



Università degli Studi di Bari
Facoltà di Scienze Matematiche, Fisiche e Naturali

Nicola Cufaro Petroni

STATISTICA
CON ELEMENTI DI PROBABILITÀ

anno accademico 2004/05

Il testo di queste lezioni è liberamente disponibile all'indirizzo
<http://www.ba.infn.it/~cufaro/didactic/StatisticaMatematica.pdf>
e può anche essere richiesto scrivendo a
cufaro@ba.infn.it

Copyright © 2004 Nicola Cufaro Petroni
Università degli Studi di Bari
Facoltà di Scienze Matematiche, Fisiche e Naturali
via E.Orabona 4, 70125 Bari

Prefazione

La struttura di queste lezioni riflette una duplicità di significato della parola *Statistica*. Per essere più chiari cominceremo con un esempio: supponiamo di voler conoscere l'orientamento politico generale dei cittadini di un determinato paese. È ben noto che in questo caso si organizzano delle *elezioni* che consistono nel raccogliere il voto di *tutti* gli elettori. Una volta esaurite le operazioni di voto si passerà allo spoglio delle schede e alla registrazione dei risultati. Tali risultati si presentano in generale come una grande quantità di dati numerici che possono essere esaminati, combinati e rappresentati in diverse maniere in modo da estrarre l'informazione rilevante ai fini elettorali. Le elezioni generali sono però tipicamente delle operazioni complesse e costose, e per questo motivo spesso si preferisce affidarsi a dei *sondaggi* per avere delle informazioni, almeno approssimative e provvisorie, sulla volontà dei cittadini. Questi sondaggi consistono nella registrazione delle opinioni di un *piccolo numero* di soggetti, a partire dal quale si ricavano delle indicazioni sulla volontà generale della popolazione intera. Ovviamente i sondaggi non possono essere sostitutivi delle elezioni, e non solo perchè bisogna dare a tutti i cittadini la possibilità di esprimere la propria opinione, ma anche per una profonda differenza fra i dati delle due operazioni. Il risultato del sondaggio, infatti, è *aleatorio*: siccome il campione di cittadini intervistato è scelto casualmente, una ripetizione del sondaggio – per quanto eseguita con i medesimi criteri – porterebbe inevitabilmente a dei risultati numerici diversi, anche se di poco. Viceversa nel caso delle elezioni l'indagine esaurisce l'intera popolazione degli elettori: una eventuale ripetizione del voto – supponendo per semplicità che non vi siano ripensamenti o errori – non modificherebbe il risultato. Noi diremo che l'esame dei risultati elettorali complessivi è compito della *Statistica descrittiva*, mentre le tecniche per ricavare informazioni su tutta la popolazione a partire dai risultati relativi a un piccolo campione sono parte della *Statistica inferenziale*. Naturalmente, come è noto, l'uso dei sondaggi comporta dei rischi dovuti alla aleatorietà dei loro esiti, per cui diventa essenziale per la Statistica inferenziale essere in grado di misurare l'affidabilità dei risultati: in questo giocheranno un ruolo essenziale i concetti e le tecniche del *Calcolo delle probabilità*.

Si noti che nel caso dell'esempio elettorale citato la possibilità di registrare il voto di tutti i cittadini esiste comunque: pertanto, in linea di principio, è sempre possibile confrontare i risultati dei sondaggi con quelli delle elezioni generali e verificarne

quindi l'attendibilità. Questa possibilità, però, non sussiste sempre: in molti casi infatti un'indagine che esaurisca l'intera popolazione semplicemente non è possibile, e ci si deve accontentare invece di esaminare le misure eseguite su un campione tentando di dedurre le caratteristiche generali del fenomeno studiato. Ad esempio in linea di principio la misura della massa di una particella elementare può essere eseguita infinite volte, e data la delicatezza della misura i risultati variano sempre in maniera aleatoria. In pratica il numero delle nostre misure sarà sempre finito, e d'altra parte, per quanto grande sia questo numero, non potremo mai dire di aver esaurito l'intera popolazione teoricamente disponibile. Allo stesso modo la determinazione della lunghezza media degli insetti di una determinata specie non potrà che essere effettuata su un campione casuale, visto che l'intera popolazione di insetti resta comunque praticamente inaccessibile. In queste occasioni, ovviamente, il raffinamento delle tecniche probabilistiche diventa essenziale.

Nasce da queste osservazioni la struttura – ormai classica – di queste lezioni divise in tre parti. La prima esamina gli strumenti principali della Statistica descrittiva: tabelle, grafici, indici di centralità e dispersione, correlazioni. La seconda introduce i concetti più rilevanti del Calcolo delle probabilità, e funge da introduzione alla terza che infine studia le tecniche più note della Statistica inferenziale: stime, intervalli di fiducia, test di ipotesi. Per ovvie ragioni di spazio e tempo gli argomenti non saranno trattati in maniera esaustiva. In particolare la parte di Calcolo delle probabilità è ridotta a quanto è strettamente necessario per la comprensione della parte di Statistica inferenziale. Lo scopo del corso rimane quello di mettere gli studenti in grado di usare gli strumenti più semplici e più noti della statistica, ma anche di stabilire alcuni pilastri concettuali che consentano loro, qualora se ne presentasse l'occasione, di estendere le loro capacità in maniera autonoma. Alcuni completamente sono comunque già previsti per le prossime edizioni di questo testo.

Devo infine ringraziare il collega Paolo Baldi dell'Università di Roma "Tor Vergata" per aver messo gentilmente a mia disposizione il testo non pubblicato di un suo corso di Statistica dal quale ho liberamente attinto idee, osservazioni ed esempi.

Bari, Novembre 2004

NICOLA CUFARO PETRONI

Indice

Prefazione	I
I Statistica descrittiva	1
1 Statistica univariata	3
1.1 Dati e frequenze	3
1.2 Tabelle e grafici	6
1.3 Moda, media e varianza	9
1.4 Mediana e quantili	15
1.5 Momenti, asimmetria e curtosi	18
1.6 Medie generalizzate	19
2 Statistica multivariata	23
2.1 Dati multidimensionali	23
2.2 Covarianza, correlazione e regressione	25
2.3 Componenti principali	28
II Probabilità	37
3 Spazi di probabilità	39
3.1 Spazio dei campioni	39
3.2 Eventi	41
3.3 Probabilità	43
4 Condizionamento e indipendenza	45
4.1 Probabilità condizionata	45
4.2 Indipendenza	47
5 Variabili aleatorie	51
5.1 Variabili aleatorie	51
5.2 Variabili aleatorie discrete	54
5.3 Variabili aleatorie continue	59

5.4	Quantili	65
5.5	Vettori aleatori	68
6	Attesa e varianza	71
6.1	Attesa e varianza	71
6.2	Esempi di attese e varianze	75
7	Teoremi limite	79
7.1	Legge dei Grandi Numeri	79
7.2	Teorema Limite Centrale	82
III	Statistica inferenziale	87
8	Stima di parametri	89
8.1	Stima puntuale	89
8.2	Stima per intervalli	98
8.2.1	Intervallo di fiducia per l'attesa μ	99
8.2.2	Intervallo di fiducia per la varianza σ^2	101
8.3	Stima di Massima Verosimiglianza	103
9	Test di Ipotesi	109
9.1	Ipotesi ed errori	109
9.2	Test sulla media	116
9.2.1	Test di Gauss	117
9.2.2	Test di Student	120
9.3	Test per il confronto delle medie	121
9.3.1	Campioni accoppiati	122
9.3.2	Campioni indipendenti	126
9.4	Test di Fisher sulla varianza	128
9.5	Test del χ^2 di adattamento	129
9.6	Test del χ^2 di indipendenza	133
IV	Appendici	137
A	Esercizi	139
A.1	Esercizi di Statistica Descrittiva	141
A.2	Esercizi di Probabilità	145
A.3	Esercizi di Statistica Inferenziale	148
B	Schemi	157
B.1	Formulario di Statistica Inferenziale	159

C	Notazioni	163
C.1	Notazioni vettoriali	165
D	Tavole Numeriche	167
D.1	Legge Normale standard $\mathcal{N}(0, 1)$	169
D.2	Legge di Student $t(n)$	170
D.3	Legge del Chi Quadro $\chi^2(n)$	171
D.4	Legge di Fisher $F(n, m)$	172
D.5	Valori di $e^{-\lambda}$	174
	Indice analitico	175

Parte I
Statistica descrittiva

Capitolo 1

Statistica univariata

1.1 Dati e frequenze

Distingueremo innanzitutto i dati in due categorie: quantitativi e qualitativi. I *dati quantitativi* sono risultati di misure che forniscono valori numerici: ad esempio posizioni, velocità, masse di particelle; peso, altezza, età di individui di una specie di animali; reddito dei cittadini di un paese. I *dati qualitativi* viceversa non sono in genere rappresentati da numeri: ad esempio i colori delle palline estratte da un'urna; il gruppo sanguigno A , B , AB e 0 di un insieme di persone; il partito politico votato dagli elettori di un paese. Si noti però che la differenza principale fra i due tipi di dati non consiste nel fatto formale di essere rappresentati o meno da numeri: in fondo potremmo convenzionalmente rappresentare i colori, i gruppi sanguigni e i partiti degli esempi precedenti con dei numeri. Quel che è profondamente diverso è il significato di questi numeri: ad esempio il colore rosso delle palline dell'urna può essere rappresentato indifferentemente con 1, o 2 o qualsiasi altro numero senza che nulla cambi nella discussione del problema. Invece i numeri che rappresentano i redditi dei cittadini di un paese non possono essere assegnati arbitrariamente senza perdere tutta l'informazione utile. Questa differenza è anche alla base del fatto che taluni indicatori statistici hanno un senso solo nel caso di dati quantitativi e non in quelli di dati qualitativi. Ad esempio è perfettamente sensato chiedersi quale è il reddito medio dei cittadini di un paese, mentre non avrebbe alcun significato il concetto di colore medio, o di partito medio: e questo anche se i dati qualitativi fossero rappresentati da numeri.

Nel linguaggio della Statistica l'insieme dei soggetti presi in considerazione nella discussione di un determinato problema (animali di una specie, palline in un'urna, cittadini di un paese) costituisce una *popolazione*, mentre le caratteristiche X, Y, \dots che si osservano (colore, gruppo sanguigno, reddito, peso \dots) prendono il nome di *caratteri*. I caratteri numerici – quelli che forniscono dati quantitativi – sono poi distinti in due categorie secondo che assumano valori discreti (ad esempio il numero di figli delle famiglie di un dato paese) o continui (il peso o l'altezza degli individui

di una popolazione). I possibili valori assunti dai caratteri numerici discreti (numero dei figli di una famiglia), o dai caratteri qualitativi (colori delle palline in un'urna) si chiamano anche *modalità*. Così ad esempio: il peso dei cittadini di un paese è un carattere numerico continuo; il gruppo sanguigno degli individui di un gruppo è un carattere qualitativo con 4 modalità (A , B , AB e 0). Come vedremo nel Capitolo 2, infine, potremo avere anche *dati multidimensionali*, nel senso che su ogni individuo si possono misurare due o più caratteri. Ad esempio se si misurano l'età e , il peso p e il reddito r dei cittadini di un dato paese, ad ogni individuo sarà associata una terna di numeri (e, p, r) .

Nell'ambito della Statistica descrittiva supporremo sempre di avere a disposizione i dati relativi a tutta la popolazione di nostro interesse che, quindi, dovrà contenere un numero finito n di individui. I nostri insiemi di dati, chiamati *campioni*, saranno pertanto del tipo x_1, \dots, x_n , ed esauriranno tutta la popolazione considerata. Si noti, però, a questo proposito che in genere il procedimento avviene in senso inverso: si parte dai dati e poi si stabilisce quale è la popolazione di riferimento. Tipicamente nella realtà il punto di partenza è il campione x_1, \dots, x_n il cui significato può cambiare secondo il punto di vista adottato. In un certo senso è lo statistico che, in base alle proprie necessità, stabilisce quale è la popolazione di riferimento: se decide che la popolazione di interesse è rappresentata solo dagli n dati a sua disposizione, allora egli si colloca nell'ambito della Statistica descrittiva; se invece considera gli n dati come un campione estratto da una popolazione più vasta sulla quale vuole ricavare delle informazioni, allora si colloca nell'ambito della Statistica inferenziale che studieremo nella Parte III. Supponiamo ad esempio di avere i risultati x_1, \dots, x_n del test d'ingresso di n studenti ad un corso di laurea universitario. Se il nostro scopo è solo quello di esaminare il livello di preparazione degli studenti che accedono al corso di laurea in quell'anno accademico, è evidente che la nostra popolazione sarà ristretta agli n individui che hanno sostenuto il test. Se invece dai risultati del test volessimo dedurre delle conclusioni, ad esempio, sulla tutta la popolazione studentesca che accede all'università in un determinato anno accademico, è altrettanto evidente che x_1, \dots, x_n dovrà ora essere considerato come un campione estratto da una popolazione più vasta. Il medesimo insieme di dati, cioè, può essere considerato da due punti di vista diversi: nel primo caso si tratta di un problema di Statistica descrittiva, nel secondo di un problema di Statistica inferenziale.

Supponiamo allora di osservare un carattere (numerico o qualitativo) X con un numero finito M di modalità su una popolazione di n individui: convenzionalmente indicheremo le M possibili modalità con i numeri interi $k = 1, \dots, M$. I dati saranno allora un campione x_1, \dots, x_n di n numeri interi con valori $k = 1, \dots, M$. Chiameremo *frequenza assoluta* della k -ma modalità il numero N_k delle x_j uguali a k . In termini formali, detto $\{j : x_j = k\}$ l'insieme degli individui della nostra popolazione che assumono il valore k , N_k è il numero di elementi di $\{j : x_j = k\}$, ovvero indicando con $\#$ la cardinalità di un dato insieme,

$$N_k = \#\{j : x_j = k\}, \quad k = 1, \dots, M. \quad (1.1)$$

Chiameremo invece *frequenza relativa* della k -ma modalità il numero

$$p_k = \frac{N_k}{n}, \quad k = 1, \dots, M \quad (1.2)$$

che rappresenta la frazione delle x_j che assume il valore k . Naturalmente sono verificate le due relazioni di *normalizzazione*

$$N_1 + \dots + N_M = n, \quad p_1 + \dots + p_M = 1. \quad (1.3)$$

Sia nel caso di frequenze assolute che in quello di frequenze relative si introduce poi il concetto di *frequenze cumulate*:

$$F_k = \sum_{i=1}^k N_i, \quad f_k = \sum_{i=1}^k p_i, \quad k = 1, \dots, M \quad (1.4)$$

Come è evidente la frequenza assoluta (rispettivamente: relativa) cumulata F_k (f_k) rappresenta il numero (la frazione) delle x_j con valore minore o uguale a k . Per le frequenze cumulate le relazioni (1.3) divengono

$$F_M = n, \quad f_M = 1.$$

Se invece il carattere X è numerico e continuo le sue modalità sono infinite e non numerabili per cui bisognerà procedere in modo diverso. Siccome anche in questo caso i valori osservati x_1, \dots, x_n che costituiscono la nostra popolazione sono in numero n finito, essi cadranno in qualche intervallo finito del tipo $[a, b]$ e noi li ripartiremo in opportune *classi* nel modo seguente: suddividiamo $[a, b]$ in M sottointervalli (non necessariamente tutti della stessa ampiezza) J_k con $k = 1, \dots, M$ e poniamo, analogamente al caso discreto,

$$N_k = \#\{j : x_j \in J_k\}, \quad p_k = \frac{N_k}{n}, \quad k = 1, \dots, M.$$

Anche ora quindi potremo parlare di frequenze assolute e relative dei ritrovamenti dei nostri dati negli intervalli J_k . Ovviamente le relazioni (1.3) continueranno a valere anche in questo caso. Si noti però che adesso i valori delle frequenze dipendono dall'ampiezza $|J_k|$ delle classi che è scelta arbitrariamente. Come vedremo in alcuni esempi successivi la determinazione delle ampiezze $|J_k|$ delle classi può rivelarsi cruciale per mettere in evidenza alcune caratteristiche dei dati. Una certa importanza riveste anche il concetto di *valore centrale* di una classe: esso coincide con la semisomma degli estremi dell'intervallo J_k . Per scopi di ulteriore analisi statistica in generale tutti i valori di una data classe vengono identificati proprio con il suo valore centrale. Infine anche nel caso di caratteri continui è possibile parlare di frequenze cumulate adottando le definizioni (1.4) per le frequenze di ritrovamenti nelle classi. Ovviamente in questo caso F_k (f_k) indica il numero (la frazione) di dati x_j che cadono all'interno dell'unione dei primi k sottointervalli, ovvero che sono minori o uguali dell'estremo destro di J_k .

3	0	3	1	1	1	2	4	1	3	2	1	0	2	1	3	3	0	2	1
3	4	3	1	3	4	1	5	0	2	0	4	1	4	2	2	2	1	2	3
2	3	2	2	3	3	2	1	2	1										

Tabella 1.1: Campione di $n = 50$ misure di un carattere con le 6 modalità $k = 0, 1, 2, 3, 4, 5$.

k	0	1	2	3	4	5
N_k	5	13	14	12	5	1
F_k	5	18	32	44	49	50
p_k	0.10	0.26	0.28	0.24	0.10	0.02
f_k	0.10	0.36	0.64	0.88	0.98	1.00

Tabella 1.2: Frequenze e frequenze cumulate, assolute e relative, per i dati riportati in Tabella 1.1.

1.2 Tabelle e grafici

L'informazione contenuta nelle frequenze assolute e relative può essere messa meglio in evidenza organizzando i dati in tabelle o anche rappresentandoli in grafici. Le tecniche di organizzazione e visualizzazione dei dati sono numerose, e la scelta di quelle più opportune dipende dal particolare problema studiato. Noi qui ne daremo solo qualche esempio senza nessuna pretesa di completezza.

Le *tabelle di frequenza* non sono altro che opportune tabelle nelle quali sono riportati in maniera organizzata i valori numerici delle varie frequenze. Per la rappresentazione grafica dei caratteri discreti lo strumento più usato è il *diagramma a barre* che consiste semplicemente nel riportare in corrispondenza di ogni singola modalità delle barre di altezza uguale ai valori delle frequenze. Su questi diagrammi possono essere rappresentate sia le frequenze assolute che quelle relative: siccome a causa di (1.2) N_k e p_k sono tutti numeri proporzionali fra loro, i diagrammi a barre dei due casi sono identici, l'unica differenza essendo la scala dei valori dell'asse verticale. Per le frequenze dei caratteri continui invece si costruiscono degli *istogrammi*. Il principio è simile a quello dei diagrammi a barre con una importante differenza: sulla classe k -ma (sottointervallo J_k dell'intervallo $[a, b]$ che contiene tutti i dati) si costruisce un rettangolo la cui *area* è uguale al valore della frequenza k -ma. Siccome le ampiezze $|J_k|$ delle varie classi (basi dei rettangoli) possono essere diverse, in generale le altezze dei rettangoli non saranno più proporzionali alle frequenze: a parità di dati contenuti, classi molto ampie tenderanno ad avere rettangoli più bassi, e viceversa. Solo nel caso in cui le ampiezze $|J_k|$ fossero scelte tutte uguali le altezze dei rettangoli sarebbero nuovamente proporzionali alle frequenze (assolute o relative) delle classi. Noteremo infine che anche le frequenze cumulate sono ovviamente suscettibili di rappresentazioni grafiche che però noi, per brevità, trascureremo limitandoci a riportare i loro valori nelle tabelle di frequenza.

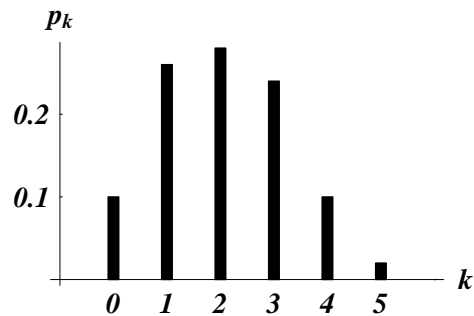


Figura 1.1: Diagramma a barre delle frequenze relative p_k dei dati della Tabella 1.1.

0.30	1.03	1.08	1.22	1.46	1.62	2.01	2.17	2.27	2.31
2.33	2.41	2.49	2.49	2.57	2.58	2.59	2.63	2.75	2.75
2.84	2.93	2.95	3.08	3.09	3.23	3.27	3.27	3.28	3.37
3.39	3.42	3.47	3.49	3.56	3.60	3.78	3.78	3.79	3.87
3.91	3.91	3.95	3.95	3.96	4.02	4.11	4.12	4.12	4.22
4.31	4.35	4.58	4.69	4.76	4.89	5.12	5.18	5.20	5.34
5.34	5.37	5.40	5.46	5.54	5.62	5.64	5.64	5.68	5.71
5.73	5.94	6.10	6.19	6.24	6.28	6.31	6.33	6.35	6.40
6.44	6.44	6.55	6.56	6.63	6.68	6.73	6.75	6.89	6.99
7.01	7.08	7.11	7.15	7.26	7.44	7.47	7.93	8.21	8.44

Tabella 1.3: Campione di $n = 100$ misure di un carattere continuo X . Per comodità i dati sono stati riportati in ordine crescente. La coincidenza di alcuni dei valori – particolarmente improbabile nel caso di caratteri continui – è dovuta agli arrotondamenti effettuati.

Esempio 1.1. Supponiamo di aver raccolto $n = 50$ misure di un carattere con $M = 6$ modalità che qui per comodità rappresenteremo senz'altro con i numeri $k = 0, 1, 2, 3, 4, 5$. Come esempio concreto possiamo pensare di aver esaminato 50 famiglie con 5 figli e di aver registrato per ciascuna di esse il numero dei figli maschi che ovviamente è un numero intero da 0 a 5; alternativamente potremmo pensare di aver lanciato 50 volte 5 monete e di aver registrato in ogni lancio il numero delle teste. I dati di partenza del nostro esempio sono mostrati nella Tabella 1.1. È facile a questo punto calcolare le frequenze assolute e relative da (1.1) e (1.2): i risultati sono riportati nella Tabella 1.2. Le frequenze possono poi essere rappresentate in un diagramma a barre come quello di Figura 1.1. Per evitare ripetizioni abbiamo scelto di riportare solo il diagramma a barre delle frequenze relative: quello delle frequenze assolute sarebbe identico, tranne che per la scala dei valori dell'asse verticale.

Esempio 1.2. Supponiamo di avere le $n = 100$ misure di un carattere continuo X riportate nella Tabella 1.3. Ad esempio potrebbero essere – in una opportuna unità di misura – le misure di una dimensione fisica di un gruppo di insetti; ovvero le misure della mas-

J_k	N_k	F_k	p_k	f_k
[0.0, 2.0]	6	6	0.06	0.06
[2.0, 4.0]	39	45	0.39	0.45
[4.0, 6.0]	27	72	0.27	0.72
[6.0, 8.0]	26	98	0.26	0.98
[8.0, 10.0]	2	100	0.02	1.00

Tabella 1.4: Frequenze dei dati di Tabella 1.3 per 5 classi di ampiezza 2.0 sull'intervallo $[0, 10]$.

sa delle particelle elementari presenti in un determinato esperimento. Ovviamente nella realtà i valori non si ottengono nell'ordine crescente nel quale li abbiamo riportati; noi però abbiamo riordinato il campione perché questo facilita il calcolo delle frequenze senza modificarne il valore. La tabella delle frequenze dipende ora dalle classi scelte. Si vede subito che i dati cadono tutti fra 0.30 e 8.44, ma per rendere più simmetriche le classi possiamo, ad esempio, considerare un intervallo un po' più ampio del tipo $[0, 10]$. Per semplicità sceglieremo per ora classi tutte della stessa ampiezza, e cominceremo con il dividere $[0, 10]$ in sottointervalli di ampiezza 2.0. In questo caso le frequenze sono quelle della Tabella 1.4. Se invece avessimo scelto come ampiezza delle classi 0.5 o 0.1 avremmo ottenuto frequenze piuttosto diverse: per brevità non ne riportiamo le tabelle limitandoci solo alla loro successiva rappresentazione grafica. Si può passare a questo punto a costruire gli istogrammi corrispondenti a ciascuna scelta delle classi, ricordando che su ogni classe dovrà essere disegnato un rettangolo di *area* uguale alla rispettiva frequenza. Gli istogrammi ottenuti con le tre scelte delle classi (ampiezze rispettivamente 2.0, 0.5 e 0.1) sono riportati nella Figura 1.2. Si noterà che l'aspetto dei tre istogrammi è piuttosto diverso: quello con le classi più ampie (ampiezza 2.0, in alto a sinistra) fornisce una rappresentazione piuttosto grossolana, mentre quello con le classi meno ampie (ampiezza 0.1, in alto a destra) dà una rappresentazione piuttosto confusa. Viceversa l'istogramma con classi di ampiezza 0.5 (in basso a sinistra) sembra avere un aspetto più equilibrato, e mostra alcune caratteristiche dei dati che non appaiono negli altri due: in particolare esso indica che le frequenze presentano due massimi relativi in corrispondenza delle classi $[3.0, 3.5]$, $[3.5, 4.0]$ e $[6.0, 6.5]$. Questa struttura dell'istogramma è interessante da un punto di vista statistico in quanto potrebbe indicare che la nostra popolazione è in realtà composta della sovrapposizione di due popolazioni con proprietà differenti: una con valori del carattere prevalentemente compresi fra 3 e 4, e l'altra con valori del carattere prevalentemente vicini a 6. Infine sempre nella stessa Figura 1.2 è riportato un istogramma dello stesso campione costruito con classi di ampiezze diverse fra loro. Si vede quindi come la scelta delle classi modifichi l'aspetto dell'istogramma, volta a volta mettendo in evidenza o nascondendo alcune caratteristiche dei dati. Non ci sono però delle regole per scegliere le classi nella maniera migliore, e d'altra parte non è detto che quel che viene messo in evidenza da un particolare istogramma sia poi in realtà statisticamente significativo. Il ricercatore avveduto, guidato dalla sua esperienza, farà diversi tentativi, e cercherà successivamente delle conferme per le conclusioni suggerite dalle diverse rappresentazioni dei suoi dati.

