)-w_0]}$$

Il processo $Y(t)$ risulta Gaussiano se $Y(t_0)$ lo è, e in particolare se la condizione iniziale è degenera $Y(t_0) = y_0$, \mathbf{P} -qo; $X(t)$ invece è Markoviano ma non Gaussiano. Potremo comunque calcolare le ddp di transizione per $Y(t)$ e per $X(t)$, e successivamente anche le altre distribuzioni mediante le equazioni di Chapman-Kolmogorov

Proposizione 8.21. *La legge al tempo t delle soluzioni dell'EDS (8.35) con condizione iniziale degenera*

$$dX(t) = cX(t) dW(t) \quad X(t_0) = x_0 > 0, \quad \mathbf{P}\text{-qo} \quad (8.38)$$

è la ddp di transizione log-normale

$$\ln \mathfrak{N} \left(\ln x_0 - \frac{Dc^2}{2} (t - t_0), Dc^2(t - t_0) \right) \quad (8.39)$$

e inoltre

$$\mathbf{cov} [X(s), X(t)] = x_0^2 \left(e^{Dc^2(s-t_0) \wedge (t-t_0)} - 1 \right) \quad (8.40)$$

Dimostrazione: Converrà partire da un esame del processo ausiliario $Y(t)$ perchè questo, con condizioni iniziali Gaussiane, è a sua volta Gaussiano e la sua legge è completamente determinata da $\mathbf{E} [Y(t)]$ e $\mathbf{cov} [Y(s), Y(t)]$ che si possono calcolare facilmente. Infatti se $Y_0 \sim \mathfrak{N}(y_0, \sigma_0^2)$ è la v -a iniziale indipendente da $W(t)$, da (8.37) si ha innanzitutto

$$\mathbf{E} [Y(t)] = y_0 - \frac{Dc^2}{2} (t - t_0) \quad (8.41)$$

Posto poi per comodità $\tilde{Y}_0 = Y_0 - y_0$, e

$$\tilde{Y}(t) = Y(t) - \mathbf{E} [Y(t)] = Y_0 - y_0 + c[W(t) - w_0] = \tilde{Y}_0 + c\tilde{W}(t)$$

si ottiene dalle proprietà del processo di Wiener centrato

$$\begin{aligned} \mathbf{cov} [Y(s), Y(t)] &= \mathbf{E} [\tilde{Y}(s)\tilde{Y}(t)] = \mathbf{E} [(\tilde{Y}_0 + c\tilde{W}(s))(\tilde{Y}_0 + c\tilde{W}(t))] \\ &= \mathbf{E} [\tilde{Y}_0^2] + c^2 \mathbf{E} [\tilde{W}(s)\tilde{W}(t)] \\ &= \sigma_0^2 + Dc^2 \min\{t - t_0, s - t_0\} \end{aligned} \quad (8.42)$$

da cui anche $\mathbf{V}[Y(t)] = \sigma_0^2 + Dc^2(t - t_0)$ e quindi al tempo t

$$Y(t) \sim \mathfrak{N} \left(y_0 - \frac{Dc^2}{2}(t - t_0), \sigma_0^2 + Dc^2(t - t_0) \right)$$

Pertanto, tornando al processo $X(t) = e^{Y(t)}$, con la condizione iniziale $X_0 = e^{Y_0} \sim \ln\mathfrak{N}(y_0, \sigma_0^2) = \ln\mathfrak{N}(\ln x_0, \sigma_0^2)$, risulterà al tempo t

$$X(t) \sim \ln\mathfrak{N} \left(\ln x_0 - \frac{Dc^2}{2}(t - t_0), \sigma_0^2 + Dc^2(t - t_0) \right)$$

Il processo $X(t)$ quindi non è più Gaussiano, ma si presenta come una generalizzazione di un processo di Wiener geometrico. Ponendo poi $\sigma_0 = 0$ si ottiene la *ddp* di transizione (8.39) e – con le equazioni di Chapman-Kolmogorov e con la regola di moltiplicazione a catena – la legge completa del processo. Bisogna osservare che, diversamente dal processo $Y(t)$ che è Gaussiano, questa legge globale del processo $X(t)$ non è invece deducibile dalla semplice conoscenza di $\mathbf{E}[X(t)]$ e $\mathbf{cov}[X(s), X(t)]$. Ciononostante calcoleremo per completezza queste quantità, che comunque contengono importanti informazioni sul processo, sempre per il caso $X_0 \sim \ln\mathfrak{N}(y_0, \sigma_0^2)$ con $x_0 = e^{y_0}$. Usando a questo scopo le formule (3.68) per le leggi log-normali, abbiamo innanzitutto

$$\mathbf{E}[X(t)] = x_0 e^{\sigma_0^2/2} \quad \mathbf{V}[X(t)] = x_0^2 e^{\sigma_0^2} \left(e^{\sigma_0^2 + Dc^2(t-t_0)} - 1 \right) \quad (8.43)$$

Per l'autocorrelazione invece osserviamo che $X(s)X(t) = e^{Y(s)+Y(t)}$, e che in base alla Proposizione 4.20 le *v-a* $Y(s) + Y(t)$ risultano comunque sempre normali con medie e varianze facilmente calcolabili. Infatti da (8.41) si ha innanzitutto

$$\mathbf{E}[Y(s) + Y(t)] = 2y_0 - \frac{Dc^2}{2}(s + t - 2t_0)$$

mentre dalla Proposizione 3.29 e da (8.42)

$$\begin{aligned} \mathbf{V}[Y(s) + Y(t)] &= \mathbf{V}[Y(s)] + \mathbf{V}[Y(t)] + 2\mathbf{cov}[Y(s), Y(t)] \\ &= 4\sigma_0^2 + Dc^2[(s + t - 2t_0) + 2\min\{t - t_0, s - t_0\}] \end{aligned}$$

Complessivamente quindi risulta

$$\begin{aligned} X(s)X(t) \sim \ln\mathfrak{N} \left(2y_0 - \frac{Dc^2}{2}(s + t - 2t_0), \right. \\ \left. 4\sigma_0^2 + Dc^2[(s + t - 2t_0) + 2(t - t_0) \wedge (s - t_0)] \right) \end{aligned}$$

per cui da (3.68) si ha

$$\mathbf{E}[X(s)X(t)] = e^{2y_0 + 2\sigma_0^2 + Dc^2(t-t_0) \wedge (s-t_0)} = x_0^2 e^{2\sigma_0^2} e^{Dc^2(t-t_0) \wedge (s-t_0)}$$

e finalmente da (8.43)

$$\mathbf{cov}[X(s), X(t)] = x_0^2 e^{2\sigma_0^2} \left(e^{Dc^2(t-t_0) \wedge (s-t_0)} - 1 \right)$$

Il risultato (8.40) si ottiene infine con condizione iniziale degenera, cioè $\sigma_0 = 0$ \square

8.5.4 **EDS con trascinamento lineare in x**

Assumendo $a(x, t) = -\alpha x$ con $\alpha > 0$, e $b(x, t) = 1$: la nostra *EDS* sarà

$$dX(t) = -\alpha X(t)dt + dW(t) \quad X(t_0) = X_0 \quad (8.44)$$

e, con la solita identificazione dei coefficienti

$$A(x, t) = a(x, t) = -\alpha x \quad B(x, t) = Db^2(x, t) = D$$

corrisponderà all'equazione di Fokker-Planck per processo di Ornstein-Uhlenbeck

$$\partial_t f(x, t) = \alpha \partial_x [xf(x, t)] + \frac{D}{2} \partial_x^2 f(x, t) \quad f(x, t_0) = f_0(x)$$

già introdotta nella Proposizione 7.39, e della quale conosciamo già le soluzioni che ora ricaveremo nuovamente con l'applicazione del calcolo stocastico. Per determinare la soluzione di (8.44) converrà anche in questo caso introdurre un processo ausiliario

$$Y(t) = X(t) e^{\alpha(t-t_0)} \quad Y(t_0) = X_0$$

la cui equazione si determina da (8.28) osservando che per $g(x, t) = xe^{\alpha(t-t_0)}$ si ha

$$g_x(x, t) = e^{\alpha(t-t_0)} \quad g_{xx}(x, t) = 0 \quad g_t(x, t) = \alpha xe^{\alpha(t-t_0)}$$

Ne segue allora che $Y(t)$ soddisfa l'equazione

$$dY(t) = e^{\alpha(t-t_0)} dW(t) \quad (8.45)$$

che è del tipo con coefficienti dipendenti solo dal tempo (8.31) con soluzione

$$Y(t) = X_0 + \int_{t_0}^t e^{\alpha(s-t_0)} dW(s)$$

Ricordando allora che $X(t) = e^{-\alpha(t-t_0)} Y(t)$, la soluzione dell'equazione (8.44) sarà

$$X(t) = X_0 e^{-\alpha(t-t_0)} + \int_{t_0}^t e^{-\alpha(t-s)} dW(s) \quad (8.46)$$

e sarà quindi Gaussiana se X_0 lo è, in particolare se $X_0 = x_0$, **P**-qo. Se quindi supponiamo $X_0 \sim \mathfrak{N}(x_0, \sigma_0^2)$, potremo ricavare tutte le proprietà del processo $X(t)$ calcolandone l'attesa e l'autocovarianza: la legge globale del processo seguirà poi dalle equazioni di Chapman-Kolmogorov e dalla regola di moltiplicazione a catena

Proposizione 8.22. *La soluzione $X(t)$ dell'EDS*

$$dX(t) = -\alpha X(t)dt + dW(t) \quad X(t_0) = x_0, \quad \mathbf{P}\text{-qo} \quad (8.47)$$

è un processo Gaussiano con $(\beta^2 = D/2\alpha)$

$$\mathbf{E}[X(t)] = x_0 e^{-\alpha(t-t_0)} \quad \mathbf{cov}[X(s), X(t)] = \beta^2 (-e^{-\alpha(s+t-2t_0)} + e^{-\alpha|t-s|}) \quad (8.48)$$

e quindi la *ddp* di transizione è $\mathfrak{N}(x_0 e^{-\alpha(t-t_0)}, \beta^2(1 - e^{-2\alpha(t-t_0)}))$

Dimostrazione: Siccome la soluzione con condizione iniziale Gaussiana è ancora Gaussiana, basterà calcolarne attesa e covarianza, e cominceremo supponendo che $X_0 \sim \mathfrak{N}(x_0, \sigma_0^2)$: dalla (8.17) si ha innanzitutto

$$\mathbf{E}[X(t)] = \mathbf{E}[X_0] e^{-\alpha(t-t_0)} = x_0 e^{-\alpha(t-t_0)} \quad (8.49)$$

Quanto all'autocovarianza invece, da (8.17), (8.18) e dalle indipendenze dei processi su intervalli non sovrapposti si ha

$$\begin{aligned} \mathbf{cov}[X(s), X(t)] &= \mathbf{E}[(X(t) - \mathbf{E}[X(t)])(X(s) - \mathbf{E}[X(s)])] \\ &= \mathbf{E}\left[\left((X_0 - x_0)e^{-\alpha(t-t_0)} + \int_{t_0}^t e^{-\alpha(t-t')} dW(t')\right) \cdot \right. \\ &\quad \left. \left((X_0 - x_0)e^{-\alpha(s-t_0)} + \int_{t_0}^s e^{-\alpha(s-s')} dW(s')\right)\right] \\ &= \mathbf{V}[X_0] e^{-\alpha(s+t-2t_0)} \\ &\quad + \mathbf{E}\left[\int_{t_0}^t e^{-\alpha(t-t')} dW(t') \int_{t_0}^s e^{-\alpha(s-s')} dW(s')\right] \\ &= \sigma_0^2 e^{-\alpha(s+t-2t_0)} + D \int_{t_0}^{s \wedge t} e^{-\alpha(t+s-2t')} dt' \\ &= \sigma_0^2 e^{-\alpha(s+t-2t_0)} + D e^{-\alpha(t+s)} \frac{e^{2\alpha(s \wedge t)} - e^{2\alpha t_0}}{2\alpha} \\ &= \left(\sigma_0^2 - \frac{D}{2\alpha}\right) e^{-\alpha(s+t-2t_0)} + \frac{D}{2\alpha} e^{-\alpha|t-s|} \end{aligned}$$

essendo facilmente verificato che $s+t-2(s \wedge t) = |t-s|$. In questo modo il processo $X(t)$ è completamente determinato e in particolare la sua varianza è

$$\mathbf{V}[X(t)] = \mathbf{cov}[X(t), X(t)] = \sigma_0^2 e^{-2\alpha(t-t_0)} + \frac{D}{2\alpha} (1 - e^{-2\alpha(t-t_0)}) \quad (8.50)$$

Questi risultati permettono di ricavare anche la ddp di transizione scegliendo come condizione iniziale $X_0 = x_0$, \mathbf{P} -qo, cioè $\sigma_0^2 = 0$: infatti ponendo $\beta^2 = D/2\alpha$ facilmente si ottiene una legge $\mathfrak{N}(x_0 e^{-\alpha(t-t_0)}, \beta^2(1 - e^{-2\alpha(t-t_0)}))$ coincidente con la ben nota espressione (7.54) date in precedenza per il processo di Ornstein-Uhlenbeck. Si noti la relativa facilità con la quale si ricava questo risultato dall'*EDS* (8.44) invece che da una soluzione meno elementare della corrispondente equazione di Fokker-Planck 7.39. \square

Capitolo 9

Teoria dinamica del moto Browniano

Nel 1930 L.S. Ornstein e G.F. Uhlenbeck affrontarono nuovamente il problema del moto Browniano sviluppando in maniera più dettagliata l'equazione dinamica di Langevin per studiare il fenomeno a scale di tempi più piccole di quelle ipotizzate da Einstein e Smoluchowski nel 1905-6. Noi daremo ora, adattandola alle nostre notazioni, un resoconto della teoria di Ornstein-Uhlenbeck, ed esamineremo sotto quali condizioni la teoria di Einstein-Smoluchowski è una buona approssimazione.

9.1 Particella Browniana libera

Nella teoria di Ornstein-Uhlenbeck la posizione della particella Browniana è rappresentata mediante un processo $X(t)$ che viene supposto derivabile in modo che esista sempre il processo della velocità $V(t) = \dot{X}(t)$. Riprendendo allora la discussione della Sezione 6.4.2 potremo scrivere le equazioni di Newton di una particella Browniana libera sferica, di massa m e diametro a , come un sistema di equazioni differenziali del tipo

$$\dot{X}(t) = V(t) \tag{9.1}$$

$$m\dot{V}(t) = -6\pi\eta aV(t) + B(t) \tag{9.2}$$

dove, come visto nella Sezione 8.1, $B(t)$ è un rumore bianco Wieneriano, mentre η è la viscosità del mezzo. In particolare la (9.2) indica che sulla nostra particella agiscono due tipi di forze: una forza di attrito proporzionale alla velocità $V(t)$, e una forza aleatoria rappresentata da un rumore bianco. Data la natura singolare di $B(t)$ sappiamo però che il nostro sistema deve più correttamente essere scritto in termini di *EDS* come

$$dX(t) = V(t) dt \tag{9.3}$$

$$dV(t) = -\alpha V(t) dt + dW(t) \tag{9.4}$$

dove abbiamo posto

$$\alpha = \frac{6\pi\eta a}{m} \quad (9.5)$$

mentre $W(t)$ è ora un processo di Wiener con un opportuno coefficiente di diffusione D . Si noti inoltre che il sistema (9.3), (9.4) è essenzialmente disaccoppiato perchè la $X(t)$ non compare nella seconda equazione: questo ci consentirà di trattare le due equazioni separatamente, risolvendo prima la (9.4) e utilizzandone poi la soluzione in (9.3). D'altra parte con un banale adeguamento della notazione si vede subito che l'EDS (9.4) coincide con la (8.44) trattata nella Sezione 8.5.4 per cui, scegliendo per semplicità $t_0 = 0$, e assegnando le condizioni iniziali $X(0) = x_0$, e $V(0) = v_0$, le soluzioni del nostro sistema si scrivono immediatamente come

$$X(t) = x_0 + \int_0^t V(s) ds \quad (9.6)$$

$$V(t) = v_0 e^{-\alpha t} + \int_0^t e^{-\alpha(t-s)} dW(s) \quad (9.7)$$

Proposizione 9.1. *La velocità $V(t)$ di un moto Browniano libero è un processo di Ornstein-Uhlenbeck con*

$$\mathbf{E}[V(t)] = v_0 e^{-\alpha t} \quad (9.8)$$

$$\mathbf{cov}[V(s), V(t)] = \beta^2 (e^{-\alpha|s-t|} - e^{-\alpha(s+t)}) \quad (9.9)$$

dove $\beta^2 = D/2\alpha$ e, con k costante di Boltzmann e T temperatura assoluta, si ha

$$\beta^2 = \frac{kT}{m} \quad (9.10)$$

La posizione $X(t)$ è invece un processo Gaussiano, ma non Markoviano con

$$\mathbf{E}[X(t)] = x_0 + \frac{v_0}{\alpha} (1 - e^{-\alpha t}) \quad (9.11)$$

$$\mathbf{cov}[X(s), X(t)] = \frac{\beta^2}{\alpha^2} \left[2\alpha(s \wedge t) - 2 + 2e^{-\alpha s} + 2e^{-\alpha t} - e^{-\alpha|s-t|} - e^{-\alpha(s+t)} \right] \quad (9.12)$$

Dimostrazione: Che il processo $V(t)$ in (9.7) sia un processo di Ornstein-Uhlenbeck con attesa (9.8) e autocovarianza (9.9) è già stato mostrato nella Sezione 8.5.4: anche le proprietà generali di tale processo sono già state completamente esposte nelle Sezioni 7.1.9 e 7.2.4. Quanto alla (9.10) bisogna ricordare che secondo la Proposizione 7.27 per $t \rightarrow +\infty$ la distribuzione della velocità del processo di Ornstein-Uhlenbeck tende verso la legge stazionaria $\mathfrak{N}(0, \beta^2)$. Pertanto in questa condizione di equilibrio termodinamico potremo applicare il principio di equipartizione dell'energia nella forma

$$\frac{1}{2} kT = \frac{1}{2} m\beta^2$$

da cui la (9.10) segue immediatamente. Dalla (9.6) inoltre si ricava che il processo della posizione $X(t)$ è Gaussiano per cui, come abbiamo mostrato nella Sezione 7.1.10, la sua legge è completamente determinata dalla conoscenza dell'attesa e dell'autocovarianza. Per provare allora che per la posizione $X(t)$ valgono (9.11) e (9.12), osserviamo innanzitutto che da (9.6) si ha

$$\mathbf{E}[X(t)] = x_0 + \int_0^t \mathbf{E}[V(s)] ds = x_0 + v_0 \int_0^t e^{-\alpha s} ds = x_0 + \frac{v_0}{\alpha} (1 - e^{-\alpha t})$$

cioè la (9.11). Per l'autocovarianza invece si nota prima che

$$\mathbf{cov}[X(s), X(t)] = \int_0^s \int_0^t \mathbf{cov}[V(s'), V(t')] ds' dt' \quad (9.13)$$

Infatti, introducendo per comodità i processi centrati,

$$\tilde{V}(t) = V(t) - \mathbf{E}[V(t)] = V(t) - v_0 e^{-\alpha t} \quad (9.14)$$

$$\tilde{X}(t) = X(t) - \mathbf{E}[X(t)] = \int_0^t V(s) ds - \frac{v_0}{\alpha} (1 - e^{-\alpha t}) = \int_0^t \tilde{V}(s) ds \quad (9.15)$$

si ha facilmente che

$$\begin{aligned} \mathbf{cov}[X(s), X(t)] &= \mathbf{E}[\tilde{X}(s)\tilde{X}(t)] = \int_0^s \int_0^t \mathbf{E}[\tilde{V}(s')\tilde{V}(t')] ds' dt' \\ &= \int_0^s \int_0^t \mathbf{cov}[V(s'), V(t')] ds' dt' \end{aligned}$$

A questo punto da (9.13) e (9.9) risulta

$$\mathbf{cov}[X(s), X(t)] = \beta^2 \int_0^s \int_0^t \left(e^{-\alpha|s'-t'|} - e^{-\alpha(s'+t')} \right) ds' dt'$$

e la (9.12) si ricava con un'integrazione laboriosa ma elementare. Resta da mostrare che $X(t)$ non è un processo di Markov, e noi lo faremo calcolando esplicitamente la ddp di transizione di $X(t)$, e verificando poi che essa non soddisfa le equazioni di Chapman-Kolmogorov. Per calcolare esplicitamente le ddp Gaussianhe congiunte di $X(t)$ chiamiamo per brevità $b(t)$, $a^2(t)$ ed $r(s, t)$ rispettivamente l'attesa, la varianza e il coefficiente di correlazione che si ricavano da (9.11) e (9.12): si vede allora che le ddp a un tempo $f(x, t)$ sono $\mathfrak{N}(b(t), a^2(t))$, e quelle congiunte a due tempi $f(x, t; y, s)$ sono $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ con

$$\mathbf{b} = \begin{pmatrix} b(s) \\ b(t) \end{pmatrix} \quad \mathbb{A} = \begin{pmatrix} a^2(s) & a(s)a(t)r(s, t) \\ a(s)a(t)r(s, t) & a^2(t) \end{pmatrix}$$

A questo punto le ddp di transizione $f(x, t | y, s)$ con $s < t$ si ricavano dalla Proposizione 3.40 e sono $\mathfrak{N}(A(s, t)y + B(s, t), C^2(s, t))$ dove abbiamo posto

$$\begin{aligned} A(s, t) &= r(s, t) \frac{a(t)}{a(s)} & B(s, t) &= b(t) - r(s, t) \frac{a(t)}{a(s)} b(s) \\ C^2(s, t) &= a^2(t)[1 - r^2(s, t)] \end{aligned}$$