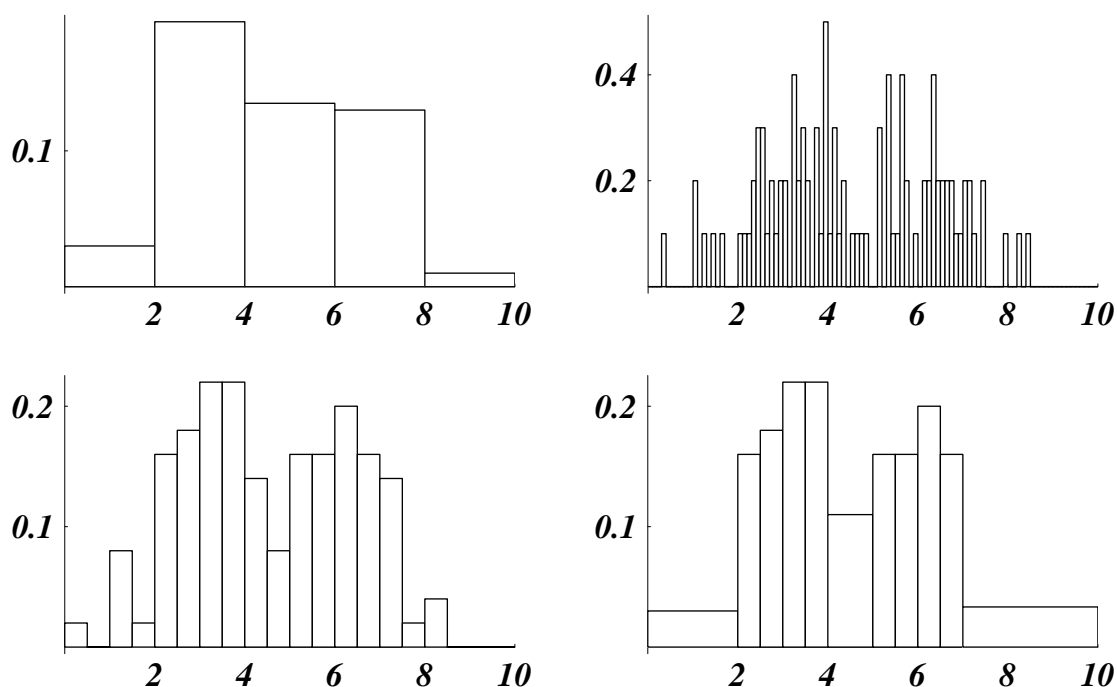


Figura 1.2: Istogrammi dei dati riportati in Tabella 1.3. I due istogrammi in alto si riferiscono a classi di ampiezze rispettivamente 2.0 e 0.1. L'istogramma in basso a sinistra è invece costruito con classi di ampiezza 0.5, mentre quello in basso a destra è costruito con classi di ampiezza variabile.

1.3 Moda, media e varianza

L'analisi statistica non si esaurisce nella rappresentazione delle frequenze dei dati: un altro importante aspetto consiste nella ricerca di opportuni indici che permettano concentrare in pochi numeri le caratteristiche più rilevanti dei dati. Sono di particolare importanza gli *indici di centralità* e quelli di *dispersione*. I primi forniscono un'idea dei valori attorno ai quali sono prevalentemente concentrati i dati; i secondi misurano la dispersione dei dati attorno ai valori centrali. In questa e nelle successive sezioni esamineremo, senza nessuna pretesa di completezza, alcuni dei principali indici statistici. Cominceremo con alcuni indici di centralità

Definizione 1.1. *Data la distribuzione di frequenze di un carattere discreto chiameremo **moda** la modalità corrispondente alla frequenza più grande. Nel caso di caratteri continui la moda è la classe (o il suo valore centrale) corrispondente al rettangolo più alto dell'istogramma.*

Ad esempio nel diagramma a barre di Figura 1.1 la moda è 2. Nel caso dei dati dell'Esempio 1.2, invece, l'identificazione della moda è un po' più delicata. Intanto è chiaro dagli istogrammi di Figura 1.2 che la moda dipende dalla scelta delle clas-

si. In secondo luogo questi grafici mettono in evidenza che può capitare di avere istogrammi (o anche diagrammi a barre) con più di un massimo locale. In realtà il concetto di moda coincide piuttosto con quello di massimo locale che con quello di massimo assoluto. Pertanto un insieme di dati può avere anche più di una moda. Tornando alla Figura 1.2 vediamo allora che per l'istogramma in alto a sinistra la moda è la classe $[2.0, 4.0]$ ovvero il suo valore centrale 3; per i due istogrammi in basso, invece, ci sono due mode: la classe (unione di due classi) $[3.0, 4.0]$ ovvero il suo valore centrale 3.5, e la classe $[6.0, 6.5]$ ovvero 6.25. Infine l'istogramma in alto a destra è caratterizzato da un eccessivo numero di massimi locali che lo rendono confuso e poco adatto ad un'analisi statistica. Ancora una volta, come notato nell'Esempio 1.2, vediamo che classi eccessivamente larghe conducono ad un'analisi troppo grossolana che può far perdere dell'informazione, mentre classi più ristrette (ma non eccessivamente) possono mettere in evidenza alcune caratteristiche importanti.

Definizione 1.2. Si chiama **media** del campione x_1, \dots, x_n del carattere X la quantità

$$m_X = \bar{x} = \frac{x_1 + \dots + x_n}{n} = \frac{1}{n} \sum_{j=1}^n x_j$$

La media è l'indice di centralità più noto e usato anche per le sue proprietà che noi esamineremo brevemente nel seguito. Essa rappresenta in un certo senso il baricentro del campione, se si immagina la densità dei dati come una specie di densità di massa.

Teorema 1.1. Se i dati x_1, \dots, x_n sono misure di un carattere numerico discreto X con modalità w_1, \dots, w_M , e se p_k sono le frequenze relative di tali modalità, si ha

$$m_X = \bar{x} = \sum_{k=1}^M p_k w_k.$$

Dimostrazione: Basterà osservare che per (1.2) $np_k = N_k$ è il numero dei dati che assume il valore w_k , e che quindi

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j = \frac{np_1 w_1 + \dots + np_M w_M}{n} = \sum_{k=1}^M p_k w_k$$

come affermato nel Teorema. □

Teorema 1.2. Assegnato il campione x_1, \dots, x_n , due numeri reali a e b , e costruito il nuovo campione $y_j = ax_j + b$, si ha

$$\bar{y} = a\bar{x} + b.$$

Dimostrazione: Si ha infatti

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j = \frac{1}{n} \sum_{j=1}^n (ax_j + b) = a \left(\frac{1}{n} \sum_{j=1}^n x_j \right) + \frac{1}{n} \sum_{j=1}^n b = a\bar{x} + b$$

che completa la dimostrazione. \square

Esempio 1.3. Supponiamo di sapere che un campione x_1, \dots, x_n di misure di temperatura in gradi Fahrenheit ha media $\bar{x} = 50^\circ F$: come possiamo convertire questa misura in gradi centigradi? Teoricamente dovremmo convertire ogni misura x_j in gradi centigradi con la nota relazione

$$y_j = \frac{100}{180} (x_j - 32) \quad (1.5)$$

e poi calcolare la media \bar{y} . Il calcolo potrebbe essere lungo, e d'altra parte il problema ci fornisce solo il valore di \bar{x} , e non quello delle singole misure. Possiamo però usare il Teorema 1.2 visto che la relazione (1.5) è proprio del tipo $y_j = ax_j + b$. Un semplice calcolo conduce allora al valore

$$\bar{y} = \frac{100}{180} (\bar{x} - 32) = \frac{100}{180} (50 - 32) = 10^\circ C.$$

Teorema 1.3. *Dati due campioni x_1, \dots, x_ℓ e y_1, \dots, y_m con medie \bar{x} e \bar{y} , e detto $z_1, \dots, z_n = x_1, \dots, x_\ell, y_1, \dots, y_m$ il campione ottenuto unificando i primi due con $n = \ell + m$, si ha*

$$\bar{z} = \frac{\ell \bar{x} + m \bar{y}}{n}.$$

Dimostrazione: La media \bar{z} si esprime facilmente come

$$\bar{z} = \frac{1}{n} \sum_{j=1}^n z_j = \frac{1}{n} \left(\sum_{j=1}^{\ell} x_j + \sum_{j=1}^m y_j \right) = \frac{\ell \bar{x} + m \bar{y}}{n}$$

che è il risultato richiesto. \square

Si osservi come il risultato del Teorema 1.3 possa essere riformulato dicendo che la media di campioni combinati è la *media pesata* delle medie dei due campioni separati: in questo caso i pesi rispettivi sono ℓ/n e m/n . In generale si parla di media pesata quando i dati che si mediano non contano tutti allo stesso modo come nella Definizione 1.2 nella quale tutti i dati hanno lo stesso peso $1/n$. Un altro esempio di media pesata è il risultato del Teorema 1.1 nel quale le singole modalità w_k sono mediate pesandole con le frequenze relative p_k . I pesi rappresentano una misura dell'importanza relativa dei dati all'interno di una media.

Definizione 1.3. *Assegnati i numeri x_1, \dots, x_n , e i pesi q_1, \dots, q_n tali che*

$$0 \leq q_k \leq 1, \quad k = 1, \dots, n \quad e \quad q_1 + \dots + q_n = 1$$

\hat{w}_k	1	3	5	7	9
p_k	0.06	0.39	0.27	0.26	0.02

Tabella 1.5: Tabella dei dati raggruppati del campione riportato in Tabella 1.3 per 5 classi di ampiezza 2.0 sull'intervallo $[0, 10]$.

si chiama **media pesata** il numero

$$\sum_{k=1}^n q_k x_k .$$

A volte, nel caso di caratteri numerici continui, non viene fornita l'intera tabella dei dati, ma ci si limita a fornire la tabella delle frequenze in certe determinate classi. In questo caso, in linea di principio, non si potrebbe calcolare la media del campione. C'è però modo di ottenere un valore approssimato identificando tutti i valori x_j che cadono nella classe k —ma con il valore centrale di quella classe che chiameremo \hat{w}_k . Potremo allora eseguire una media dei *dati raggruppati* così ottenuti pesando ogni \hat{w}_k con la frequenza relativa p_k della classe k —ma:

$$\bar{x} \simeq \sum_{k=1}^M p_k \hat{w}_k . \tag{1.6}$$

Esempio 1.4. Nell'Esempio 1.2 potrebbe essere nota solo la Tabella 1.4 (o una analoga con un'altra scelta delle classi), e non l'intera Tabella 1.3 dei dati. Usando allora la Tabella 1.5 dei dati raggruppati da (1.6) si ottiene la media 4.58. D'altra parte utilizzando i dati originali della Tabella 1.3 e la Definizione 1.2 si ottiene il valore esatto 4.56. Come si può vedere l'approssimazione è piuttosto buona anche se le classi scelte sono molto ampie; è intuitivo, comunque, che il valore approssimato è tanto più affidabile quanto più le classi sono strette.

Definizione 1.4. Chiameremo **varianza** di un campione x_1, \dots, x_n di X con media \bar{x} la quantità

$$s_X^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 = \overline{(x - \bar{x})^2} ,$$

e **scarto quadratico** (o **deviazione standard**) la radice quadrata s_X della varianza. Infine si chiama **coefficiente di variazione** il rapporto $s_X/|\bar{x}|$.

Le quantità introdotte nella precedente Definizione sono tutte misure della dispersione dei dati attorno al loro baricentro \bar{x} . In particolare grandi valori della varianza s_X^2 indicano che ci sono delle x_j anche molto lontane da \bar{x} , mentre piccoli valori di s_X^2 indicano che il campione è piuttosto concentrato attorno a \bar{x} . Il caso limite $s_X^2 = 0$, poi, implica che tutti i valori x_j coincidono con \bar{x} .

Teorema 1.4. Se i dati x_1, \dots, x_n sono misure di un carattere numerico discreto con modalità w_1, \dots, w_M , e se p_k sono le frequenze relative di tali modalità, si ha

$$s_X^2 = \sum_{k=1}^M p_k (w_k - \bar{x})^2.$$

Dimostrazione: La dimostrazione è analoga a quella del Teorema 1.1. □

Teorema 1.5. Dato un campione x_1, \dots, x_n con media \bar{x} , si ha

$$s_X^2 = \overline{x^2} - \bar{x}^2 = \frac{1}{n} \sum_{j=1}^n x_j^2 - \left(\frac{1}{n} \sum_{j=1}^n x_j \right)^2,$$

dove $\overline{x^2}$ indica la media dei quadrati del campione, e \bar{x}^2 il quadrato della sua media.

Dimostrazione: Infatti si ha

$$\begin{aligned} s_X^2 &= \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 = \frac{1}{n} \sum_{j=1}^n (x_j^2 + \bar{x}^2 - 2x_j\bar{x}) \\ &= \frac{1}{n} \sum_{j=1}^n x_j^2 + \frac{1}{n} \sum_{j=1}^n \bar{x}^2 - 2\bar{x} \frac{1}{n} \sum_{j=1}^n x_j = \frac{1}{n} \sum_{j=1}^n x_j^2 + \bar{x}^2 - 2\bar{x}^2 \\ &= \overline{x^2} - \bar{x}^2 \end{aligned}$$

che dimostra il Teorema. □

Il teorema precedente è particolarmente usato per semplificare il calcolo della varianza: una volta calcolata \bar{x} , infatti, è in genere più conveniente calcolare la media del campione dei quadrati e usare il Teorema 1.5, piuttosto che calcolare direttamente la varianza dalla definizione.

Teorema 1.6. Dato un campione x_1, \dots, x_n di X con media \bar{x} e varianza s_X^2 , e due numeri a e b , definito il nuovo campione $y_j = ax_j + b$ di Y si ha

$$s_Y^2 = a^2 s_X^2.$$

Dimostrazione: Infatti si ha dalle definizioni e dal Teorema 1.2 che

$$s_Y^2 = \frac{1}{n} \sum_{j=1}^n (y_j - \bar{y})^2 = \frac{1}{n} \sum_{j=1}^n (ax_j + b - a\bar{x} - b)^2 = \frac{a^2}{n} \sum_{j=1}^n (x_j - \bar{x})^2 = a^2 s_X^2,$$

come volevasi dimostrare. □

Definizione 1.5. Chiameremo **errore quadratico medio (eqm)** di un campione x_1, \dots, x_n rispetto al numero a la quantità

$$\mathcal{E}(a) = \frac{1}{n} \sum_{j=1}^n (x_j - a)^2 = \overline{(x - a)^2}.$$

Teorema 1.7. La media \bar{x} di un campione x_1, \dots, x_n è il valore di a che rende minimo l'eqm $\mathcal{E}(a)$ del campione.

Dimostrazione: Per determinare il punto di minimo dell'eqm bisogna imporre che si annulli la derivata prima $\mathcal{E}'(a)$, cioè

$$\mathcal{E}'(a) = -\frac{2}{n} \sum_{j=1}^n (x_j - a) = -2(\bar{x} - a) = 0,$$

da cui si deduce immediatamente il risultato. □

Definizione 1.6. Diremo che x_1, \dots, x_n è un **campione standardizzato** quando $m_X = \bar{x} = 0$ e $s_X^2 = 1$.

Teorema 1.8. Dato il campione x_1, \dots, x_n con media \bar{x} e varianza s_X^2 , il campione

$$y_j = \frac{x_j - \bar{x}}{s_X}$$

è **standardizzato**.

Dimostrazione: Infatti dai Teoremi 1.2 e 1.6 con $a = 1/s_X$ e $b = -\bar{x}/s_X$ si ha

$$\begin{aligned} \bar{y} &= \frac{\bar{x}}{s_X} - \frac{\bar{x}}{s_X} = 0, \\ s_Y^2 &= \frac{s_X^2}{s_X^2} = 1, \end{aligned}$$

il che prova la tesi. □

Va ricordato infine, nel caso di caratteri numerici continui, che anche per la varianza si può eseguire un calcolo approssimato con *dati raggruppati* in classi. Data infatti la tabella delle frequenze relative p_k in certe determinate classi si identificano i valori x_j che cadono nella classe k -ma con il valore centrale di quella classe \hat{w}_k . Potremo allora calcolare prima di tutto un'approssimazione della media con (1.6), e poi un'approssimazione della varianza dei dati raggruppati:

$$s_X^2 \simeq \sum_{k=1}^M p_k \hat{w}_k^2 - \left(\sum_{k=1}^M p_k \hat{w}_k \right)^2.$$

Esempio 1.5. Riprendiamo i dati dell'Esempio 1.2 riportati in Tabella 1.3 per i quali abbiamo già calcolato la media (che vale 4.56) nell'Esempio 1.4, e calcoliamone la varianza. Un'applicazione diretta della Definizione 1.4 ai dati della Tabella 1.3 fornisce un valore di 3.40, ma il calcolo è abbastanza laborioso. Può essere conveniente allora calcolare la media dei quadrati dei dati (che è 24.22) e usare poi il Teorema 1.5 per ottenere più rapidamente lo stesso valore: $24.22 - 4.56^2 = 3.40$. Usando invece i dati raggruppati di Tabella 1.5 i calcoli sono molto più veloci, ma approssimati. Ricordando dall'Esempio 1.4 che la media approssimata vale 4.58, e calcolando la media approssimata dei quadrati 24.68, si ottiene per la varianza con dati raggruppati: $24.68 - 4.58^2 = 3.70$.

1.4 Mediana e quantili

Abbiamo già osservato che è talora utile riordinare un campione x_1, \dots, x_n in ordine crescente: per distinguere i due tipi di campioni indicheremo i *campioni ordinati* con la notazione $x_{[1]}, \dots, x_{[n]}$ in modo tale che da ora in poi

$$x_{[1]} \leq x_{[2]} \leq \dots \leq x_{[n]}.$$

Definizione 1.7. Chiameremo **quantile di ordine** α ($0 < \alpha < 1$) di un campione x_1, \dots, x_n un numero q_α maggiore o uguale di una frazione α degli elementi del campione ordinato $x_{[1]}, \dots, x_{[n]}$, nel senso che il numero delle $x_{[j]}$ che risulta minore o uguale di q_α non deve superare $\alpha(n+1)$. Per evitare ambiguità dovute al carattere intero degli indici del campione adotteremo la seguente procedura per la determinazione di q_α : si calcola $\alpha(n+1)$,

- se $\alpha(n+1)$ è intero, si considera l'indice $j = \alpha(n+1)$ e si pone $q_\alpha = x_{[j]}$;
- se $\alpha(n+1)$ non è intero, si considera l'indice j tale che $j < \alpha(n+1) < j+1$ e si pone

$$q_\alpha = \frac{x_{[j]} + x_{[j+1]}}{2}. \quad (1.7)$$

Definizione 1.8. Il quantile di ordine $\alpha = \frac{1}{2}$ prende il nome di **mediana**; i quantili di ordini $\alpha = \frac{k}{4}$ con $k = 1, 2, 3$ si chiamano rispettivamente primo, secondo e terzo **quartile**: è evidente che il secondo quartile coincide con la mediana. I quantili con ordini $\alpha = \frac{k}{10}$ con $k = 1, \dots, 9$ si chiamano **decili**, e infine quelli con ordini $\alpha = \frac{k}{100}$ con $k = 1, \dots, 99$ si chiamano **percentili**.

Esempio 1.6. Riprendendo la Tabella 1.3 dei dati dell'Esempio 1.2 abbiamo $n = 100$: per calcolare la mediana osserviamo allora che $\alpha(n+1) = 101/2 = 50.5$ non è intero. Pertanto sceglieremo $i = 50$ e useremo (1.7): dai dati si ha allora per la mediana

$$q_{\frac{1}{2}} = \frac{x_{[50]} + x_{[51]}}{2} = \frac{4.22 + 4.31}{2} = 4.265.$$

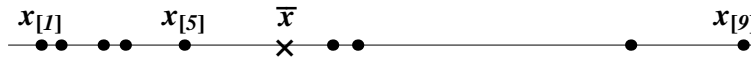


Figura 1.3: Media \bar{x} e mediana $x_{[5]}$ di un campione di $n = 9$ dati rappresentati su un asse numerico: diversamente dalla media il valore della mediana non è influenzato da eventuali variazioni nei valori degli altri dati.

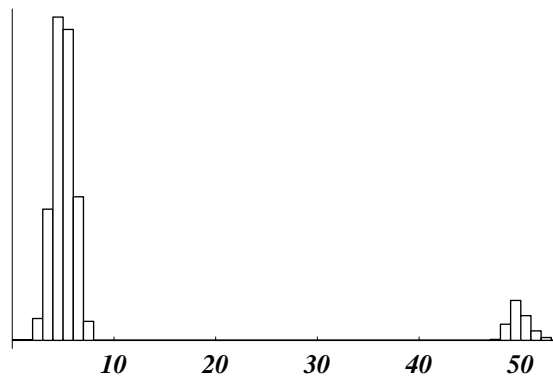


Figura 1.4: Distribuzione ipotetica del reddito dei dipendenti di un'azienda con 1 000 impiegati e operai, e 100 dirigenti. Data la differenza di reddito dei due gruppi la mediana costituisce un indice più conveniente della media per rappresentare il reddito del tipico dipendente.

Analogamente si ha $\alpha(n + 1) = 101/4 = 25.25$, per cui $i = 25$ e i due quartili sono

$$q_{\frac{1}{4}} = \frac{x_{[25]} + x_{[26]}}{2} = \frac{3.09 + 3.23}{2} = 3.16, \quad q_{\frac{3}{4}} = \frac{x_{[75]} + x_{[76]}}{2} = \frac{6.24 + 6.28}{2} = 6.26.$$

La mediana è un altro indice di centralità come la media e la moda: i loro valori sono in generale differenti, e la scelta dell'opportuno indice dipende dal particolare problema trattato. Anche le proprietà dei diversi indici sono differenti: ad esempio per la mediana non ci sono formule semplici come quelle dei teoremi sulla media richiamati nella Sezione 1.3. Per altri versi invece la mediana presenta il vantaggio di essere un indice più *robusto* della media nel senso che il suo valore è meno sensibile a variazioni o errori nei dati del campione.

Esempio 1.7. Si consideri il campione di $n = 9$ numeri rappresentato graficamente in Figura 1.3: dalla Definizione 1.7 si vede subito che in questo caso la mediana coincide con il dato $x_{[5]}$. Sull'asse è riportata anche la posizione della media \bar{x} . Supponiamo ora di aumentare o diminuire il valore di uno dei dati, ad esempio $x_{[9]}$: è evidente che, finché $x_{[9]}$ rimane a destra di $x_{[5]}$, la mediana mantiene lo stesso valore $x_{[5]}$. Non avviene invece la

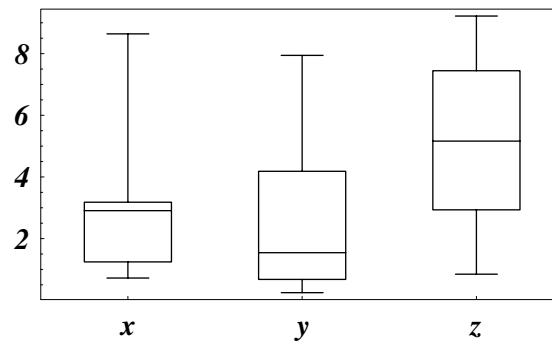


Figura 1.5: Esempi di boxplot costruiti sui tre campioni riportati nella Tabella 1.6

$x_{[i]}$	0.72	1.10	1.24	1.98	2.82	2.99	3.01	3.18
	3.31	8.64						
$y_{[j]}$	0.25	0.66	0.68	1.07	1.09	1.15	1.94	3.11
	4.18	4.79	6.18	7.94				
$z_{[k]}$	0.85	1.49	2.19	2.93	4.46	4.61	4.62	5.16
	5.67	6.41	6.46	7.45	7.66	8.65	9.22	

Tabella 1.6: Campioni (ordinati) utilizzati per i boxplot della Figura 1.5.

stessa cosa per la media \bar{x} il cui valore dato dalla Definizione 1.2 è ovviamente influenzato da eventuali variazioni di $x_{[9]}$.

Esempio 1.8. La mediana è un indice utile soprattutto nei casi in cui la media rischia di non essere significativa. Supponiamo di considerare un'azienda con 1000 impiegati e operai, e 100 dirigenti, e supponiamo che l'istogramma dei redditi di tutti i dipendenti sia quello di Figura 1.4: i redditi dei 1000 impiegati e operai sono concentrati attorno a 5, mentre quelli dei dirigenti si distribuiscono attorno a 50. D'altra parte si ottiene per calcolo diretto dai valori del campione – qui non riportati – che la mediana è 5.13 mentre la media è 9.08: un valore quasi doppio. La media, quindi, risente molto della presenza di un piccolo numero di redditi alti, ed è meno rappresentativa del reddito tipico di quanto non sia invece la mediana.

Definizione 1.9. Chiameremo *range* del campione il numero $x_{[n]} - x_{[1]}$, ovvero l'ampiezza dell'intervallo $[x_{[1]}, x_{[n]}]$ che contiene tutti i dati; chiameremo invece *differenza interquartile* il numero $q_{\frac{3}{4}} - q_{\frac{1}{4}}$, ovvero l'ampiezza dell'intervallo $[q_{\frac{1}{4}}, q_{\frac{3}{4}}]$ delimitato dai due quartili.

Il range e la differenza interquartile sono ovviamente degli indici di dispersione. Assieme alla mediana essi possono essere rappresentati su un grafico noto come *boxplot* a causa della sua tipica forma: una volta determinati gli intervalli $[x_{[1]}, x_{[n]}]$ e $[q_{\frac{1}{4}}, q_{\frac{3}{4}}]$ si disegna un rettangolo, o scatola (*box*), i cui lati inferiore e superiore sono rispettivamente il primo e il terzo quartile. L'altezza di tale scatola è quindi

pari alla differenza interquartile. All'interno della scatola si traccia una linea in corrispondenza della mediana. All'esterno, infine, si riportano due segmenti orizzontali corrispondenti ai valori $x_{[1]}$ e $x_{[n]}$ (la loro distanza è ovviamente il range) e due segmenti verticali che li congiungono ai lati della scatola. Nella Figura 1.5 sono disegnati i boxplot dei tre campioni di Tabella 1.6. In particolare questi grafici mettono bene in evidenza il contrasto fra la simmetria del campione $z_{[k]}$ e la asimmetria dei campioni $x_{[i]}$ e $y_{[j]}$: in questi due casi infatti la mediana è lontana dal centro della scatola (quindi è più vicina a uno dei due quartili), e inoltre i due dati estremi sono a distanze piuttosto diverse dai rispettivi quartili. Il grafico mette anche bene in evidenza in che senso il range e la distanza interquartile sono due misure differenti della dispersione del campione: ad esempio le $z_{[k]}$, pur avendo approssimativamente lo stesso range delle $x_{[i]}$ hanno una differenza interquartile sensibilmente più elevata.

1.5 Momenti, asimmetria e curtosi

Definizione 1.10. Chiameremo rispettivamente **momento di ordine k** e **momento centrato di ordine k** di un campione x_1, \dots, x_n le quantità

$$m_k = \frac{1}{n} \sum_{j=1}^n x_j^k, \quad \mu_k = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^k.$$

Ovviamente $m_1 = \bar{x}$, e $\mu_2 = s_X^2$. Si chiama inoltre **asimmetria** del campione la quantità

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{s_X^3},$$

e **curtosi** la quantità

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{s_X^4}.$$

I momenti sono indici che generalizzano medie e varianze e forniscono ulteriori informazioni sulla dispersione, la simmetria e in generale la forma della distribuzione del campione. In particolare l'indice di *asimmetria* γ_1 prende valori prossimi a zero se i dati si distribuiscono in maniera simmetrica attorno alla media, mentre prende valori apprezzabilmente diversi da zero se la distribuzione è asimmetrica (vedi Figura 1.6). Il valore di γ_1 può essere positivo o negativo: valori positivi indicano la presenza di code verso destra; valori negativi sono invece associati a code verso sinistra. La *curtosi* γ_2 invece assume solo valori positivi perché coinvolge solo medie di potenze pari dei dati: essa è legata alla velocità con cui l'istogramma tende a zero allontanandosi dal valore medio. In particolare la curtosi ha valori vicini a zero quando le code dell'istogramma sono corte, cioè quando l'istogramma si annulla rapidamente; viceversa assume valori grandi e positivi quando ci sono code lunghe, cioè quando sono presenti dati anche molto lontani dalla media (vedi Figura 1.7).

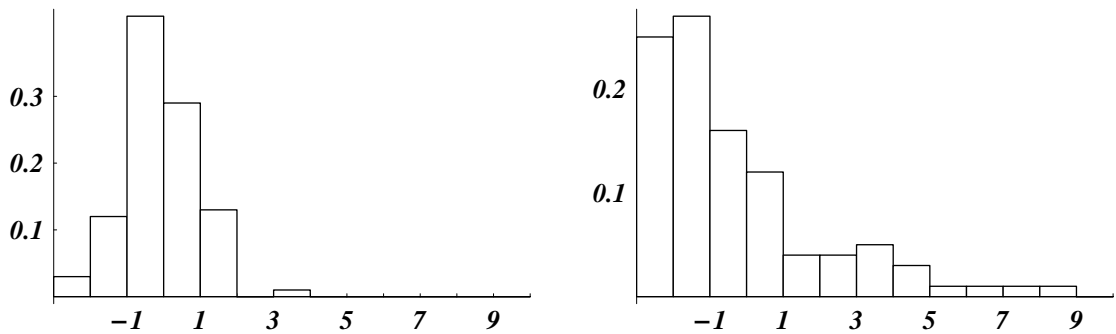


Figura 1.6: Istogrammi di dati con diversa asimmetria: $\gamma_1 = 0.02$ per il primo, e $\gamma_1 = 1.56$ per il secondo.

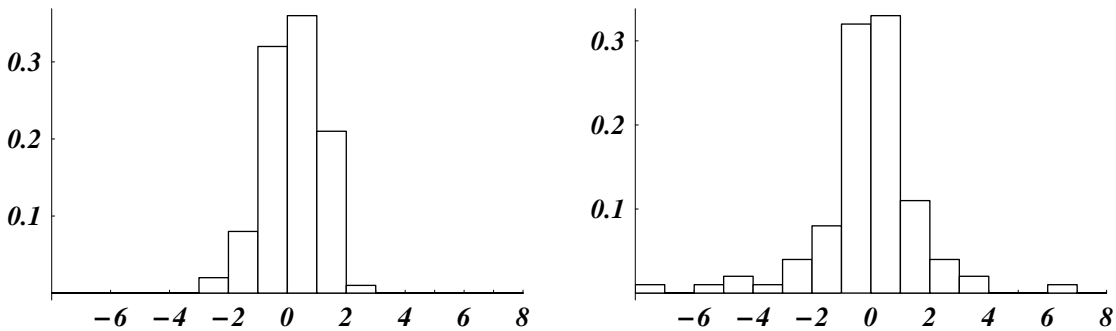


Figura 1.7: Istogrammi di dati con diversa curtosi: $\gamma_2 = 2.59$ per il primo, e $\gamma_2 = 7.76$ per il secondo.

1.6 Medie generalizzate

Il concetto di media aritmetica introdotto nella Sezione 1.3 può essere opportunamente generalizzato secondo le esigenze del problema in discussione. Chiariremo con alcuni esempi in che senso può essere necessario fare ricorso a questi nuovi concetti.

Esempio 1.9. Supponiamo che una certa quantità di capitale C sia stata investita a un tasso di interesse che viene aggiornato ogni mese, e supponiamo di indicare con p_1, \dots, p_n i tassi di interesse in un periodo di n mesi: quale valore dovremmo considerare come *tasso medio* p di interesse? È intuitivo che il criterio dovrebbe essere il seguente: p è il tasso di interesse costante che applicato per n mesi produce lo stesso aumento di capitale prodotto dalla applicazione successiva dei tassi p_1, \dots, p_n . Siccome nel caso di tasso variabile dopo il primo mese il capitale è $(1+p_1)C$, dopo il secondo $(1+p_2)(1+p_1)C$ e così via, è evidente che il nostro criterio impone la relazione

$$(1+p)^n C = (1+p_1) \cdot \dots \cdot (1+p_n) C,$$

e quindi in definitiva

$$1 + p = [(1 + p_1) \cdot \dots \cdot (1 + p_n)]^{\frac{1}{n}}.$$

Il valore di $1 + p$ così ottenuto si chiama *media geometrica* dei dati $1 + p_j$ con $j = 1, \dots, n$.

Esempio 1.10. Supponiamo che una ditta produttrice di automobili svolga la sua attività in n stabilimenti ciascuno dei quali ha un suo tempo di produzione, nel senso che esse producono una automobile rispettivamente nei tempi T_1, \dots, T_n : quale valore dovremmo considerare come *tempo medio* T di produzione della ditta? In questo caso adotteremo il seguente criterio: T è il tempo di produzione con il quale la ditta produrrebbe complessivamente nell'unità di tempo un numero di auto uguale a quello prodotto dagli stabilimenti con i tempi T_1, \dots, T_n . Siccome ogni ditta produce $1/T_j$ automobili nell'unità di tempo, il nostro criterio impone che

$$\frac{n}{T} = \frac{1}{T_1} + \dots + \frac{1}{T_n}$$

ovvero

$$T = \frac{1}{\frac{1}{n} \left(\frac{1}{T_1} + \dots + \frac{1}{T_n} \right)}.$$

Il valore di T così ottenuto si chiama *media armonica* dei tempi T_j con $j = 1, \dots, n$.

Esempio 1.11. I batteri di una determinata specie si organizzano in colonie di forma circolare, e il numero di batteri è proporzionale alla superficie delle colonie. Si osservano n colonie con diametri d_1, \dots, d_n : che valore possiamo attribuire al diametro medio d delle colonie? Anche in questo caso ci facciamo guidare da un criterio ragionevole: richiederemo che n colonie tutte con lo stesso diametro d abbiano la stessa superficie totale delle n colonie con diametri differenti d_1, \dots, d_n . In tal caso dovremo imporre che

$$n\pi d^2 = \pi(d_1^2 + \dots + d_n^2)$$

e quindi avremo

$$d = \sqrt{\frac{d_1^2 + \dots + d_n^2}{n}}.$$

Il valore di d così ottenuto si chiama *media quadratica* dei diametri d_j con $j = 1, \dots, n$.

Definizione 1.11. Dato un campione x_1, \dots, x_n chiameremo rispettivamente *media geometrica, armonica e quadratica* le espressioni

$$(x_1 \cdot \dots \cdot x_n)^{\frac{1}{n}}, \quad \left(\frac{x_1^{-1} + \dots + x_n^{-1}}{n} \right)^{-1}, \quad \left(\frac{x_1^2 + \dots + x_n^2}{n} \right)^{\frac{1}{2}}.$$

Le medie così definite sono esempi di medie generalizzate che si ottengono tutte con la seguente procedura: data una funzione reale dotata di inversa h^{-1} si costruisce un nuovo campione $h(x_1), \dots, h(x_n)$, se ne calcola la media aritmetica e infine si applica h^{-1} al risultato:

$$h^{-1} \left[\frac{h(x_1) + \dots + h(x_n)}{n} \right].$$

Questo in pratica vuol dire che, per ragioni derivanti dal particolare problema discusso, può essere più significativo eseguire la media non direttamente sui dati x_j , ma sui dati trasformati $h(x_j)$. Le medie geometrica, armonica e quadratica si ottengono rispettivamente con le seguenti scelte della funzione $h(x)$:

$$\log(x) \qquad \frac{1}{x} \qquad x^2$$

come si vede facilmente applicando la Definizione 1.11.

Capitolo 2

Statistica multivariata

2.1 Dati multidimensionali

Sugli individui di una popolazione possono essere eseguite osservazioni e misure di due o più caratteri con lo scopo di metterne anche in evidenza gli eventuali legami statistici. Ad esempio possiamo misurare altezza e peso dei cittadini di una determinata comunità per mettere in evidenza una relazione fra le due misure. In questo caso gli elementi del nostro campione non saranno più dei semplici numeri, ma vettori con due o più componenti. In questa sezione ci limiteremo ad esaminare il caso di due caratteri (X, Y) , sicché il nostro campione sarà del tipo $(x_1, y_1), \dots, (x_n, y_n)$.

Se i caratteri sono qualitativi o numerici discreti con un numero finito di modalità A_1, \dots, A_r del carattere X e B_1, \dots, B_s del carattere Y , una prima maniera di rappresentare il campione sarà quella di costruire una *tabella di contingenza* di frequenze assolute come quella riportata in Tabella 2.1. In essa si riportano innanzitutto le *frequenze congiunte* $N_{j,k}$, cioè il numero delle volte in cui si presenta la coppia di modalità (A_j, B_k) ; sui margini della tabella si riportano poi le *frequenze marginali* $N_{j\cdot}$ e $N_{\cdot,k}$, cioè il numero di volte in cui si presentano separatamente le modalità A_j e B_k ; nell'angolo destro in basso si riporta infine la numerosità totale n del campione. Si noti che, per un dato j la marginale $N_{j\cdot}$ è la somma delle $N_{j,k}$ della sua riga, mentre per un dato k la marginale $N_{\cdot,k}$ è la somma delle $N_{j,k}$ della sua colonna; infine anche n totale è la somma delle marginali sia della sua riga che

	B_1	\dots	B_s	
A_1	$N_{1,1}$	\dots	$N_{1,s}$	$N_{1\cdot}$
\vdots	\vdots	\ddots	\vdots	\vdots
A_r	$N_{r,1}$	\dots	$N_{r,s}$	$N_{r\cdot}$
	$N_{\cdot,1}$	\dots	$N_{\cdot,s}$	n

Tabella 2.1: Tabella di contingenza per due caratteri X e Y rispettivamente con modalità A_j e B_k .

	GIU	ECO	LET	SCI	MED	FAR	ALTRO	
<i>Proprietario</i>	80	36	134	99	65	28	69	511
<i>Contadino</i>	6	2	15	6	4	1	5	39
<i>Imprenditore</i>	168	74	312	137	208	53	83	1 035
<i>Professionista</i>	470	191	806	400	876	164	124	3 031
<i>Dirigente</i>	236	99	493	264	281	56	123	1 552
<i>Impiegato</i>	145	52	281	133	135	30	74	850
<i>Operaio</i>	166	64	401	193	127	23	157	1 131
<i>Altro</i>	321	121	651	258	309	49	142	1 851
	1 592	639	3 093	1 490	2 005	404	777	10 000

Tabella 2.2: Tabella di contingenza per la scelta della Facoltà universitaria di $n = 10\,000$ studenti, secondo l'attività lavorativa del padre (dati relativi all'a.a. 1975/76; INSEE, Paris 1978).

della sua colonna. In maniera del tutto analoga si costruisce anche la tabella di contingenza delle frequenze relative congiunte e marginali

$$p_{j,k} = \frac{N_{j,k}}{n}; \quad p_{j,\cdot} = \frac{N_{j,\cdot}}{n}, \quad p_{\cdot,k} = \frac{N_{\cdot,k}}{n}.$$

In questo caso però, a causa della normalizzazione (1.3) delle frequenze relative, nell'angolo destro in basso comparirà 1 invece di n . Una tabella di contingenza può essere redatta anche per modalità numeriche continue, ma in questo caso – come per gli istogrammi – bisognerà raggruppare i dati in classi con una opportuna suddivisione in intervalli.

Esempio 2.1. Nella Tabella 2.2 sono riportati in forma di tabella di contingenza i dati relativi alla scelta della facoltà universitaria di $n = 10\,000$ studenti secondo l'attività lavorativa del padre. La tabella mette in evidenza la composizione sociale degli studenti universitari (marginali verticali), il gradimento delle diverse facoltà universitarie (marginali orizzontali), e infine l'eventuale relazione che intercorre fra l'estrazione socio-professionale della famiglia degli studenti e la scelta della facoltà universitaria (frequenze congiunte).

Nel caso in cui le modalità osservate siano numeriche e continue è utile rappresentare graficamente in un piano x, y i punti con coordinate $(x_1, y_1), \dots, (x_n, y_n)$. La conformazione della nuvola di punti disegnata fornisce una prima indicazione sulla eventuale relazione intercorrente fra i due caratteri. Nella Figura 2.1 sono riportati alcuni esempi con $n = 100$ punti: nel caso *A* i punti sono disposti in modo da non suggerire nessun tipo di dipendenza funzionale tra i due caratteri X e Y . Invece in *B* si nota che i valori di Y tendono ad crescere (decrescere) quando anche i valori di X crescono (decregono); anzi la conformazione della nuvola indica una approssimativa dipendenza funzionale lineare $Y = aX + b$ con $a > 0$. Anche nel caso *C* i dati mostrano una dipendenza approssimativamente lineare, ma questa volta con $a < 0$: infatti ora i valori di Y tendono a crescere (decrescere) quando i valori di

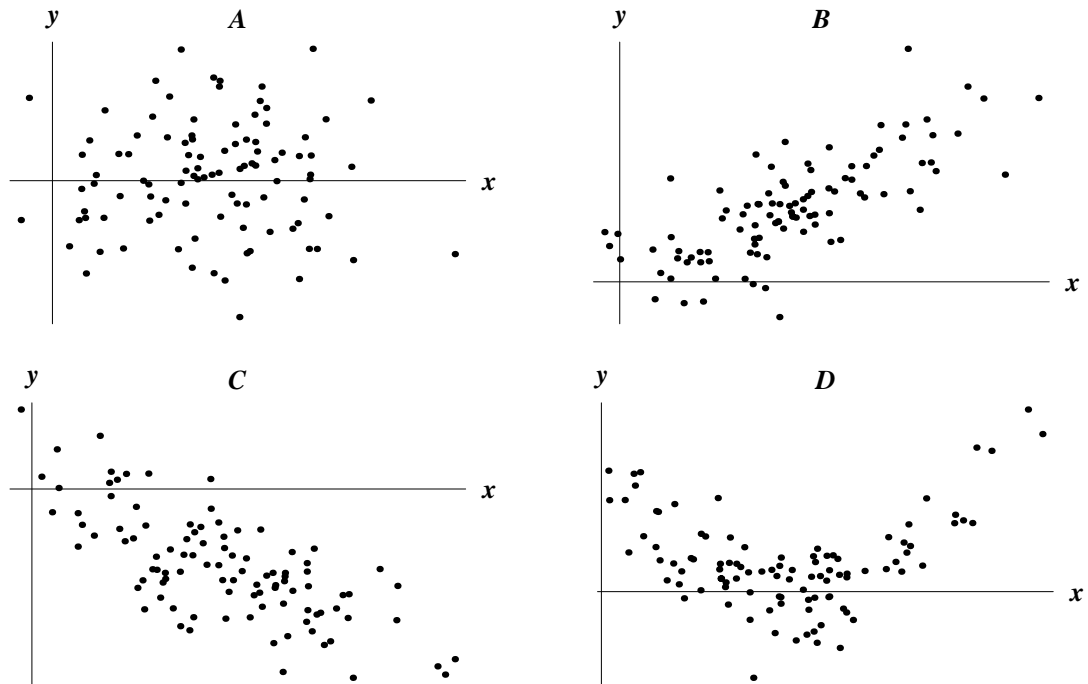


Figura 2.1: Esempi di conformazioni delle nuvole di punti di dati bidimensionali.

X decrescono (crescono). Infine il caso D suggerisce una dipendenza non lineare, approssimativamente parabolica, tra i due caratteri dato che i valori di Y crescono quando i valori di X si allontanano – nei due versi – dal centro della nuvola.

2.2 Covarianza, correlazione e regressione

Definizione 2.1. Dato un campione $(x_1, y_1), \dots, (x_n, y_n)$ di due caratteri numerici X e Y , si chiama **covarianza** di X e Y la quantità

$$s_{XY} = \overline{(x - \bar{x})(y - \bar{y})} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

dove \bar{x} e \bar{y} sono le medie delle x_i e delle y_i . Si chiama poi **coefficiente di correlazione** la quantità

$$r_{XY} = \frac{s_{XY}}{s_X s_Y}$$

dove s_X e s_Y sono le deviazioni standard delle x_i e delle y_i . Se infine $s_{XY} = 0$ (e quindi anche $r_{XY} = 0$) diremo che x e y sono **non correlati**, mentre parleremo di **correlazione positiva (negativa)** se $s_{XY} > 0$ ($s_{XY} < 0$).

La covarianza e il coefficiente di correlazione sono indicatori numerici importanti nell'analisi della relazione che intercorre fra due caratteri X e Y . In particolare, come vedremo, essi entrano nella valutazione quantitativa della dipendenza lineare di un carattere dall'altro, cioè nella determinazione dei coefficienti a e b di una retta $Y = aX + b$ che descriva (almeno approssimativamente) l'andamento dei dati.

Teorema 2.1. *Dato un campione $(x_1, y_1), \dots, (x_n, y_n)$ con medie \bar{x} e \bar{y} si ha*

$$s_{XY} = \overline{xy} - \bar{x}\bar{y} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \left(\frac{1}{n} \sum_{j=1}^n y_j \right),$$

dove \overline{xy} indica la media dei prodotti $x_i y_i$, e $\bar{x}\bar{y}$ il prodotto delle due medie separate.

Teorema 2.2. *Dato un campione $(x_1, y_1), \dots, (x_n, y_n)$ il coefficiente di correlazione r_{XY} soddisfa sempre la relazione $|r_{XY}| \leq 1$; in particolare se $|r_{XY}| = 1$, allora esistono due numeri a e b tali che $y_i = ax_i + b$ per $i = 1, \dots, n$, e il segno di a coincide con il segno di r_{XY} . Infine il valore di r_{XY} resta invariato per cambiamenti di unità di misura; cioè, se moltiplichiamo i dati per due numeri arbitrari A e B ottenendo i nuovi dati $(Ax_1, By_1), \dots, (Ax_n, By_n)$, il valore di r_{XY} non cambia.*

Torniamo ora al problema dell'analisi delle nuvole di punti di dati bidimensionali come quelle di Figura 2.1, e domandiamoci se non sia possibile trovare una relazione analitica che descriva – almeno approssimativamente – la dipendenza delle y_i dalle x_i . L'ipotesi più semplice è che ci sia una relazione di tipo lineare $Y = aX + b$, ma un semplice sguardo ai grafici di Figura 2.1 ci convince del fatto che in generale sarà impossibile trovare due numeri a e b tali che $y_i = ax_i + b$ per tutte le $i = 1, \dots, n$; cioè che è impossibile trovare una retta che passi per tutti i punti della nuvola. Potremo invece provare a determinare a e b in modo che la retta $y = ax + b$ approssimi nel modo migliore l'andamento della nuvola di punti. Il senso in cui parliamo di approssimazione ottimale è precisato nella Definizione seguente.