La conoscenza delle $f(x, t | y, s)$ ci mette in condizione di verificare direttamente – ma noi non eseguiremo questo calcolo che si presenta piuttosto elaborato – che le equazioni di Chapman-Kolmogorov (7.17) sarebbero soddisfatte solo se risultasse

$$r(s, u)r(u, t) = r(s, t) \quad s < u < t$$

Siccome in base alle (9.12) questo non accade, possiamo concludere che nella teoria di Ornstein-Uhlenbeck il processo della posizione $X(t)$ non è Markoviano in evidente contrasto con quanto avviene nella teoria di Einstein e Smoluchowski nella quale, come rilevato nella Sezione 6.4.1, la posizione è invece descritta da un processo di Wiener che è Markoviano \square

9.2 Confronto con Einstein-Smoluchowski

Per paragonare correttamente la trattazione di Einstein-Smoluchowski descritta nel Capitolo 6.4 con quella di Ornstein-Uhlenbeck qui introdotta bisogna notare subito che nelle due teorie il simbolo D assume due significati differenti per cui saremo obbligati ora a introdurre anche due notazioni distinte:

- nella teoria di Einstein-Smoluchowski chiameremo D_X il coefficiente di diffusione del processo di Wiener $W_X(t)$ che rappresenta direttamente la posizione della particella Browniana; inoltre, dal fatto che la varianza di tale posizione cresce come $D_X t$, ricaviamo che le sue dimensioni fisiche sono

$$[D_X] = \frac{\text{mt}^2}{\text{sec}}$$

mentre da (6.76) sappiamo che il suo valore in termini di costanti fisiche è

$$D_X = \frac{kT}{3\pi\eta a}$$

- nella teoria di Ornstein-Uhlenbeck, invece, chiameremo D_V il coefficiente di diffusione del rumore Wieneriano $W_V(t)$ che disturba l'equazione della velocità, per cui ora $D_V t$ rappresenterà l'andamento della varianza di una velocità e le sue dimensioni fisiche saranno

$$[D_V] = \frac{\text{mt}^2}{\text{sec}^3}$$

mentre da (9.10) sappiamo che il suo valore è

$$D_V = 2\alpha\beta^2 = \frac{2kT\alpha}{m}$$

Ricordando allora che, come osservato in (9.5), nell'equazione (9.4) il coefficiente di attrito è $\alpha = 6\pi\eta a/m$, è immediato verificare che fra i due coefficienti di diffusione sussiste la relazione

$$D_X = \frac{D_V}{\alpha^2}$$

Possiamo a questo punto istituire un confronto fra le due teorie osservando innanzitutto che nel modello di Ornstein-Uhlenbeck la varianza della posizione si ricava da (9.12) e vale

$$\mathbf{V}[X(t)] = \frac{\beta^2}{\alpha^2} (2\alpha t - 3 + 4e^{-\alpha t} - e^{-2\alpha t})$$

e quindi che per $t \gg 1/\alpha$ avremo

$$\mathbf{V}[X(t)] \simeq \frac{2\beta^2}{\alpha} t = \frac{D_V}{\alpha^2} t = D_X t$$

come nella teoria di Einstein-Smoluchowski. Questo suggerisce quindi di considerare quest'ultima come una buona approssimazione della teoria di Ornstein-Uhlenbeck per grandi tempi, o equivalentemente per grandi valori del coefficiente di attrito α

Proposizione 9.2. *Nell'ambito delle notazioni fin qui adottate, se $\alpha \rightarrow +\infty$ (regime sovra-smorzato) e se $2\beta^2/\alpha$ converge verso un valore finito che chiamiamo D_X , allora il processo della posizione $X(t)$ di Ornstein-Uhlenbeck con condizione iniziale $X(0) = x_0$ converge in distribuzione – nel senso della Definizione 5.4 – verso un processo di Wiener $W_X(t)$ con coefficiente di diffusione D_X e $W_X(0) = x_0$*

Dimostrazione: Da (9.11) e (9.12) si vede facilmente che, nel limite indicato nell'enunciato, per ogni fissato t l'attesa e la covarianza del processo Gaussiano $X(t)$ convergono verso

$$\mathbf{E}[X(t)] \rightarrow x_0 \quad \mathbf{cov}[X(s), X(t)] \rightarrow D_X(s \wedge t)$$

ovvero $X(t)$ converge in distribuzione verso un processo di Wiener $W_X(t)$ con coefficiente di diffusione D_X e condizione iniziale $W_X(0) = x_0$ \square

In conclusione potremo ritenere da ora in poi che in regime sovra-smorzato, o comunque per tempi $t \gg 1/\alpha$, la posizione del processo che descrive il moto Browniano libero sia un processo di Wiener e soddisfi un'EDS del tipo

$$dX(t) = dW_X(t) \tag{9.16}$$

Si noti in particolare che asintoticamente la posizione $X(t)$ diffonde in maniera isotropa dato che, come si vede da (9.8), la velocità iniziale v_0 viene assorbita dal rumore di fondo e, dopo un breve transitorio, $\mathbf{E}[V(t)] \rightarrow 0$ per $t \gg 1/\alpha$

9.3 Markovianità di Ornstein-Uhlenbeck

Abbiamo visto nella Proposizione 9.1 che nella teoria di Ornstein-Uhlenbeck il processo della velocità $V(t)$ è un processo di Markov, mentre la posizione $X(t)$ non lo è. Da un punto di vista formale questo deriva dal fatto che $V(t)$ soddisfa l'*EDS* di Langevin (9.4) e quindi è Markoviana in base alla Proposizione 8.18, mentre la (9.3) dice solo che $X(t)$ ammette un differenziale stocastico dipendente però da un altro processo, autonomo rispetto alla posizione. Dalla discussione svolta nella Sezione 7.1.1 sappiamo anche però che in generale è possibile recuperare la Markovianità di un processo aggiungendo l'informazione necessaria a questo scopo: in questo caso tipicamente dobbiamo prendere in considerazione processi vettoriali con più di una componente proprio per poter accomodare tutta l'informazione necessaria. Nel nostro caso la via da seguire è indicata dal ben noto fatto che in un sistema dinamico Newtoniano lo stato completo non è definito dalla sola posizione $x(t)$, ma deve essere descritto in uno spazio delle fasi dalla coppia $x(t), v(t)$ di posizione e velocità. Questo ci suggerisce quindi di prendere in considerazione il processo vettoriale nello spazio delle fasi

$$\mathbf{Z}(t) = \begin{pmatrix} X(t) \\ V(t) \end{pmatrix}$$

mediante il quale potremo scrivere il sistema di equazioni (9.3) e (9.4) come un'unica *EDS*

$$d\mathbf{Z}(t) = \mathbf{a}(\mathbf{Z}(t)) dt + \mathbb{C} d\mathbf{W}(t) \quad (9.17)$$

per la quale abbiamo posto

$$\mathbf{a}(\mathbf{z}) = \mathbf{a}(x, v) = \begin{pmatrix} v \\ -\alpha v \end{pmatrix} \quad \mathbb{C} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad (9.18)$$

mentre $\mathbf{W}(t)$ è un processo di Wiener vettoriale con

$$\mathbf{W}(t) = \begin{pmatrix} W_X(t) \\ W_V(t) \end{pmatrix}$$

Per ragioni di brevità noi non abbiamo trattato in precedenza il caso di ***EDS vettoriali*** come la (9.17), che più in generale possono essere scritte nella forma

$$d\mathbf{Z}(t) = \mathbf{a}(\mathbf{Z}(t), t) dt + \mathbb{C}(\mathbf{Z}(t), t) d\mathbf{W}(t) \quad (9.19)$$

ma daremo intuitivamente per acquisito che – con qualche complicazione della notazione – la maggior parte dei risultati enunciati nelle sezioni precedenti valga anche per le *EDS* del tipo (9.19). La soluzione di (9.17) con la condizione iniziale

$$\mathbf{Z}(0) = \mathbf{z}_0 = \begin{pmatrix} x_0 \\ v_0 \end{pmatrix} \quad (9.20)$$

è ovviamente costituita dal vettore $\mathbf{Z}(t)$ le cui componenti sono le soluzioni (9.6) e (9.7) già determinate in precedenza, ma in questa nuova formulazione viene messo

in luce un aspetto trascurato nella discussione della Sezione 9.1: la necessità di determinare anche la *correlazione* incrociata fra i due processi $X(t)$ e $V(t)$, e più in generale le loro *leggi congiunte*, e non solo le loro marginali separate

Proposizione 9.3. *La covarianza incrociata dei processi $X(t), V(t)$ di Ornstein-Uhlenbeck è*

$$\mathbf{cov}[X(s), V(t)] = \frac{\beta^2}{\alpha} \left[1 + \frac{|t-s|}{t-s} (e^{-\alpha|t-s|} - 1) - 2e^{-\alpha t} + e^{-\alpha(t+s)} \right] \quad (9.21)$$

Dimostrazione: Utilizzando di nuovo i processi centrati (9.14) e (9.15) si vede prima di tutto che

$$\begin{aligned} \mathbf{cov}[X(s), V(t)] &= \mathbf{E} \left[\tilde{X}(s) \tilde{V}(t) \right] = \mathbf{E} \left[\tilde{V}(t) \int_0^s \tilde{V}(t') dt' \right] \\ &= \int_0^s \mathbf{E} \left[\tilde{V}(t) \tilde{V}(t') \right] dt' = \int_0^s \mathbf{cov}[V(t), V(t')] dt' \end{aligned}$$

e quindi da (9.9) potremo scrivere

$$\mathbf{cov}[X(s), V(t)] = \beta^2 \int_0^s (e^{-\alpha|t-t'|} - e^{-\alpha(t+t')}) dt'$$

da cui il risultato (9.21) segue con un'integrazione elementare □

Proposizione 9.4. *La soluzione $\mathbf{Z}(t)$ dell'EDS (9.17) con condizione iniziale (9.20) è un processo di Markov vettoriale e Gaussiano con legge congiunta $\mathfrak{N}(\mathbf{b}, \mathbb{A})$ all'istante t , dove*

$$\mathbf{b} = \begin{pmatrix} \mathbf{E}_0[X(t)] \\ \mathbf{E}_0[V(t)] \end{pmatrix} = \begin{pmatrix} x_0 + v_0 (1 - e^{-\alpha t}) / \alpha \\ v_0 e^{-\alpha t} \end{pmatrix} \quad (9.22)$$

$$\begin{aligned} \mathbb{A} &= \begin{pmatrix} \mathbf{V}_0[X(t)] & \mathbf{cov}_0[X(t), V(t)] \\ \mathbf{cov}_0[X(t), V(t)] & \mathbf{V}_0[V(t)] \end{pmatrix} \\ &= \frac{\beta^2}{\alpha^2} \begin{pmatrix} 2\alpha t - 3 + 4e^{-\alpha t} - e^{-2\alpha t} & \alpha(1 - 2e^{-\alpha t} + e^{-2\alpha t}) \\ \alpha(1 - 2e^{-\alpha t} + e^{-2\alpha t}) & \alpha^2(1 - e^{-2\alpha t}) \end{pmatrix} \end{aligned} \quad (9.23)$$

Dimostrazione: Se $\mathbf{Z}(t)$ soddisfa la (9.17), una generalizzazione della Proposizione 8.18 ci permette di affermare che tale soluzione è anche un processo di Markov vettoriale. Sappiamo anche dalla Proposizione 9.1 che le componenti $X(t)$ e $V(t)$ di $\mathbf{Z}(t)$ sono separatamente processi Gaussiani, ma questo come è noto non è sufficiente per affermare che tali componenti siano anche *congiuntamente* Gaussiane: noi per il momento lo ammetteremo senza dimostrazione rimandando alla proposizione successiva l'indicazione di una procedura per verificarlo. Da (9.21) e dalle (9.9) e (9.12) si ricava allora che il vettore delle medie e la matrice delle covarianze del vettore $\mathbf{Z}(t)$ sono (9.22) e (9.23) □

Anche la Proposizione 8.19 che associa *EDS* ed equazioni di Fokker-Planck può essere generalizzata al caso di processi vettoriali del tipo (9.19), ma naturalmente ora l'equazione di Fokker-Planck avrà la forma di una equazione multivariata del tipo (7.77) per la quale – almeno per il caso in cui sia presente un solo rumore Wieneriano – i coefficienti saranno dati con le regole seguenti che generalizzano le (8.29)

$$\mathbf{A}(\mathbf{x}, t) = \mathbf{a}(\mathbf{x}, t) \quad \mathbb{B}(\mathbf{x}, t) = D \mathbb{C}(\mathbf{x}, t) \mathbb{C}^T(\mathbf{x}, t) \quad (9.24)$$

dove \mathbb{C}^T è la trasposta di \mathbb{C} . Questa osservazione permette di dimostrare facilmente la proposizione che segue

Proposizione 9.5. *Le ddp congiunte del vett-a $\mathbf{Z}(t)$ soluzione dell'EDS (9.17) con condizioni iniziali (9.20) soddisfano nello spazio delle fasi l'equazione di Fokker-Planck*

$$\begin{aligned} \partial_t f(x, v, t) &= -v \partial_x f(x, v, t) + \alpha \partial_v [v f(x, v, t)] + \frac{D}{2} \partial_v^2 f(x, v, t) \quad (9.25) \\ f(x, v, 0) &= \delta(x - x_0) \delta(v - v_0) \end{aligned}$$

Dimostrazione: Basterà scrivere l'equazione di Fokker-Planck multivariata (7.77) tenendo conto di (9.24) e di (9.18) □

Si noti che viceversa le *ddp* del processo $X(t)$ isolato non soddisfano nessuna equazione differenziale alle derivate parziali di questo tipo dato che $X(t)$ non è un processo di Markov. L'equazione (9.25) permette infine una conferma diretta del risultato enunciato nella Proposizione 9.4 secondo il quale il vettore $\mathbf{Z}(t)$ è Gaussiano: sarebbe infatti sufficiente scrivere esplicitamente le *ddp* di $\mathbf{Z}(t)$ dalle (9.22) e (9.23) e poi controllare che esse sono soluzioni di (9.25). Noi però per brevità ci asterremo da questa verifica

9.4 Particella Browniana in un campo di forze

Supponiamo ora che la nostra particella Browniana sia immersa in un campo di forze esterno, e quindi che il sistema di equazioni (9.3) e (9.4) sia modificato in

$$dX(t) = V(t) dt \quad (9.26)$$

$$dV(t) = \gamma(X(t), t) dt - \alpha V(t) dt + dW_V(t) \quad (9.27)$$

dove γ è un nuovo termine con le dimensioni di un'accelerazione che descrive il nostro campo di forze. Il nuovo sistema potrà essere ancora scritto come un'unica *EDS* vettoriale della forma (9.17) per $\mathbf{Z}(t)$ con coefficienti

$$\mathbf{a}(\mathbf{z}) = \mathbf{a}(x, v) = \begin{pmatrix} v \\ \gamma(x, t) - \alpha v \end{pmatrix} \quad \mathbb{C} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad (9.28)$$

e pertanto $\mathbf{Z}(t)$, in quanto soluzione di (9.17), risulta ancora essere un processo di Markov vettoriale, ma questa volta le due equazioni che compongono il sistema sono accoppiate in una maniera che non consente più di risolverle separatamente una dopo l'altra. Una trattazione generale del problema presenterebbe quindi maggiori difficoltà rispetto al caso libero trattato in precedenza, ma noi, più che alle soluzioni generali, siamo interessati a un'indagine sulla possibilità di estendere – sotto opportune condizioni – la formulazione approssimata che abbiamo discusso nella Sezione 9.2. Dalla Proposizione 9.2 sappiamo infatti che per il moto Browniano libero un processo di Wiener sullo spazio delle configurazioni (posizioni x) è – sotto opportune condizioni – una buona approssimazione per la posizione del processo di Markov vettoriale $\mathbf{Z}(t)$ sullo spazio delle fasi x, v . Nel caso del moto Browniano in un campo di forze tale approssimazione Markoviana sullo spazio delle configurazioni è stata trovata da Smoluchowski e noi ne descriveremo nel seguito le caratteristiche principali.

Partiamo prima di tutto dal caso più elementare in cui il campo di forze

$$\gamma(x, t) = \gamma_0$$

è costante in modo che le due equazioni del nostro sistema

$$dX(t) = V(t) dt \tag{9.29}$$

$$dV(t) = [\gamma_0 - \alpha V(t)] dt + dW_V(t) \tag{9.30}$$

siano di nuovo disaccoppiate e possano essere risolte facilmente come nel caso libero. L'unica differenza è ora la presenza della costante γ_0 che però può essere riassorbita definendo un nuovo processo per la velocità

$$V_\gamma(t) = V(t) - \frac{\gamma_0}{\alpha}$$

per il quale sia soddisfatta la nuova equazione

$$dV_\gamma(t) = -\alpha V_\gamma(t) dt + dW_V(t)$$

che sostituisce (9.30) e che è formalmente identica all'equazione di Ornstein-Uhlenbeck (9.4) per $V(t)$ nel caso libero. La soluzione per $V_\gamma(t)$ è quindi ancora della forma (9.7), e pertanto da (9.8) si deduce che, con qualunque condizione iniziale e per tempi $t \gg 1/\alpha$, la $V_\gamma(t)$ tenderà in media ad annullarsi, mentre corrispondentemente la velocità $V(t)$ tenderà verso la costante γ_0/α . A questo punto potremo anche affermare che – dopo un breve transitorio iniziale – la posizione $X(t)$ del processo di Markov $\mathbf{Z}(t)$ soddisferà l'equazione

$$dX(t) = \frac{\gamma_0}{\alpha} dt + dW_X(t) \tag{9.31}$$

che generalizza la (9.16) del caso libero, e che ha come soluzione un processo di Wiener sovrapposto ad un trascinamento costante $\gamma_0 t/\alpha$.

Il passo successivo consiste nell'osservare che questa discussione suggerisce di estendere il risultato anche al caso in cui il campo $\gamma(X(t), t)$ varia *lentamente* sulle scale di tempi $1/\alpha$ caratteristiche del modello, e può quindi essere ritenuto approssimativamente costante. In questo modo si ottiene l'**equazione di Smoluchowski**

$$dX(t) = \frac{\gamma(X(t), t)}{\alpha} dt + dW_X(t) \quad (9.32)$$

che è quindi alla base di una teoria approssimata nella quale la dinamica compare solo come un termine di trascinamento, e la posizione si comporta come un processo di Markov. L'equazione di Smoluchowski descrive così nello spazio delle configurazioni una teoria dinamica con molte caratteristiche interessanti

Esempio 9.6. Forza di richiamo elastica: *L'approssimazione di Smoluchowski fornisce delle soluzioni del tutto accettabili nel caso in cui il campo di forze ha la forma lineare di una forza di richiamo elastica*

$$\gamma(x, t) = -\omega^2 x \quad (9.33)$$

per cui le equazioni della teoria di Ornstein-Uhlenbeck sono

$$dX(t) = V(t) dt \quad (9.34)$$

$$dV(t) = -\omega^2 X(t) dt - \alpha V(t) dt + dW_V(t) \quad (9.35)$$

ovvero

$$d\mathbf{Z}(t) = \mathbf{a}(\mathbf{Z}(t), t) dt + \mathbb{C}(\mathbf{Z}(t), t) d\mathbf{W}(t)$$

con

$$\mathbf{a}(\mathbf{z}) = \mathbf{a}(x, v) = \begin{pmatrix} v \\ -\omega^2 x - \alpha v \end{pmatrix} \quad \mathbb{C} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Le soluzioni di (9.34) e (9.35) possono essere calcolate esplicitamente¹, ma sono piuttosto complicate e noi eviteremo di discuterle. È più interessante invece osservare che la soluzione della corrispondente equazione di Smoluchowski

$$dX(t) = -\frac{\omega^2}{\alpha} X(t) dt + dW_X(t) \quad (9.36)$$

è piuttosto semplice e realistica². Infatti la (9.36) – con un'opportuna ridefinizione dei coefficienti – si presenta come un'equazione di Ornstein-Uhlenbeck (8.44) per la posizione $X(t)$ la quale sarà ora un processo di Markov, Gaussiano con legge $\mathfrak{N}(x_0 e^{-\omega^2 t/\alpha}, \beta^2(1 - e^{-2\omega^2 t/\alpha}))$ e con

$$\beta^2 = \frac{\alpha D_X}{2\omega^2} = \frac{kT}{m\omega^2}$$

¹S. Chandrasekhar, *Rev. Mod. Phys.* **15** (1943) 1

²Il comportamento di un moto Browniano soggetto ad una forza di richiamo elastica è anche stato investigato sperimentalmente con degli ingegnosi esperimenti di E. Kappler, *Ann. Phys.* **11** (1931) 233, che confermano l'idea che l'approssimazione di Smoluchowski è valida quando l'attrito α è grande

dove come al solito k è la costante di Boltzmann e T la temperatura. Asintoticamente tale processo ammette poi una distribuzione stazionaria $\mathfrak{N}(0, \beta^2)$ che descrive una situazione nella quale – trascorso il transitorio, o comunque in regime sovrasmorzato – il nostro processo non diffonde più in maniera illimitata a causa del contrasto esercitato dalla forza elastica che lo lega