Definizione 2.2. *Dato un campione $(x_1, y_1), \dots, (x_n, y_n)$, chiameremo **retta di regressione** la retta $y = ax + b$ i cui coefficienti a e b rendono minimo l'**errore quadratico medio (eqm)**, cioè la quantità*

$$\mathcal{E}(a, b) = \overline{[y - (ax + b)]^2} = \frac{1}{n} \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Teorema 2.3. *Dato il campione $(x_1, y_1), \dots, (x_n, y_n)$, i coefficienti a e b della retta di regressione sono*

$$a = \frac{s_{XY}}{s_X^2}, \quad b = \bar{y} - a\bar{x} = \bar{y} - \frac{s_{XY}}{s_X^2} \bar{x}.$$

Dimostrazione: Per determinare le a e b che rendono minimo l'eqm calcoliamo le derivate di $\mathcal{E}(a, b)$

$$\frac{\partial \mathcal{E}}{\partial a} = -\frac{2}{n} \sum_{i=1}^n x_i [y_i - (ax_i + b)], \quad \frac{\partial \mathcal{E}}{\partial b} = -\frac{2}{n} \sum_{i=1}^n [y_i - (ax_i + b)]$$

e poniamole uguali a 0 ottenendo il sistema di equazioni in a e b

$$\begin{aligned} \sum_{i=1}^n x_i [y_i - (ax_i + b)] &= 0, \\ \sum_{i=1}^n [y_i - (ax_i + b)] &= 0. \end{aligned}$$

La seconda equazione del sistema si scrive anche come

$$\sum_{i=1}^n (y_i - ax_i) - nb = 0$$

da cui si ricava subito

$$b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i) = \frac{1}{n} \sum_{i=1}^n y_i - \frac{a}{n} \sum_{i=1}^n x_i = \bar{y} - a\bar{x}. \quad (2.1)$$

Sostituendo nella prima equazione si ottiene allora

$$\sum_{i=1}^n x_i [y_i - (ax_i + b)] = \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - (\bar{y} - a\bar{x}) \sum_{i=1}^n x_i = 0$$

e dividendo per n

$$\frac{1}{n} \sum_{i=1}^n x_i y_i - \frac{a}{n} \sum_{i=1}^n x_i^2 - (\bar{y} - a\bar{x}) \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}\bar{y} - a\bar{x}^2 - (\bar{y} - a\bar{x})\bar{x} = 0.$$

Ricordando ora i risultati dei Teoremi 1.5 e 2.1 l'equazione diviene

$$s_{XY} + \bar{x}\bar{y} - a(s_X^2 + \bar{x}^2) - \bar{x}\bar{y} + a\bar{x}^2 = s_{XY} - as_X^2 = 0$$

da cui si ricava immediatamente

$$a = \frac{s_{XY}}{s_X^2}. \quad (2.2)$$

Le soluzioni (2.2) e (2.1) del nostro sistema di equazioni rendono quindi minimo l'eqm e sono, come richiesto nel Teorema, i coefficienti della retta di regressione. \square

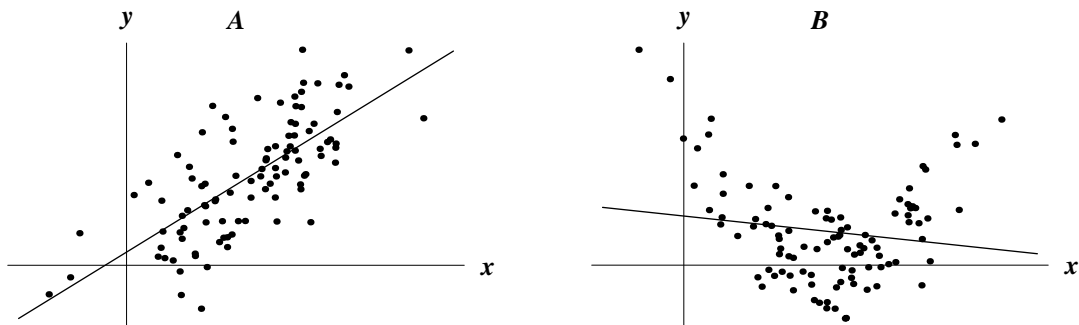


Figura 2.2: Esempi di rette di regressione per dati bidimensionali. Nel caso A il coefficiente di correlazione è $r_{XY} = 0.72$, mentre nel caso B è $r_{XY} = -0.14$.

Esempi di rette di regressione sono riportati nella Figura 2.2. Nella parte A la retta, calcolata a partire dal Teorema 2.3, offre una descrizione approssimata ma significativa della relazione che intercorre fra i dati del campione. Bisogna però evitare di credere che sia significativo in ogni caso supporre una relazione *lineare* fra X e Y . Nella parte B della Figura, ad esempio, si può vedere un campione in cui la relazione fra X e Y è presumibilmente non lineare (parabolica). Anche in questo caso si può determinare la retta di regressione, ma ora l'approssimazione dei punti con una retta è poco significativa. Si noti che per il Teorema 2.3 se X e Y sono non correlati (cioè se $s_{XY} = 0$, e $r_{XY} = 0$) allora $a = 0$, cioè la retta di regressione è orizzontale e i valori di X non mostrano nessuna dipendenza dai valori di Y . Inoltre il coefficiente angolare a ha lo stesso segno di s_{XY} , e quindi la retta avrà un andamento crescente (decrescente) se vi è correlazione positiva (negativa). Un ovvio esempio di correlazione positiva è quello delle coppie (*peso*, *altezza*) di un campione di n persone. Il coefficiente di correlazione r_{XY} è una misura della linearità della relazione fra X e Y . In base al Teorema 2.2 infatti, mentre la covarianza s_{XY} può assumere ogni valore positivo e negativo, r_{XY} cade sempre in $[-1, 1]$, e se $r_{XY} = \pm 1$ allora $y_i = ax_i + b$, cioè la retta di regressione passa attraverso tutti i punti.

2.3 Componenti principali

Quando ad ogni individuo della popolazione sono associati $p \geq 3$ caratteri numerici X_1, \dots, X_p , gli n elementi del campione diventano vettori con p componenti $\mathbf{x}_j = (x_{j1}, \dots, x_{jp}) \in \mathbf{R}^p$, $j = 1, \dots, n$, e i dati si presentano come una matrice $p \times n$

$$\|x_{jk}\| = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ x_{21} & \dots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix}$$

nella quale la riga j -ma è il vettore \mathbf{x}_j , mentre la colonna k -ma è l'insieme dei valori assunti dal carattere X_k . Si consiglia di consultare l'Appendice C.1 per gli opportuni richiami di Notazione vettoriale. Teoricamente, come nel caso $p = 2$ studiato nella Sezione 2.1, i vettori \mathbf{x}_j potrebbero essere rappresentati come una nuvola di n punti nello spazio p -dimensionale \mathbf{R}^p , ma in pratica una simile rappresentazione è impossibile e quindi saremo obbligati a sviluppare degli altri strumenti di analisi.

Definizione 2.3. Chiameremo **baricentro** dei dati il vettore $\bar{\mathbf{x}} = (\bar{x}_{.1}, \dots, \bar{x}_{.p}) \in \mathbf{R}^p$ le cui componenti sono le medie dei valori di ciascun carattere, ossia le medie lungo le colonne di $\|x_{jk}\|$

$$\bar{x}_{.k} = \frac{1}{n} \sum_{j=1}^n x_{jk}, \quad k = 1, \dots, p.$$

Chiameremo poi **matrice di covarianza** $p \times p$ la matrice $\mathbb{S} = \|s_{kl}\|$ i cui elementi sono le covarianze dei caratteri X_k e X_ℓ , ossia delle colonne k -ma e ℓ -ma di $\|x_{jk}\|$

$$s_{kl} = \frac{1}{n} \sum_{j=1}^n (x_{jk} - \bar{x}_{.k})(x_{j\ell} - \bar{x}_{.\ell}) = \frac{1}{n} \sum_{j=1}^n x_{jk}x_{j\ell} - \bar{x}_{.k}\bar{x}_{.\ell}, \quad k, \ell = 1, \dots, p.$$

Analogamente si chiama **matrice di correlazione** $p \times p$ la matrice $\mathbb{R} = \|r_{kl}\|$ i cui elementi sono i coefficienti di correlazione dei caratteri X_k e X_ℓ . Si chiama infine **dispersione totale** dei dati la quantità

$$\Delta = \frac{1}{n} \sum_{j=1}^n |\mathbf{x}_j - \bar{\mathbf{x}}|^2.$$

Il baricentro e le matrici di covarianza \mathbb{S} e di correlazione \mathbb{R} sono strumenti fondamentali nello studio di dati p -dimensionali. Val la pena notare a questo punto che le matrici \mathbb{S} e \mathbb{R} sono matrici simmetriche nel senso che

$$s_{kl} = s_{\ell k}, \quad r_{kl} = r_{\ell k};$$

infatti è ovvio che la correlazione di X_k e X_ℓ coincide con la correlazione di X_ℓ e X_k . Inoltre gli elementi diagonali della matrice di correlazione sono tutti uguali a 1 dato che si tratta delle correlazioni di ciascun carattere con se stesso.

Torniamo ora al problema di rappresentare graficamente i nostri dati p -dimensionali $\|x_{jk}\|$. Dato che non possiamo disegnare grafici in \mathbf{R}^p , proveremo a utilizzare delle *proiezioni* (si veda l'Appendice 2.1 per le definizioni di base) dei dati su rette o su piani bi-dimensionali passanti per l'origine di \mathbf{R}^p . Ma siccome è intuitivo che in questo modo si perde dell'informazione, e siccome la scelta della retta o del piano di proiezione è in linea di principio arbitraria, dovremo anche trovare in \mathbf{R}^p le direzioni lungo le quali la proiezione risulta più fedele. Infatti punti che in una

proiezione cadono vicini possono anche essere proiezioni di punti che nello spazio p -dimensionale sono molto lontani.

Iniziamo con il discutere il caso della *proiezione su una retta* individuata da un versore \mathbf{v} , ricordando che intuitivamente la proiezione di un punto $\mathbf{x} \in \mathbf{R}^p$ su una retta è determinata dal punto di questa retta che si trova alla minima distanza da \mathbf{x} , e che (vedi Appendice 2.1) il modulo del vettore proiettato è il prodotto scalare $\mathbf{x} \cdot \mathbf{v}$. Pertanto la proiezione riduce il campione di n vettori $\mathbf{x}_1, \dots, \mathbf{x}_n$ ad un campione di n numeri $\mathbf{x}_1 \cdot \mathbf{v}, \dots, \mathbf{x}_n \cdot \mathbf{v}$. In pratica possiamo anche considerare questi numeri

$$y_j = \mathbf{x}_j \cdot \mathbf{v} = \sum_{k=1}^p x_{jk} v_k, \quad j = 1, \dots, n$$

come i valori di un nuovo carattere $Y = v_1 X_1 + \dots + v_p X_p$ ottenuto come combinazione lineare dei caratteri originali X_1, \dots, X_p mediante dei coefficienti che sono le componenti del versore $\mathbf{v} = (v_1, \dots, v_p)$. Si noti che per un dato campione $\mathbf{x}_1, \dots, \mathbf{x}_n$ i valori di Y dipendono dalla scelta del versore \mathbf{v} , e il nostro compito sarà quello di determinare tale versore in modo che la rappresentazione proiettata sia la più fedele possibile, ovvero in modo che nell'unico carattere Y sia riassunta la più grande quantità possibile dell'informazione contenuta nei p caratteri X_1, \dots, X_p . Per determinare la direzione ottimale adotteremo il seguente criterio: *la proiezione su \mathbf{v} sarà tanto più fedele quanto più grande sarà la **dispersione** (cioè la varianza) dei valori y_j del nuovo carattere Y* . Infatti, siccome in una proiezione il rischio che si corre è quello di sovrapporre punti che nella realtà sono lontani fra loro, richiedere che la varianza dei punti proiettati sia la più grande possibile significa richiedere che queste proiezioni siano il più possibile lontane e distinte. Per mettere in pratica questo principio ci serviremo di alcuni risultati che ora enunceremo. Per il seguito (vedi anche Appendice 2.1) indicheremo con

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

gli autovalori (eventualmente anche coincidenti) della matrice di covarianza \mathbb{S} dei dati, e con $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ i corrispondenti autovettori ortonormali (vedi anche Appendice 2.1), ricordando che questo vuol dire che sono verificate le equazioni

$$\mathbb{S}\mathbf{v}_k = \lambda_k \mathbf{v}_k, \quad k = 1, \dots, p.$$

Teorema 2.4. *Con le notazioni precedenti, la media e la dispersione (varianza) del carattere Y sono*

$$m_Y(\mathbf{v}) = \bar{\mathbf{x}} \cdot \mathbf{v} = \sum_{k=1}^p \bar{x}_{.k} v_k, \quad s_Y^2(\mathbf{v}) = \mathbf{v} \cdot \mathbb{S}\mathbf{v} = \sum_{k,\ell=1}^p v_k s_{k\ell} v_\ell, \quad (2.3)$$

dove $\bar{\mathbf{x}}$ è il baricentro e \mathbb{S} è la matrice di covarianza dei caratteri originali X_1, \dots, X_p . Inoltre, se proiettiamo i dati lungo la direzione di uno degli autovettori \mathbf{v}_k , allora la dispersione di Y vale

$$s_Y^2(\mathbf{v}_k) = \lambda_k. \quad (2.4)$$

Dimostrazione: Per brevità dimostreremo solo la prima delle (2.3) e la (2.4): tenendo conto dell'equazione (C.1) si ha infatti

$$m_Y(\mathbf{v}) = \frac{1}{n} \sum_{j=1}^n y_j = \frac{1}{n} \sum_{j=1}^n (\mathbf{x}_j \cdot \mathbf{v}) = \left(\frac{1}{n} \sum_{j=1}^n \mathbf{x}_j \right) \cdot \mathbf{v} = \bar{\mathbf{x}} \cdot \mathbf{v},$$

che prova la prima delle (2.3). Inoltre, dalla definizione di autovettore e per le proprietà del prodotto scalare (vedi Appendice 2.1), applicando la seconda delle (2.3) si ha

$$s_Y^2(\mathbf{v}_k) = \mathbf{v}_k \cdot \mathbb{S} \mathbf{v}_k = \mathbf{v}_k \cdot (\lambda_k \mathbf{v}_k) = \lambda_k (\mathbf{v}_k \cdot \mathbf{v}_k) = \lambda_k |\mathbf{v}_k|^2 = \lambda_k,$$

il che prova anche la (2.4). Si noti che siccome per definizione la varianza $s_Y^2(\mathbf{v}_k)$ è positiva, questa relazione mostra anche che *gli autovalori della matrice di covarianza sono sempre positivi*. \square

Teorema 2.5. *Il vettore \mathbf{v} di modulo 1 per il quale la dispersione $s_Y^2(\mathbf{v})$ è massima coincide con l'autovettore \mathbf{v}_1 relativo all'autovalore più grande. Successivamente, limitandosi ai vettori \mathbf{v} di modulo 1 ortogonali a \mathbf{v}_1 , il vettore per il quale la dispersione è massima è \mathbf{v}_2 ; e così via per gli altri autovettori. Infine per la dispersione totale dei dati vale la relazione*

$$\Delta = \sum_{k=1}^p \lambda_k.$$

Il Teorema 2.5 dice in pratica che per rendere massima la dispersione del carattere Y bisogna calcolare la matrice di covarianza \mathbb{S} dei dati, determinarne autovalori ed autovettori, e infine scegliere di proiettare i dati lungo la direzione dell'autovettore \mathbf{v}_1 associato al più grande degli autovalori. Per conservare una maggiore quantità di informazione, poi, si possono proiettare i dati su un piano definito da due direzioni, e sempre il Teorema 2.5 ci dice di scegliere come seconda direzione l'autovettore \mathbf{v}_2 . Pertanto la proiezione più fedele ai dati è quella eseguita nel piano definito da \mathbf{v}_1 e \mathbf{v}_2 , cioè considerando i nuovi caratteri (Y_1, Y_2) con valori $(y_{j1}, y_{j2}) = (\mathbf{x}_j \cdot \mathbf{v}_1, \mathbf{x}_j \cdot \mathbf{v}_2)$ per $j = 1, \dots, n$. Ulteriore informazione si ottiene esaminando le proiezioni lungo gli altri autovettori, sempre privilegiando quelli con gli autovalori più grandi. La seconda parte del Teorema 2.5 infine ci suggerisce che ogni autovalore λ_k contribuisce alla dispersione totale Δ in proporzione al suo valore: osservazione coerente con il fatto che le direzioni privilegiate per la proiezione sono proprio quelle degli autovettori relative agli autovalori più grandi.

Definizione 2.4. *Chiameremo **direzioni o componenti principali** quelle degli autovettori ortonormali \mathbf{v}_k della matrice di covarianza \mathbb{S} , e ordinatamente diremo **prima direzione principale** quella di \mathbf{v}_1 , **seconda direzione principale** quella di \mathbf{v}_2 , e così via. I piani individuati dalle coppie di autovettori $(\mathbf{v}_k, \mathbf{v}_\ell)$ si chiamano poi **piani principali**, e in particolare il piano $(\mathbf{v}_1, \mathbf{v}_2)$ sarà il **primo piano principale**.*

Definizione 2.5. Chiameremo **fedeltà** della proiezione dei dati sul piano principale $(\mathbf{v}_k, \mathbf{v}_\ell)$ il rapporto

$$\frac{\lambda_k + \lambda_\ell}{\lambda_1 + \dots + \lambda_p};$$

siccome λ_1 e λ_2 sono gli autovalori più grandi, dal Teorema 2.5 segue allora che la massima fedeltà si ottiene proiettando i dati sul primo piano principale.

Bisogna notare che i dati originari del nostro problema $\|x_{jk}\|$ possono essere disomogenei per i loro ordini di grandezza. Supponiamo ad esempio di voler compilare una statistica relativa alle condizioni meteorologiche di una località registrando pressione atmosferica (in *mmHg*), temperatura (in $^{\circ}C$), velocità del vento (in *Km/h*) e copertura nuvolosa (in *ottavi* di cielo coperto). Per le unità di misura scelte le misure di pressione saranno numeri dell'ordine di 10^3 , mentre ad esempio la copertura nuvolosa sarà un numero intero da 1 a 8, e la temperatura un numero dell'ordine delle decine. La velocità del vento infine potrà variare da 0 fino a circa 10^2 . In queste condizioni le quantità rappresentate dai numeri più grandi assumerebbero un peso sproporzionato rispetto alle altre e senza una ragionevole motivazione. D'altra parte, siccome le unità di misura sono arbitrarie, questa osservazione mette anche in evidenza il fatto che è possibile modificare l'importanza relativa delle quantità osservate in un modo altrettanto arbitrario. Per eliminare questi effetti della scelta delle unità di misura di solito si preferisce *standardizzare* i dati originali $\|x_{jk}\|$, cioè li si sostituisce con

$$z_{jk} = \frac{x_{jk} - \bar{x}_{.k}}{s_{.k}}$$

dove ovviamente abbiamo posto

$$\bar{x}_{.k} = \frac{1}{n} \sum_{j=1}^n x_{jk}, \quad s_{.k}^2 = \frac{1}{n} \sum_{j=1}^n (x_{jk} - \bar{x}_{.k})^2.$$

I dati $\|z_{jk}\|$, infatti, avendo ora media nulla e varianza 1 sono stati ridotti ad una scala in cui sono tutti rappresentabili con numeri di grandezza comparabile; inoltre essi sono anche insensibili ai cambiamenti di unità di misura. Se ora ripetessimo la nostra analisi delle componenti principali a partire dai nuovi dati standardizzati ci accorgeremmo che la matrice di *covarianza* delle $\|z_{jk}\|$ non è nient'altro che la matrice di *correlazione* delle $\|x_{jk}\|$ che, come è noto, è insensibile ai cambiamenti di scala. In conclusione, per evitare i problemi derivanti dall'arbitrarietà delle unità di misura, è *sempre consigliabile eseguire l'analisi delle componenti principali basandosi sulla matrice di correlazione invece che su quella di covarianza*. In questo caso si ottengono risultati analoghi a quelli dei Teoremi 2.4 e 2.5, con la differenza che ora bisognerà calcolare autovalori e autovettori della matrice di correlazione invece che quelli della matrice di covarianza. Si può dimostrare, infine, che la somma degli autovalori della matrice di correlazione è sempre uguale al numero p dei caratteri X_1, \dots, X_p , e quindi anche che, in base al Teorema 2.5, la dispersione totale dei dati standardizzati è sempre uguale a p .

X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4
3.061	2.417	3.924	3.361	2.558	3.502	2.548	3.616
3.189	3.696	1.514	4.073	2.839	1.095	2.667	3.061
3.433	3.560	2.820	5.040	3.408	3.244	2.129	3.762
3.249	2.806	2.528	1.544	2.070	2.269	4.173	2.251
3.400	3.198	3.236	4.241	3.058	2.531	3.351	3.729
2.147	1.087	1.659	3.518	3.026	3.096	2.107	3.238
1.838	1.384	1.977	2.199	3.437	3.896	2.235	3.295
2.891	3.343	4.174	4.100	2.818	2.941	3.660	3.680
3.603	3.306	2.906	3.035	3.695	3.188	3.286	3.088
3.725	1.099	3.179	2.964	3.836	3.378	2.965	3.595
2.687	2.823	2.134	2.476	0.992	3.124	1.138	4.959
2.404	3.475	2.457	3.559	3.927	3.153	1.099	1.753
3.159	2.699	2.680	2.523	4.113	1.713	2.669	2.624
2.182	2.359	3.184	3.992	2.774	2.714	3.324	2.532
4.071	3.024	2.443	3.937	2.965	2.352	2.154	1.980
3.351	4.206	2.377	2.232	1.875	4.419	3.043	3.156
0.935	3.531	3.954	1.215	2.876	2.437	2.661	3.543
3.579	3.852	2.307	3.235	3.314	3.848	2.957	2.125
2.086	3.428	3.129	4.731	2.390	3.892	2.768	3.288
0.765	3.760	3.036	2.454	2.859	2.689	2.538	2.518
3.853	1.755	2.898	2.604	3.166	3.625	2.679	2.307
4.767	3.575	1.736	2.690	2.925	3.647	3.179	3.342
3.138	2.528	2.438	4.704	1.927	4.173	3.250	2.178
1.429	2.864	3.256	2.436	3.529	4.558	2.532	3.071
3.558	3.411	3.341	1.656	2.363	3.697	2.946	2.422
5.739	4.882	4.442	5.697	3.909	5.353	5.358	4.472
4.722	3.856	5.223	5.300	6.166	6.079	4.190	5.167
5.366	5.293	6.676	3.362	4.701	5.506	4.473	4.999
4.223	5.348	5.197	6.689	3.683	5.229	3.216	5.201
4.669	5.667	7.106	5.797	4.689	4.948	5.699	5.261
5.119	6.221	3.844	5.445	4.655	4.616	4.471	5.130
4.894	5.768	5.779	5.298	4.268	5.178	6.439	4.327
4.775	5.016	3.917	5.770	4.215	7.500	4.981	4.983
5.643	3.663	5.926	5.561	4.666	4.568	5.605	3.760
4.128	3.485	4.394	4.232	4.493	5.253	3.842	6.306
5.640	4.501	5.438	4.808	4.793	5.769	5.136	5.434
3.546	6.051	5.467	6.610	5.937	4.383	5.171	6.327
6.504	5.075	6.572	5.937	4.753	6.663	3.348	5.095
4.532	4.019	5.422	3.788	4.905	5.107	4.997	5.624
4.884	5.052	5.072	4.963	5.467	4.798	4.651	4.980
4.666	5.672	5.527	5.346	4.629	4.459	5.378	4.685
4.630	3.929	4.952	4.814	3.480	4.244	4.542	4.206
5.785	5.280	5.260	3.721	3.469	7.792	5.108	3.423
4.171	5.004	5.074	4.813	5.926	5.510	4.978	5.144
5.020	4.721	6.992	4.161	4.541	3.735	4.427	4.340
3.856	5.492	5.111	4.547	3.891	4.352	3.805	4.663
5.521	4.918	4.869	3.736	5.418	4.546	4.485	5.366
5.743	4.291	3.891	5.352	5.327	4.709	4.195	5.736
4.317	4.597	5.968	4.831	6.966	5.292	4.989	5.437
4.133	5.867	5.258	5.699	4.891	4.513	5.264	5.354

Tabella 2.3: Campione di $n = 100$ misure di quattro caratteri continui.

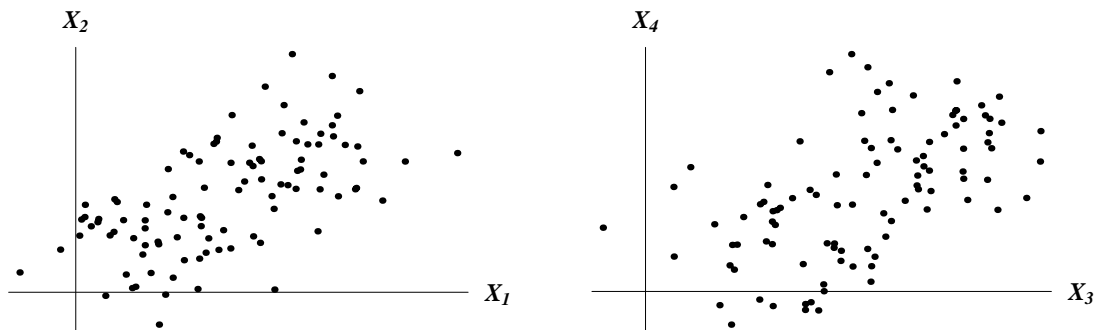


Figura 2.3: Rappresentazione di coppie di componenti dai dati della Tabella 2.3.