L'approssimazione di Smoluchowski (9.32) ha una validità che non si limita solo al caso delle forze di richiamo elastiche: una formulazione più precisa di questa affermazione è contenuta nell'enunciato che segue nel quale introduciamo anche la velocità

$$a(x, t) = \frac{\gamma(x, t)}{\alpha}$$

Proposizione 9.7. *Se la funzione $a(x, t)$ soddisfa ragionevoli condizioni di regolarità, dette $X(t)$ e $V(t)$ le soluzioni del sistema di EDS*

$$\begin{aligned} dX(t) &= V(t) dt & X(0) &= x_0 \\ dV(t) &= \alpha a(X(t), t) dt - \alpha V(t) dt + \alpha dW(t) & V(0) &= v_0 \end{aligned}$$

e detta $Y(t)$ la soluzione dell'EDS

$$dY(t) = a(Y(t), t) dt + dW(t) \quad Y(0) = x_0$$

per ogni fissato v_0 si ha

$$\lim_{\alpha \rightarrow \infty} X(t) = Y(t) \quad \mathbf{P}\text{-}qo$$

uniformemente in t in tutti i compatti di $[0, +\infty)$

Dimostrazione: Omessa³ □

9.5 Distribuzione di Boltzmann

Nella equazione di Smoluchowski di un moto Browniano in un campo di forze

$$dX(t) = \frac{\gamma(X(t), t)}{\alpha} dt + dW(t) \tag{9.37}$$

la dinamica esterna rappresentata da $\gamma(x, t)$ si riflette solo nella forma della velocità di trascinamento $a = \gamma/\alpha$, mentre è completamente assente nel termine di diffusione $b = 1$. In questa sezione prenderemo in considerazione il caso di forze esterne $m\gamma(x)$ indipendenti dal tempo per le quali potremo supporre l'esistenza di un'energia potenziale $\phi(x)$ tale che

$$m\gamma(x) = -\phi'(x) \tag{9.38}$$

³E. Nelson, DYNAMICAL THEORIES OF BROWNIAN MOTION, Princeton UP (Princeton, 1967)

in modo che (9.37) diviene

$$dX(t) = -\frac{\phi'(X(t))}{\alpha m} dt + dW(t)$$

D'altra parte da (6.76) e (9.5) otteniamo che

$$\frac{1}{\alpha m} = \frac{D}{2kT} = \frac{D\beta}{2}$$

dove il parametro termodinamico $1/kT$ tradizionalmente indicato con il simbolo β (che abbiamo preferito conservare) non deve essere confuso con l'omonimo parametro del processo di Ornstein-Uhlenbeck trattato nelle sezioni precedenti. Potremo pertanto scrivere l'equazione di Smoluchowski nella forma

$$dX(t) = -\frac{D}{2} \beta \phi'(X(t)) dt + dW(t)$$

in modo che dalla Proposizione 8.19 con $a(x, t) = -\frac{D}{2} \beta \phi'(x)$, e $b(x, t) = 1$ si ottiene per il nostro moto Browniano immerso in un potenziale $\phi(x)$ la seguente equazione di Fokker-Planck

$$\partial_t f(x, t) = \frac{D}{2} \partial_x [\beta \phi'(x) f(x, t)] + \frac{D}{2} \partial_x^2 f(x, t) \quad (9.39)$$

Proposizione 9.8. *La soluzione stazionaria dell'equazione (9.39) – se esiste – è la **distribuzione di Boltzmann***

$$f(x) = \frac{e^{-\beta \phi(x)}}{Z(\beta)} \quad (9.40)$$

dove la costante di normalizzazione

$$Z(\beta) = \int_{-\infty}^{+\infty} e^{-\beta \phi(x)} dx \quad (9.41)$$

prende il nome di **funzione di partizione**

Dimostrazione: Innanzitutto verifichiamo che la distribuzione di Boltzmann soddisfa l'equazione (9.39): a questo scopo basterà osservare da (9.40) che $\partial_t f = 0$, e che $\partial_x f = -\beta \phi' f$. Viceversa, se $f(x)$ è soluzione stazionaria, si ha $\partial_t f = 0$ e quindi da (9.39) si ricava che $f(x)$ deve soddisfare l'equazione del primo ordine

$$\beta \phi'(x) f(x) + f'(x) = C$$

dove C è una costante di integrazione. D'altra parte, per essere una *ddp* normalizzabile, $f(x)$ deve essere infinitesima per $x \rightarrow \pm\infty$. Supponendo allora che f si annulli abbastanza rapidamente all'infinito con la sua derivata prima, in modo tale

che il primo membro dell'equazione sia complessivamente infinitesimo per $x \rightarrow \pm\infty$, otterremo che $C = 0$ e quindi che la f stazionaria deve soddisfare l'equazione

$$\beta\phi'(x)f(x) + f'(x) = 0$$

la cui soluzione si calcola con metodi elementari e, una volta normalizzata, coincide con la distribuzione di Boltzmann (9.40) \square

Esempio 9.9. Forza di richiamo elastica: Riprendendo l'Esempio 9.6 con γ data da (9.33), si vede da (9.38) che ϕ è il potenziale dell'oscillatore armonico

$$\phi(x) = \frac{1}{2} m\omega^2 x^2 \quad (9.42)$$

per cui con la notazione adottata in questa sezione l'equazione di Smoluchowski (9.36) assume la forma

$$dX(t) = -\frac{D}{2} \beta m\omega^2 X(t) dt + dW(t) \quad (9.43)$$

Dalla Proposizione 9.8 si ottiene allora la distribuzione di Boltzmann

$$\begin{aligned} Z(\beta) &= \sqrt{\frac{2\pi}{\beta m\omega^2}} = \sqrt{\frac{2\pi kT}{m\omega^2}} \\ f(x) &= \frac{e^{-\frac{1}{2}\beta m\omega^2 x^2}}{\sqrt{\frac{2\pi}{\beta m\omega^2}}} = \frac{e^{-m\omega^2 x^2/2kT}}{\sqrt{\frac{2\pi kT}{m\omega^2}}} \end{aligned}$$

che ovviamente coincide con la soluzione stazionaria $\mathfrak{N}(0, \frac{kT}{m\omega^2})$ dell'equazione di Ornstein-Uhlenbeck (9.36) esaminata nella sezione precedente.

Esempio 9.10. Forza peso: Supporremo ora che l'accelerazione sia costante $\gamma(x) = -g$ e che il processo avvenga solo sul semiasse positivo $x \geq 0$. Se immaginiamo che x rappresenti l'altezza della particella al di sopra del suolo posto in $x = 0$ questo modello descriverà la distribuzione delle particelle Browniane soggette alla forza peso. In questo caso da (9.38) si ottiene per il potenziale $\phi(x) = mgx$, mentre l'equazione di Smoluchowski (9.37) diviene

$$dX(t) = -\frac{D}{2} \beta mg dt + dW(t)$$

ed è del tipo a coefficienti costanti già discusso nella Sezione 8.5.1, ma con la condizione aggiuntiva $x \geq 0$, cioè $f = 0$ per $x < 0$, per cui ora la soluzione non potrà più essere Gaussiana. La corrispondente equazione di Fokker-Planck (9.39) è

$$\partial_t f(x, t) = \frac{D}{2} \beta mg \partial_x f(x, t) + \frac{D}{2} \partial_x^2 f(x, t) \quad x \geq 0$$

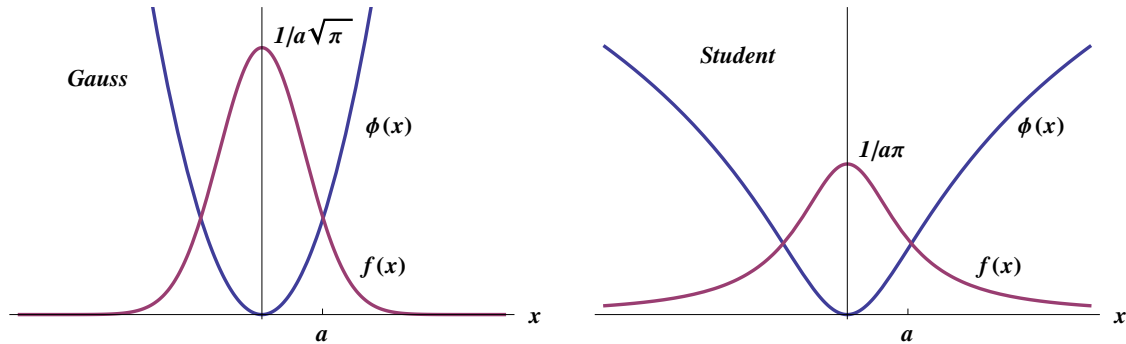


Figura 9.1: Distribuzioni stazionarie di Gauss e Student rispettivamente per le equazioni di Smoluchowski (9.43) e (9.46). La temperatura T è scelta in modo che $2kT = m\omega^2 a^2$: ciò implica in particolare che la legge di Student coincide con la legge di Cauchy. La scala delle energie è invece fissata in maniera convenzionale per rendere paragonabili le curve sovrapposte.

e dalla *Proposizione 9.8* si ottiene per la soluzione stazionaria

$$\begin{aligned} Z(\beta) &= \frac{1}{\beta mg} = \frac{kT}{mg} \\ f(x) &= \beta mg e^{-\beta mg x} \vartheta(x) = \frac{mg}{kT} e^{-mgx/kT} \vartheta(x) \end{aligned}$$

dove ϑ è la *funzione di Heaviside* (2.13): pertanto la legge stazionaria è un'esponenziale $\mathfrak{E}(\beta mg) = \mathfrak{E}\left(\frac{mg}{kT}\right)$.

La *Proposizione 9.8* permette infine di formulare e risolvere il seguente **problema inverso**: *determinare la dinamica (il potenziale ϕ) di un moto Browniano che ammetta come soluzione stazionaria un'assegnata distribuzione di Boltzmann (9.40)*

Esempio 9.11. Leggi di Student: *La famiglia $\mathfrak{T}(\beta)$ di leggi di Boltzmann*

$$f(x) = \frac{1}{a B\left(\frac{1}{2}, \frac{\beta m \omega^2 a^2 - 1}{2}\right)} \left(\frac{a^2}{a^2 + x^2}\right)^{\frac{1}{2} \beta m \omega^2 a^2} = \frac{e^{-\beta \phi(x)}}{Z(\beta)} \quad (9.44)$$

è una generalizzazione delle distribuzioni di Student \mathfrak{T}_n introdotte alla fine della *Sezione 3.5.2*: qui $a > 0$ è una lunghezza caratteristica, $\omega > 0$ è un parametro che rappresenta l'intensità del potenziale esterno e

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)}$$

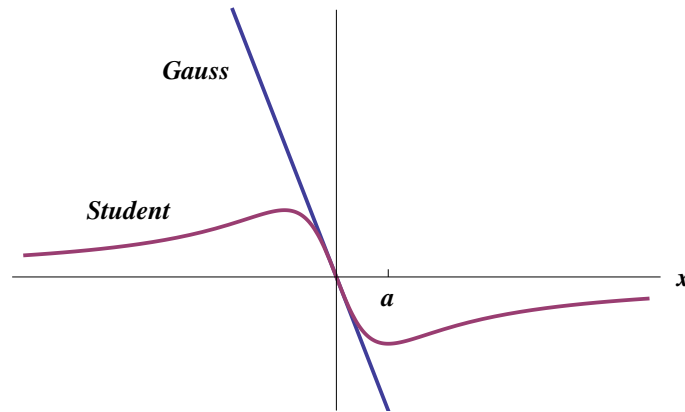


Figura 9.2: Velocità di trascinamento (9.47) rispettivamente per le equazioni di Smoluchowski (9.43) e (9.46). I valori dei parametri coincidono con quelli usati nella Figura 9.1.

è la funzione Beta di Riemann. Queste distribuzioni sono ben definite quando $\beta m \omega^2 a^2 > 1$, ovvero quando $m \omega^2 a^2 > kT$: ciò indica che le nostre soluzioni stazionarie esistono solo se viene rispettata questa relazione fra l'intensità del potenziale ω e la temperatura T . Si vede subito da (9.44) che

$$\phi(x) = \frac{1}{2} m \omega^2 a^2 \ln \left(1 + \frac{x^2}{a^2} \right) \quad Z(\beta) = a B \left(\frac{1}{2}, \frac{\beta m \omega^2 a^2 - 1}{2} \right) \quad (9.45)$$

per cui l'equazione di Smoluchowski assume ora la forma

$$dX(t) = -\frac{D}{2} \beta m \omega^2 X(t) \frac{a^2}{a^2 + X^2(t)} dt + dW(t) \quad (9.46)$$

Sarà istruttivo esaminare analogie e differenze fra le soluzioni stazionarie di Student dell'equazione di Smoluchowski (9.46), e le soluzioni stazionarie Gaussiane dell'equazione di Smoluchowski (9.43). Nella Figura 9.1 sono rappresentati due esempi di queste distribuzioni stazionarie assieme ai potenziali $\phi(x)$ che le producono: le curve sono tracciate usando gli stessi valori dei parametri a, ω e T per i due casi. I due potenziali (9.42) e (9.45), che coincidono nelle vicinanze di $x = 0$, divergono ambedue per $x \rightarrow \pm\infty$, ma con diverse velocità: nel caso armonico (9.42) cresce come x^2 , mentre (9.45) cresce solo come $\ln x$. Da un punto di vista fisico è proprio questo che produce la differenza fra le due distribuzioni stazionarie: il potenziale armonico (9.42) infatti, essendo più intenso e vincolante, produce delle distribuzioni stazionarie Gaussiane che sono visibilmente più concentrate attorno a $x = 0$ delle leggi di Student (si osservi nella Figura 9.1 il diverso comportamento delle code). Siccome infine nell'approssimazione di Smoluchowski l'effetto della dinamica compare solo nella forma della velocità di trascinamento $a(x)$, sarà utile anche

paragonarne le due forme

$$-\frac{D}{2} \beta m \omega^2 x \qquad -\frac{D}{2} \beta m \omega^2 x \frac{a^2}{a^2 + x^2} \qquad (9.47)$$

ottenute rispettivamente dalle equazioni di Smoluchowski (9.43) e (9.46). La Figura 9.2 mostra i due comportamenti: ambedue i campi trascinano la particella Browniana verso il centro $x = 0$ da ogni parte dell'asse x ; ma mentre nel caso Gaussiano di (9.43) il trascinamento è sempre lo stesso a qualunque distanza da $x = 0$, per le leggi di Student di (9.46) esso raggiunge un massimo a distanza a dal centro e poi si spegne asintoticamente. Ancora una volta il confronto mostra in che senso il potenziale armonico (9.42) deve essere considerato più forte del potenziale (9.45) che genera distribuzioni di Student.

Parte III
Appendici

Appendice A

Consistenza (Sez. 2.3.4)

La condizione di *consistenza* gioca un ruolo determinante nei due Teoremi di Kolmogorov 2.35 e 2.37, ma noi la abbiamo incontrata anche prima, in un contesto più elementare, alla fine della Sezione 2.3.4 sulle copule. Nella discussione che segue mostreremo che il rispetto di tale condizione non è affatto scontato neanche nei casi elementari, e per far questo ci limiteremo al caso di distribuzioni discrete su insiemi finiti di numeri interi.

Per cominciare supporremo che sia stata assegnata una distribuzione discreta trivariata sull'insieme $\{0, 1\} \times \{0, 1\} \times \{0, 1\}$ di \mathbf{N}^3 che per brevità (con notazioni analoghe a quelle della Sezione 2.1) denoteremo con i simboli

$$p_{ijk} = \mathbf{P}\{i, j, k\} \quad i, j, k \in \{0, 1\}$$

Tale distribuzione sarà sempre ben definita se gli 8 numeri p_{ijk} soddisfano le relazioni

$$0 \leq p_{ijk} \leq 1 \quad \sum_{i,j,k} p_{ijk} = 1 \quad (\text{A.1})$$

Qui e nel seguito è sempre sottinteso che gli indici delle somme variano sui due valori 0 e 1. Dalla p_{ijk} sarà poi sempre possibile ricavare – con notazioni simili a quelle della Sezione 2.3.3 – le tre distribuzioni marginali bivariate su $\{0, 1\} \times \{0, 1\}$

$$p_{jk}^{(1)} = \sum_i p_{ijk} \quad p_{ik}^{(2)} = \sum_j p_{ijk} \quad p_{ij}^{(3)} = \sum_k p_{ijk}$$

e le tre distribuzioni marginali univariate (di Bernoulli) su $\{0, 1\}$

$$p_k^{(1,2)} = \sum_{i,j} p_{ijk} \quad p_i^{(2,3)} = \sum_{j,k} p_{ijk} \quad p_j^{(1,3)} = \sum_{i,k} p_{ijk}$$

Le precedenti procedure di calcolo garantiscono inoltre la *consistenza* dei tre livelli di distribuzioni nel senso che anche le altre relazioni di marginalizzazione

$$\begin{aligned} p_k^{(1,2)} &= \sum_j p_{jk}^{(1)} = \sum_i p_{ik}^{(2)} \\ p_i^{(2,3)} &= \sum_k p_{ik}^{(2)} = \sum_j p_{ij}^{(3)} \\ p_j^{(1,3)} &= \sum_k p_{jk}^{(1)} = \sum_i p_{ij}^{(3)} \end{aligned}$$

risultano banalmente verificate. Noi siamo interessati ora a vedere in che misura è possibile ricostruire questo schema consistente a tre livelli partendo però *dal basso*, e cioè dalle marginali univariate

Supponiamo allora di partire con tre arbitrarie distribuzioni univariate di Bernoulli su $\{0, 1\}$ (gli indici in alto ora sono scomparsi perché non stiamo più supponendo a priori di aver ricavato queste distribuzioni da altre distribuzioni multivariate)

$$\begin{aligned} p_i &= \begin{cases} P & i = 1 \\ 1 - P & i = 0 \end{cases} & 0 \leq P \leq 1 \\ q_j &= \begin{cases} Q & j = 1 \\ 1 - Q & j = 0 \end{cases} & 0 \leq Q \leq 1 \\ r_k &= \begin{cases} R & k = 1 \\ 1 - R & k = 0 \end{cases} & 0 \leq R \leq 1 \end{aligned}$$

e di domandarci innanzitutto se è possibile determinare tre distribuzioni bivariate p_{ij} , q_{jk} e r_{ik} tali che le univariate assegnate siano le loro marginali nel senso che

$$\sum_j p_{ij} = \sum_k r_{ik} = p_i \quad \sum_i p_{ij} = \sum_k q_{jk} = q_j \quad \sum_j q_{jk} = \sum_i r_{ik} = r_k \quad (\text{A.2})$$

Si tratta quindi di risolvere un sistema di 12 equazioni nelle 12 incognite p_{ij} , q_{jk} e r_{ik} , ma bisogna anche aggiungere che per essere accettabili le soluzioni dovranno consistere di numeri che cadono in $[0, 1]$ e che rispettano le condizioni

$$\sum_{ij} p_{ij} = \sum_{jk} q_{jk} = \sum_{ik} r_{ik} = 1$$

Si può vedere ora che delle 12 equazioni (A.2) solo 9 sono indipendenti (il rango della matrice dei coefficienti è infatti 9, e coincide con il rango della matrice dei coefficienti aumentata della colonna dei termini noti) e quindi ci attendiamo ∞^3 soluzioni, classificate da tre nuovi parametri p, q, r che dovranno essere scelti – se possibile – in modo opportuno per ottenere delle soluzioni accettabili. È facile allora constatare che, assegnati i tre numeri P, Q, R che fissano le marginali univariate

iniziali, le distribuzioni bivariate potranno essere messe nella la forma

$$\begin{cases} p_{11} = p \\ p_{10} = P - p \\ p_{01} = Q - p \\ p_{00} = 1 - P - Q + p \end{cases} \quad \begin{cases} q_{11} = q \\ q_{10} = Q - q \\ q_{01} = R - q \\ q_{00} = 1 - R - Q + q \end{cases} \quad \begin{cases} r_{11} = r \\ r_{10} = P - r \\ r_{01} = R - r \\ r_{00} = 1 - P - R + r \end{cases}$$

e risulteranno accettabili purché i parametri P, Q, R, p, q, r rispettino le seguenti limitazioni

$$0 \leq P \leq 1 \quad 0 \leq Q \leq 1 \quad 0 \leq R \leq 1 \quad (\text{A.3})$$

$$0 \leq p \leq P \wedge Q \quad 0 \leq q \leq Q \wedge R \quad 0 \leq r \leq P \wedge R \quad (\text{A.4})$$

che possono essere sempre facilmente soddisfatte (qui $x \wedge y = \min\{x, y\}$). Conseguentemente, comunque scelti in $[0, 1]$ i numeri P, Q, R (ovvero: comunque scelte le distribuzioni univariate), potremo sempre trovare (infinite) distribuzioni bivariate consistenti con le univariate assegnate.