Esempio 2.2. Nella Tabella 2.3 sono riportate $n = 100$ misure di quattro caratteri numerici continui ($p = 4$) ottenute con una simulazione: esse potrebbero rappresentare le misure di quattro dimensioni fisiche di 100 animali di una data specie (altezza, lunghezza, ...), o rilevazioni di quattro parametri economici relativi a 100 paesi (popolazione, reddito pro capite, ...), o altro ancora. Ovviamente è impossibile rappresentare graficamente i punti corrispondenti perché questi si trovano in uno spazio a 4 dimensioni; si potrebbe però pensare di rappresentarne due componenti per volta sul corrispondente piano: un primo scopo di questa rappresentazione potrebbe essere quello di vedere se i dati mostrano la tendenza a raggrupparsi in due o più classi (*clusters*), indicando in questo modo una *classificazione* dei nostri 100 soggetti in base alle misure effettuate. Ad esempio gli animali della specie considerata potrebbero essere classificati in due o più razze sulla base delle quattro dimensioni fisiche considerate; oppure i 100 paesi potrebbero essere classificati in diversi livelli di sviluppo economico secondo i valori dei quattro indicatori rilevati.

Nella Figura 2.3 sono riportati come esempio i punti che si ottengono considerando prima le coordinate X_1, X_2 , e poi le altre due coordinate X_3, X_4 della Tabella 2.3. Queste immagini, pur mettendo in evidenza una certa correlazione fra i vari caratteri, non mostrano però nessuna evidente tendenza dei punti a raggrupparsi in classi con caratteristiche diverse. Altri grafici si potrebbero ottenere scegliendo altre coppie di coordinate, ed altri ancora se si considerano proiezioni su piani passanti per l'origine di \mathbf{R}^4 ma non coincidenti con gli originari piani coordinati: è possibile che con particolari scelte di questi piani si possano mettere in evidenza due o più classi di punti, ma è evidente che la scelta del piano migliore per una prima classificazione non può essere eseguita per tentativi, e deve avvalersi di una opportuna strategia di ricerca. Siccome il nostro problema è quello di separare delle classi, il miglior criterio sarà quello di scegliere la proiezione sul piano che rende massima la dispersione totale dei punti, e quindi faremo uso dell'analisi in componenti principali che abbiamo esposto nella presente Sezione.

L'analisi parte con il calcolo della matrice di correlazione \mathbb{R} dei dati della Tabella 2.3:

$$\mathbb{R} = \|r_{k\ell}\| = \begin{pmatrix} 1.000 & 0.606 & 0.719 & 0.620 \\ 0.606 & 1.000 & 0.599 & 0.600 \\ 0.719 & 0.599 & 1.000 & 0.560 \\ 0.620 & 0.600 & 0.560 & 1.000 \end{pmatrix}$$

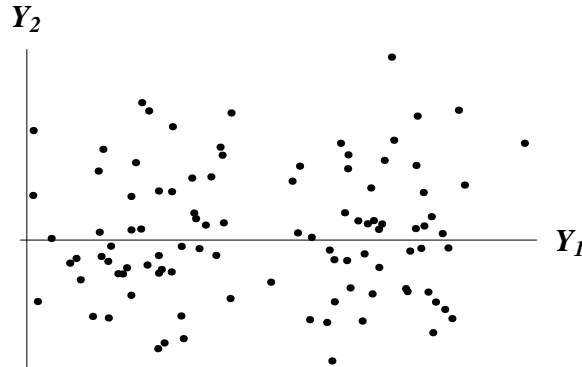


Figura 2.4: Rappresentazione dei dati della Tabella 2.3 nel primo piano principale.

e prosegue con il calcolo degli autovalori (ordinati):

$$\lambda_1 = 2.854, \quad \lambda_2 = 0.471, \quad \lambda_3 = 0.403, \quad \lambda_4 = 0.273,$$

e dei relativi autovettori (ortonormali):

$$\mathbf{v}_1 = \begin{pmatrix} -0.519 \\ -0.490 \\ -0.506 \\ -0.485 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} -0.369 \\ 0.372 \\ -0.580 \\ 0.624 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 0.224 \\ -0.784 \\ -0.024 \\ 0.578 \end{pmatrix}, \quad \mathbf{v}_4 = \begin{pmatrix} -0.738 \\ -0.079 \\ 0.638 \\ 0.204 \end{pmatrix}.$$

Sebbene per piccoli valori di p questi calcoli possano essere effettuati a mano, in generale la determinazione di autovalori e autovettori è affidata ad un computer. I valori dei λ_k mostrano innanzitutto che già la prima componente principale \mathbf{v}_1 ha una fedeltà di 0.713, e che il primo piano principale ha una fedeltà di 0.831; la restante quantità di informazione, pari a 0.169, si trova nelle altre due componenti \mathbf{v}_3 e \mathbf{v}_4 . Nel primo piano principale le coordinate (y_{j1}, y_{j2}) con $j = 1, \dots, n$ dei nostri n punti saranno allora

$$\begin{aligned} y_{j1} &= \mathbf{x}_j \cdot \mathbf{v}_1 = x_{j1}v_{11} + x_{j2}v_{12} + x_{j3}v_{13} + x_{j4}v_{14} \\ &= -0.519 x_{j1} - 0.490 x_{j2} - 0.506 x_{j3} - 0.485 x_{j4} \\ y_{j2} &= \mathbf{x}_j \cdot \mathbf{v}_2 = x_{j1}v_{21} + x_{j2}v_{22} + x_{j3}v_{23} + x_{j4}v_{24} \\ &= -0.369 x_{j1} + 0.372 x_{j2} - 0.580 x_{j3} + 0.624 x_{j4} \end{aligned}$$

La rappresentazione grafica delle (y_{j1}, y_{j2}) è riportata nella Figura 2.4 e mostra che in effetti è possibile separare i punti in due gruppi abbastanza distinti: in particolare è proprio la coordinata y_{j1} che maggiormente contribuisce a tale classificazione.

I valori e i segni delle componenti degli autovettori \mathbf{v}_k indicano quanto e in che verso i caratteri originari X_k contribuiscono alla combinazione che definisce i nuovi caratteri Y_k . Supponiamo ad esempio che le X_k siano misure di dimensioni fisiche di animali: il fatto che le componenti di \mathbf{v}_1 abbiano valori abbastanza vicini e tutti dello stesso segno indica che Y_1 è un carattere che distingue gli n individui in base al valore di tutte le dimensioni

fisiche considerate. In pratica Y_1 è una misura complessiva della *grandezza* dell'animale, e distingue gli individui in animali grandi e piccoli. Negli altri autovettori, invece, le componenti hanno segni differenti: questo indica che gli altri tre caratteri mettono in contrasto i valori delle diverse dimensioni misurate e sono quindi indicatori della *forma* dell'animale. In pratica essi distingueranno ad esempio gli individui in alti e corti, bassi e lunghi e così via. L'importanza che i diversi caratteri Y_k assumono nella classificazione è poi stabilita dal valore relativo degli autovalori λ_k .

Parte II
Probabilità

Capitolo 3

Spazi di probabilità

3.1 Spazio dei campioni

Il calcolo delle probabilità diviene uno strumento essenziale della statistica quando si considerano campioni estratti da una popolazione mediante procedure casuali. In questo caso, infatti, i calcoli non sono più effettuati su tutta la popolazione esistente, e le stime saranno soggette a variazioni aleatorie quando il campionamento viene ripetuto. Consideriamo inizialmente degli esempi di esperimenti che diano luogo solo ad un numero finito di possibili **risultati (o eventi elementari)** casuali.

Esempio 3.1. Il caso più semplice è quello del lancio di una moneta nel quale si osserva il verificarsi di uno dei due risultati possibili: 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** vuol dire che essa è non truccata, per cui nessuno dei due risultati è favorito rispetto all'altro ed è possibile attribuire loro le medesime possibilità di verificarsi; in tal caso diremo anche che i due eventi elementari T e C sono **equiprobabili**. Per dare una veste quantitativa a queste considerazioni si usa attribuire ad ogni evento elementare una **probabilità** intesa come frazione dell'unità, sicché nel nostro caso avremo:

$$p = \mathbf{P}(T) = \frac{1}{2}; \quad q = \mathbf{P}(C) = \frac{1}{2}.$$

Osserviamo che $p + q = 1$, dato che *con certezza* (ossia con probabilità eguale ad 1) uno dei due casi, T oppure C , si verifica, e non vi sono altre possibilità.

Esempio 3.2. Considerazioni analoghe a quelle dell'Esempio precedente applicate al caso di un dado *equo* conducono alla seguente attribuzione di probabilità per le sei facce che qui indicheremo con le cifre romane I, II, \dots, VI :

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

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

Da quanto precede si ricava che, almeno per casi semplici, si possono attribuire delle probabilità mediante una *enumerazione*. Questa idea è alla base della cosiddetta

definizione classica della probabilità: per attribuire una probabilità ad un evento A (in generale *non* elementare, cioè non ridotto ad un solo risultato) si enumerano i risultati *possibili* (ritenuti, in base a qualche ipotesi, equiprobabili), e quelli *favorevoli* all'evento A (quelli, cioè, che danno luogo al verificarsi di A), e si attribuisce ad A la probabilità:

$$\mathbf{P}(A) = \frac{\text{numero dei casi favorevoli}}{\text{numero dei casi possibili}}.$$

Notiamo che anche in questo caso la probabilità assegnata ad A è un numero positivo compreso fra 0 ed 1.

Esempio 3.3. Nel lancio di un dado equo consideriamo gli eventi (*non* elementari) $A =$ “appare una faccia contrassegnata da un numero pari”, $B =$ “appare una faccia contrassegnata da un multiplo di tre”, $C =$ “appare una faccia diversa da VI ”. Una semplice enumerazione in base alla definizione classica ci porta a concludere che, essendo 6 i casi possibili, e rispettivamente 3, 2 e 5 i casi favorevoli ad A, B e C , si avrà:

$$\mathbf{P}(A) = \frac{1}{2}; \quad \mathbf{P}(B) = \frac{1}{3}; \quad \mathbf{P}(C) = \frac{5}{6}.$$

Consideriamo ora un lancio di *due* dadi non truccati. È facile verificare che i risultati elementari possibili sono ora 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 si ha

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

Sempre per enumerazione si può verificare allora ad esempio che all'evento $A =$ “non appare la coppia (VI, VI) ” si può attribuire una probabilità $\mathbf{P}(A) = 35/36$.

Dalla discussione precedente segue che la probabilità di un evento può essere pensata come un numero compreso tra 0 ed 1: il valore 1 indica la *certezza* del verificarsi, e il valore 0 la sua *impossibilità*; i valori intermedi rappresentano tutti gli altri casi. Queste probabilità possono essere calcolate nei casi semplici mediante una enumerazione di risultati equiprobabili, ma questo metodo non può in nessun modo essere considerato come generale.

Definizione 3.1. *Chiameremo spazio dei campioni o spazio degli eventi elementari l'insieme Ω (finito o infinito) costituito da tutti i possibili risultati ω del nostro esperimento.*

Negli esempi precedenti lo spazio dei campioni $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$ era costituito da un numero finito di elementi. Ad esempio nel caso di un solo lancio di una moneta lo spazio dei campioni è composto di soli due elementi:

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

mentre nel caso di un solo lancio di un dado si ha

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

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

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

e così via. 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} .

3.2 Eventi

Definizione 3.2. Chiameremo *evento* ogni sottinsieme $A \subseteq \Omega$ del quale è possibile calcolare la probabilità.

Nel caso di tre lanci di una moneta lo spazio dei campioni è composto di $N = 2^3 = 8$ elementi:

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

e il sottinsieme

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

rappresenterà l'evento “ T appare almeno due volte su tre lanci”. Le osservazioni della Sezione 3.1 mostrano come calcolare la probabilità di tale evento. Gli eventi così definiti possono essere considerati come rappresentazioni di *proposizioni logiche*, e le corrispondenti operazioni tra eventi (intese come operazioni tra insiemi) possono essere considerate come un modello per i *connettivi logici* che uniscono delle proposizioni. Così, ad esempio, i connettivi *oppure* (OR) ed *e* (AND) sono rappresentati rispettivamente dalle operazioni di *unione* ed *intersezione*:

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

mentre il significato logico delle seguenti operazioni è facilmente deducibile tenendo presenti i diagrammi di Venn della Figura 3.1:

$$\begin{aligned} \bar{A} &= \{\omega : \omega \notin A\}; \\ A - B &= A \cap \bar{B} = \{\omega : \omega \in A, \text{ ma } \omega \notin B\}. \end{aligned}$$

Si noti che Ω rappresenta l'evento *certo* (nel senso che qualunque risultato cade per definizione in Ω), e \emptyset rappresenta l'evento *impossibile* (dato che nessun risultato appartiene a \emptyset). Diremo inoltre che i due eventi A e B sono *disgiunti* (o anche *incompatibili*) quando $A \cap B = \emptyset$ (cioè quando un risultato ω non può mai verificare contemporaneamente gli eventi A e B). Un evento può anche ridursi ad un solo elemento $A = \{\omega\}$, nel qual caso parleremo di evento *elementare*.

In generale non saremo interessati a considerare come eventi tutti i possibili sottinsiemi di Ω ; piuttosto si preferisce selezionare opportune famiglie di tali sottinsiemi da considerare come eventi. Bisogna però, per ragioni di coerenza, garantire che tali

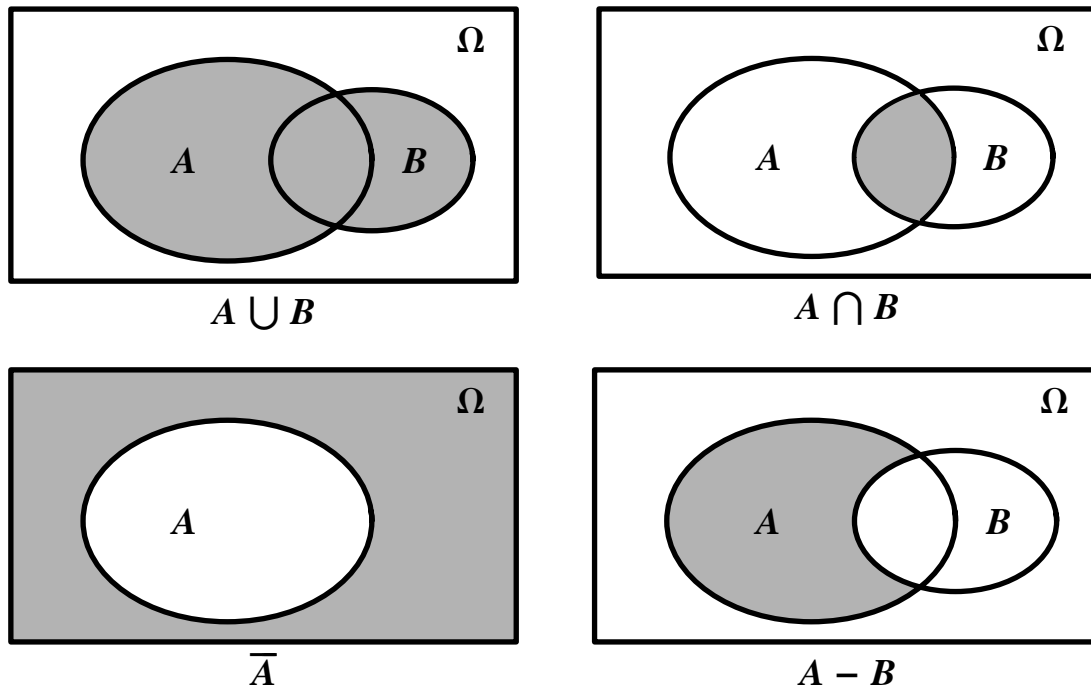


Figura 3.1: Le zone ombreggiate rappresentano i risultati delle operazioni insiemistiche indicate.

famiglie siano chiuse sotto le varie operazioni insiemistiche (logiche): ad esempio, se A e B sono due sottinsiemi della nostra famiglia degli eventi, anche la loro unione o intersezione deve appartenere alla famiglia degli eventi.

Definizione 3.3. Diremo che una famiglia \mathfrak{F} di parti di Ω costituisce un'algebra quando essa è chiusa sotto tutte le operazioni insiemistiche.

In particolare \bar{A} , $A \cap B$, $A \cup B$ e $A - B$ saranno tutti elementi di \mathfrak{F} se $A, B \in \mathfrak{F}$. Si vede facilmente, ad esempio, che dato un Ω arbitrario, e $A \in \Omega$ la seguente famiglia di parti di Ω

$$\mathfrak{F} = \{A, \bar{A}, \Omega, \emptyset\}.$$

è un'algebra detta algebra generata da A .

Definizione 3.4. Diremo che una famiglia \mathcal{D} di parti di Ω è una decomposizione di Ω se i suoi elementi D_k sono parti di Ω disgiunte e tali che $\bigcup_k D_k = \Omega$.

Una decomposizione non è un'algebra: essa, ad esempio, non contiene le unioni dei suoi elementi. In particolare, se $A \in \Omega$, la famiglia

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

è una semplice decomposizione di Ω . Le decomposizioni giocheranno un ruolo rilevante nel capitolo sul condizionamento.

3.3 Probabilità

La probabilità \mathbf{P} è una regola che consente di attribuire un peso probabilistico $\mathbf{P}(A)$ (un numero fra 0 e 1) ad ogni evento $A \in \mathfrak{F}$. Il modo in cui tale regola viene assegnata varia secondo la natura del problema considerato. In particolare, se Ω è un insieme finito di cardinalità $\#\Omega = N$ (numero dei *casi possibili*) e se i suoi elementi ω_k possono essere considerati equiprobabili, si può far ricorso alla *definizione classica* (vedi Sezione 3.1): si assegna ad ogni evento elementare la probabilità $\mathbf{P}\{\omega_k\} = 1/N$, e ad ogni evento $A \in \mathfrak{F}$ la probabilità

$$\mathbf{P}(A) = \frac{N_A}{N} \quad (3.1)$$

dove $N_A = \#A$ è la cardinalità di A , ossia il numero di elementi ω_k appartenenti ad A (numero dei *casi favorevoli*).

Esempio 3.4. (Problema delle coincidenze) 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. Il nostro spazio dei campioni Ω sarà allora formato dagli $N = M^n$ eventi elementari $\omega = (a_1, \dots, a_n)$ costituiti dalle n -uple di numeri estratti (con possibili *ripetizioni*). Supporremo che tali ω siano tutti equiprobabili. Consideriamo ora l'evento:

$$\begin{aligned} A &= \{\omega : \text{i valori delle } a_k \text{ sono tutti diversi}\} \\ &= \text{“nelle } n \text{ estrazioni non ci sono ripetizioni”} \end{aligned}$$

e calcoliamone la probabilità secondo la definizione classica. Un momento di riflessione ci convincerà del fatto che

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

per cui la probabilità richiesta è

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

Questo risultato permette 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? Il modello discusso in questo esempio ci permette di dare una risposta ponendo $M = 365$; in tal caso, essendo $\mathbf{P}(A)$ la probabilità che *tutti* i compleanni cadano in giorni differenti, si ha

$$p_n = 1 - \mathbf{P}(A) = 1 - \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \dots \left(1 - \frac{n-1}{365}\right)$$

In particolare si ottengono i seguenti sorprendenti risultati:

n	4	16	22	23	40	64	...
p_n	0.016	0.284	0.476	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. Inoltre se $n \geq 366$ avremo $p_n = 1$ e $\mathbf{P}(A) = 0$ dato che nel prodotto comparirà un fattore nullo: infatti con un numero di persone superiore al numero di date disponibili (365) le coincidenze diventano inevitabili. Osserviamo comunque che questi risultati sono meno sorprendenti se si riflette al fatto che essi sarebbero ben diversi se la domanda posta fosse la seguente: supponendo che io sia una delle n persone considerate nel problema precedente, quale è la probabilità q_n che almeno una celebri il suo compleanno nello stesso giorno in cui lo celebri io? Non entreremo nel dettaglio della soluzione di questo secondo problema, e ci limiteremo a riferire che nel secondo caso le probabilità delle coincidenze sono decisamente più piccole. Inoltre, per sottolinearne la differenza fra i due casi, noteremo che nel secondo q_n è sempre diversa da 1 (anche per $n \geq 366$) 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.

La formula (3.1) può anche essere generalizzata al caso in cui le ω_k non sono equiprobabili, ma hanno ognuna una probabilità $\mathbf{P}\{\omega_k\} = p_k$: la probabilità di un evento A sarà allora la somma delle p_k di tutti i risultati ω_k contenuti in A , cioè

$$\mathbf{P}(A) = \sum_{\omega_k \in A} p_k. \quad (3.2)$$

Le formule (3.1) e (3.2), nonostante la loro semplicità, consentono di trattare anche problemi di una certa sofisticazione, ma non possono essere adottate in situazioni più generali. I modelli *finiti* di probabilità si rivelano infatti ben presto insufficienti perché gli spazi dei campioni sono spesso insiemi infiniti e addirittura non numerabili. In questi casi la $\mathbf{P}(A)$ non può essere costruita secondo la definizione classica, ma deve essere data per altra via. Noi qui ricorderemo solo le proprietà generali che una probabilità deve sempre avere, riservandoci di discutere nei prossimi capitoli il modo in cui essa viene effettivamente calcolata nei casi di nostro interesse.

Definizione 3.5. *Data un'algebra \mathfrak{F} di eventi di Ω , chiameremo **probabilità** ogni applicazione $\mathbf{P} : \mathfrak{F} \rightarrow [0, 1]$ che sia **additiva**, cioè tale che, comunque scelti A e B eventi disgiunti di \mathfrak{F} , risulta*

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B), \quad \text{se } A \cap B = \emptyset \quad (3.3)$$

Elencheremo infine, senza dimostrazione, le proprietà più note delle probabilità

1. $\mathbf{P}(\emptyset) = 0, \quad \mathbf{P}(\Omega) = 1;$
2. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B), \quad \forall A, B \in \mathfrak{F};$
3. $\mathbf{P}(B) \leq \mathbf{P}(A) \quad \text{se } B \subseteq A, \quad \text{con } A, B \in \mathfrak{F};$

Definizione 3.6. *Chiameremo **spazio di probabilità** una terna $(\Omega, \mathfrak{F}, \mathbf{P})$ in cui Ω è un insieme detto spazio dei campioni, \mathfrak{F} è un'algebra di eventi di Ω , e \mathbf{P} è una probabilità su \mathfrak{F} .*

Capitolo 4

Condizionamento e indipendenza

4.1 Probabilità condizionata

Il condizionamento risponde all'esigenza di fondere una certa quantità di nuova informazione con l'informazione già contenuta in un dato spazio di probabilità $(\Omega, \mathfrak{F}, \mathbf{P})$. L'acquisizione di nuova informazione, infatti, modifica le nostre conoscenze e quindi ci permette di valutare la probabilità degli eventi in una maniera diversa da quella suggerita dalle nostre informazioni iniziali.

Esempio 4.1. Supponiamo di considerare una scatola contenente M palline delle quali m sono bianche ed $M - m$ nere ed eseguiamo due estrazioni successive. Se le palline sono estratte tutte con la medesima probabilità, e se la prima estrazione è effettuata *con rimessa*, è facile convincersi del fatto che l'evento $B =$ “alla seconda estrazione viene estratta una pallina bianca” si verifica con una probabilità $\frac{m}{M}$. Diversa sarebbe invece la nostra valutazione se la prima estrazione venisse effettuata *senza rimessa*: la probabilità di estrarre una pallina bianca *sapendo che* in precedenza ne è stata estratta un'altra bianca sarebbe $\frac{m-1}{M-1}$; se invece in precedenza fosse stata estratta una pallina nera si avrebbe $\frac{m}{M-1}$.

Definizione 4.1. Dato uno spazio di probabilità $(\Omega, \mathfrak{F}, \mathbf{P})$ e due eventi $A, B \in \mathfrak{F}$ con $\mathbf{P}(A) \neq 0$, chiameremo **probabilità condizionata** di B rispetto ad A (cioè probabilità che si verifichi B sapendo che si è verificato A) la quantità

$$\mathbf{P}(B | A) = \frac{\mathbf{P}(B \cap A)}{\mathbf{P}(A)}.$$

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

Si controlla facilmente che l'applicazione $\mathbf{P}(\cdot | A) : \mathfrak{F} \rightarrow [0, 1]$ così definita non è altro che una nuova probabilità. Va inoltre ricordato che il simbolo $\mathbf{P}(\cdot | \cdot)$ non è simmetrico nei suoi due argomenti: in generale $\mathbf{P}(B | A) \neq \mathbf{P}(A | B)$.

Teorema 4.1. (Formula della Probabilità Totale): *Dati un evento A e una decomposizione $\mathcal{D} = \{D_1, \dots, D_n\}$ con $\mathbf{P}(D_i) \neq 0$, $i = 1, \dots, n$, risulta sempre*

$$\mathbf{P}(A) = \sum_{i=1}^n \mathbf{P}(A \mid D_i) \mathbf{P}(D_i).$$

Dimostrazione: Basterà osservare che

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^n D_i \right) = \bigcup_{i=1}^n (A \cap D_i),$$

che gli eventi $A \cap D_i$ sono tutti disgiunti, e che quindi per l'additività di \mathbf{P} e per la Definizione 4.1

$$\mathbf{P}(A) = \mathbf{P} \left(\bigcup_{i=1}^n (A \cap D_i) \right) = \sum_{i=1}^n \mathbf{P}(A \cap D_i) = \sum_{i=1}^n \mathbf{P}(A \mid D_i) \mathbf{P}(D_i).$$

Osserviamo che in particolare, se la decomposizione si riduce a $\mathcal{D} = \{B, \bar{B}\}$, la formula della Probabilità Totale diviene

$$\mathbf{P}(A) = \mathbf{P}(A \mid B) \mathbf{P}(B) + \mathbf{P}(A \mid \bar{B}) \mathbf{P}(\bar{B}),$$

espressione particolarmente facile da usare ed interpretare. \square

Esempio 4.2. Riprendiamo in considerazione la scatola di palline dell'Esempio 4.1, estraiamo in successione e senza rimessa due palline e, senza guardare la prima, chiediamoci quale è la probabilità che la seconda sia bianca. Definiamo, a questo scopo, gli eventi

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

si ha ovviamente che

$$\mathbf{P}(A) = \frac{m}{M}, \quad \mathbf{P}(\bar{A}) = \frac{M-m}{M}.$$

Inoltre, nel caso di due estrazioni successive, enumerando i casi possibili e i casi favorevoli agli eventi presi in considerazione, otteniamo facilmente che

$$\mathbf{P}(B \mid A) = \frac{m-1}{M-1}, \quad \mathbf{P}(B \mid \bar{A}) = \frac{m}{M-1},$$

per cui si ha, tenendo conto del Teorema 4.1,

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

La probabilità di B è dunque diversa secondo le informazioni disponibili, e in particolare essa non è influenzata dal risultato della prima estrazione quando questo è sconosciuto: infatti $\mathbf{P}(B) = \mathbf{P}(A)$, mentre $\mathbf{P}(B \mid A)$ e $\mathbf{P}(B \mid \bar{A})$ sono diversi da $\mathbf{P}(A)$.

Teorema 4.2. (Formula di Bayes): Dati due eventi A, B con $\mathbf{P}(A) \neq 0, \mathbf{P}(B) \neq 0$, 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 di Ω con $\mathbf{P}(D_i) \neq 0, i = 1, \dots, n$, risulta anche

$$\boxed{\mathbf{P}(D_i | B) = \frac{\mathbf{P}(B | D_i) \mathbf{P}(D_i)}{\sum_{j=1}^n \mathbf{P}(B | D_j) \mathbf{P}(D_j)}}. \quad (4.1)$$

Dimostrazione: La dimostrazione della prima relazione si basa sul fatto che per definizione di probabilità condizionata si ha

$$\mathbf{P}(B | A) \mathbf{P}(A) = \mathbf{P}(A \cap B) = \mathbf{P}(A | B) \mathbf{P}(B);$$

la seconda relazione si ottiene poi dalla prima tramite il Teorema 4.1. \square

Nelle applicazioni statistiche gli eventi D_i del Teorema di Bayes sono spesso chiamati *ipotesi* e $\mathbf{P}(D_i)$ *probabilità a priori* di tali ipotesi, mentre le probabilità condizionate $\mathbf{P}(D_i | B)$ si chiamano *probabilità a posteriori*. Per comprendere il significato di questa terminologia con la discussione di un esempio (vedi Esempio 4.4) converrà introdurre prima un altro importante concetto: quello di *indipendenza*.

4.2 Indipendenza

Due eventi sono indipendenti quando il verificarsi di uno di essi non ha alcun effetto sul valore della probabilità che viene attribuita all'altro. Sulla base del concetto di probabilità condizionata introdotto prima diremo quindi che l'evento A è indipendente dall'evento B quando $\mathbf{P}(A | B) = \mathbf{P}(A)$ e quindi, per la Definizione 4.1, se $\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B)$. È facile inoltre, data la simmetria di queste relazioni, verificare che se A è indipendente da B , anche B è indipendente da A .

Definizione 4.2. Dato uno spazio di probabilità $(\Omega, \mathfrak{F}, \mathbf{P})$, diremo che A e B sono *eventi indipendenti* quando

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B).$$

Il concetto di indipendenza può essere esteso anche al caso in cui il numero di eventi è maggiore di due, ma bisogna notare che sarà ora possibile parlare di indipendenza *due a due*, nel senso di $\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B)$, di indipendenza *tre a tre*, nel senso di $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A) \mathbf{P}(B) \mathbf{P}(C)$, e così via. Questi diversi *livelli di indipendenza*, però, non si implicano l'uno con l'altro: infatti, ad esempio, tre eventi possono essere indipendenti due a due senza esserlo tre a tre e viceversa.

Definizione 4.3. n eventi si dicono **indipendenti** se essi sono indipendenti in tutti i modi possibili, cioè due a due, tre a tre, \dots , n a n .

Esempio 4.3. Consideriamo una moneta con

$$\mathbf{P}(T) = p, \quad \mathbf{P}(C) = 1 - p$$

e calcoliamo la probabilità che in n lanci esca un numero fissato k di teste in un ordine qualsiasi. I risultati ω del nostro esperimento saranno delle n -ple di simboli T e C , ed è ragionevole fare l'ipotesi che i lanci siano indipendenti. Pertanto se ω è una n -pla che contiene k teste e $n - k$ croci, dall'indipendenza avremo

$$\mathbf{P}\{\omega\} = \underbrace{\mathbf{P}(T) \cdot \dots \cdot \mathbf{P}(T)}_{k \text{ volte}} \cdot \underbrace{\mathbf{P}(C) \cdot \dots \cdot \mathbf{P}(C)}_{n-k \text{ volte}} = p^k (1 - p)^{n-k}. \quad (4.2)$$

Siccome però le k teste possono essere disposte in vario modo nella n -pla ω , l'evento $A =$ "escono k teste su n lanci" sarà composto di un certo numero di eventi elementari tutti con la stessa probabilità (4.2). Si può dimostrare – ma noi trascureremo di farlo – che il numero di queste diverse combinazioni è

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

e quindi con una banale applicazione della formula (3.2) si ottiene il risultato cercato

$$\mathbf{P}(A) = \binom{n}{k} p^k (1 - p)^{n-k}. \quad (4.3)$$

Si comprende facilmente che questo risultato resta invariato quale che sia il significato concreto dell'evento A ; in altri termini potremo dire che (4.3) rappresenta la *probabilità di avere k successi in n tentativi indipendenti di verifica di un generico evento A con $\mathbf{P}(A) = p$.*

Esempio 4.4. Supponiamo ora che la probabilità p della moneta dell'Esempio 4.3 sia sconosciuta e che, per semplificare il problema, i possibili valori di p siano solo due: $\frac{1}{2}$ e $\frac{2}{3}$. In assenza di altre informazioni riterremo che questi due valori siano equiprobabili, ma è facile capire che l'osservazione dei risultati di un certo numero di lanci della moneta potrà darci delle indicazioni preziose. In particolare l'osservazione di un eccessivo numero di teste ci farebbe intuitivamente propendere per il valore più grande di p , e viceversa nel caso contrario. La Formula di Bayes (4.1) ci permette ora di dare veste quantitativa precisa a queste considerazioni qualitative. Per fissare le idee supponiamo di aver lanciato la moneta $n = 10$ volte e di aver osservato $k = 4$ teste e $n - k = 6$ croci; definiamo poi, con le notazioni del Teorema 4.2, i seguenti eventi

$$\begin{aligned} D_1 &= \text{“la probabilità che esca testa è } p = \frac{1}{2}\text{”} \\ D_2 &= \text{“la probabilità che esca testa è } p = \frac{2}{3}\text{”} \\ B &= \text{“su } n = 10 \text{ lanci escono } k = 4 \text{ teste”} \end{aligned}$$

È evidente che $\mathcal{D} = \{D_1, D_2\}$ è una decomposizione di Ω , e che in assenza di altre informazioni le probabilità *a priori* delle due ipotesi sono

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

D'altra parte da (4.3) è facile vedere che

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

per cui dalla formula di Bayes (4.1) otteniamo:

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

Come si noterà, l'aver osservato un numero relativamente scarso di teste favorisce l'ipotesi con il valore di p più piccolo. Ulteriori osservazioni, comunque, condurranno a modificare questa valutazione anche se, a lungo andare, ci si attende intuitivamente una stabilizzazione del risultato.

Capitolo 5

Variabili aleatorie

5.1 Variabili aleatorie

In generale lo spazio dei campioni Ω non è necessariamente un insieme numerico nel senso che i suoi elementi sono oggetti astratti: ad esempio nel caso della moneta gli elementi di Ω sono T e C . D'altra parte nelle applicazioni statistiche gli aspetti più rilevanti sono legati ai valori numerici che si ottengono dalle misure. È estremamente importante quindi introdurre delle procedure che consentano di associare dei numeri ai risultati dei nostri esperimenti aleatori:

Definizione 5.1. Dato uno spazio di probabilità $(\Omega, \mathfrak{F}, \mathbf{P})$ si dice **variabile aleatoria (v.a.)** una applicazione $X : \Omega \rightarrow \mathbf{R}$ tale che tutti gli insiemi di Ω del tipo

$$\{X \in J\} = \{\omega : X(\omega) \in J\}$$

con J arbitrario intervallo di \mathbf{R} , sono eventi di \mathfrak{F} (vedi anche Figura 5.1).

Esempio 5.1. Si consideri un dado con le facce colorate con sei colori diversi: in questo caso Ω è costituito dall'insieme dei sei colori scelti. Si supponga poi di stabilire le regole di

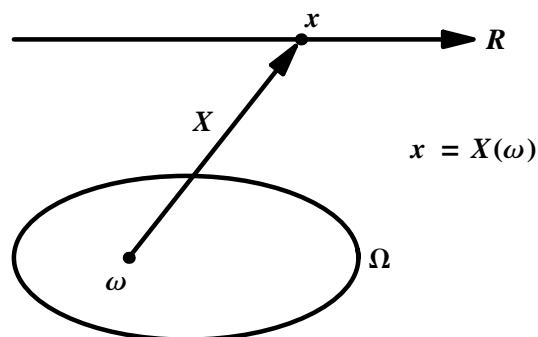


Figura 5.1: Illustrazione grafica della definizione di variabile aleatoria.

un gioco nel quale ad ogni colore è associata una vincita in denaro: la regola che attribuisce la vincita ad ogni colore è una v.a. Un altro esempio semplice, ma molto rilevante di v.a. è costituito dall'**indicatore** $I_A(\omega)$ di un evento $A \in \mathfrak{F}$:

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

In pratica si tratta della v.a. che vale 1 per tutti i risultati che verificano A , e 0 in tutti gli altri casi.

La Definizione 5.1 è fondamentale perché consente di associare una probabilità agli intervalli J di \mathbf{R} : in pratica la v.a. X proietta sull'insieme \mathbf{R} un'immagine P_X della probabilità \mathbf{P} definita su Ω ; questa nuova probabilità è definita da

$$P_X(J) = \mathbf{P}\{X \in J\} \quad (5.1)$$

e permette di attribuire una probabilità agli *insiemi di valori* (intervalli di \mathbf{R}) di X .

Definizione 5.2. Chiameremo **legge o distribuzione** della v.a. X la probabilità P_X da essa definita su \mathbf{R} tramite la (5.1).

Si noti che ogni v.a. attribuisce al medesimo intervallo una probabilità diversa, cioè v.a. diverse hanno, in generale, leggi diverse. Niente vieta però che ci siano v.a. diverse (nel senso della Definizione 5.1) con la medesima legge: in questo caso si parla di **v.a. identicamente distribuite (i.d.)**

Esempio 5.2. Sia $A \in \mathfrak{F}$ un evento con $\mathbf{P}(A) = p$, e sia I_A il suo indicatore. La v.a. I_A assume solo i due valori 0 e 1: pertanto eventi del tipo $\{I_A \in [2, 4]\}$, oppure $\{I_A \in (-\infty, -3]\}$ non capitano mai, cioè $\{I_A \in [2, 4]\} = \{I_A \in (-\infty, -3]\} = \emptyset$, e quindi

$$P_{I_A}[2, 4] = P_{I_A}(-\infty, -3] = 0.$$

Viceversa si vede facilmente ad esempio che $\{I_A \in [\frac{1}{2}, 2]\} = A$, $\{I_A \in [-\frac{1}{2}, \frac{1}{2}]\} = \bar{A}$, e infine $\{I_A \in [-2, 2]\} = \Omega$, per cui

$$P_{I_A}\left[\frac{1}{2}, 2\right] = p, \quad P_{I_A}\left[-\frac{1}{2}, \frac{1}{2}\right] = 1 - p, \quad P_{I_A}[-2, 2] = 1.$$

Esempio 5.3. Per costruire un esempio di v.a. distinte ma i.d. si consideri un dado equo e si definiscano le due v.a.

$$X = \begin{cases} 1 & \text{se esce una faccia } \textit{pari}, \\ 0 & \text{altrimenti.} \end{cases} \quad Y = \begin{cases} 1 & \text{se esce } 1, 2 \text{ oppure } 3, \\ 0 & \text{altrimenti.} \end{cases}$$

X ed Y sono ovviamente v.a. diverse: ad esempio se esce 1 X prende valore 0, mentre Y vale 1. Ciononostante esse sono i.d. In effetti X ed Y sono gli indicatori di due eventi, rispettivamente $A = \text{"esce una faccia } \textit{pari}"}$ e $B = \text{"esce } 1, 2 \text{ oppure } 3"$, che pur essendo diversi hanno la stessa probabilità $\frac{1}{2}$. Esse pertanto assumono gli stessi valori, 1 e 0, su due eventi equiprobabili. Ragionando come nell'Esempio 5.2 si può allora mostrare che esse attribuiscono la stessa probabilità a tutti gli intervalli di \mathbf{R} .

Definizione 5.3. Diremo che X ed Y sono **v.a. indipendenti** se tutti gli eventi del tipo $\{X \in J_1\}$ e $\{Y \in J_2\}$ sono indipendenti comunque scelti gli intervalli J_1 e J_2 ; cioè se

$$\mathbf{P}\{X \in J_1, Y \in J_2\} = \mathbf{P}(\{X \in J_1\} \cap \{Y \in J_2\}) = \mathbf{P}\{X \in J_1\} \mathbf{P}\{Y \in J_2\}.$$

comunque scelti gli intervalli J_1 e J_2 .

Nelle sezioni seguenti esamineremo gli strumenti matematici che ci permetteranno di costruire le distribuzioni delle v.a. di uso più comune, e a questo scopo distingueremo le v.a. in due grandi categorie, le *v.a. discrete* e le *v.a. continue*, che esauriscono tutti i casi di maggiore interesse pratico. Per il momento ci limiteremo ad introdurre un concetto comune ad ambedue le categorie che sarà molto utile in seguito.

Definizione 5.4. Chiameremo **funzione di distribuzione (FD)** di una v.a. X la funzione definita come

$$F_X(x) = \mathbf{P}\{X \leq x\} = \mathbf{P}\{X \in (-\infty, x]\} = P_X(-\infty, x], \quad x \in \mathbf{R}.$$

Teorema 5.1. La FD $F_X(x)$ di una v.a. X gode delle seguenti proprietà:

- $0 \leq F_X(x) \leq 1$, per ogni $x \in \mathbf{R}$;
- $F_X(x)$ è una funzione non decrescente di x ;
- $F_X(x)$ tende a 0 per $x \rightarrow -\infty$, e tende a 1 per $x \rightarrow +\infty$;
- la probabilità attribuita da P_X agli intervalli semi-aperti del tipo $(a, b]$ si calcola dalla FD mediante la formula

$$\boxed{\mathbf{P}\{a < X \leq b\} = P_X(a, b] = F_X(b) - F_X(a)}. \quad (5.2)$$

In alcuni problemi sarà necessario associare ad ogni $\omega \in \Omega$ non un solo numero, ma un intero vettore di m numeri, per cui X sarà in realtà una applicazione da Ω in \mathbf{R}^m . In pratica questo equivale a definire m v.a. X_1, \dots, X_m che costituiscono le componenti del vettore X : sarà opportuno pertanto introdurre anche la seguente Definizione:

Definizione 5.5. Chiameremo **variabile aleatoria m -dimensionale (o vettore aleatorio)** un vettore $X = (X_1, \dots, X_m)$ le cui componenti X_j sono v.a. nel senso della Definizione 5.1.

La Definizione 5.5 consente di associare una probabilità ai rettangoli $B = J_1 \times \dots \times J_m$ di \mathbf{R}^m (dove J_i sono intervalli di \mathbf{R}): in pratica la v.a. m -dimensionale X proietta su \mathbf{R}^m una probabilità P_X definita da

$$P_X(B) = \mathbf{P}\{X_1 \in J_1, \dots, X_m \in J_m\}, \quad (5.3)$$

dove $\{X_1 \in J_1, \dots, X_m \in J_m\}$ è una abbreviazione che indica l'evento $\{X_1 \in J_1\} \cap \dots \cap \{X_m \in J_m\}$.

Definizione 5.6. Chiameremo **legge o distribuzione congiunta** del vettore X la probabilità P_X da essa definita su \mathbf{R}^m tramite la (5.3). Le leggi delle singole componenti X_i si chiamano invece **leggi o distribuzioni marginali**

Definizione 5.7. Chiameremo **funzione di distribuzione congiunta** di un vettore $X = (X_1, \dots, X_m)$ la funzione

$$F_X(x_1, \dots, x_m) = \mathbf{P}\{X_1 \leq x_1, \dots, X_m \leq x_m\},$$

e **funzioni di distribuzione marginali** le funzioni

$$F_{X_j}(x_j) = \mathbf{P}\{X_j \leq x_j\}, \quad j = 1, \dots, m$$

cioè le FD delle singole componenti X_j .

Teorema 5.2. Le componenti di un vettore aleatorio $X = (X_1, \dots, X_m)$ sono indipendenti se e solo se

$$F_X(x_1, \dots, x_m) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_m}(x_m)$$

cioè se la FD congiunta si fattorizza nel prodotto delle FD marginali.

Nel seguito saremo spesso interessati a considerare v.a. ottenute come *funzioni, somme o altre combinazioni di v.a.* : ad esempio, data la v.a. X , possiamo considerare X^2 , oppure $\cos X$, o altre funzioni di X ; analogamente date due (o più) v.a. X e Y potremo essere interessati alla v.a. $X + Y$, oppure XY e così via. Le nuove v.a. in quanto funzioni da Ω in \mathbf{R} sono definite con opportune composizioni di funzioni, e vi sono tecniche particolari che consentono di ricavarne le leggi a partire dalle leggi delle v.a. iniziali. Noi non entreremo in questi dettagli, ma ci limiteremo, ove necessario, a ricordare i risultati più importanti senza dimostrazioni.

5.2 Variabili aleatorie discrete

Definizione 5.8. Chiameremo **v.a. discrete** le v.a. X che assumono solo un insieme finito (o infinito numerabile) di valori che indicheremo con x_k con $k = 1, \dots, n$ (oppure $n \in \mathbf{N}$).

Spesso i valori x_k coincideranno proprio con i numeri interi $1, 2, \dots$. L'indicatore I_A di un evento A è ovviamente una v.a. discreta che assume solo i due valori 0 e 1. È facile capire che la legge di una v.a. discreta X è determinata non appena siano assegnate le probabilità (con $p(x) \neq 0$ solo per $x = x_k$)

$$p_k = p(x_k) = \mathbf{P}\{X = x_k\}, \quad k = 1, 2, \dots$$

dove l'evento $\{X = x_k\}$ è il sottinsieme delle $\omega \in \Omega$ per le quali $X(\omega) = x_k$. Adotteremo la notazione più opportuna secondo i casi esaminati. L'assegnazione dei numeri p_k può essere arbitraria purché essi soddisfino le due seguenti proprietà:

$$p_k \geq 0, \quad k = 1, 2, \dots; \quad \sum_k p_k = 1. \quad (5.4)$$

Si noti che la somma nella relazione precedente è in realtà una serie se la v.a. X assume un insieme infinito di valori. Quando le p_k sono note anche la legge P_X è nota: dato un arbitrario intervallo J , la probabilità $P_X(J) = \mathbf{P}\{X \in J\}$ si otterrà sommando le p_k relative ai valori x_k che cadono in J .

La FD di una v.a. discreta X è una *funzione a scalini*: essa presenta delle discontinuità nei valori x_k assunti da X , e rimane costante fra due valori consecutivi x_k e x_{k+1} . Inoltre l'altezza del salto effettuato in ogni discontinuità x_k coincide proprio con la probabilità che X assuma il valore x_k per cui si ha

$$p_k = p(x_k) = \mathbf{P}\{X = x_k\} = F_X(x_k) - F_X(x_{k-1}). \quad (5.5)$$

Negli esempi seguenti esamineremo alcune importanti leggi di v.a. discrete che assumono solo valori interi $0, 1, 2, \dots$. In questi esempi mostreremo anche come si possono rappresentare graficamente le leggi delle v.a. Si noti che in realtà ogni esempio tratterà non una sola legge, ma una intera *famiglia di leggi* caratterizzate da distribuzioni che differiscono fra loro solo per il valore di uno o più parametri: ad esempio le leggi Binomiali $B(n, p)$ sono classificate dai due parametri n intero e $p \in [0, 1]$; le leggi di Poisson $\mathcal{P}(\lambda)$ sono invece classificate da un solo parametro $\lambda > 0$, e così via. La stessa osservazione si applicherà al caso delle v.a. continue nella successiva Sezione 5.3.

Esempio 5.4. (Legge di Bernoulli) Si dice che una v.a. X è distribuita secondo la legge di Bernoulli (o che è una v.a. di Bernoulli) quando essa assume i seguenti valori

$$X = \begin{cases} 1 & \text{con probabilità } p, \\ 0 & \text{con probabilità } 1 - p. \end{cases}$$

con $0 \leq p \leq 1$. In altri termini si ha

$$p_0 = \mathbf{P}\{X = 0\} = 1 - p, \quad p_1 = \mathbf{P}\{X = 1\} = p. \quad (5.6)$$

È evidente che ogni indicatore I_A di un evento A con $\mathbf{P}(A) = p$ è una v.a. di Bernoulli. Infatti

$$\mathbf{P}\{I_A = 0\} = \mathbf{P}(\bar{A}) = 1 - p, \quad \mathbf{P}\{I_A = 1\} = \mathbf{P}(A) = p.$$

Nella Figura 5.2 è mostrato prima di tutto il grafico della FD di una legge di Bernoulli: esso presenta due discontinuità in 0 e 1; inoltre le altezze dei due salti coincidono proprio con le probabilità $1 - p$ e p che X prenda rispettivamente i valori 0 e 1. Sempre nella Figura 5.2 sono rappresentati con un grafico a barre i valori delle probabilità p e $1 - p$ che X prenda rispettivamente i valori 1 e 0.

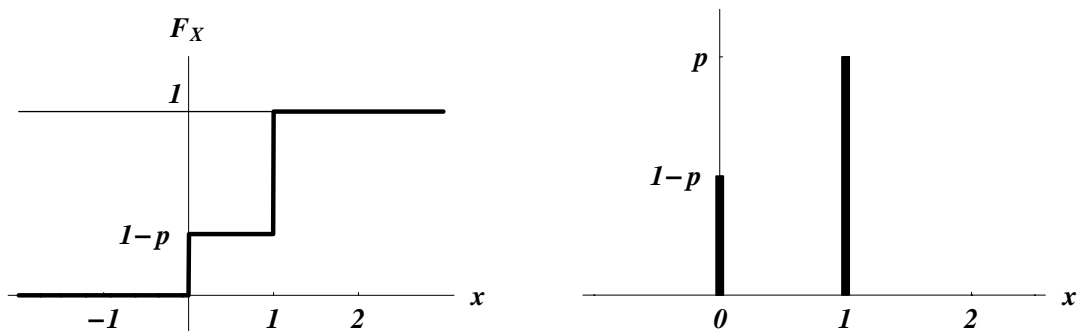


Figura 5.2: FD e grafico a barre di una legge di Bernoulli.