Passiamo ora al livello successivo: supponendo di aver preso i 6 numeri P, Q, R, p, q, r in modo da rispettare le condizioni (A.3) e (A.4) – supponendo cioè di aver assegnato in maniera consistente le distribuzioni univariate p_i, q_j, r_k e bivariate p_{ij}, q_{jk}, r_{ik} – domandiamoci se è sempre possibile determinare anche una distribuzione trivariata p_{ijk} consistente con le univariate e bivariate assegnate. Domandiamoci, in altri termini, se è possibile trovare 8 numeri p_{ijk} che rispettino le limitazioni (A.1) e soddisfino le 12 equazioni

$$\sum_k p_{ijk} = p_{ij} \quad \sum_i p_{ijk} = q_{jk} \quad \sum_j p_{ijk} = r_{ik} \quad (\text{A.5})$$

Il sistema (A.5) è chiaramente sovradeterminato (12 equazioni per 8 incognite), ma si può verificare che la matrice dei coefficienti, e quella dei coefficienti aumentata della colonna dei termini noti

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & p \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & P - p \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & Q - p \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 - P - Q + p \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & q \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & Q - q \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & R - q \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 - Q - R + q \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & r \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & R - r \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & P - r \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 - P - R + r \end{pmatrix}$$

hanno ambedue rango 7. Pertanto – a norma del Teorema di Rouché-Capelli – il sistema (A.5), pur essendo sovradeterminato, risulta compatibile, e anzi ammetterà

infinite soluzioni classificate da un parametro s . Si trova allora che le soluzioni del sistema (A.5) possono essere messe nella forma

$$\left\{ \begin{array}{l} p_{111} = 1 - P - Q - R + p + q + r - s \\ p_{110} = P + Q + R - 1 - q - r + s \\ p_{101} = P + Q + R - 1 - p - q + s \\ p_{100} = 1 - Q - R + q - s \\ p_{011} = P + Q + R - 1 - p - r + s \\ p_{010} = 1 - P - R + r - s \\ p_{001} = 1 - P - Q + p - s \\ p_{000} = s \end{array} \right. \quad (\text{A.6})$$

e noi dobbiamo domandarci se, comunque fissati i numeri P, Q, R, p, q, r nel rispetto delle condizioni (A.3) e (A.4), sia possibile determinare qualche valore di $s \in [0, 1]$ in modo tale che le (A.6) risultino accettabili in base alle limitazioni (A.1). Sorprendentemente la risposta a questa domanda è negativa, e noi lo mostreremo con un controesempio. Siccome è facile verificare che la somma delle p_{ijk} in (A.6) è sempre uguale a 1, resta solo da controllare se è possibile che tutti questi 8 numeri cadano in $[0, 1]$. Supponiamo allora di prendere in particolare

$$P = Q = R = \frac{1}{2} \quad p = q = \frac{2 + \sqrt{2}}{8} \approx 0.426777 \quad r = \frac{1}{4} \quad (\text{A.7})$$

numeri che soddisfano le condizioni (A.3) e (A.4) e producono la seguente famiglia consistente di distribuzioni bivariate e univariate

$$\left\{ \begin{array}{l} p_{11} = \frac{2+\sqrt{2}}{8} \\ p_{10} = \frac{2-\sqrt{2}}{8} \\ p_{01} = \frac{2-\sqrt{2}}{8} \\ p_{00} = \frac{2+\sqrt{2}}{8} \end{array} \right. \quad \left\{ \begin{array}{l} q_{11} = \frac{2+\sqrt{2}}{8} \\ q_{10} = \frac{2-\sqrt{2}}{8} \\ q_{01} = \frac{2-\sqrt{2}}{8} \\ q_{00} = \frac{2+\sqrt{2}}{8} \end{array} \right. \quad \left\{ \begin{array}{l} r_{11} = 1/4 \\ r_{10} = 1/4 \\ r_{01} = 1/4 \\ r_{00} = 1/4 \end{array} \right. \quad (\text{A.8})$$

$$p_1 = p_0 = 1/2 \quad q_1 = q_0 = 1/2 \quad r_1 = r_0 = 1/2$$

Con questa scelta però le (A.6) diventano

$$\left\{ \begin{array}{l} p_{111} = \frac{1+\sqrt{2}}{4} - s \approx 0.603553 - s \\ p_{110} = -\frac{\sqrt{2}}{8} + s \approx -0.176777 + s \\ p_{101} = -\frac{\sqrt{2}}{4} + s \approx -0.353553 + s \\ p_{100} = \frac{2+\sqrt{2}}{8} - s \approx 0.426777 - s \\ p_{011} = -\frac{\sqrt{2}}{8} + s \approx -0.176777 + s \\ p_{010} = \frac{1}{4} - s = 0.25 - s \\ p_{001} = \frac{2+\sqrt{2}}{8} - s \approx 0.426777 - s \\ p_{000} = s \end{array} \right.$$

ed è evidente che nessun valore di $s \in [0, 1]$ consente di avere tutte le p_{ijk} in $[0, 1]$. Infatti, ad esempio, bisognerebbe scegliere $s \geq 0.353553$ per avere $p_{101} \geq 0$, ma in

questo caso avremmo anche $p_{010} \leq 0.25 - 0.353553 = -0.103553$. Si danno pertanto casi di famiglie consistenti di distribuzioni univariate e bivariate che non ammettono distribuzioni trivariate altrettanto consistenti

Bisogna osservare a questo punto che il fenomeno messo in evidenza riguarda solo le *famiglie di leggi*, e non gli *spazi di probabilità*. Se infatti supponessimo *a priori* di muoverci all'interno di un *unico* spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ assegnato, e se supponessimo inoltre di considerare solo le famiglie di leggi associate ad arbitrari eventi $A, B, C \in \mathcal{F}$ da relazioni del tipo

$$\begin{aligned} p_{111} &= \mathbf{P}\{ABC\} & p_{110} &= \mathbf{P}\{ABC\bar{C}\} & \dots \\ p_{11} &= \mathbf{P}\{AB\} & p_{10} &= \mathbf{P}\{A\bar{B}\} & \dots & q_{11} &= \mathbf{P}\{BC\} & \dots \\ p_1 &= \mathbf{P}\{A\} & p_0 &= \mathbf{P}\{\bar{A}\} & q_1 &= \mathbf{P}\{B\} & \dots \end{aligned}$$

si potrebbe facilmente mostrare che esse risulterebbero sempre tutte perfettamente consistenti. L'impossibilità di costruire leggi trivariate consistenti con un dato sistema di leggi univariate e bivariate si manifesta invece solo quando consideriamo le nostre distribuzioni *senza associarle preventivamente a spazi di probabilità*. Anzi, come conseguenza dell'esempio presentato in questa Appendice, potremmo affermare che famiglie di leggi univariate e bivariate determinate da parametri simili a quelli da noi scelti in (A.7) – pur essendo perfettamente consistenti fra loro – non sono derivabili da un'unica legge trivariata, e non possono quindi essere descritte per mezzo di eventi presi in un unico spazio di probabilità. D'altra parte è bene ricordare anche che, mentre le leggi sono entità direttamente connesse con le osservazioni empiriche, gli spazi di probabilità (pur essendo estremamente importanti per dare basi solide alla teoria) non sono che costruzioni concettuali introdotte per descrivere il modo in cui le probabilità vengono combinate, e in linea di principio il modello per tali combinazioni potrebbe non essere quello degli spazi di probabilità Kolmogoroviani introdotti nel Capitolo 1

L'importanza di questa ultima osservazione si coglie però nella sua intrezza solo se si riflette ad un aspetto che finora abbiamo trascurato: potrebbe infatti risultare naturale chiedersi perché mai dovremmo preoccuparci dei comportamenti paradossali di particolari famiglie di distribuzioni come la (A.8): è mai capitato di osservare nella realtà dei fatti qualche sistema che si comporti in questo modo eccezionale? Non potrebbe trattarsi, insomma, solo di situazioni sì anomale, ma in fondo sprovviste di rilevanza pratica? Anche la risposta a questa domanda è piuttosto sorprendente: le distribuzioni (A.8) non sono state affatto scelte capziosamente, e rivestono invece un interesse pratico considerevole. Mostriamo infatti che le (A.8) corrispondono alle probabilità che la *meccanica quantistica* attribuisce (mediante le solite procedure basate su prodotti scalari e moduli quadri) alle misure delle tre osservabili $\hat{\alpha} \cdot \mathbf{S}$, $\hat{\beta} \cdot \mathbf{S}$, $\hat{\gamma} \cdot \mathbf{S}$ proiezioni dello spin $\mathbf{S} = (\sigma_x, \sigma_y, \sigma_z)$ delle matrici di Pauli su tre versori $\hat{\alpha}, \hat{\beta}$ e $\hat{\gamma}$ giacenti nel piano x, z ad angoli $0, \frac{\pi}{4}, \frac{\pi}{2}$ con l'asse z , e partendo in uno stato iniziale che sia autostato di σ_y .

Le componenti cartesiane di un generico versore $\hat{\mathbf{v}} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$ in uno spazio tridimensionale sono determinate dall'angolo $\theta \in [0, \pi]$ che $\hat{\mathbf{v}}$ forma con l'asse z , e dall'angolo $\phi \in [0, 2\pi]$ che la sua proiezione sul piano x - y forma con l'asse x . Pertanto i versori del nostro esempio hanno componenti

$$\hat{\boldsymbol{\alpha}} = (0, 0, 1) \quad \hat{\boldsymbol{\beta}} = \left(\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2} \right) \quad \hat{\boldsymbol{\gamma}} = (1, 0, 0)$$

le proiezioni dello spin sono

$$\begin{aligned} \hat{\boldsymbol{\alpha}} \cdot \mathbf{S} &= \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ \hat{\boldsymbol{\beta}} \cdot \mathbf{S} &= \frac{\sqrt{2}}{2}(\sigma_x + \sigma_z) = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ \hat{\boldsymbol{\gamma}} \cdot \mathbf{S} &= \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{aligned}$$

e lo stato iniziale sarà un autostato di

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

Si controlla facilmente che tutte e quattro le osservabili coinvolte hanno come autovalori ± 1 , che i sistemi ortonormali di autovettori delle proiezioni dello spin sono

$$\begin{aligned} |\alpha+\rangle &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} & |\beta+\rangle &= \frac{\sqrt{2-\sqrt{2}}}{2} \begin{pmatrix} 1 \\ \sqrt{2}-1 \end{pmatrix} & |\gamma+\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ |\alpha-\rangle &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} & |\beta-\rangle &= \frac{\sqrt{2-\sqrt{2}}}{2} \begin{pmatrix} 1 \\ -\sqrt{2}-1 \end{pmatrix} & |\gamma-\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \end{aligned}$$

mentre i due possibili stati iniziali, autovettori ortonormali di σ_y , sono

$$|y+\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \quad |y-\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

Se allora scegliamo $|y+\rangle$ come stato iniziale, e identificando p_i, q_j e r_k rispettivamente con le distribuzioni di $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$, avremo innanzitutto coerentemente con (A.8)

$$\begin{aligned} p_1 &= |\langle \alpha+ | y+\rangle|^2 = 1/2 & q_1 &= |\langle \beta+ | y+\rangle|^2 = 1/2 & r_1 &= |\langle \gamma+ | y+\rangle|^2 = 1/2 \\ p_0 &= |\langle \alpha- | y+\rangle|^2 = 1/2 & q_0 &= |\langle \beta- | y+\rangle|^2 = 1/2 & r_0 &= |\langle \gamma- | y+\rangle|^2 = 1/2 \end{aligned}$$

Per ricavare le distribuzioni bivariate, poi, bisogna ricordare che, ad esempio, il valore quantistico $|\langle \alpha+ | \beta+\rangle|^2$ è identificata con la probabilità *condizionata* $p_{1|1}$, per cui la corrispondente probabilità congiunta p_{11} si otterrà come prodotto

$$p_{11} = p_{1|1} q_1 = |\langle \alpha+ | \beta+\rangle|^2 |\langle \beta+ | y+\rangle|^2$$

Sulla base di questa osservazione – usando le espressioni esplicite degli autostati, e le distribuzioni univariate p_i, q_j e r_k calcolate in precedenza – si ottiene allora il seguente sistema di distribuzioni congiunte bivariate

$$\begin{cases} p_{11} = |\langle \alpha + |\beta+\rangle|^2 |\langle \beta + |y+\rangle|^2 = \frac{2+\sqrt{2}}{8} \\ p_{10} = |\langle \alpha + |\beta-\rangle|^2 |\langle \beta - |y+\rangle|^2 = \frac{2-\sqrt{2}}{8} \\ p_{01} = |\langle \alpha - |\beta+\rangle|^2 |\langle \beta + |y+\rangle|^2 = \frac{2-\sqrt{2}}{8} \\ p_{00} = |\langle \alpha - |\beta-\rangle|^2 |\langle \beta - |y+\rangle|^2 = \frac{2+\sqrt{2}}{8} \\ q_{11} = |\langle \beta + |\gamma+\rangle|^2 |\langle \gamma + |y+\rangle|^2 = \frac{2+\sqrt{2}}{8} \\ q_{10} = |\langle \beta + |\gamma-\rangle|^2 |\langle \gamma - |y+\rangle|^2 = \frac{2-\sqrt{2}}{8} \\ q_{01} = |\langle \beta - |\gamma+\rangle|^2 |\langle \gamma + |y+\rangle|^2 = \frac{2-\sqrt{2}}{8} \\ q_{00} = |\langle \beta - |\gamma-\rangle|^2 |\langle \gamma - |y+\rangle|^2 = \frac{2+\sqrt{2}}{8} \\ r_{11} = |\langle \alpha + |\gamma+\rangle|^2 |\langle \gamma + |y+\rangle|^2 = 1/4 \\ r_{10} = |\langle \alpha + |\gamma-\rangle|^2 |\langle \gamma - |y+\rangle|^2 = 1/4 \\ r_{01} = |\langle \alpha - |\gamma+\rangle|^2 |\langle \gamma + |y+\rangle|^2 = 1/4 \\ r_{00} = |\langle \alpha - |\gamma-\rangle|^2 |\langle \gamma - |y+\rangle|^2 = 1/4 \end{cases}$$

che coincide esattamente con (A.8). Pertanto le distribuzioni consistenti che si ottengono con la scelta (A.7) sono proprio quelle che la meccanica quantistica assegna alle proiezioni dello spin che abbiamo considerato

In conclusione il nostro esempio mostra che esistono sistemi quantistici che non sono suscettibili di una descrizione coerente in termini di spazi di probabilità Kolmogoroviani, e quindi mette in evidenza quelle che potremmo chiamare le radici probabilistiche dei *paradossi* della meccanica quantistica. È noto, peraltro, che il modello di probabilità adottato dalla teoria quantistica non fa uso di spazi $(\Omega, \mathcal{F}, \mathbf{P})$ come quelli classici, ma fa riferimento piuttosto a stati come vettori in spazi di Hilbert con tutte le conseguenze che conosciamo. Questa discussione, d'altra parte, suggerisce che la possibilità di incontrare famiglie non consistenti di leggi è apertamente ammessa in probabilità anche indipendentemente dall'introduzione di modelli quantistici (sebbene questi sembrano essere per il momento gli unici esempi concreti). In altri termini, c'è di più nelle famiglie consistenti di leggi di quanto non sia contenuto negli spazi di probabilità Kolmogoroviani, e la possibilità di leggi congiunte che si comportino in maniera *quantistica* è già contemplata per le distribuzioni *classiche*. Le inconsistenze ricordate in questa appendice, infatti, sono note da tempo e hanno anche stimolato molte ricerche per la determinazione di condizioni generali che garantiscano l'esistenza di modelli Kolmogoroviani per date famiglie di leggi: in questa prospettiva le celebri *disuguaglianze di Bell* (scoperte negli anni '60 nel contesto di una discussione sul paradosso quantistico di Einstein-Podolski-Rosen) possono essere considerate proprio come un esempio di questo tipo di condizioni che, ovviamente, non sono in generale rispettate dai sistemi quantistici

Appendice B

Disuguaglianze (Sez. 3.3.2)

In questa appendice richiameremo – nella loro formulazione probabilistica – alcune importanti disuguaglianze tipiche della teoria dell'integrazione che saranno applicate in alcune parti del testo

Proposizione B.1. Disuguaglianza di Jensen: *Se $g(x)$ è una funzione di Borel convessa (verso il basso) e se X è una v-a integrabile, si ha*

$$g(\mathbf{E}[X]) \leq \mathbf{E}[g(X)]$$

Dimostrazione: La disuguaglianza di Jensen è una proprietà molto generale, utile nella deduzione delle successive proposizioni. Se $g(x)$ è convessa (verso il basso), comunque scelto $x_0 \in \mathbf{R}$ esisterà un numero $\lambda(x_0)$ tale che

$$g(x) \geq g(x_0) + (x - x_0)\lambda(x_0), \quad \forall x \in \mathbf{R}$$

Sostituendo allora $\mathbf{E}[X]$ a x_0 , e calcolando le funzioni in X si ha

$$g(X) \geq g(\mathbf{E}[X]) + (X - \mathbf{E}[X])\lambda(\mathbf{E}[X])$$

sicché il risultato si ottiene prendendo il valore d'attesa di ambedue i membri. \square

Corollario B.2. Disuguaglianza di Lyapunov: *Se X è una v-a si ha*

$$\mathbf{E}[|X|^s]^{1/s} \leq \mathbf{E}[|X|^t]^{1/t} \quad 0 < s \leq t$$

In particolare si ha la seguente catena di disuguaglianze

$$\mathbf{E}[|X|] \leq \mathbf{E}[|X|^2]^{1/2} \leq \dots \leq \mathbf{E}[|X|^n]^{1/n} \leq \dots$$

Dimostrazione: Posto $r = t/s \geq 1$, e $Y = |X|^s$, si applica la disuguaglianza di Jensen alla funzione convessa $g(x) = |x|^r$ e si ottiene $|\mathbf{E}[Y]|^r \leq \mathbf{E}[|Y|^r]$, cioè

$$\mathbf{E}[|X|^{st/s}] \leq \mathbf{E}[|X|^t]$$

da cui segue immediatamente la tesi. La successiva catena di disuguaglianze non è che un caso particolare. Una conseguenza importante della disuguaglianza di Lyapunov è che se una v -a X ha un momento assoluto di ordine r finito ($\mathbf{E}[|X|^r] < +\infty$), allora anche tutti i momenti assoluti di ordine inferiore a r sono finiti; non altrettanto si può dire in generale sui momenti assoluti di ordine superiore a r . \square

Proposizione B.3. Disuguaglianza di Hölder: *Dati due numeri p, q con*

$$1 < p < +\infty \quad 1 < q < +\infty \quad \frac{1}{p} + \frac{1}{q} = 1$$

e date le v -a X e Y con $\mathbf{E}[|X|^p] < +\infty$ e $\mathbf{E}[|Y|^q] < +\infty$, allora il prodotto XY risulta integrabile e si ha

$$\mathbf{E}[|XY|] \leq \mathbf{E}[|X|^p]^{1/p} \mathbf{E}[|Y|^q]^{1/q}$$

*Si noti che la ben nota **disuguaglianza di Schwarz***

$$\mathbf{E}[|XY|]^2 \leq \mathbf{E}[X^2] \mathbf{E}[Y^2]$$

è un caso particolare di questa disuguaglianza per $p = q = 2$.

Dimostrazione: Omessa¹: riprodurremo per completezza solo la dimostrazione della disuguaglianza di Schwarz. Consideriamo dapprima il caso in cui $\mathbf{E}[X^2] \neq 0$ e $\mathbf{E}[Y^2] \neq 0$ e poniamo

$$\tilde{X} = \frac{X}{\sqrt{\mathbf{E}[X^2]}} \quad \tilde{Y} = \frac{Y}{\sqrt{\mathbf{E}[Y^2]}}$$

in modo che risulti $\mathbf{E}[\tilde{X}^2] = 1$ e $\mathbf{E}[\tilde{Y}^2] = 1$. Siccome $(|\tilde{X}| - |\tilde{Y}|)^2 \geq 0$, e quindi

$$2|\tilde{X}\tilde{Y}| \leq \tilde{X}^2 + \tilde{Y}^2$$

abbiamo che

$$2\mathbf{E}[|\tilde{X}\tilde{Y}|] \leq \mathbf{E}[\tilde{X}^2] + \mathbf{E}[\tilde{Y}^2] = 2$$

ovvero

$$\mathbf{E}[|\tilde{X}\tilde{Y}|]^2 \leq 1 = \mathbf{E}[\tilde{X}^2] \cdot \mathbf{E}[\tilde{Y}^2]$$

e la tesi si ottiene sostituendo le definizioni di \tilde{X} e \tilde{Y} in termini di X e Y . Quando invece almeno uno dei due valore d'attesa è nullo, ad esempio se $\mathbf{E}[X^2] = 0$, da 5 di Proposizione 3.26 risulta $X = 0$ \mathbf{P} -qo, e quindi da 3 della medesima proposizione si ha anche $\mathbf{E}[|XY|] = 0$. È facile vedere allora che la tesi è banalmente verificata anche in questo caso. \square

¹N. Cufaro Petroni, **CALCOLO DELLE PROBABILITÀ**, Edizioni dal Sud (Bari, 1996)

Proposizione B.4. Disuguaglianza di Minkowski: *Dato il numero p con*

$$1 \leq p < +\infty$$

e due v-a X e Y tali che $\mathbf{E}[|X|^p] < +\infty$ e $\mathbf{E}[|Y|^p] < +\infty$, allora $\mathbf{E}[|X+Y|^p] < +\infty$ e si ha

$$\mathbf{E}[|X+Y|^p]^{1/p} \leq \mathbf{E}[|X|^p]^{1/p} + \mathbf{E}[|Y|^p]^{1/p}$$

Dimostrazione: Omessa²

□

²N. Cufaro Petroni, CALCOLO DELLE PROBABILITÀ, Edizioni dal Sud (Bari, 1996)

Appendice C

Paradosso di Bertrand (Sez. 3.5.1)

Nel primo capitolo del suo classico trattato *Calcul des Probabilités* (Paris, 1889) Joseph Bertrand si sofferma a lungo sulla definizione di probabilità, e in particolare osserva che i modelli aleatori con un numero infinito non numerabile di possibili risultati sono soggetti a equivoci particolarmente insidiosi. Ad esempio, se chiediamo quale è la probabilità di avere un numero reale maggiore di 50 scegliendolo a caso fra 0 e 100 la risposta naturale sembra essere $\frac{1}{2}$; ma siccome i numeri reali tra 0 e 100 sono anche biunivocamente associati ai loro quadrati fra 0 e 10 000, la nostra domanda potrebbe essere equivalentemente riformulata chiedendo quale è la probabilità che il nostro numero estratto a caso abbia un quadrato maggiore di 2 500. Se però immaginiamo di prendere un numero a caso fra 0 e 10 000, intuitivamente valutiamo $\frac{3}{4}$, e non $\frac{1}{2}$ la probabilità di trovarlo maggiore di 2 500. I due problemi sono equivalenti, ma le due risposte (ambedue legittime) sono differenti: da cosa nasce questo paradosso? Bertrand afferma, con ragione, che la spiegazione risiede nel fatto che i due enunciati sono imprecisi perché l'espressione *a caso* è usata troppo genericamente, e che gli esempi potrebbero moltiplicarsi all'infinito: egli ne elenca molti, ma in particolare quello che segue è oggi universalmente noto come *Paradosso di Bertrand*.