Esempio 5.5. (Legge Binomiale) Diremo che una v.a. X segue la legge Binomiale $B(n, p)$ con $n = 1, 2, \dots$ e $p \geq 0$ quando essa assume gli $n + 1$ valori $0, 1, \dots, n$ con le seguenti probabilità

$$p_k = \mathbf{P}\{X = k\} = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n \quad (5.7)$$

È facile verificare – ma noi trascureremo di farlo – che la somma di queste p_k vale esattamente 1 quale che sia il valore di n e p . Nella Figura 5.3 è rappresentato il grafico della FD della legge $B(n, p)$: esso presenta $n + 1$ discontinuità nei punti $0, 1, \dots, n$, rimane costante fra due successive discontinuità, vale 0 per $x < 0$ e 1 per $x \geq n$; inoltre l'altezza di ogni salto nel punto k coincide con la probabilità p_k . Nella medesima figura è rappresentato anche il grafico a barre dei valori p_k . L'andamento di questi grafici ovviamente cambia al variare dei valori di n e p : in particolare il grafico a barre è simmetrico quando $p = \frac{1}{2}$; viceversa esso presenta le p_k più grandi verso i valori più (rispettivamente meno) elevati di k se p è prossimo a 1 (rispettivamente a 0). Da un confronto con la formula (4.3) dell'Esempio 4.3 discende quindi che *il numero di successi in n tentativi indipendenti di verifica di un evento A con $\mathbf{P}(A) = p$ è una v.a. binomiale $B(n, p)$* . Si verifica anche facilmente che la legge di Bernoulli per un dato p dell'Esempio 5.4 non è altro che la legge Binomiale nel caso di $n = 1$; pertanto nel seguito indicheremo le leggi di Bernoulli con il simbolo $B(1, p)$. In un esperimento consistente in n tentativi indipendenti di verifica di un dato evento A con $\mathbf{P}(A) = p$ potremo definire da un lato le n v.a. di Bernoulli indipendenti X_1, \dots, X_n che assumono valore 1 se A si verifica e 0 in caso contrario, e dall'altro la v.a. X con legge Binomiale $B(n, p)$ che rappresenta il numero di successi sugli n tentativi. È intuitivo che fra queste v.a. sussista la relazione $X = X_1 + \dots + X_n$: questo risultato è richiamato nel seguente Teorema che non dimosteremo.

Teorema 5.3. *Se n v.a. indipendenti X_1, \dots, X_n sono tutte identicamente distribuite secondo la legge di Bernoulli $B(1, p)$, la loro somma $X = X_1 + \dots + X_n$ è distribuita secondo la legge Binomiale $B(n, p)$. Viceversa, ogni v.a. X Binomiale $B(n, p)$ può essere considerata come somma di opportune v.a. X_1, \dots, X_n indipendenti e di Bernoulli $B(1, p)$.*

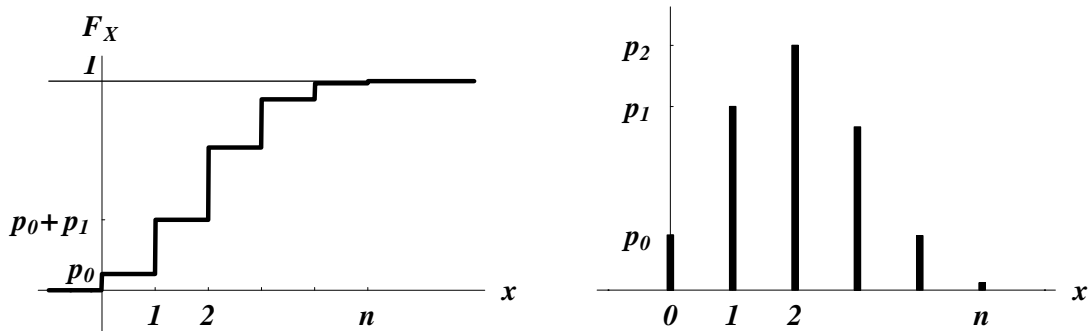


Figura 5.3: FD e grafico a barre di una legge Binomiale $B(n, p)$.

Esempio 5.6. (Legge di Poisson) Diremo che una v.a. X segue la legge di Poisson $\mathcal{P}(\lambda)$ con $\lambda > 0$ quando essa assume tutti i valori interi $k \in \mathbf{N}$ con le seguenti probabilità

$$p_k = \mathbf{P}\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots \quad (5.8)$$

È facile verificare – ma noi trascureremo di farlo – che la somma di queste infinite p_k vale esattamente 1 quale che sia il valore di λ . Nella Figura 5.4 è rappresentato il grafico della FD della legge $\mathcal{P}(\lambda)$: esso presenta infinite discontinuità nei punti $0, 1, 2, \dots$, rimane costante fra due successive discontinuità, vale 0 per $x < 0$ e tende verso 1 per $x \rightarrow +\infty$; inoltre l'altezza di ogni salto nel punto k coincide con la probabilità p_k . Nella medesima figura è rappresentato anche il grafico a barre di alcuni dei valori p_k . L'andamento di questi grafici ovviamente cambia al variare del valore di λ : in particolare al crescere di λ il massimo del grafico a barre si sposta verso valori più elevati di k . La legge di Poisson è particolarmente adatta a descrivere v.a. che rappresentano conteggi e che possono assumere un numero illimitato di valori: numero di telefonate che arrivano ad un centralino in un dato periodo di tempo; numero di clienti che si presentano allo sportello di un ufficio durante una giornata; numero di stelle presenti in una determinata regione di cielo. Il motivo per cui questo avviene è chiarito dal successivo Teorema e dalla discussione dell'Esempio 5.7.

Teorema 5.4. (Teorema di Poisson) Date le leggi Binomiali $B(n, \frac{\lambda}{n})$

$$p_k(n) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}, \quad k = 0, 1, \dots, n, \quad n = 1, 2, \dots$$

risulta

$$\lim_n p_k(n) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

cioè $p_k(n)$ converge verso la distribuzione di Poisson $\mathcal{P}(\lambda)$ per ogni valore di k .

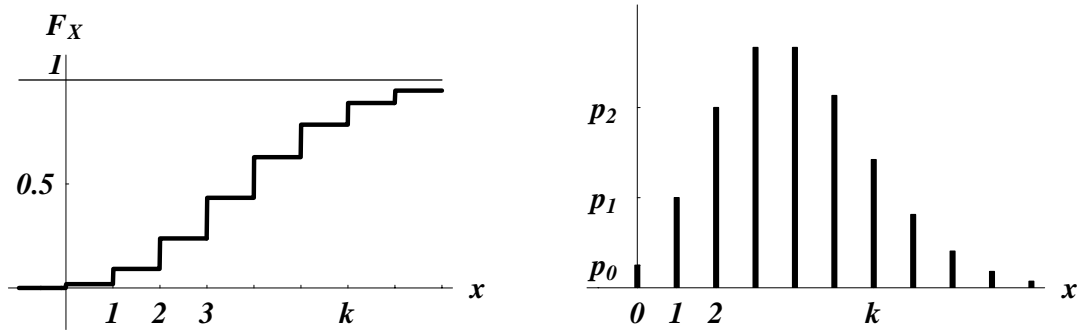


Figura 5.4: FD e grafico a barre di una legge di Poisson $\mathcal{P}(\lambda)$.

Dimostrazione: Infatti si ha

$$\begin{aligned}
 p_k(n) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
 &= \frac{\lambda^k n(n-1)\dots(n-k+1)}{k! n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
 &= \frac{\lambda^k}{k!} \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}
 \end{aligned}$$

Siccome è noto che

$$\lim_n \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) = 1, \quad \lim_n \left(1 - \frac{\lambda}{n}\right)^{-k} = 1, \quad \lim_n \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

il risultato segue immediatamente. □

Il Teorema 5.4 afferma in pratica che se n è molto grande e p molto piccola una legge Binomiale $B(n, p)$ è ben approssimata dalla legge di Poisson $\mathcal{P}(np)$. Illustreremo questo risultato con un esempio.

Esempio 5.7. Supponiamo di voler studiare la legge del numero aleatorio X di telefonate che arrivano ad un centralino telefonico in un intervallo di tempo T . Naturalmente X sarà una v.a. discreta che può assumere tutti i valori interi $k = 0, 1, 2, \dots$ dato che non vi è nessun limite *a priori* per il numero di telefonate che arrivano nel tempo T . Possiamo cominciare costruendo un modello approssimato per il nostro problema: dividiamo l'intervallo T in n parti uguali di lunghezza $\frac{T}{n}$, con n abbastanza grande da poter supporre che in ogni intervallo $\frac{T}{n}$ arrivi al più una telefonata (ovviamente questa ipotesi diviene sempre più realistica al limite per $n \rightarrow \infty$). Definiamo poi n v.a. indipendenti X_1, \dots, X_n in modo tale che X_i valga 1 oppure 0 secondo che nell'intervallo i -mo arrivi o meno una telefonata. Faremo inoltre l'ipotesi che esista un numero positivo λ tale che $\mathbf{P}(X_i = 1) = \frac{\lambda}{n}$ con $i = 1, 2, \dots, n$ in modo che le X_i siano indipendenti e tutte Bernoulli $B(1, \frac{\lambda}{n})$. In pratica questo vuol dire che la probabilità che arrivi una telefonata nell' i -mo intervallo diminuisce

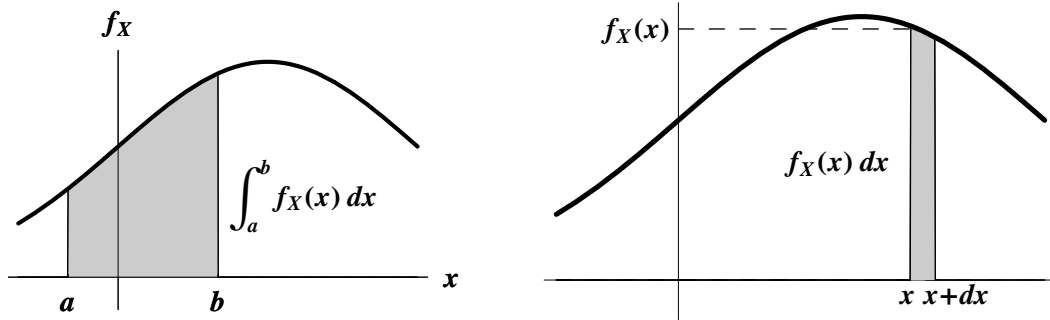


Figura 5.5: L'area fra a e b al di sotto della curva $f_X(x)$ è la probabilità che X assuma valori fra a e b (vedi equazione (5.11)); inoltre $f_X(x) dx$ rappresenta la probabilità infinitesima che X stia nell'intervallo $[x, x + dx]$.

con il crescere n ; il valore di λ invece è fisso e dipende dalla lunghezza dell'intervallo T e dalla intensità con la quale arrivano le telefonate in quel periodo della giornata. Da quanto detto deriva che per n grande il numero totale X di telefonate che arriva sarà approssimato da $X_1 + \dots + X_n$ che, per il Teorema 5.3, è Binomiale $B(n, \frac{\lambda}{n})$. Il Teorema di Poisson 5.4 ci garantisce allora che al limite per $n \rightarrow \infty$ la legge di $X_1 + \dots + X_n$ convergerà verso la legge di Poisson $\mathcal{P}(\lambda)$. In conclusione possiamo affermare che il numero X di telefonate che arrivano al nostro centralino telefonico è na v.a. di Poisson $\mathcal{P}(\lambda)$. Naturalmente resta da studiare in che modo si possa stimare il valore del parametro λ a partire da osservazioni empiriche reali.

5.3 Variabili aleatorie continue

Definizione 5.9. *Chiameremo v.a. continue le v.a. X che assumono tutti i valori di un intervallo J di numeri reali, non escluso l'intero insieme \mathbf{R} .*

Abbiamo visto nella Sezione 5.2 che la legge delle v.a. discrete è determinata dall'assegnazione di numeri p_k che soddisfino le proprietà (5.4). Nel caso di v.a. continue questa procedura elementare non è più possibile e bisognerà passare all'uso degli strumenti del calcolo differenziale e integrale.

Definizione 5.10. *Chiameremo **funzione di densità (fd)** ogni funzione $f(x)$ definita su \mathbf{R} che goda delle seguenti proprietà:*

$$f(x) \geq 0, \quad \int_{-\infty}^{+\infty} f(x) dx = 1. \tag{5.9}$$

Si può mostrare che la legge di una v.a. continua X è caratterizzata dall'assegnazione di una opportuna fd $f_X(x)$ mediante la quale è possibile eseguire tutti i calcoli necessari. In particolare le relazioni fra la fd f_X e la FD F_X sono riassunte da

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \quad f_X(x) = F_X'(x) \tag{5.10}$$

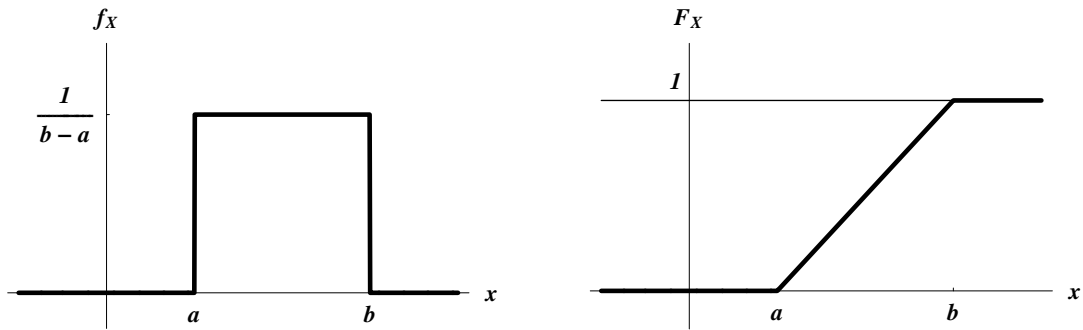


Figura 5.6: fd e FD della legge Uniforme $\mathcal{U}(a, b)$.

cioè F_X è la primitiva di f_X che si annulla per $x \rightarrow -\infty$, mentre a sua volta f_X è la derivata di F_X . Inoltre (vedi Figura 5.5) avremo che, comunque scelto un intervallo $[a, b]$ (per v.a. continue non è importante che sia *chiuso o aperto*) risulterà

$$\mathbf{P}\{a \leq X \leq b\} = P_X[a, b] = \int_a^b f_X(x) dx = F_X(b) - F_X(a) \quad (5.11)$$

Si noti che $f_X(x)$ non è la probabilità che la v.a. X assuma il valore x : per le v.a. continue la probabilità di assumere un singolo valore è zero. D'altra parte si potrebbe far vedere con dei banali esempi che una fd può assumere anche valori maggiori di 1, e quindi non può in nessun modo essere una probabilità. Solo la quantità infinitesima $f_X(x) dx$ può essere interpretata come la probabilità che X prenda valori nell'intervallo infinitesimo $[x, x + dx]$ (vedi Figura 5.5).

Definizione 5.11. Chiameremo **moda** di una v.a. continua X il valore x per il quale la sua fd f_X assume il valore massimo. La definizione si applica anche nel caso di massimi relativi.

Il calcolo delle probabilità mediante l'integrale in (5.11) non è sempre un'operazione elementare: per questo motivo si usano delle apposite Tavole nelle quali sono elencati i valori delle FD F_X delle leggi più usuali in modo che il calcolo di $\mathbf{P}\{a \leq X \leq b\}$ possa essere effettuato direttamente mediante la differenza $F_X(b) - F_X(a)$.

Esempio 5.8. (Legge uniforme) Il caso più semplice di leggi di v.a. continue è quello delle leggi Uniformi nell'intervallo $[a, b]$ (con $a, b \in \mathbf{R}$) indicate con il simbolo $\mathcal{U}(a, b)$. Esse sono caratterizzate dalla fd

$$f_X(x) = \begin{cases} 1/(b-a) & \text{se } a \leq x \leq b, \\ 0 & \text{altrimenti.} \end{cases} \quad (5.12)$$

La FD si calcola poi in maniera elementare:

$$F_X(x) = \begin{cases} 0 & \text{se } x < a, \\ (x-a)/(b-a) & \text{se } a \leq x \leq b, \\ 1 & \text{se } x > b. \end{cases}$$

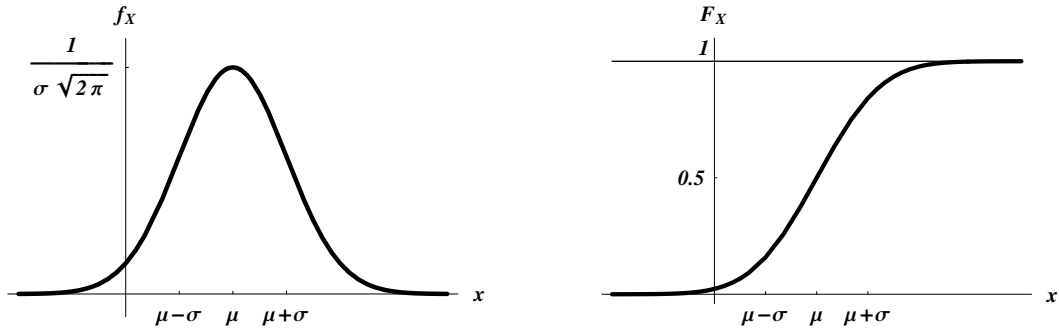


Figura 5.7: fd e FD della legge Normale $N(\mu, \sigma^2)$.

Queste due funzioni sono rappresentate nella Figura 5.6. Ovviamente le relazioni (5.9) sono sempre soddisfatte dato che l'area di un rettangolo di base $b - a$ e altezza $1/(b - a)$ è sempre 1. Si vede inoltre dall'equazione (5.11) che per $a \leq c \leq c + \Delta \leq b$ si ha $\mathbf{P}\{c \leq X \leq c + \Delta\} = \Delta/(b - a)$ indipendentemente dal valore di c ; ovvero: ad intervalli di larghezza Δ interni ad $[a, b]$ viene attribuita sempre la stessa probabilità $\Delta/(b - a)$ indipendentemente dalla loro collocazione in $[a, b]$. Questo è il significato della *uniformità* della distribuzione.

Esempio 5.9. (Legge Normale o Gaussiana) Il caso più noto di leggi di v.a. continue è quello delle cosiddette leggi Normali o Gaussiane $\mathcal{N}(\mu, \sigma^2)$ con $\mu \in \mathbf{R}$ e $\sigma > 0$. Esse sono caratterizzate dalla fd

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (5.13)$$

Queste funzioni soddisfano le relazioni (5.9) per ogni valore di μ e σ , ma noi non lo verificheremo. Ci limiteremo qui a dare solo una descrizione qualitativa del comportamento di queste funzioni che sono rappresentate nella Figura 5.7. La fd f_X è una curva a campana simmetrica attorno ad un massimo nel punto $x = \mu$ (moda). La funzione va rapidamente verso zero allontanandosi dal centro della curva e la larghezza della campana è regolata dal valore di σ : grandi valori di σ corrispondono a curve larghe e piatte; piccoli valori di σ corrispondono a curve strette e alte. Si può mostrare che la curva presenta due flessi proprio in $x = \mu \pm \sigma$. La FD

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-(t-\mu)^2/2\sigma^2} dt \quad (5.14)$$

non ha un'espressione analitica elementare, ma il suo grafico è molto semplice e regolare, tipico delle FD: ha una forma di S allungata che varia da 0 verso 1 con un punto di flesso in $x = \mu$. La FD di una Normale diviene sempre più ripida (e al limite approssima un gradino di altezza 1) quando $\sigma \rightarrow 0$; viceversa si allunga sempre di più con il crescere di σ . La legge $\mathcal{N}(0, 1)$ la cui fd è

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (5.15)$$

è detta anche **legge Normale standard** e riveste una importanza particolare perchè, come vedremo, il calcolo delle probabilità relative a leggi Normali generiche può sempre essere facilmente ricondotto all'uso delle Tavole dell'Appendice D per la FD della legge Normale standard

$$\boxed{\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt} \quad (5.16)$$

La fd e la FD della Normale standard presentano lo stesso andamento qualitativo di Figura 5.7, ma la moda si trova in $x = 0$ e i due flessi in $x = \pm 1$. Inoltre, data la evidente simmetria di queste due funzioni è anche facile verificare che

$$\varphi(-x) = \varphi(x), \quad \Phi(-x) = 1 - \Phi(x) \quad (5.17)$$

relazioni che risulteranno utili in seguito.

Teorema 5.5.

1. Se X è una v.a. $\mathcal{N}(\mu, \sigma^2)$, e se a e b sono due numeri, allora la v.a. $aX + b$ è $\mathcal{N}(a\mu + b, a^2\sigma^2)$.
2. Se X e Y sono v.a. indipendenti, rispettivamente $\mathcal{N}(\mu, \sigma^2)$ e $\mathcal{N}(\nu, \tau^2)$, allora la v.a. $X + Y$ è $\mathcal{N}(\mu + \nu, \sigma^2 + \tau^2)$.

Una conseguenza di questo Teorema è che se X è una v.a. Normale standard $\mathcal{N}(0, 1)$, allora $Y = \sigma X + \mu$ è una v.a. Normale $\mathcal{N}(\mu, \sigma^2)$. In altre parole: ogni v.a. $\mathcal{N}(\mu, \sigma^2)$ è del tipo $Y = \sigma X + \mu$ con X Normale standard. Pertanto avremo che, se Y è $\mathcal{N}(\mu, \sigma^2)$, allora comunque scelti i numeri a e b

$$\mathbf{P}\{a \leq Y \leq b\} = \mathbf{P}\{a \leq \sigma X + \mu \leq b\} = \mathbf{P}\left\{\frac{a - \mu}{\sigma} \leq X \leq \frac{b - \mu}{\sigma}\right\}$$

e quindi da (5.11) e (5.16) si ha

$$\boxed{\mathbf{P}\{a \leq Y \leq b\} = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)} \quad (5.18)$$

calcolo che si riduce alla consultazione delle Tavole della FD Normale standard in Appendice D. L'uso delle Tavole dell'Appendice D sarà molto utile anche per le altre leggi che di solito vengono introdotte in statistica. Per questo motivo, e per ragioni di brevità, non daremo la forma esplicita della fd e della FD delle leggi che discuteremo nei prossimi esempi, ma ci limiteremo piuttosto ad una loro descrizione qualitativa e all'enunciato delle principali relazioni che le legano.

Teorema 5.6. Se Y_1, \dots, Y_n sono v.a. indipendenti e tutte Normali standard $\mathcal{N}(0, 1)$, allora la v.a.

$$X = Y_1^2 + \dots + Y_n^2$$

segue la legge $\chi^2(n)$ detta del chi quadro con n gradi di libertà.

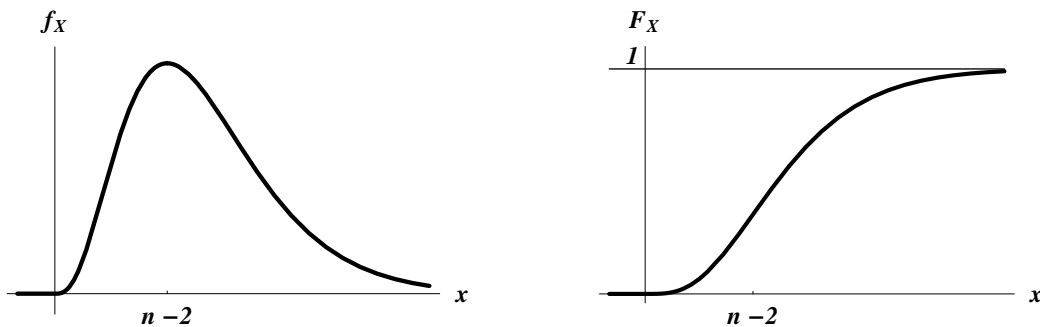


Figura 5.8: fd e FD della legge del *chi quadro* $\chi^2(n)$ con $n > 2$.

La **legge del Chi quadro** con $n = 1, 2, \dots$ gradi di libertà $\chi^2(n)$ ha una fd e una FD del tipo mostrato nella Figura 5.8 per $n > 2$. La fd è diversa da zero solo per $x \geq 0$ mentre è rigorosamente nulla per $x < 0$; sul semiasse reale positivo il grafico è asimmetrico e presenta una una lunga coda che si annulla asintoticamente per $x \rightarrow +\infty$. La moda si trova in $x = n - 2$, e tende ad allontanarsi dall'origine per n crescenti. I valori della FD di $\chi^2(n)$ possono essere trovati sulle opportune Tavole e vengono usati nel calcolo tramite l'equazione (5.11).

Teorema 5.7. *Se X e Y sono due v.a. di legge rispettivamente $\mathcal{N}(0, 1)$ e $\chi^2(n)$, allora la v.a.*

$$T = \frac{X}{\sqrt{Y/n}}$$

segue la legge $t(n)$ detta di Student con n gradi di libertà.