Facendo riferimento alla Figura C.1, si prenda *a caso* una corda su un cerchio Γ di raggio 1: quale è la probabilità che la sua lunghezza sia maggiore di quella (pari a $\sqrt{3}$) del lato del triangolo equilatero inscritto? Sono possibili tre soluzioni (tutte legittime) che forniscono tre risultati numericamente differenti e che discuteremo facendo sempre riferimento alla Figura C.1:

1. Scegliere una corda a caso è come sceglierne il punto di mezzo: affinché la corda sia più lunga del lato del triangolo equilatero è necessario e sufficiente che tale punto di mezzo si trovi all'interno del cerchio concentrico γ di raggio $\frac{1}{2}$ inscritto nel triangolo equilatero. La probabilità che ciò avvenga sarà data quindi dal rapporto fra l'area $\frac{\pi}{4}$ di γ , e l'area π di Γ : si ottiene quindi che la probabilità cercata è $p_1 = \frac{1}{4}$.

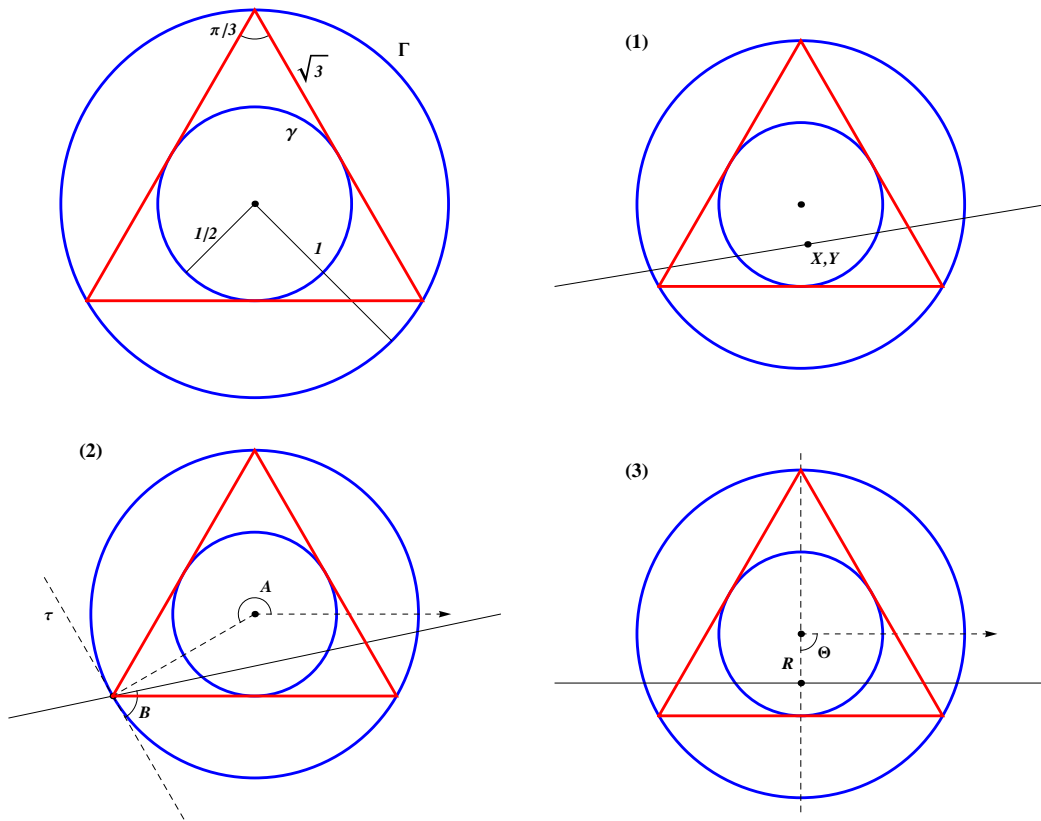


Figura C.1: Paradosso di Bertrand.

2. La posizione di un'estremità della corda non influenza la probabilità a causa della simmetria del cerchio. Presa quindi una delle due estremità, la lunghezza della corda dipenderà dall'angolo compreso fra 0 e π che essa forma con la tangente nell'estremità fissata. Se allora si traccia il triangolo equilatero inscritto con uno dei vertici nel punto considerato, la corda sarà più grande del suo lato se l'angolo che essa forma con la tangente si trova fra $\frac{\pi}{3}$ e $\frac{2\pi}{3}$, e pertanto la probabilità che ciò accada è $p_2 = \frac{1}{3}$.
3. La conoscenza della direzione della corda presa a caso non influenza il valore della probabilità richiesta a causa della simmetria del cerchio. Assegnata allora tale direzione, la corda sarà più lunga di $\sqrt{3}$ se la sua intersezione con il raggio ad essa perpendicolare si troverà ad una distanza dal centro inferiore a $\frac{1}{2}$, e questo si verificherà con probabilità $p_3 = \frac{1}{2}$.

Per cogliere la radice del paradosso dobbiamo ricordare che quando parliamo di prendere *a caso* un numero, in genere intendiamo che esso è *uniformemente* distribuito in qualche intervallo. Ora è possibile mostrare che ciò che è uniformemente distribuito in ciascuna delle tre soluzioni, non è affatto altrettanto uniformemente

distribuito nelle altre due soluzioni: in pratica nelle tre soluzioni noi – scegliendo diversamente ciò che consideriamo uniformemente distribuito – surrettiziamente adottiamo tre diverse misure di probabilità, e quindi non c'è da meravigliarsi se alla fine le risposte sono differenti.

Per discutere più precisamente il problema dobbiamo introdurre (come in Figura C.1) tre coppie di v - a che rappresentano le coordinate necessarie per determinare la posizione della nostra corda nelle tre soluzioni proposte:

1. le coordinate Cartesiane (X, Y) del centro della corda;
2. gli angoli (A, B) che individuano rispettivamente la posizione dell'estremo fissato della corda e l'orientamento di quest'ultima rispetto alla tangente;
3. le coordinate polari (R, Θ) dell'intersezione fra la corda e il raggio ad essa perpendicolare.

In ognuna delle tre soluzioni è nascosta l'ipotesi che la corrispondente coppia di coordinate sia distribuita uniformemente, ma queste tre assunzioni, come vedremo, non sono affatto compatibili fra di loro, anzi sottintendono – sullo spazio (Ω, \mathcal{F}) sul quale tutte le nostre v - a sono definite – tre diverse misure di probabilità \mathbf{P}_1 , \mathbf{P}_2 e \mathbf{P}_3 . Più precisamente le tre soluzioni prospettate presuppongono le seguenti tre distribuzioni congiunte uniformi ($\chi_{[a,b]}(x)$ è come al solito un indicatore):

1. la ddp congiunta e uniforme di (X, Y) su $(\Omega, \mathcal{F}, \mathbf{P}_1)$ è

$$f_{XY}(x, y) = \frac{1}{\pi} \chi_{[0,1]}(x^2 + y^2) \quad (\text{C.1})$$

e le due v - a non sono indipendenti;

2. la ddp congiunta e uniforme di (A, B) su $(\Omega, \mathcal{F}, \mathbf{P}_2)$ è

$$f_{AB}(\alpha, \beta) = \frac{1}{2\pi^2} \chi_{[0,2\pi]}(\alpha) \chi_{[0,\pi]}(\beta) \quad (\text{C.2})$$

con componenti indipendenti;

3. la ddp congiunta e uniforme di (R, Θ) su $(\Omega, \mathcal{F}, \mathbf{P}_3)$ è infine

$$f_{R\Theta}(r, \theta) = \frac{1}{2\pi} \chi_{[0,1]}(r) \chi_{[-\pi,\pi]}(\theta) \quad (\text{C.3})$$

anch'essa con componenti indipendenti.

Naturalmente se si adottasse un unico spazio di probabilità per tutte e tre le soluzioni i tre risultati sarebbero perfettamente coincidenti, ma in questo caso solo una delle tre coppie di coordinate potrebbe essere uniformemente distribuita, mentre le leggi congiunte delle altre due coppie dovrebbero essere ricavate con la tecnica sviluppata nella sezione 3.47 per le funzioni di v - a . Infatti è chiaro che ci sono delle

precise trasformazioni di variabili che permettono di passare da una coppia all'altra: utilizzando queste trasformazioni si potrebbe allora verificare che se una delle tre coppie ha legge uniforme, le altre due hanno leggi diverse da quella uniforme.

Senza sviluppare completamente i calcoli per tutti e tre i casi ci limiteremo a illustrare queste affermazioni discutendo il rapporto fra la soluzione (1) e la soluzione (3). Le trasformazioni che legano le coordinate Cartesiane (X, Y) a quelle polari (R, Θ) sono ben note:

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases} \quad \begin{cases} r = \sqrt{x^2 + y^2} \\ \theta = \arctan \frac{y}{x} \end{cases} \quad \begin{matrix} r > 0 \\ -\pi < \theta \leq \pi \end{matrix}$$

con determinante Jacobiano

$$J(r, \theta) = \begin{vmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{vmatrix} = \begin{vmatrix} \cos \theta & \sin \theta \\ -\frac{1}{r} \sin \theta & \frac{1}{r} \cos \theta \end{vmatrix} = \frac{1}{r}$$

Conseguentemente da (3.66) otteniamo che se (X, Y) hanno ddp uniforme (C.1), la coppia (R, Θ) avrà ora ddp non uniforme

$$f_{R\Theta}^{(1)}(r, \theta) = \frac{r}{\pi} \chi_{[0,1]}(r) \chi_{[-\pi,\pi]}(\theta)$$

diversa dalla $f_{R\Theta}$ in (C.3) e che rappresenterà la sua legge nello spazio $(\Omega, \mathcal{F}, \mathbf{P}_1)$. Con questa nuova distribuzione si vede facilmente ora che la probabilità calcolata nel quadro della soluzione (3) è

$$p_3 = \int_0^{\frac{1}{2}} \frac{r}{\pi} dr \int_{-\pi}^{\pi} d\theta = \frac{1}{4}$$

in perfetto accordo con la soluzione (1).

Sarà importante precisare, per concludere, che – come già osservato all'inizio di questa appendice – i paradossi di Bertrand si manifestano solo quando consideriamo delle probabilità attribuite a insiemi non numerabili. Per chiarire questo punto basterà riprendere l'esempio, ricordato all'inizio, della probabilità che un numero scelto *a caso* fra i numeri reali in $[0, 100]$ risulti maggiore di 50: qui il paradosso nasce dalla constatazione che si otterrebbe una risposta diversa (ma altrettanto legittima) se la domanda fosse riformulata considerando i *quadrati* dei nostri numeri scelti ora *a caso* in $[0, 10\,000]$, e con il numero 50 sostituito dal suo quadrato 2\,500. Il paradosso scompare invece se si richiede la probabilità di scegliere un numero maggiore di 50 prendendolo *a caso* fra gli *interi* (equiprobabili) da 1 a 100. Infatti questa volta otterremmo la stessa risposta (banalmente: $\frac{1}{2}$) anche passando ai quadrati 1, 4, 9, ..., 10\,000, perché ovviamente ora i nostri quadrati sono ancora 100 ed equiprobabili,

Appendice D

Spazi L^p di v -a (Sez. 4.1)

Con il simbolo $L^p(\Omega, \mathcal{F}, \mathbf{P})$, ovvero anche semplicemente con L^p , si denotano gli insiemi delle v -a definite su $(\Omega, \mathcal{F}, \mathbf{P})$ e tali che $\mathbf{E}[|X|^p] < +\infty$, con $p > 0$. Questi insiemi possono essere dotati di strutture geometriche particolarmente adatte alle applicazioni. Prima di tutto osserviamo che, quale che sia il valore di $p > 0$, è sempre possibile dotarli di una **metrica**, cioè una distanza fra v -a definita come

$$d(X, Y) = \mathbf{E}[|X - Y|^p]^{1/p}$$

per cui tutti i nostri L^p saranno *spazi metrici*. Se poi $p \geq 1$, la disuguaglianza di Minkowski (Proposizione B.4) consente di affermare che i corrispondenti L^p sono anche *spazi vettoriali* per i quali combinazioni lineari di elementi di L^p sono ancora in L^p . Inoltre su questi spazi vettoriali L^p è possibile introdurre una **norma**, cioè una lunghezza dei vettori $X \in L^p$, mediante la definizione

$$\|X\|_p = \mathbf{E}[|X|^p]^{1/p}$$

e quindi anche il concetto di convergenza verso X delle successioni $(X_n)_{n \in \mathbf{N}}$ mediante la convergenza verso zero delle successioni numeriche $\|X_n - X\|_p$. Siccome tali *spazi normati* sono anche *completi*¹, essi sono **spazi di Banach**, e in essi la distanza viene assegnata tramite la norma:

$$d(X, Y) = \|X - Y\|_p$$

Si noti che dalla disuguaglianza di Lyapunov (Corollario B.2) si deduce immediatamente che

$$\|X\|_1 \leq \|X\|_p \leq \|X\|_q \quad 1 \leq p \leq q < +\infty$$

¹In uno spazio normato $(\mathcal{E}, \|\cdot\|)$ una successione $(x_n)_{n \in \mathbf{N}}$ si chiama *successione di Cauchy* quando

$$\lim_{n, m} \|x_n - x_m\| = 0$$

Si dice che lo spazio normato è *completo* se ogni successione di Cauchy di elementi di \mathcal{E} converge verso un altro elemento di \mathcal{E} . In tal caso si dice anche che $(\mathcal{E}, \|\cdot\|)$ è uno *spazio di Banach*.

Ne segue che, se $1 \leq p \leq q$, e $X \in L^q$, allora anche $X \in L^p$, sicché

$$L^1 \supseteq L^p \supseteq L^q \quad 1 \leq p \leq q < +\infty$$

Fra gli spazi di Banach L^p con $p \geq 1$, un ruolo particolarmente rilevante è giocato dal caso $p = 2$, cioè dallo spazio $L^2(\Omega, \mathcal{F}, \mathbf{P})$: è facile controllare infatti che in questo caso la norma $\| \cdot \|_2$ può essere realizzata tramite il **prodotto scalare**

$$\langle X, Y \rangle = \mathbf{E}[XY]$$

nel senso che in L^2 si ha

$$\|X\|_2 = \sqrt{\langle X, X \rangle} = \sqrt{\mathbf{E}[X^2]}$$

Gli spazi dotati di prodotto scalare, quando sono completi, prendono anche il nome di **spazi di Hilbert**. L'introduzione del concetto di prodotto scalare consente di usare in probabilità, non solo i ben noti metodi dell'analisi funzionale, ma anche idee e concetti mutuati dalla geometria. Diremo ad esempio che due v.a. $X, Y \in L^2$ sono **ortogonali** quando $\langle X, Y \rangle = \mathbf{E}[XY] = 0$ e diremo che una famiglia di v.a. di L^2 forma un **sistema ortogonale** quando comunque prese in essa due v.a. distinte esse risultano ortogonali. Se inoltre, in un sistema ortogonale, risulta anche $\|X\|_2 = 1$ per ogni elemento, si dice che la famiglia forma un **sistema ortonormale**. Notiamo infine che, se due v.a. sono non correlate si ha

$$\langle X, Y \rangle = \mathbf{E}[XY] = \mathbf{E}[X] \mathbf{E}[Y]$$

per cui esse risulteranno ortogonali se almeno una delle due ha valore d'attesa nullo.

Appendice E

Momenti e cumulanti (Sez. 4.2.1)

Se tutti i momenti $m_n = \mathbf{E}[X^n]$ di una v -a X esistono e sono finiti, il Teorema 4.11 ci dice che essi ci permettono di scrivere lo sviluppo in serie di potenze della *fc* $\varphi(u)$ della nostra v -a; inoltre i Teoremi 4.12 e 4.13 ci dicono che la *fc* $\varphi(t)$ determina in modo unico la *ddp* $f(x)$ della X (che per semplicità supporremo *ac*). Ha quindi senso porsi la seguente domanda che prende il nome di **problema dei momenti**: è possibile risalire in modo unico dalla conoscenza dei momenti $(m_n)_{n \in \mathbf{N}}$ di una v -a X alla sua *ddp* $f(x)$? In particolare il problema dell'unicità può essere posto in questa forma: se fossero date due *ddp* $f(x)$ e $g(x)$ tali che

$$\int_{-\infty}^{+\infty} x^n f(x) dx = \int_{-\infty}^{+\infty} x^n g(x) dx, \quad n \geq 1$$

è possibile affermare che $f(x) = g(x)$ in ogni x ? In realtà si fa vedere con dei controesempi¹ che la risposta è in generale negativa: è possibile cioè costruire esplicitamente leggi diverse che danno luogo alla medesima successione di momenti. Sarà allora importante determinare sotto quali condizioni sufficienti il problema dei momenti ammette invece una ed una sola soluzione

Teorema E.1. *Data una v -a X e posto $m_n = \mathbf{E}[X^n]$ e $\mu_n = \mathbf{E}[|X|^n]$, se i momenti assoluti μ_n risultano tutti finiti e se*

$$\overline{\lim}_n \frac{\mu_n^{1/n}}{n} < +\infty$$

allora i momenti m_n determinano la legge di X in maniera unica. In particolare queste condizioni sufficienti sono soddisfatte se la distribuzione di X è concentrata su un intervallo limitato.

La formula 4.18 del Teorema 4.11 sullo sviluppo in serie della *fc* di una v -a X può essere estesa anche al caso in cui $\varphi(\mathbf{u})$ è la *fc* di un *vett*-a $\mathbf{X} = (X_1, \dots, X_n)$

¹A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

nella forma

$$\varphi(u_1, \dots, u_n) = \sum_{\{\mathbf{k}\}} \frac{i^{|\mathbf{k}|}}{k_1! \dots k_n!} u_1^{k_1} \cdot \dots \cdot u_n^{k_n} m_n(k_1, \dots, k_n)$$

dove per brevità abbiamo posto $\{\mathbf{k}\} = \{k_1, \dots, k_n\}$ e $|\mathbf{k}| = k_1 + \dots + k_n$, mentre

$$m_n(\mathbf{u}) = m_n(k_1, \dots, k_n) = \mathbf{E} [X_1^{k_1} \cdot \dots \cdot X_n^{k_n}]$$

sono i *momenti misti* delle componenti di \mathbf{X} . Naturalmente lo sviluppo in serie degenera in una somma finita con resto infinitesimo (formula di Taylor) se i momenti da un certo ordine in poi non esistono.

È utile introdurre a questo punto anche la cosiddetta *caratteristica logaritmica* della *v-a* X

$$\eta(u_1, \dots, u_n) = \ln \varphi(u_1, \dots, u_n)$$

che viene a volte usata al posto della *fc*. Essa infatti è spesso più semplice della φ e le sue proprietà possono essere studiate più facilmente. Ad esempio per una *v-a* Gaussiana $\mathfrak{N}(b, a^2)$ si ha

$$\eta(u) = ibu - \frac{a^2 u^2}{2}$$

per una Cauchy $\mathfrak{C}(a, b)$ è

$$\eta(u) = ibu - a|u|$$

e per una Poisson $\mathfrak{P}(\alpha)$ è

$$\eta(u) = \alpha(e^{iu} - 1)$$

Anche la caratteristica logaritmica ammette (con le precisazioni dovute all'esistenza dei momenti) uno sviluppo in serie della la forma

$$\eta(u_1, \dots, u_n) = \sum_{\{\mathbf{k}\}} \frac{i^{|\mathbf{k}|}}{k_1! \dots k_n!} u_1^{k_1} \cdot \dots \cdot u_n^{k_n} c_n(k_1, \dots, k_n)$$

ma i coefficienti $c_n(k_1, \dots, k_n)$, che prendono il nome di **cumulanti**, non sono semplici valori d'attesa di prodotti di *v-a*. Confrontando opportunamente i due sviluppi in serie è però possibile dedurre delle relazioni che connettono i cumulanti con i momenti misti delle componenti di \mathbf{X} , e ad esempio si ha (qui la scelta della collocazione degli indici diversi da zero è arbitraria e solo esemplificativa)

$$\begin{aligned} c_n(1, 0, 0, \dots, 0) &= m_n(1, 0, 0, \dots, 0) \\ c_n(1, 1, 0, \dots, 0) &= m_n(1, 1, 0, \dots, 0) - m_n(1, 0, 0, \dots, 0)m_n(0, 1, 0, \dots, 0) \\ c_n(1, 1, 1, \dots, 0) &= m_n(1, 1, 1, \dots, 0) - m_n(1, 1, 0, \dots, 0)m_n(0, 0, 1, \dots, 0) \\ &\quad - m_n(1, 0, 1, \dots, 0)m_n(0, 1, 0, \dots, 0) \\ &\quad - m_n(0, 1, 1, \dots, 0)m_n(1, 0, 0, \dots, 0) \\ &\quad + 2m_n(1, 0, 0, \dots, 0)m_n(0, 1, 0, \dots, 0)m_n(0, 0, 1, \dots, 0). \end{aligned}$$

Le relazioni generali sono piuttosto complicate e non saranno qui riportate², noteremo però che il valore dei cumulanti in cui più di un indice è diverso da zero costituisce una misura della correlazione delle corrispondenti componenti X_k . Ad esempio, se le X_k sono tutte indipendenti la fc si fattorizza e quindi $\eta(u_1, \dots, u_n)$ si riduce alla somma di n termini ciascuno dei quali dipende da una sola u_k . In questo caso si vede facilmente dallo sviluppo in cumulanti che le c_n con più di un indice diverso da zero si annullano.

Inoltre, mentre a causa della disuguaglianza di Lyapunov $\mathbf{E}[X^n]^2 \leq \mathbf{E}[X^{2n}]$ non possiamo annullare tutti i momenti da un certo ordine in poi (cioè tutti i momenti contengono informazione rilevante), per i cumulanti questo è possibile in casi particolari. A questo proposito si dimostra in particolare³ che se $\eta(\mathbf{u})$ è un polinomio, il suo grado non può essere superiore a 2: si veda ad esempio la caratteristica logaritmica di $\mathfrak{N}(b, a^2)$. Ne segue quindi che o tutti i cumulanti si annullano eccetto i primi due, oppure il numero dei cumulanti non nulli è infinito.

²Per una procedura di calcolo dei cumulanti vedi ad esempio **C.W. Gardiner**, HANDBOOK OF STOCHASTIC METHODS, Springer (Berlin, 1997)

³**A.N. Shiryaev**, PROBABILITY, Springer (New York, 1996)

Appendice F

Teoremi limite binomiali (Sez. 4.3)

Le forme più antiche di Teoremi limite (inizio del XVIII secolo) riguardavano fondamentalmente successioni di v -a binomiali ed erano dimostrati utilizzando le proprietà analitiche delle loro distribuzioni. Le formulazioni più moderne riportate nel testo, confermando e generalizzando questi risultati, si riferiscono invece a casi molto più generali e usano tecniche di dimostrazione più evolute. In questa appendice richiameremo brevemente queste forme arcaiche di teoremi limite che conservano un loro potere suggestivo

Il teorema più antico dovuto a J. Bernoulli¹ parte dall'osservazione che se le v -a della successione $(X_n)_{n \in \mathbf{N}}$ sono *iid* $\mathfrak{B}(1; p)$, e rappresentano gli esiti delle estrazioni di palline bianche e nere secondo il modello di Bernoulli delle Sezioni 2.1.2 e 3.2.4, le somme $S_n = X_1 + \dots + X_n$ risultano binomiali $\mathfrak{B}(n; p)$ e per esse sappiamo che

$$\mathbf{E}[S_n] = np \qquad \mathbf{V}[S_n] = np(1 - p)$$

Ne segue allora che il valore d'attesa della v -a *frequenza empirica* S_n/n coincide anche con la *probabilità* p di trovare una pallina bianca in ogni singolo tentativo:

$$\mathbf{E}\left[\frac{S_n}{n}\right] = p$$

Ma la frequenza S_n/n è una v -a, non un numero come p , e quindi in generale il suo valore non coinciderà con p quando eseguiamo una n -pla di tentativi. Sarà allora interessante valutare di quanto la v -a *frequenza* S_n/n si discosta dal suo valore d'attesa, cioè dalla *probabilità* p , per sapere con quale affidabilità possiamo stimare p (che in generale non è nota) tramite una misura della frequenza empirica S_n/n su n tentativi di verifica. L'unica quantità accessibile alle nostre osservazioni, infatti, è sempre e soltanto il conteggio delle frequenze e non il valore p delle probabilità *a priori*. Si potrebbe in breve dire che il problema originario della **statistica** è proprio quello di stabilire le condizioni sotto le quali un'osservazione della frequenza empirica S_n/n consente di ricavare una *stima affidabile* di p . Mostriamo ora in che

¹J. Bernoulli, ARS CONIECTANDI, Thurneysen (Basilea, 1713)

senso preciso si possa ritenere piccolo lo scarto fra la frequenza e la probabilità a priori p prendendo n abbastanza grande.

Teorema F.1. Legge dei grandi numeri di Bernoulli: *Se S_n è una successione di v -a binomiali $\mathfrak{B}(n; p)$, si ha*

$$\frac{S_n}{n} \xrightarrow{\mathbf{P}} p$$

Dimostrazione: Dalla disuguaglianza di Chebyshev (3.42), e dalle proprietà delle v -a binomiali $\mathfrak{B}(n; p)$ si ha

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \geq \epsilon\right\} \leq \frac{1}{\epsilon^2} \mathbf{V}\left[\frac{S_n}{n}\right] = \frac{1}{n^2\epsilon^2} \mathbf{V}[S_n] = \frac{np(1-p)}{n^2\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$$

da cui discende facilmente il risultato in base alla Definizione 4.1. □

Anche la prima versione del **Teorema limite centrale** di De Moivre² si limitava a considerare successioni di v -a binomiali $\mathfrak{B}(n; \frac{1}{2})$, e le formulazioni successive di Laplace³ si riferivano pur sempre a successioni di v -a $\mathfrak{B}(n; p)$ con $0 < p < 1$. Questi teoremi limite si presentavano sotto molteplici vesti, e qui per brevità riporteremo informalmente solo il contenuto degli enunciati più classici. Se le X_n sono v -a iid di Bernoulli $\mathfrak{B}(1; p)$, sappiamo che $S_n = X_1, \dots, X_n \sim \mathfrak{B}(n; p)$, e tenendo conto di (3.35) e (3.36) la somma standardizzata

$$S_n^* = \frac{S_n - np}{\sqrt{npq}} \tag{F.1}$$

assumerà gli $n + 1$ valori (non interi)

$$x_k = \frac{k - np}{\sqrt{npq}} \quad k = 0, 1, \dots, n$$

Si avrà quindi da (2.1)

$$\mathbf{P}\{S_n^* = x_k\} = \mathbf{P}\{S_n = k\} = p_n(k) = p_n(np + x_k\sqrt{npq}) = \binom{n}{k} p^k q^{n-k}$$

La forma classica dei teoremi limite per il caso binomiale consiste nella formulazione di risultati asintotici ($n \rightarrow \infty$) che consentono di esprimere probabilità di eventi relativi alla v -a S_n^* in termini di funzioni di Gauss. Noi non li enunceremo nella loro forma rigorosa che è piuttosto complicata⁴, ma ci limiteremo solo a ricordarne i contenuti essenziali.

Un primo enunciato che porta il nome di **Teorema Limite Locale (TLL)** rappresenta la formulazione esatta dell'affermazione secondo la quale, per grandi

²A. De Moivre, THE DOCTRINE OF CHANCES, Woodfall (London, 1738)

³P.S. de Laplace, THÉORIE ANALYTIQUE DES PROBABILITÉS, Courcier (Paris, 1812)

⁴A.N. Shiryaev, PROBABILITY, Springer (New York, 1996)

valori di n , i valori della distribuzione binomiale $p_n(k) = p_n(np + x_k\sqrt{npq})$ sono ben approssimati dalla funzione gaussiana

$$\frac{e^{-(k-np)^2/2npq}}{\sqrt{2\pi npq}} = \frac{1}{\sqrt{npq}} \frac{e^{-x_k^2/2}}{\sqrt{2\pi}}$$

Va però precisato che questa approssimazione è buona solo se k non è troppo lontano dal valore d'attesa np di S_n , ovvero se x_k non è troppo lontano da 0. In pratica il *TLL* stabilisce che per grandi valori di n esistono degli opportuni numeri positivi A_n e B_n tali che

$$\begin{aligned} \mathbf{P}\{S_n = k\} &\simeq \frac{e^{-(k-np)^2/2npq}}{\sqrt{2\pi npq}} && \text{se } |k - np| \leq A_n, \\ \mathbf{P}\{S_n^* = x_k\} &\simeq \frac{1}{\sqrt{npq}} \frac{e^{-x_k^2/2}}{\sqrt{2\pi}} && \text{se } |x_k| \leq B_n. \end{aligned}$$

L'approssimazione non è invece altrettanto buona se ci spostiamo troppo verso le code della distribuzione, cioè se k si allontana troppo da np , ovvero se x_k si allontana troppo da 0.

Se vogliamo eliminare queste limitazioni si deve passare ad una seconda formulazione nota come **Teorema Limite Integrale (TLI)**. Per far questo osserviamo innanzitutto che, per p ed n dati, i numeri x_k sono tutti equidistanti, e poniamo

$$\Delta x_k = x_{k+1} - x_k = \frac{1}{\sqrt{npq}}$$

Per $n \rightarrow \infty$ e con x_k non troppo lontano da 0 il *TLL* ci autorizza allora a scrivere

$$\mathbf{P}\{S_n^* = x_k\} = p_n(np + x_k\sqrt{npq}) \simeq \frac{e^{-x_k^2/2}}{\sqrt{2\pi}} \Delta x_k$$

Siccome per $n \rightarrow \infty$ risulta anche $\Delta x_k \rightarrow 0$, l'insieme dei punti x_k tenderà a ricoprire tutta la retta reale, e quindi potremo attenderci che, in qualche senso opportunamente definito, per n molto grande e con $a < b$ arbitrari, il valore di

$$\mathbf{P}\{a < S_n^* \leq b\} = \sum_{k:a < x_k \leq b} p_n(np + x_k\sqrt{npq}) \simeq \sum_{k:a < x_k \leq b} \frac{e^{-x_k^2/2}}{\sqrt{2\pi}} \Delta x_k$$

possa essere approssimato mediante l'integrale

$$\int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \Phi(b) - \Phi(a)$$

dove $\Phi(x)$ è la funzione errore standard (2.16). Il *TLI* stabilisce infatti che, comunque scelti $-\infty \leq a < b \leq +\infty$, per $n \rightarrow \infty$ otterremo sempre

$$\mathbf{P}\{a < S_n^* \leq b\} \rightarrow \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \Phi(b) - \Phi(a)$$

ovvero, prendendo $\alpha = np + a\sqrt{npq}$ e $\beta = np + b\sqrt{npq}$,

$$\mathbf{P}\{\alpha < S_n \leq \beta\} \rightarrow \Phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha - np}{\sqrt{npq}}\right)$$

Da un punto di vista tecnico la differenza fra le due formulazioni dei Teoremi limite binomiali sta nel fatto che, mentre nel *TLL* si confrontano i singoli valori della distribuzione (discreta) binomiale standardizzata con i valori di una funzione (continua) normale standard, nel *TLI* si confrontano somme della suddetta distribuzione binomiale con integrali della normale standard su intervalli anche illimitati.

Le dimostrazioni tradizionali di questi due teoremi fanno ricorso ad argomenti analitici piuttosto complicati che noi tralascieremo⁵. Per mostrare invece ancora una volta l'utilità delle *fc* daremo una semplice dimostrazione della convergenza in distribuzione delle binomiali standard (F.1) S_n^* verso una legge normale standard $\mathfrak{N}(0, 1)$ facendo uso delle *fc*. Infatti, se X_1, \dots, X_n sono *v-a iid* di Bernoulli $\mathfrak{B}(1; p)$, posto

$$Y_k = \frac{X_k - p}{\sqrt{npq}} = \frac{X_k}{\sqrt{npq}} - \sqrt{\frac{p}{nq}}$$

potremo scrivere

$$S_n^* = \sum_{k=1}^n Y_k$$

e siccome da (4.3) e (4.8) si ha

$$\varphi_{Y_k}(u) = \mathbf{E}[e^{iuY_k}] = e^{-iu\sqrt{p/nq}} \varphi_{X_k}\left(\frac{u}{\sqrt{npq}}\right) = p e^{iu\sqrt{q/np}} + q e^{-iu\sqrt{p/nq}}$$

la *fc* di S_n^* risulta essere

$$\varphi_{S_n^*}(u) = \mathbf{E}[e^{iuS_n^*}] = \prod_{k=1}^n \mathbf{E}[e^{iuY_k}] = \left(p e^{iu\sqrt{q/np}} + q e^{-iu\sqrt{p/nq}}\right)^n$$

Dallo sviluppo in serie di potenze degli esponenziali si ha allora

$$\begin{aligned} \varphi_{S_n^*}(u) &= \left[p \left(1 + iu\sqrt{\frac{q}{np}} - \frac{u^2}{2} \frac{q}{np} \right) + q \left(1 - iu\sqrt{\frac{p}{nq}} - \frac{u^2}{2} \frac{p}{nq} \right) + o\left(\frac{1}{n}\right) \right]^n \\ &= \left[1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right) \right]^n \rightarrow e^{-u^2/2} \end{aligned}$$

e quindi in base al Teorema 4.16 si ottiene $S_n^* \xrightarrow{d} \mathfrak{N}(0, 1)$.

⁵A.N. Shiryaev, **PROBABILITY**, Springer (New York, 1996)

Appendice G

Processi di punto con intensità non uniforme (Sez. 6.1.1)

Nella procedura di limite che ci ha condotto a definire i processi di punto nella Sezione 6.1.1 abbiamo supposto che la distribuzione dei punti lanciati su ogni intervalli finito restasse sempre uniforme $\mathcal{U}[-\frac{\tau}{2}, \frac{\tau}{2}]$. Questa ipotesi però non è essenziale e può essere opportunamente modificata immaginando che l'intensità della pioggia di punti possa variare da zona a zona.

Per esaminare questa idea osserviamo innanzitutto che, nel caso uniforme studiato finora, la *v-a* N che conta il numero di punti caduti in un arbitrario intervallo di larghezza Δt risulta seguire una legge di Poisson $\mathfrak{P}(\alpha)$ con $\alpha = \lambda\Delta t$, per cui $\mathbf{E}[N] = \alpha = \lambda\Delta t$. Se allora teniamo conto del fatto che

$$\alpha = \mathbf{E}[N] \rightarrow 0 \quad \text{per} \quad \Delta t \rightarrow 0$$

e adottiamo la notazione $\Delta\nu = \alpha = \mathbf{E}[N] =$ *numero medio di punti in un intervallo di larghezza Δt* , potremo anche scrivere

$$\lambda = \frac{\Delta\nu}{\Delta t} \longrightarrow \frac{d\nu}{dt} \quad \Delta t \rightarrow 0$$

coerentemente con l'idea che λ rappresenta il *numero medio di punti per unità di tempo*. Una λ costante, come da noi ipotizzato in precedenza, corrisponderà all'idea di una densità uniforme di punti, ma possiamo anche supporre che $\lambda(t)$ sia una densità variabile con il tempo. In questo caso avremmo

$$\lambda(t) = \frac{d\nu(t)}{dt} \quad \text{ossia} \quad d\nu(t) = \lambda(t) dt$$

e quindi anche

$$\alpha = \Delta\nu = \int_t^{t+\Delta t} \lambda(s) ds \quad (\text{G.1})$$

Se ora N è il numero di punti aleatori che cadono in $[t, t + \Delta t]$, ripercorrendo gli stessi argomenti esposti in precedenza per il caso uniforme potremo mostrare ancora

una volta che N segue una distribuzione di Poisson $\mathfrak{P}(\alpha)$, ma in questo caso il valore di α sarà (G.1) e non dipenderà solo dall'ampiezza Δt dell'intervallo considerato, ma anche dalla collocazione temporale t del suo estremo sinistro. Si noti infine che $\lambda(t)$ è una densità (numero medio di punti per unità di tempo), ma non è una *ddp*. Infatti tipicamente si ha

$$\int_{-\infty}^{+\infty} \lambda(t) dt = +\infty$$

coerentemente con il fatto che tale integrale rappresenta il numero totale (infinito, quindi) di punti lanciati su *tutto* l'asse dei tempi. Nel testo abbiamo supposto sempre che l'intensità λ sia costante, ma le eventuali generalizzazioni possono comunque essere ottenute senza troppe difficoltà sulla base delle precedenti osservazioni.

Appendice H

Paradossi del calcolo stocastico (Sez. 6.4.2)

Per mostrare quali errori si possono commettere adattando in maniera eccessivamente sbrigativa l'usuale calcolo differenziale ai processi stocastici, proviamo ad estendere il metodo euristico di Langevin esposto nella Sezione 6.4.2 a un problema un po' diverso: il rumore granulare prodotto nei tubi a vuoto dalle correnti aleatorie generate dall'arrivo di elettroni isolati.

Per rappresentare la corrente aleatoria $I(t)$ prodotta dall'arrivo degli elettroni useremo un rumore granulare con $h(t) = \vartheta(t)qe^{-at}$, per cui, tenendo conto anche di (6.67) con rumore bianco di Poisson, la nostra corrente sarà un processo del tipo

$$\begin{aligned} I(t) &= \sum_{k=1}^{\infty} h(t - T_k) = [h * \dot{N}](t) = \int_{-\infty}^{+\infty} h(t - s)\dot{N}(s) ds \\ &= qe^{-at} \int_{-\infty}^t e^{as}\dot{N}(s) ds \end{aligned}$$

Dalle usuali regole del calcolo differenziale, e introducendo anche il rumore bianco (6.66) del processo di Poisson compensato, si ha

$$\begin{aligned} \dot{I}(t) &= -qae^{-at} \int_{-\infty}^t e^{as}\dot{N}(s) ds + q\dot{N}(t) \\ &= -aI(t) + q\dot{N}(t) = [\lambda q - aI(t)] + q\tilde{\dot{N}}(t) \end{aligned} \quad (\text{H.1})$$

Si tratta di una equazione differenziale del primo ordine simile all'equazione di Langevin (6.78) nella quale il ruolo della forza fluttuante a media nulla $B(t)$ è giocato da $q\tilde{\dot{N}}(t)$ che sarà anche qui supposto non correlato con $I(t)$. Per esaminare le fluttuazioni di $I(t)$ ne studieremo la varianza

$$\mathbf{V} [I(t)] = \mathbf{E} [I^2(t)] - \mathbf{E} [I(t)]^2$$

e a questo scopo prenderemo innanzitutto l'attesa di (H.1)

$$\frac{d}{dt} \mathbf{E} [I(t)] = \lambda q - a \mathbf{E} [I(t)]$$

da cui si ottiene facilmente

$$\mathbf{E} [I(t)] = \frac{\lambda q}{a} + C e^{-at} \quad (\text{H.2})$$

dove C è una costante di integrazione. Per ottenere la varianza dovremo ora calcolare $\mathbf{E} [I^2(t)]$: moltiplicando (H.1) per $I(t)$, dalle regole del calcolo differenziale abbiamo innanzitutto

$$\frac{1}{2} \frac{dI^2(t)}{dt} = I(t) \dot{I}(t) = \lambda q I(t) - a I^2(t) + q I(t) \tilde{N}(t) \quad (\text{H.3})$$

e quindi prendendo il valor d'attesa

$$\frac{1}{2} \frac{d}{dt} \mathbf{E} [I^2(t)] = \lambda q \mathbf{E} [I(t)] - a \mathbf{E} [I^2(t)] \quad (\text{H.4})$$

Dalla (H.2) si ha allora

$$\frac{d}{dt} \mathbf{E} [I^2(t)] + 2a \mathbf{E} [I^2(t)] = 2\lambda q \mathbf{E} [I(t)] = 2\lambda q \left(\frac{\lambda q}{a} + C e^{-at} \right)$$

e con un'altra costante di integrazione A

$$\mathbf{E} [I^2(t)] = \left(\frac{\lambda q}{a} \right)^2 + C \frac{2\lambda q}{a} e^{-at} + A e^{-2at}$$

In conclusione la varianza della corrente aleatoria sarà

$$\mathbf{V} [I(t)] = \mathbf{E} [I^2(t)] - \mathbf{E} [I(t)]^2 = \left(\frac{\lambda q}{a} \right)^2 + C \frac{2\lambda q}{a} e^{-at} + A e^{-2at} - \left(\frac{\lambda q}{a} + C e^{-at} \right)^2$$

sicché per grandi tempi paradossalmente avremo

$$\lim_{t \rightarrow +\infty} \mathbf{V} [I(t)] = 0 \quad (\text{H.5})$$

cioè, dopo un breve transitorio, le fluttuazioni semplicemente spariscono, mentre ci saremmo attesi di veder convergere la loro varianza verso una costante positiva non nulla. Questo risultato merita di essere esaminato in maggior dettaglio

Ripartiamo dalla relazione – apparentemente ovvia – utilizzata nell'equazione (H.3):

$$\frac{dI^2(t)}{dt} = 2I(t) \dot{I}(t)$$

e rivisitiamo, alla luce delle considerazioni svolte nella Sezione 6.3, i classici argomenti che la giustificano. Tipicamente, con un dt infinitesimo, si considera

$$d [I^2(t)] = I^2(t+dt) - I^2(t) = [I(t) + dI(t)]^2 - I^2(t) = 2I(t)dI(t) + [dI(t)]^2 \quad (\text{H.6})$$

ma ora – a causa della non derivabilità di $I(t)$ – non possiamo più affermare che il termine $[dI(t)]^2$ coincide con $[\dot{I}(t)dt]^2$, cioè con un infinitesimo di ordine superiore a dt . Piuttosto dobbiamo riprendere l'equazione (H.1) riscritta – per evitare di introdurre derivate – nella forma

$$dI(t) = \lambda q dt - aI(t)dt + qd\tilde{N}(t) \quad (\text{H.7})$$

e sostituirla in (H.6):

$$\begin{aligned} d[I^2(t)] &= 2I(t)[\lambda q dt - aI(t)dt + qd\tilde{N}(t)] + [\lambda q dt - aI(t)dt + qd\tilde{N}(t)]^2 \\ &= [2\lambda q I(t) - 2aI^2(t)]dt + [\lambda q - aI(t)]^2(dt)^2 \\ &\quad + 2qI(t)d\tilde{N}(t) + 2q[\lambda q - aI(t)]d\tilde{N}(t)dt + q^2[d\tilde{N}(t)]^2 \end{aligned}$$

Passando poi alle attese, si trascurano i termini di ordine superiore a dt ricordando però che in base a (6.73) $\mathbf{E}[d\tilde{N}^2] = \lambda dt$ è ora dell'ordine di dt per cui

$$d\mathbf{E}[I^2(t)] = (2\lambda q \mathbf{E}[I(t)] - 2a \mathbf{E}[I^2(t)]) dt + \lambda q^2 dt$$

In definitiva invece di (H.4) avremo

$$\frac{1}{2} \frac{d}{dt} \mathbf{E}[I^2(t)] = \lambda q \mathbf{E}[I(t)] - a \mathbf{E}[I^2(t)] + \frac{\lambda q^2}{2}$$

nella quale compare il nuovo termine aggiuntivo $\lambda q^2/2$. Utilizzando allora ancora (H.2) e ripercorrendo i medesimi passaggi che hanno condotto al risultato paradossale (H.5) si perviene alla seguente soluzione con i corretti andamenti asintotici per $t \rightarrow +\infty$

$$\begin{aligned} \mathbf{E}[I(t)] &= \frac{\lambda q}{a} + C e^{-at} \longrightarrow \frac{\lambda q}{a} \\ \mathbf{E}[I^2(t)] &= \left(\frac{\lambda q}{a}\right)^2 + \frac{\lambda q^2}{2a} + C \frac{2\lambda q}{a} e^{-at} + A e^{-2at} \longrightarrow \left(\frac{\lambda q}{a}\right)^2 + \frac{\lambda q^2}{2a} \\ \mathbf{V}[I(t)] &= \mathbf{E}[I^2(t)] - \mathbf{E}[I(t)]^2 \longrightarrow \frac{\lambda q^2}{2a} > 0 \end{aligned}$$

Le osservazioni sugli ordini degli infinitesimi introdotte nella Sezione 6.3 – ancorché imprecise e intuitive – giocano quindi un ruolo cruciale per ottenere un risultato accettabile

Merita infine un breve esame il modo in cui Langevin – pur basandosi su una formulazione matematica poco rigorosa – è riuscito ad evitare gli errori prima esposti per le correnti, arrivando invece a risultati corretti. Mostriamo infatti che, a differenza della relazione (H.3), le due relazioni (6.79) e (6.80), per quanto solo simboliche, sono sostanzialmente corrette; e a questo scopo osserveremo innanzitutto che la distinzione fra le due formulazioni (6.77) e (6.78) dell'equazione dinamica

nasconde alcuni dettagli importanti. La possibilità di tale distinzione indica infatti che $X(t)$ è considerata derivabile, cioè che il processo $\dot{X}(t) = V(t)$ esiste, e che in qualche senso risulta

$$X(t) = \int_0^t V(s) ds$$

mentre l'equazione di Newton

$$m\ddot{X}(t) = -6\pi\eta a\dot{X}(t) + B(t) \tag{H.8}$$

equivale al sistema

$$\begin{aligned} \dot{X}(t) &= V(t) \\ m\dot{V}(t) &= -6\pi\eta aV(t) + B(t) \end{aligned} \tag{H.9}$$

Siccome la forza aleatoria $B(t)$ disturba essenzialmente solo la $V(t)$, questa – diversamente da $X(t)$ – risulterà non derivabile, sicché l'equazione di Langevin (H.9), con $V(t)$ al posto di $I(t)$, sarà dello stesso tipo dell'equazione (H.1) usata per discutere il problema di rumore granulare. Da queste osservazioni discende prima di tutto che se Langevin avesse usato la (H.9) assieme a una formula del tipo

$$\frac{dV^2(t)}{dt} = 2V(t)\dot{V}(t) \tag{H.10}$$

avrebbe raggiunto per $V(t)$ le stesse conclusioni paradossali dell'equazione (H.5) per il rumore granulare: dopo un breve transitorio la particella Browniana si arresterebbe, con $V(t) = 0$ non solo in media, ma addirittura **P**-qo, cioè con varianza nulla. Il suo ragionamento parte invece dall'equazione di Newton (H.8) per il processo $X(t)$ e usa le relazioni, simboliche ma sostanzialmente corrette,

$$\frac{d}{dt} [X^2(t)] = 2X(t)\dot{X}(t) \tag{H.11}$$

$$\frac{d^2}{dt^2} [X^2(t)] = 2\dot{X}^2(t) + 2X(t)\ddot{X}(t) = 2V^2(t) + 2X(t)\ddot{X}(t) \tag{H.12}$$

per ricavare il risultato di Einstein (6.82). Dobbiamo allora spiegare perché le equazioni (H.11) e (H.12) possono essere usate senza problemi, mentre la (H.10), come abbiamo visto, conduce a dei paradossi.

Prima di tutto bisogna osservare che, essendo $X(t)$ derivabile, risulta $dX(t) = \dot{X}(t)dt = V(t)dt$, per cui $dX(t)$ è un infinitesimo dell'ordine di dt , e quindi la (H.11) è corretta e potremo sicuramente scrivere

$$\frac{d}{dt} [X^2(t)] = 2X(t)\dot{X}(t) = 2X(t)V(t) \tag{H.13}$$

La (H.12), invece, pur essendo sostanzialmente corretta, resta puramente simbolica perché contiene una derivata $\dot{X}(t) = \dot{V}(t)$ che non esiste. Per vedere allora in che

senso essa resta accettabile dovremo osservare che

$$\begin{aligned}
d[X(t)V(t)] &= X(t+dt)V(t+dt) - X(t)V(t) \\
&= [X(t) + dX(t)] [V(t) + dV(t)] - X(t)V(t) \\
&= [X(t) + V(t)dt] [V(t) + dV(t)] - X(t)V(t) \\
&= V^2(t)dt + X(t)dV(t) + V(t)dV(t)dt
\end{aligned}$$

D'altra parte $dV(t)$ è certamente un infinitesimo, anzi più precisamente un infinitesimo dell'ordine $O(dt^{1/2})$: infatti, come vedremo meglio nella Sezione 8.1, la forza aleatoria $B(t)$ che compare nell'equazione di Langevin (H.9) è proprio un rumore bianco Wieneriano $\dot{W}(t)$, per cui riscrivendo (H.9) nella forma

$$m dV(t) = -6\pi\eta a V(t)dt + dW(t)$$

si ricava che $dV(t)$ è un infinitesimo dello stesso ordine di $dW(t)$, cioè $O(dt^{1/2})$. In definitiva potremo ritenere che $dV(t)dt$ sia comunque un infinitesimo di ordine superiore al primo, cioè $O(dt^{3/2})$, e al primo ordine potremo scrivere

$$d[X(t)V(t)] = V^2(t)dt + X(t)dV(t)$$

Pertanto, almeno simbolicamente, possiamo dire che

$$\frac{d}{dt}[X(t)V(t)] = V^2(t) + X(t)\dot{V}(t) = V^2(t) + X(t)\ddot{X}(t)$$

e possiamo quindi giustificare la (H.12) tramite la (H.13).

Appendice I

Processi pseudo-Markoviani (Sez. 7.1.2)

Daremo in questa appendice un semplice esempio¹ di processo non Markoviano le cui probabilità di transizione soddisfano comunque l'equazione di Chapman-Kolmogorov: processi di questo genere sono anche chiamati pseudo-Markoviani. Si consideri il processo $X(t)$ che assume solo i due valori 0, 1 ai tempi $t = 1, 2, 3$: si tratta di un processo estremamente semplice le cui traiettorie si riducono alle $8 = 2^3$ possibili terne di simboli 0, 1, e quindi la sua legge può essere fissata scegliendo in modo coerente le probabilità per questi 8 elementi. In particolare – adottando per brevità notazioni del tipo $p_{1,2,3}(x_1, x_2, x_3) = \mathbf{P}\{X(1) = x_1, X(2) = x_2, X(3) = x_3\}$, oppure $p_{3|2,1}(x_3|x_2, x_1) = \mathbf{P}\{X(3) = x_3 | X(2) = x_2, X(1) = x_1\}$ e così via – noi fisseremo i seguenti valori per la distribuzione congiunta $p_{1,2,3}(x_1, x_2, x_3)$

x_1	x_2	x_3	$p_{1,2,3}$
0	0	0	0
0	0	1	$1/4$
0	1	0	$1/4$
0	1	1	0
1	0	0	$1/4$
1	0	1	0
1	1	0	0
1	1	1	$1/4$

Questa tabella definisce completamente la legge del processo perché tutte le altre leggi congiunte possono poi da essa essere ricavate per marginalizzazione, come negli

¹Questo esempio è riportato come esercizio alla p. 79 di **N.G. van Kampen**, STOCHASTIC PROCESSES IN PHYSICS AND CHEMISTRY, North-Holland (Amsterdam, 1992), ma la sua origine risale a una nota di **P. Lévy** (C. R. Acad. Sci. Paris **228** (1949) 2204) ripresa prima da **W. Feller** (Ann. Math. Stat. **30** (1959) 1252) e poi da **E. Parzen** (STOCHASTIC PROCESSES, Holden-Day (San Francisco, 1962) p. 203)

esempi seguenti

$$\begin{aligned}
 p_{1,2}(0,0) &= p_{1,2,3}(0,0,0) + p_{1,2,3}(0,0,1) = 1/4 \\
 p_{1,2}(1,0) &= p_{1,2,3}(1,0,0) + p_{1,2,3}(1,0,1) = 1/4 \\
 p_{2,3}(0,0) &= p_{1,2,3}(0,0,0) + p_{1,2,3}(1,0,0) = 1/4 \\
 p_2(0) &= p_{1,2,3}(0,0,0) + p_{1,2,3}(1,0,0) + p_{1,2,3}(0,0,1) + p_{1,2,3}(1,0,1) = 1/2 \\
 p_2(1) &= p_{1,2,3}(0,1,0) + p_{1,2,3}(1,1,0) + p_{1,2,3}(0,1,1) + p_{1,2,3}(1,1,1) = 1/2
 \end{aligned}$$

dando luogo ai risultati riportati in dettaglio nelle successive tabelle

x_1	x_2	$p_{1,2}$
0	0	$1/4$
0	1	$1/4$
1	0	$1/4$
1	1	$1/4$

x_2	x_3	$p_{2,3}$
0	0	$1/4$
0	1	$1/4$
1	0	$1/4$
1	1	$1/4$

x_1	x_3	$p_{1,3}$
0	0	$1/4$
0	1	$1/4$
1	0	$1/4$
1	1	$1/4$

x_1	p_1
0	$1/2$
1	$1/2$

x_2	p_2
0	$1/2$
1	$1/2$

x_3	p_3
0	$1/2$
1	$1/2$

Siamo pertanto in grado di calcolare anche le distribuzioni condizionate e di verificare innanzitutto che il processo $X(t)$ non è Markoviano: infatti si ha che

$$\begin{aligned}
 p_{3|2,1}(0|0,0) &= \frac{p_{1,2,3}(0,0,0)}{p_{1,2}(0,0)} = \frac{0}{1/4} = 0 \\
 p_{3|2}(0|0) &= \frac{p_{2,3}(0,0)}{p_2(0)} = \frac{1/4}{1/2} = 1/2
 \end{aligned}$$

per cui, almeno in un caso, risulta $p_{3|2,1} \neq p_{3|2}$, e quindi il processo *non è Markoviano*. Ciononostante è facile vedere che le probabilità di transizione $p_{2|1}, p_{3|2}$ e $p_{3|1}$ soddisfano le equazioni di Chapman-Kolmogorov, cioè – in base alla Definizione 7.6 – sono *leggi di transizione Markoviane*. Infatti dalle tabelle precedenti si ricava facilmente che in tutti i casi

$$p_{2|1}(x_2 | x_1) = p_{3|2}(x_3 | x_2) = p_{3|1}(x_3 | x_1) = 1/2 \quad x_1, x_2, x_3 = 0, 1$$

per cui le equazioni di Chapman-Kolmogorov

$$\sum_{x_2=0}^1 p_{3|2}(x_3 | x_2) p_{2|1}(x_2 | x_1) = 1/2 \cdot 1/2 + 1/2 \cdot 1/2 = 1/4 + 1/4 = 1/2 = p_{3|1}(x_3 | x_1)$$

sono sempre soddisfatte.

Questo paradosso è discusso in generale nella Sezione 7.1.2, dove si fa notare che esso può essere risolto ricordando che – con lo stesso spazio delle traiettorie –

possono esistere processi dotati di leggi differenti, ma che condividono la medesima famiglia di leggi di transizione Markoviane: di questi processi solo uno può essere Markoviano. Nel nostro esempio – sullo spazio delle 8 traiettorie (x_1, x_2, x_3) – il processo $X(t)$ la cui legge p è definita nella tabella iniziale non è Markoviano, ma le sue leggi di transizione lo sono: a partire da queste ultime, mediante la regola della moltiplicazione a catena (Proposizione 7.4), sarà quindi possibile definire un *altro processo* $\tilde{X}(t)$ (sullo stesso spazio delle traiettorie, ma con una *legge differente* \tilde{p}) che ovviamente risulterà Markoviano. Per le leggi congiunte di questo secondo processo avremo infatti per tutti i valori $x_1, x_2, x_3 = 0, 1$

$$\tilde{p}_{1,2,3}(x_1, x_2, x_3) = p_{3|2}(x_3 | x_2)p_{2|1}(x_2 | x_1)p_1(x_1) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

cioè la tabella

x_1	x_2	x_3	$\tilde{p}_{1,2,3}$
0	0	0	$\frac{1}{8}$
0	0	1	$\frac{1}{8}$
0	1	0	$\frac{1}{8}$
0	1	1	$\frac{1}{8}$
1	0	0	$\frac{1}{8}$
1	0	1	$\frac{1}{8}$
1	1	0	$\frac{1}{8}$
1	1	1	$\frac{1}{8}$

Le distribuzioni bivariate e univariate, invece, restano uguali a quelle del processo $X(t)$

Appendice J

Moto Browniano frazionario (Sez. 7.1.10)

Bisogna ricordare che esistono casi anche piuttosto semplici di ddp di transizione che non soddisfano l'equazione di Chapman-Kolmogorov (7.17): in tal caso la nostra ddp – pur potendo legittimamente essere la ddp condizionata a due tempi di un qualche ipotetico processo stocastico – non può giocare il ruolo di ddp di transizione di un processo di Markov, e quindi non permetterà da sola di ricostruire la legge del processo mediante la regola di moltiplicazione a catena. Abbiamo però già notato nella Sezione 7.1.10 che per la determinazione della legge del processo si può supplire alla mancanza di Markovianità se il processo è Gaussiano. Ricorderemo allora che un importante caso di processo Gaussiano non Markoviano è rappresentato dal cosiddetto **moto Browniano frazionario** $Y(t)$ che può essere considerato come una generalizzazione dell'usuale processo di Wiener. Partiamo dall'osservazione (facilmente verificata con un calcolo diretto) che per $s, t > 0$ risulta

$$\min\{s, t\} = \frac{t + s - |t - s|}{2} = \frac{|t| + |s| - |t - s|}{2}$$

e poi per definire le leggi Gaussiane di $Y(t)$ prendiamo $m_Y(t) = 0$ e la nuova funzione di autocovarianza

$$C_Y(t, s) = \frac{D}{2} (|t|^{2H} + |s|^{2H} - |t - s|^{2H})$$

dove H , detto *indice di Hurst*, è un numero reale con $0 < H < 1$. Da quanto detto prima è evidente che l'autocovarianza dell'usuale processo di Wiener si ottiene nel caso particolare $H = \frac{1}{2}$, mentre per tutti gli altri valori di H si prova che C_Y è definita non negativa, e quindi che può essere legittimamente utilizzata per costruire il processo Gaussiano $Y(t)$ che prende il nome di moto Browniano frazionario. Ovviamente le specifiche proprietà del processo $Y(t)$ cambiano secondo il valore di H fissato, e possono essere esaminate in dettaglio – anche se noi eviteremo di farlo – visto che le leggi finito-dimensionali del processo sono esplicitamente conosciute. In

particolare il valore dell'indice di Hurst H è legato alla dipendenza fra gli incrementi di $Y(t)$, e conseguentemente alla irregolarità delle traiettorie. Si potrebbe mostrare infatti che – escludendo il caso Wieneriano $H = \frac{1}{2}$ per il quale gli incrementi sono indipendenti – un valore $H > \frac{1}{2}$ indica una *correlazione positiva* fra gli incrementi, mentre un valore $H < \frac{1}{2}$ è segno di una *correlazione negativa*. Intuitivamente potremmo dire che l'andamento precedente (crescente o decrescente) della traiettoria ne influenza quello successivo: se $H > \frac{1}{2}$ la correlazione positiva tende a rendere più regolari le traiettorie (l'andamento precedente tende ad essere conservato), mentre se $H < \frac{1}{2}$ la correlazione negativa produce l'effetto opposto (l'andamento precedente viene costantemente smentito da quello successivo).

Appendice K

Equazioni del processo di Ornstein-Uhlenbeck (Sez. 7.2.4)

Riportiamo in questa appendice un calcolo esplicito dei coefficienti dell'equazione in avanti per processi di Ornstein-Uhlenbeck $X(t)$, per mostrare che le loro leggi soddisfano l'equazione di Fokker-Planck (7.90) della Proposizione 7.39. La continuità globale dei processi di Ornstein-Uhlenbeck discende allora dal fatto che il termine di salto si annulla.

Cominceremo infatti con il provare che il termine di salto (7.61) si annulla e quindi che $X(t)$ è continuo. Osserviamo infatti che da (7.54) si ha

$$\begin{aligned} \frac{1}{\Delta t} f(x, t + \Delta t | y, t) &= \frac{e^{-\frac{[(x-y)+y(1-e^{-\alpha\Delta t})]^2}{2\beta^2(1-e^{-2\alpha\Delta t})}}}{\Delta t \sqrt{2\pi\beta^2(1-e^{-2\alpha\Delta t})}} \\ &= \frac{\alpha e^{-\frac{(x-y)^2}{2\beta^2(1-e^{-2\alpha\Delta t})}}}{\alpha\Delta t \sqrt{2\pi\beta^2(1-e^{-2\alpha\Delta t})}} e^{-\frac{y^2(1-e^{-\alpha\Delta t})^2+2y(x-y)(1-e^{-\alpha\Delta t})}{2\beta^2(1-e^{-2\alpha\Delta t})}} \\ &= \frac{e^{-\frac{(x-y)^2\alpha\Delta t}{2\beta^2(1-e^{-2\alpha\Delta t})}} \frac{1}{\alpha\Delta t}}{(\alpha\Delta t)^{\frac{3}{2}} \sqrt{2\pi\beta^2 \frac{1-e^{-2\alpha\Delta t}}{\alpha\Delta t}}} \alpha e^{-\frac{y^2(1-e^{-\alpha\Delta t})+2y(x-y)}{2\beta^2(1+e^{-\alpha\Delta t})}} \end{aligned}$$

e siccome è facile vedere che

$$\lim_{\Delta t \rightarrow 0} \alpha e^{-\frac{y^2(1-e^{-\alpha\Delta t})+2y(x-y)}{2\beta^2(1+e^{-\alpha\Delta t})}} = \alpha e^{-\frac{y(x-y)}{2\beta^2}} \quad \lim_{\Delta t \rightarrow 0} \frac{1-e^{-2\alpha\Delta t}}{2\alpha\Delta t} = 1$$

potremo eseguire il limite in due tempi riducendoci prima a

$$\ell(x|y, t) = \alpha e^{-\frac{y(x-y)}{2\beta^2}} \lim_{\Delta t \rightarrow 0} \frac{e^{-\frac{(x-y)^2}{4\beta^2} \frac{1}{\alpha\Delta t}}}{(\alpha\Delta t)^{\frac{3}{2}} \sqrt{4\pi\beta^2}}$$

e poi, con $z = \frac{1}{\alpha\Delta t} \rightarrow +\infty$, eseguendo il limite elementare

$$\ell(x|y, t) = \frac{\alpha e^{-\frac{y(x-y)}{2\beta^2}}}{\sqrt{4\pi\beta^2}} \lim_{z \rightarrow +\infty} z^{\frac{3}{2}} e^{-\frac{(x-y)^2}{4\beta^2} z} = 0$$

Ne segue in particolare che l'equazione del processo di Ornstein-Uhlenbeck sarà un'equazione di Fokker-Planck della quale dovremo ora determinare i coefficienti A e B . Modificando opportunamente la notazione, da (7.62) abbiamo innanzitutto che

$$A(x, t) = \lim_{\epsilon \rightarrow 0^+} \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \int_{|z-x|<\epsilon} (z-x) f(z, t + \Delta t | x, t) dz$$

e quindi da (7.54), ponendo

$$y = \frac{z - xe^{-\alpha\Delta t}}{\sqrt{\beta^2(1 - e^{-2\alpha\Delta t})}} \quad a_{\pm} = \frac{x(1 - e^{-\alpha\Delta t}) \pm \epsilon}{\sqrt{\beta^2(1 - e^{-2\alpha\Delta t})}}$$

risulterà

$$A(x, t) = \lim_{\epsilon \rightarrow 0^+} \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \left[\sqrt{\beta^2(1 - e^{-2\alpha\Delta t})} \int_{a_-}^{a_+} \frac{ye^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy - x(1 - e^{-\alpha\Delta t}) \int_{a_-}^{a_+} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy \right]$$

Tenendo poi conto delle primitive Gaussiane

$$\int \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy = \Phi(y) + cost \quad \int \frac{ye^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy = -e^{-\frac{y^2}{2}} + cost$$

dove $\Phi(x)$ è la funzione errore (2.16), avremo in definitiva

$$A(x, t) = \lim_{\epsilon \rightarrow 0^+} \lim_{\Delta t \rightarrow 0^+} \left[\sqrt{\frac{\alpha\beta^2}{\pi} \frac{1 - e^{-2\alpha\Delta t}}{2\alpha\Delta t}} \frac{e^{-\frac{a_+^2}{2}} - e^{-\frac{a_-^2}{2}}}{\sqrt{\Delta t}} - \alpha x \frac{1 - e^{-\alpha\Delta t}}{\alpha\Delta t} (\Phi(a_+) - \Phi(a_-)) \right]$$

Osservando allora che per ogni $\epsilon > 0$

$$\lim_{\Delta t \rightarrow 0^+} a_{\pm} = \pm\infty \quad \lim_{\Delta t \rightarrow 0^+} (\Phi(a_+) - \Phi(a_-)) = 1 \quad \lim_{u \rightarrow 0} \frac{1 - e^{-u}}{u} = 1$$

si avrà

$$\lim_{\Delta t \rightarrow 0^+} \frac{1 - e^{-\alpha\Delta t}}{\alpha\Delta t} (\Phi(a_+) - \Phi(a_-)) = 1 \quad \lim_{\Delta t \rightarrow 0^+} \sqrt{\frac{\alpha\beta^2}{\pi} \frac{1 - e^{-2\alpha\Delta t}}{2\alpha\Delta t}} = \sqrt{\frac{\alpha\beta^2}{\pi}}$$

mentre per l'altro fattore sviluppando i calcoli si ha

$$\begin{aligned} \frac{e^{-\frac{a_+^2}{2}} - e^{-\frac{a_-^2}{2}}}{\sqrt{\Delta t}} &= e^{-\frac{x^2(1-e^{-\alpha\Delta t})^2}{2\beta^2(1-e^{-2\alpha\Delta t})}} \frac{e^{-\frac{\epsilon^2}{2\beta^2(1-e^{-2\alpha\Delta t})}}}{\sqrt{\Delta t}} \left(e^{\frac{\epsilon x(1-e^{-\alpha\Delta t})}{\beta^2(1-e^{-2\alpha\Delta t})}} - e^{-\frac{\epsilon x(1-e^{-\alpha\Delta t})}{\beta^2(1-e^{-2\alpha\Delta t})}} \right) \\ &= e^{-\frac{x^2(1-e^{-\alpha\Delta t})}{2\beta^2(1+e^{-\alpha\Delta t})}} \frac{e^{-\frac{\epsilon^2}{4\beta^2\alpha\Delta t}} \frac{2\alpha\Delta t}{1-e^{-2\alpha\Delta t}}}{\sqrt{\Delta t}} \left(e^{\frac{\epsilon x}{\beta^2(1+e^{-\alpha\Delta t})}} - e^{-\frac{\epsilon x}{\beta^2(1+e^{-\alpha\Delta t})}} \right) \end{aligned}$$

e siccome con $u = \frac{1}{\Delta t}$ risulta

$$\begin{aligned} \lim_{\Delta t \rightarrow 0^+} e^{-\frac{x^2(1-e^{-\alpha\Delta t})}{2\beta^2(1+e^{-\alpha\Delta t})}} \left(e^{\frac{\epsilon x}{\beta^2(1+e^{-\alpha\Delta t})}} - e^{-\frac{\epsilon x}{\beta^2(1+e^{-\alpha\Delta t})}} \right) &= e^{\frac{\epsilon x}{2\beta^2}} - e^{-\frac{\epsilon x}{2\beta^2}} \\ \lim_{\Delta t \rightarrow 0^+} \frac{2\alpha\Delta t}{1 - e^{-2\alpha\Delta t}} &= 1 \\ \lim_{\Delta t \rightarrow 0^+} \frac{e^{-\frac{\epsilon^2}{4\beta^2\alpha\Delta t}}}{\sqrt{\Delta t}} &= \lim_{u \rightarrow +\infty} u^{\frac{1}{2}} e^{-\frac{\epsilon^2 u}{4\alpha\beta^2}} = 0 \end{aligned}$$

in definitiva avremo per ogni $\epsilon > 0$

$$\lim_{\Delta t \rightarrow 0^+} \frac{e^{-\frac{a^2}{2}} - e^{-\frac{a^2}{2}}}{\sqrt{\Delta t}} = 0$$

Raccogliendo allora tutti i fattori si ottiene

$$A(x, t) = -\alpha x$$

Con una procedura simile, che per brevità non riporteremo, si prova infine che il coefficiente di diffusione B è costante. Più precisamente, introducendo per analogia con il caso Wieneriano la quantità $D = 2\alpha\beta^2$, si ha

$$B(x, t) = D = 2\alpha\beta^2$$

sicché in definitiva l'equazione di Fokker-Planck per il processo di Ornstein-Uhlenbeck assume la forma (7.90)

$$\partial_t f(x, t) = \alpha \partial_x [x f(x, t)] + \alpha \beta^2 \partial_x^2 f(x, t) = \alpha \partial_x [x f(x, t)] + \frac{D}{2} \partial_x^2 f(x, t)$$

La ddp di transizione (7.54) al tempo $t + \Delta t$ corrisponderà in particolare alla condizione iniziale $f(x, t) = \delta(x - y)$

Appendice L

Integrale di Stratonovich (Sez. 8.2.2)

È importante ricordare che sono state tentate delle definizioni alternative a quella di $It\bar{o}$. Una in particolare, dovuta a R.L. Stratonovich, richiede qualche precisazione: si tratta infatti di una definizione basata essenzialmente su una procedura alla Riemann del tipo (8.8), ma con i valori dell'integrando $Y(t)$ calcolati nel *punto di mezzo* dell'intervallo $[t_j, t_{j+1}]$, e non nel suo estremo sinistro t_j

$$\int_a^b Y(t) \circ dW(t) = \lim_{n, \delta \rightarrow 0} \text{-mq} \sum_{j=0}^{n-1} Y\left(\frac{t_j + t_{j+1}}{2}\right) [W(t_{j+1}) - W(t_j)]$$

Si noti l'introduzione del simbolo “ \circ ” nella notazione per distinguere questo integrale dal corrispondente integrale di $It\bar{o}$. L'interesse principale di questa definizione sta nel fatto che le conseguenti regole del calcolo stocastico restano identiche a quelle del calcolo ordinario, e questo è probabilmente il motivo per il quale l'*integrale di Stratonovich* è rimasto a lungo molto popolare fra i fisici. Purtroppo però esso non gode delle stesse proprietà matematiche dell'integrale di $It\bar{o}$, anzi la sua convergenza e la coerenza della sua definizione non sono esenti da problemi per cui la sua natura resta abbastanza incerta. Non vi è peraltro nessuna regola del tutto generale che metta in relazione le due definizioni, tranne il seguente risultato¹

Proposizione L.1. (*E. Wong, M. Zakai - 1969*) *Se $X(t)$ è soluzione dell'EDS di $It\bar{o}$ (8.26), e se $g(x, t)$ è una funzione continua e derivabile, con ipotesi di regolarità che qui trascureremo, gli integrali di $It\bar{o}$ e di Stratonovich soddisfano \mathbf{P} -qo la relazione*

$$\int_a^b g(X(t), t) \circ dW(t) = \int_a^b g(X(t), t) dW(t) + \frac{1}{2} \int_a^b g_x(X(t), t) b(X(t), t) dt$$

¹Per maggiori dettagli vedi **C.W. Gardiner**, HANDBOOK OF STOCHASTIC METHODS, Springer (Berlin, 1997); **T. Neckel, F. Rupp**, RANDOM DIFFERENTIAL EQUATIONS IN SCIENTIFIC COMPUTING, Versita (London, 2013)

nel senso che il membro sinistro esiste se e solo se esiste quello destro, e in questo caso i due coincidono. Ovviamente qui $b(x, t)$ è il coefficiente di diffusione dell'EDS di Itô (8.26)

Dimostrazione: Omessa (vedi **T. Neckel, F. Rupp**, **RANDOM DIFFERENTIAL EQUATIONS IN SCIENTIFIC COMPUTING**, Versita (London, 2013), p. 159) \square

Si noti che in particolare gli integrali di Itô e di Stratonovich coincidono se $g(x, t) = g(t)$ non dipende da x . Inoltre, se $X(t)$ è soluzione dell'EDS di Itô (8.26) (interpretata nel senso integrale (8.27)), prendendo $h(x, t) = b(x, t)$ si vede facilmente dalla Proposizione L.1 che $X(t)$ risulta anche (sempre nel senso integrale) soluzione della seguente *EDS di Stratonovich*

$$\begin{aligned} dX(t) &= \tilde{a}(X(t), t) dt + b(X(t), t) \circ dW(t) \\ \tilde{a}(x, t) &= a(x, t) - \frac{1}{2} b(x, t) b_x(x, t) \end{aligned}$$

Naturalmente tutti questi risultati possono essere usati anche per dare – almeno per questi casi particolari – una definizione coerente di integrale (e di *EDS*) di Stratonovich basata sulle analoghe definizioni di Itô che, a loro volta, sono ben poste. Noi però eviteremo di addentrarci lungo questa strada e baseremo sempre le nostre considerazioni sull'integrale di Itô – che sarà considerato come definito sulla base della procedura esposta nella Sezione 8.2.2 – con le conseguenti modifiche nelle regole del calcolo differenziale e integrale

Indice analitico

- additività, 18
- ago di Buffon, 99
- algebra, 14
- assenza di memoria, 99
- assiomi di Kolmogorov):, 21
- atomo, 15

- boreliano, 16

- cadlag, 35
- campionamento
 - con rimessa, 11
 - senza rimessa, 12
- caratteristica logaritmica, 296
- cilindro, 17
- coefficiente di correlazione, 86
- coefficiente di diffusione, 175
- combinazione, 12
- condizioni
 - di Lipschitz, 249
- consistenza, 55
- convergenza
 - P -qo, 109
 - debole, 110
 - degenere, 123
 - in L^p , 109
 - in mq , 109
 - in mq , 143
 - in distribuzione, 109
 - in generale, 110
 - in probabilità, 109
 - puntuale, 65
- convoluzione, 73, 74, 106
 - composizione di leggi, 121
- covarianza, 86
- cumulante, 296

- cumulanti, 115

- decomposizione, 15
- decomposizione di leggi, 122
- densità di probabilità, 39, 62
 - condizionata, 92, 95
 - congiunta, 69
 - del χ^2 , 107
 - di Cauchy, 42
 - di Erlang, 123
 - di Laplace, 42
 - di Student, 108
 - esponenziale, 41
 - Gaussiana (normale), 41
 - log-normale, 105, 178
 - marginale, 70
 - multivariata, 48
 - uniforme, 40
- deviazione standard, 86
- diagrammi di Venn, 13
- differenziale stocastico, 182, 247
- disposizione, 11
- distribuzione, 27
 - binomiale, 27, 38
 - congiunta, 67
 - degenere, 37
 - del χ^2 , 107
 - di Bernoulli, 28, 38
 - di Boltzmann, 270
 - di Cauchy, 42
 - di Erlang, 123, 156
 - di Laplace, 42
 - di Poisson, 28, 38
 - di probabilità, 60
 - finito-dimensionale, 68
 - di Student, 108

- discreta, 37
- esponenziale, 41
- Gaussiana (normale), 41
 - bivariata, 49
 - multivariata, 48
 - standard, 41
- log-normale, 105
- marginale, 68
- multinomiale, 33
- singolare, 43
- uniforme, 40
- disuguaglianza
 - di Chebyshev, 91
 - di Hölder, 286
 - di Jensen, 285
 - di Lyapunov, 285
 - di Minkowski, 287
 - di Schwarz, 286
- elemento aleatorio, 66
- equazione
 - di Fokker-Planck, 177, 221, 225
 - di Langevin, 187
 - di Smoluchowski, 268
 - differenziale stocastica, 248
 - soluzioni deboli, 249
 - soluzioni forti, 249
 - in avanti, 214
- equazioni
 - all'indietro, 218
 - di Chapman-Kolmogorov, 193
 - di Kolmogorov, 219
 - in avanti, 218
- equiprobabilità, 9, 19
- eventi indipendenti, 25
 - condizionatamente, 25
- evento, 13
- filtrazione
 - naturale, 235
- formula
 - della probabilità totale, 22
 - di Bayes, 24
 - di inversione, 116
 - di moltiplicazione, 23
- frequenza, 299
- funzione
 - a variazione limitata, 232
 - Beta, 212, 273
 - caratteristica, 112
 - $v-a$ $\mathfrak{N}(b, a^2)$, 114
 - $v-a$ $\mathfrak{B}(1; p)$, 113
 - $v-a$ $\mathfrak{B}(n; p)$, 113
 - $v-a$ δ_b , 113
 - $v-a$ $\mathfrak{C}(a)$, 114
 - $v-a$ $\mathfrak{E}(a)$, 114
 - $v-a$ $\mathfrak{E}_n(a)$, 123
 - $v-a$ $\mathfrak{L}(a)$, 114
 - $v-a$ $\mathfrak{P}(\alpha)$, 113
 - $v-a$ $\mathfrak{U}(a, b)$, 113
 - di $v-a$, 63
 - di Borel, 59
 - di Heaviside, 38
 - di partizione, 270
 - errore, 41
 - Gamma, 107
 - generatrice, 224
 - funzione di distribuzione, 35, 61
 - assolutamente continua, 39
 - congiunta, 68
 - generalizzata, 36
 - marginale, 49, 68
 - multivariata, 47
 - multivariata generalizzata, 47
 - funzioni
 - definite non negative, 117
- incrementi, 139
 - indipendenti, 195
 - processo degli, 139
 - stazionari, 146
- indicatore, 59
- indice di Hurst, 315
- integrale
 - di Itô, 235
 - di Lebesgue
 - esteso a un insieme, 78

- di Lebesgue–Stieltjes, 78
- di Stratonovich, 321
- stocastico, 232
- integrale di Lebesgue, 78
- integrale stocastico, 145
- integrazione per parti, 246
- legge, 27
 - Gaussiana (normale)
 - standard, 41
- legge dei grandi numeri, 101
 - debole, 124
 - di Bernoulli, 300
 - forte, 125
- legge di una v -a, 60
- leggi infinitamente divisibili, 197
- limiti di successioni
 - di v -a, 63
- marginalizzazione, 49
- master equation, 160, 218, 220
- matrice
 - definita non negativa, 48
 - delle correlazioni, 86
 - delle covarianze, 86, 91
 - di diffusione, 214
- media pesata, 76
- metrica, 293
- miscela, 43
- misura, 20
 - assolutamente continua, 38
 - di Lebesgue, 21
 - di Lebesgue–Stieltjes, 36
 - di probabilità, 20
 - di Wiener, 57
 - finita, 20
 - σ -additiva, 20
 - σ -finita, 20
- misurabilità, 59
- moda, 85
- modello di Bernoulli, 32
- modificazione, 140
- momenti, problema dei, 115, 295
- momento, 78
 - assoluto, 78
- Monte Carlo, metodo di, 101, 126
- moto Browniano, 171, 183
 - frazionario, 315
 - geometrico, 178
- norma, 293
- ortogonalità, 294
- P -qo, 21
- parte positiva (negativa), 77
- permutazione, 12
- probabilità, 9, 18, 20
 - a posteriori, 24
 - a priori, 24
 - condizionata, 22
 - congiunta, 22
 - definizione classica, 9, 19
 - spazio di, 21
 - spazio finito di, 18
- processi
 - equivalenti, 140
 - equivalenti in s.l., 140
 - indistinguibili, 140
- processo
 - ad incrementi indipendenti, 158, 195
 - canonico, 141
 - centrato, 142
 - continuo, 143, 204
 - degli impulsi, 180
 - degli incrementi, 158, 172, 182
 - derivabile, 144
 - di Cauchy, 207
 - di diffusione, 221
 - di Lévy, 197, 202
 - di Markov, 189, 191
 - di nascita, 160
 - di Ornstein-Uhlenbeck, 184, 209, 256
 - di Poisson, 158
 - compensato, 164
 - composto, 165
 - di Poisson semplice, 205
 - di punto, 155

- di salto, 220
- di Wiener, 171, 172, 206
 - geometrico, 178
 - standard, 172
- ergodico, 148, 201
- Gaussiano, 141, 177, 212
- non-anticipativo, 235
- omogeneo, 199
- pseudo-Markoviano, 311
- separabile, 141
- stazionario, 146, 198
- stazionario in senso lato, 147
- stocastico, 67, 139
- prodotto scalare, 294
- proprietà di Markov, 189
- proprietà riproduttiva, 107
 - leggi δ_b , 122
 - leggi $\mathfrak{N}(b, a^2)$, 121
 - leggi $\mathfrak{P}(\alpha)$, 122
- random walks, 171
- regola di moltiplicazione a catena, 192
- regressione, 102
- rinnovi, 156
- ripartizione, 12
- rumore bianco, 179
- rumore granulare, 169, 305
- σ -algebra, 14
 - di Borel, 16
 - generata da una v -a, 64
 - generata da una famiglia di sottoinsiemi, 15
- σ -algebre indipendenti, 25
- shot noise, 169
- spazio
 - dei campioni, 10, 21
 - di Banach, 293
 - di Hilbert, 294
 - probabilizzabile, 15
- spettro di covarianza, 152
- spettro di potenza, 150, 152
- statistica, 19, 299
- stima in mq , 101
- subadditività, 21
- successione aleatoria, 67
- successione di v -a, 67
- teorema
 - di Bayes, 24
 - di Bochner, 117
 - di continuità, 117
 - di esistenza e unicità, 249
 - di Kolmogorov su \mathbf{R}^∞ , 56
 - di Kolmogorov su \mathbf{R}^T , 57
 - di Lebesgue, 65
 - di Lebesgue–Nikodym, 44
 - di P. Lévy, 117
 - di Poisson, 133
 - per v.a. binomiali, 131
 - per vett.a. multinomiali, 132
 - di Radon–Nikodym, 39
 - di unicità, 116
 - di Wiener–Khinchin, 150
- limite
 - centrale di Lyapunov, 129
 - centrale per v.a. i.i.d., 128
 - integrale, 301
 - locale, 300
- teoremi limite, 197
- transizione
 - densità di, 174
 - probabilità di, 158
- trasformata di Fourier, 112
- valore d'attesa, 76, 77
 - v -a $\mathfrak{B}(1; p)$, 83
 - v -a $\mathfrak{B}(n; p)$, 83, 85
 - v -a δ_b , 83
 - v -a χ_n^2 , 108
 - v -a $\mathfrak{C}(a, b)$, 84
 - v -a $\mathfrak{C}(a)$, 83
 - v -a $\mathfrak{E}_n(a)$, 123
 - v -a $\mathfrak{L}(a)$, 83
 - v -a \mathfrak{T}_n , 108
 - v -a $\ln \mathfrak{N}(b, a^2)$, 105
 - v -a $\mathfrak{N}(b, a^2)$, 83
 - v -a $\mathfrak{P}(\alpha)$, 83

- v -a $\mathcal{U}(a, b)$, 83
- condizionato
 - rispetto a un $vett$ - a , 98
 - rispetto a una v - a , 96
- variabile aleatoria, 59
 - assolutamente continua, 62
 - assolutamente integrabile, 77
 - canonica, 63
 - complessa, 66
 - continua, 62
 - degenere, 60, 62
 - discreta, 62
 - estesa, 65
 - integrabile, 77
 - semplice, 60
 - standardizzata, 127
- variabili aleatorie
 - identicamente distribuite, 61, 71
 - identiche, 61
 - indipendenti, 71
 - non correlate, 86
 - uguali P -qo, 61
- varianza, 86
 - v -a $\mathfrak{B}(1; p)$, 89
 - v -a $\mathfrak{B}(n; p)$, 89
 - v -a δ_b , 89
 - v -a χ_n^2 , 108
 - v -a $\mathfrak{C}(a, b)$, 90
 - v -a $\mathfrak{C}(a)$, 89
 - v -a $\mathfrak{C}_n(a)$, 123
 - v -a $\mathfrak{L}(a)$, 90
 - v -a \mathfrak{T}_n , 108
 - v -a $\ln \mathfrak{N}(b, a^2)$, 105
 - v -a $\mathfrak{N}(b, a^2)$, 90
 - v -a $\mathfrak{P}(\alpha)$, 89
 - v -a $\mathcal{U}(a, b)$, 89
- vettore
 - di deriva, 214
 - di trascinamento, 214
- vettore aleatorio, 66
 - canonico, 63
 - discreto, 70
 - gaussiano (normale), 71, 119
 - bivariato, 90
 - vettore delle medie, 91
 - vita media, 98
 - white noise, 179