La **legge di Student** con $n = 1, 2, \dots$ gradi di libertà $t(n)$ ha una fd e una FD del tipo mostrato nella Figura 5.9. La fd di $t(n)$ è una funzione a campana con la moda in $x = 0$, simile alla $\mathcal{N}(0, 1)$. Come si vede dalla Figura 5.9 la fd di $t(n)$ si annulla però più lentamente della fd della $\mathcal{N}(0, 1)$. Quando il valore di n cresce la fd della legge $t(n)$ diviene sempre più simile alla fd Normale standard, e al limite le due funzioni coincidono. I valori della FD di $t(n)$ possono essere trovati sulle opportune Tavole e vengono usati nel calcolo tramite l'equazione (5.11).

Teorema 5.8. *Se X e Y sono due v.a. di legge rispettivamente $\chi^2(n)$ e $\chi^2(m)$, allora la v.a.*

$$F = \frac{X/n}{Y/m}$$

segue la legge $F(n, m)$ detta di Fisher con n ed m gradi di libertà.

La **legge di Fisher** con n ed m gradi di libertà $F(n, m)$ ha una fd e una FD del tipo mostrato nella Figura 5.10. La fd di $F(n, m)$ somiglia a quella di una $\chi^2(n)$: essa è diversa da zero solo per $x \geq 0$ mentre è rigorosamente nulla per $x < 0$; sul semiasse reale positivo il grafico è asimmetrico e presenta una una lunga coda che si

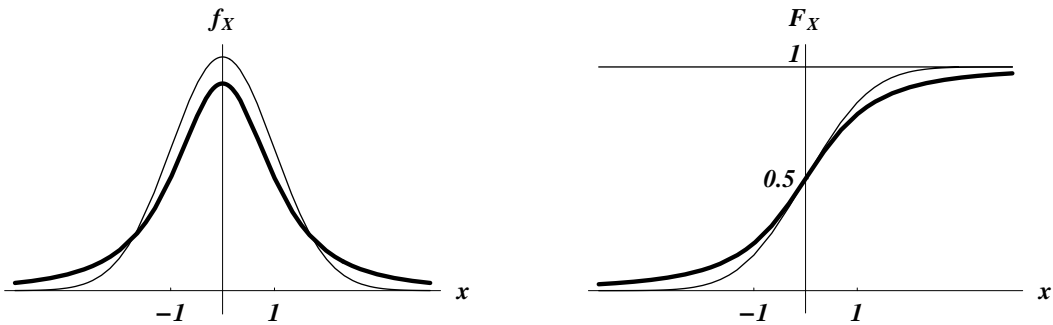


Figura 5.9: fd e FD della legge di Student con n gradi di libertà $t(n)$. Le curve più sottili sono la fd e la FD della $\mathcal{N}(0, 1)$ e sono qui riportate per confronto.

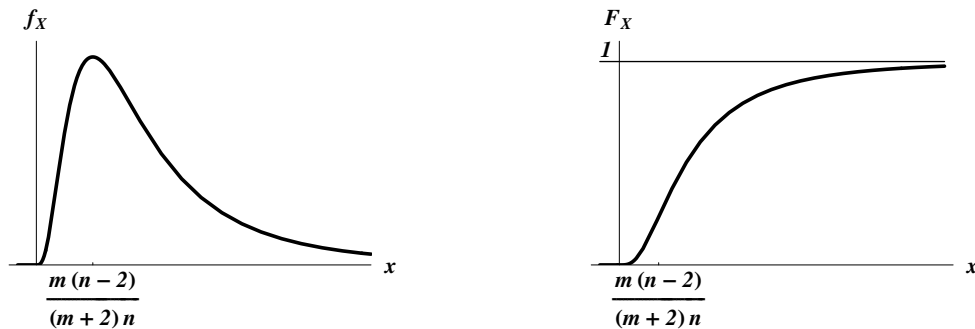


Figura 5.10: fd e FD della legge di Fisher con n ed m gradi di libertà $F(n, m)$.

annulla asintoticamente per $x \rightarrow +\infty$. La moda si trova nel punto $\frac{m(n-2)}{(m+2)n}$. I valori della FD di $F(n, m)$ possono essere trovati sulle opportune Tavole e vengono usati nel calcolo tramite l'equazione (5.11).

Teorema 5.9. *Se X_1, \dots, X_n sono v.a. indipendenti e tutte normali $\mathcal{N}(\mu, \sigma^2)$, e se poniamo*

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}, \quad S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2, \quad S = \sqrt{S^2},$$

allora le v.a.

$$Y = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}, \quad Z = (n-1) \frac{S^2}{\sigma^2}, \quad T = \sqrt{n} \frac{\bar{X} - \mu}{S}$$

seguono rispettivamente le leggi $\mathcal{N}(0, 1)$, $\chi^2(n-1)$ e $t(n-1)$.

Dimostrazione: Per semplicità discuteremo esplicitamente solo la legge di Y : dato che le X_k sono tutte $\mathcal{N}(\mu, \sigma^2)$ dal Teorema (5.5) si ha che $X_1 + \dots + X_n$ è $\mathcal{N}(n\mu, n\sigma^2)$, e che $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ è $\mathcal{N}(\mu, \frac{\sigma^2}{n})$. Utilizzando ancora ripetutamente lo stesso Teorema (5.5) si avrà allora che $\bar{X} - \mu$ è $\mathcal{N}(0, \frac{\sigma^2}{n})$ e infine che Y è $\mathcal{N}(0, 1)$. \square

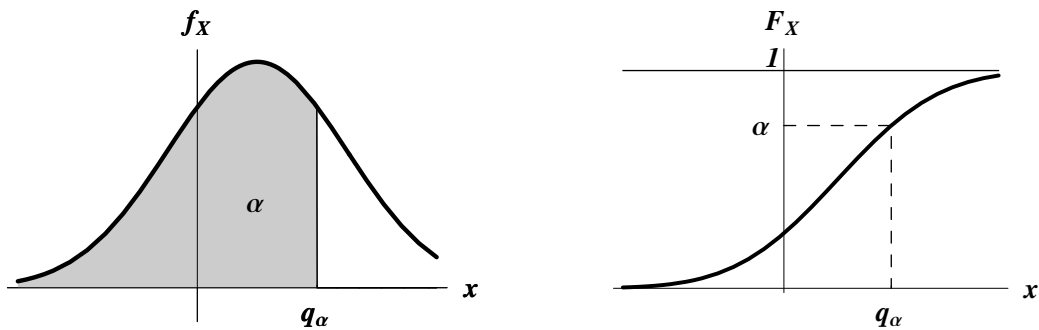


Figura 5.11: Quantile q_α di ordine α di una distribuzione con fd f_X e FD F_X .

5.4 Quantili

Supponiamo che X sia una v.a. continua con fd $f_X(x)$ e FD $F_X(x)$: abbiamo visto negli esempi della Sezione 5.3 che nei casi di nostro interesse F_X è una funzione strettamente crescente su tutto \mathbf{R} (leggi Normale e di Student), o almeno sul semiasse $x \geq 0$ (leggi del chi quadro e di Fisher). Preso allora un numero $0 < \alpha < 1$ uno sguardo ai grafici delle Figure 5.7– 5.10 ci convincerà del fatto che esiste una e una sola soluzione dell'equazione

$$F_X(x) = \alpha. \quad (5.19)$$

In analogia con quanto fatto nella parte di Statistica Descrittiva avrà allora senso introdurre la seguente Definizione:

Definizione 5.12. *Data una v.a. X continua con fd $f_X(x)$ e FD $F_X(x)$, chiameremo **quantile di ordine** α il numero q_α soluzione dell'equazione (5.19), cioè tale che $F_X(q_\alpha) = \alpha$. Il quantile di ordine $\alpha = \frac{1}{2}$ si chiama **mediana**; i quantili di ordine $\alpha = \frac{k}{4}$ con $k = 1, 2, 3$ si chiamano **quartili**.*

Il significato di questa definizione è illustrato nella Figura 5.11: nella seconda parte si vede come $F_X(q_\alpha) = \alpha$; nella prima invece, in cui è rappresentata la fd di X , si vede come q_α è il punto che lascia alla sua sinistra un'area sotto la curva pari ad α . Questo è ovviamente coerente con la Definizione 5.12 dato che dalla Definizione 5.4 e dall'equazione (5.10) si ha

$$\alpha = F_X(q_\alpha) = \mathbf{P}\{X \leq q_\alpha\} = \int_{-\infty}^{q_\alpha} f_X(x) dx.$$

Sarà utile a questo punto ricavare alcune formule di frequente uso in statistica. Dato che, come visto nell'equazione (5.16), la FD della legge $\mathcal{N}(0, 1)$ è stata indicata con il simbolo Φ , da ora in poi indicheremo il suo quantile di ordine α con il simbolo φ_α . Nella Figura 5.12 sono riportati i quantili di ordine $\frac{\alpha}{2}$ e $1 - \frac{\alpha}{2}$ di una legge Normale standard $\mathcal{N}(0, 1)$ e viene messo in evidenza come questi, per la loro definizione, delimitino due code di eguale probabilità $\frac{\alpha}{2}$. Data la simmetria della fd Normale

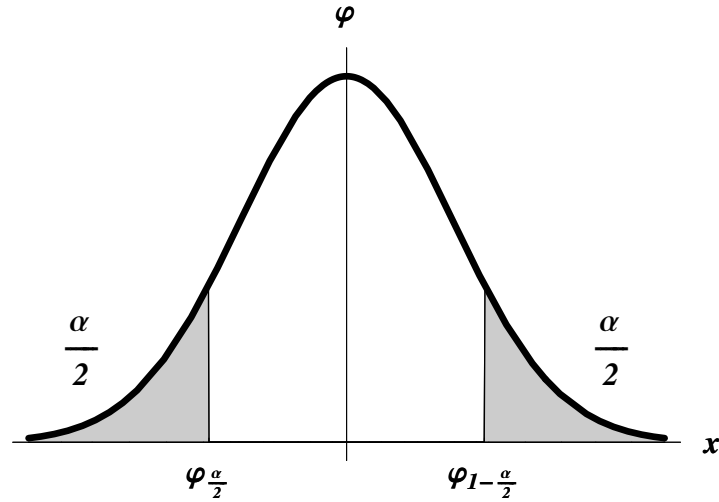


Figura 5.12: Quantili di ordine $\frac{\alpha}{2}$, e $1 - \frac{\alpha}{2}$ di una legge Normale standard $\mathcal{N}(0, 1)$.

standard $\varphi(x)$ definita in (5.15), per ogni valore di x si avrà $\varphi(-x) = \varphi(x)$, e quindi in particolare fra i due quantili mostrati in Figura 5.12 sussiste la relazione

$$\varphi_{\frac{\alpha}{2}} = -\varphi_{1-\frac{\alpha}{2}}. \quad (5.20)$$

Da (5.20) deriva che se la v.a. X segue la legge Normale standard $\mathcal{N}(0, 1)$ si ha

$$\mathbf{P}\{|X| \geq \varphi_{1-\frac{\alpha}{2}}\} = \alpha. \quad (5.21)$$

In realtà per dimostrare la (5.21) basterebbe dare un'occhiata alla Figura 5.12 e ricordare la relazione (5.20), ma come esercizio ne daremo anche una deduzione analitica esplicita. Da (5.20) si ha infatti

$$\{|X| \geq \varphi_{1-\frac{\alpha}{2}}\} = \{X \geq \varphi_{1-\frac{\alpha}{2}}\} \cup \{X \leq -\varphi_{1-\frac{\alpha}{2}}\} = \{X \geq \varphi_{1-\frac{\alpha}{2}}\} \cup \{X \leq \varphi_{\frac{\alpha}{2}}\},$$

con i due eventi del secondo membro disgiunti. La relazione (5.21) segue allora dalla additività (3.3)

$$\begin{aligned} \mathbf{P}\{|X| \geq \varphi_{1-\frac{\alpha}{2}}\} &= \mathbf{P}\{X \geq \varphi_{1-\frac{\alpha}{2}}\} + \mathbf{P}\{X \leq \varphi_{\frac{\alpha}{2}}\} \\ &= 1 - \mathbf{P}\{X \leq \varphi_{1-\frac{\alpha}{2}}\} + \mathbf{P}\{X \leq \varphi_{\frac{\alpha}{2}}\} = 1 - \left(1 - \frac{\alpha}{2}\right) + \frac{\alpha}{2} = \alpha. \end{aligned}$$

I quantili di una T di Student con n gradi di libertà con legge $t(n)$ sono indicati con il simbolo $t_\alpha(n)$ e sono definiti dalla relazione

$$\mathbf{P}\{T \leq t_\alpha(n)\} = \alpha.$$

Siccome la fd della legge $t(n)$ ha le stesse proprietà di simmetria della fd Normale standard innanzitutto si ha

$$t_{\frac{\alpha}{2}} = -t_{1-\frac{\alpha}{2}}. \quad (5.22)$$

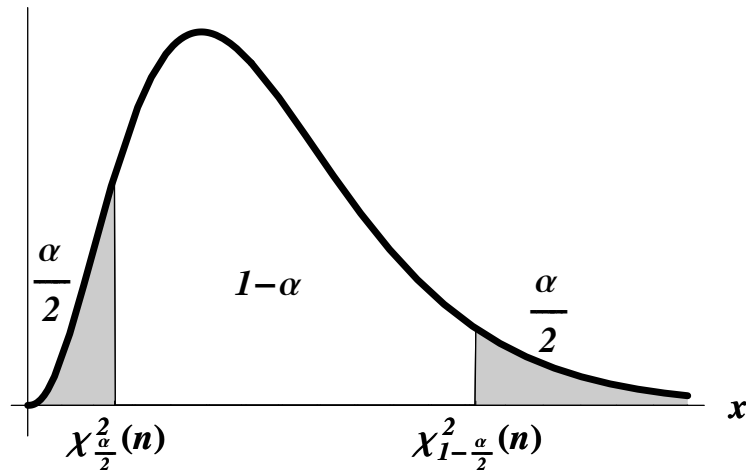


Figura 5.13: Quantili di ordine $\frac{\alpha}{2}$, e $1 - \frac{\alpha}{2}$ di una legge $\chi^2(n)$.

e in secondo luogo si dimostra una relazione del tutto simile a (5.21): se la v.a. T segue la legge di Student $t(n)$ si ha

$$\mathbf{P}\{|T| \geq t_{1-\frac{\alpha}{2}}(n)\} = \alpha. \quad (5.23)$$

I quantili di una Z di legge $\chi^2(n)$ e di una F di legge di Fisher $F(n, m)$ si indicano rispettivamente con $\chi_{\alpha}^2(n)$ e $f_{\alpha}(n, m)$ e sono definiti da

$$\mathbf{P}\{Z \leq \chi_{\alpha}^2(n)\} = \alpha, \quad \mathbf{P}\{F \leq f_{\alpha}(n, m)\} = \alpha.$$

Siccome le fd delle leggi $\chi^2(n)$ ed $F(n, m)$ non sono funzioni simmetriche le relazioni del tipo (5.21) e (5.23) assumono una forma differente. Così ad esempio, se la v.a. Z segue la legge $\chi^2(n)$ si ha

$$\mathbf{P}\{\chi_{\frac{\alpha}{2}}^2(n) \leq Z \leq \chi_{1-\frac{\alpha}{2}}^2(n)\} = 1 - \alpha. \quad (5.24)$$

La (5.24) è illustrata nella Figura 5.13 e deriva facilmente dalle definizioni stesse di quantili: dato che $\chi_{\frac{\alpha}{2}}^2(n) < \chi_{1-\frac{\alpha}{2}}^2(n)$, abbiamo innanzitutto

$$\{Z \leq \chi_{1-\frac{\alpha}{2}}^2(n)\} = \{Z \leq \chi_{\frac{\alpha}{2}}^2(n)\} \cup \{\chi_{\frac{\alpha}{2}}^2(n) \leq Z \leq \chi_{1-\frac{\alpha}{2}}^2(n)\}$$

con i due eventi del secondo membro disgiunti. Segue allora dalla additività (3.3) della probabilità che

$$\mathbf{P}\{Z \leq \chi_{1-\frac{\alpha}{2}}^2(n)\} = \mathbf{P}\{Z \leq \chi_{\frac{\alpha}{2}}^2(n)\} + \mathbf{P}\{\chi_{\frac{\alpha}{2}}^2(n) \leq Z \leq \chi_{1-\frac{\alpha}{2}}^2(n)\}$$

e quindi si ottiene la relazione (5.24) tenendo anche conto delle definizioni dei quantili

$$\begin{aligned} \mathbf{P}\{\chi_{\frac{\alpha}{2}}^2(n) \leq Z \leq \chi_{1-\frac{\alpha}{2}}^2(n)\} &= \mathbf{P}\{Z \leq \chi_{1-\frac{\alpha}{2}}^2(n)\} - \mathbf{P}\{Z \leq \chi_{\frac{\alpha}{2}}^2(n)\} \\ &= \left(1 - \frac{\alpha}{2}\right) - \frac{\alpha}{2} = 1 - \alpha \end{aligned}$$

..

I valori dei quantili delle diverse leggi possono essere ricavati dalle Tavole dell'Appendice D, ma queste per ragioni di spazio non riportano tutti i valori necessari a discutere problemi di statistica. I valori mancanti possono essere ricavati mediante relazioni come (5.20) e (5.22). A questo proposito ricorderemo allora che per la legge $\chi^2(n)$, quando i gradi di libertà sono numerosi (per le nostre Tavole: se $n > 35$), è possibile calcolare dei valori approssimati mediante i quantili della legge Normale standard tramite la relazione

$$\chi_\alpha^2(n) \simeq \frac{1}{2}(\varphi_\alpha + \sqrt{2n-1})^2. \quad (5.25)$$

Infine ricorderemo che tra i quantili della legge di Fisher vale la relazione

$$f_\alpha(n, m) = \frac{1}{f_{1-\alpha}(m, n)} \quad (5.26)$$

5.5 Vettori aleatori

In questa sezione daremo alcune sommarie indicazioni sulla maniera in cui viene assegnata la legge dei vettori aleatori $X = (X_1, \dots, X_n)$ introdotti con le Definizioni 5.5, 5.6 e 5.7. Se le m componenti X_i sono v.a. discrete, indicando con x_i i valori discreti di ciascuna di esse, la legge congiunta sarà data assegnando le quantità

$$p(x_1, \dots, x_m) = \mathbf{P}\{X_1 = x_1, \dots, X_m = x_m\}.$$

Le singole componenti X_i saranno poi dotate a loro volta di leggi marginali assegnate tramite le

$$p_i(x_i) = \mathbf{P}\{X_i = x_i\}, \quad i = 1, \dots, m$$

e dal Teorema 5.2 si può dimostrare che esse saranno indipendenti se e solo se

$$\boxed{p(x_1, \dots, x_m) = p_1(x_1) \cdot \dots \cdot p_m(x_m)}. \quad (5.27)$$

Le leggi congiunta e marginali di un dato vettore aleatorio non possono essere assegnate separatamente in maniera arbitraria: ci sono infatti semplici regole che consentono di ricavare le leggi marginali di un vettore aleatorio a partire dalla sua legge congiunta. Per brevità noi qui trascureremo di enunciare queste regole, e ci limiteremo invece ad osservare che, viceversa, non è sempre possibile ricostruire in maniera unica la legge congiunta a partire dalle marginali; l'unico caso in cui ciò è possibile è quello in cui è anche noto che le componenti sono indipendenti, per cui la congiunta si ricava dalle marginali tramite (5.27)

Esempio 5.10. (Legge Multinomiale) Per dare un esempio di legge congiunta di una v.a. m -dimensionale consideriamo n ripetizioni indipendenti di un esperimento i cui risultati casuali possono essere $m \geq 2$. Supporremo inoltre che in ogni tentativo si ottiene

il risultato 1 con probabilità q_1, \dots , e il risultato m con probabilità q_m . Naturalmente dovranno essere verificate le relazioni

$$0 \leq q_i \leq 1, \quad i = 1, \dots, m; \quad q_1 + \dots + q_m = 1.$$

Sia ora X_i la v.a. che rappresenta il numero di tentativi in cui si è ottenuto il risultato i -mo: il risultato del nostro esperimento sarà allora rappresentato dalla v.a. m -dimensionale $X = (X_1, \dots, X_m)$ che indica il numero di volte in cui si è ottenuto ciascuno degli m possibili risultati. Naturalmente deve risultare

$$X_1 + \dots + X_m = n. \quad (5.28)$$

I possibili valori di X sono i vettori di numeri interi (k_1, \dots, k_m) con $k_1 + \dots + k_m = n$ e si può dimostrare che la legge congiunta di X è

$$p(k_1, \dots, k_m) = \mathbf{P}\{X_1 = k_1, \dots, X_m = k_m\} = \frac{n!}{k_1! \dots k_m!} q_1^{k_1} \dots q_m^{k_m} \quad (5.29)$$

detta legge Multinomiale. È immediato verificare che nel caso di $m = 2$ la legge Multinomiale coincide con la legge Binomiale dell'Esempio 5.5. Un ovvio esempio concreto di legge multinomiale è fornito da n lanci di un dado con $m = 6$ facce numerate. Se in generale il dado non è ben bilanciato le probabilità q_i di ottenere il risultato i -mo saranno diverse fra loro. Detta X_i la v.a. che rappresenta il numero di volte in cui su n lanci è uscita la faccia i -ma, la legge del vettore $X = (X_1, \dots, X_6)$ è proprio (5.29) con $m = 6$. Se in particolare il dado è bilanciato, allora $q_1 = \dots = q_6 = 1/6$ e ovviamente

$$p(k_1, \dots, k_6) = \mathbf{P}\{X_1 = k_1, \dots, X_6 = k_6\} = \frac{n!}{k_1! \dots k_6!} \frac{1}{6^n}$$

Così ad esempio, la probabilità che su $n = 12$ lanci di dado ogni faccia esca esattamente due volte è

$$p(2, 2, 2, 2, 2, 2) = \frac{12!}{2! 2! 2! 2! 2! 2!} \frac{1}{6^{12}} \simeq 0.00344.$$

mentre la probabilità che tre delle sei facce escano 1 volta e le altre tre 3 volte è

$$p(1, 1, 1, 3, 3, 3) = \frac{12!}{1! 1! 1! 3! 3! 3!} \frac{1}{6^{12}} \simeq 0.00102.$$

Per le leggi multinomiali è anche possibile calcolare le leggi marginali e verificare che le componenti X_i non sono indipendenti. Per semplicità ometteremo questa verifica e ci limiteremo solo ad osservare che è intuitivamente facile giustificare la non indipendenza delle componenti se si riflette al fatto che esse devono sempre soddisfare la relazione (5.28), per cui se fissassimo arbitrariamente le prime $m - 1$ componenti, la m -a sarebbe immediatamente già determinata con valore $n - (X_1 + \dots + X_{m-1})$ e non potrebbe quindi essere indipendente dalle altre.

Se invece le componenti del vettore aleatorio $X = (X_1, \dots, X_m)$ sono v.a. continue la legge congiunta è determinata da una fd con m variabili $f(x_1, \dots, x_m)$. Le proprietà delle fd m -dimensionali sono del tutto analoghe a quelle delle fd con una

sola variabile introdotte nella Definizione 5.10, ma la loro formulazione richiede l'uso del calcolo differenziale con m variabili e noi la trascureremo. Ci limiteremo solo ad osservare che anche in questo caso dalla fd congiunta $f(x_1, \dots, x_m)$ è possibile calcolare le fd marginali $f_i(x_i)$ che determinano le leggi marginali delle singole componenti X_i , e che ancora una volta dal Teorema 5.2 si può dimostrare che le componenti sono indipendenti se e solo se

$$\boxed{f(x_1, \dots, x_m) = f_1(x_1) \cdot \dots \cdot f_m(x_m)}. \quad (5.30)$$

Capitolo 6

Attesa e varianza

6.1 Attesa e varianza

Definizione 6.1. *Data una v.a. X , se essa è discreta, per semplicità con valori interi k e legge $p_k = p(k) = \mathbf{P}\{X = k\}$, e se*

$$\sum_k |k| p_k < +\infty, \quad (6.1)$$

*chiameremo **valore d'attesa** (o semplicemente **attesa** o anche **media**) di X la quantità*

$$\mu = \mathbf{E}(X) = \sum_k k p_k; \quad (6.2)$$

se invece X è continua con fd $f_X(x)$, e se

$$\int_{-\infty}^{+\infty} |x| f_X(x) dx < +\infty, \quad (6.3)$$

l'attesa sarà definita da

$$\mu = \mathbf{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) dx. \quad (6.4)$$

Si noti che la somma in (6.2) è in realtà una serie se X assume infiniti valori. Se le condizioni (6.1) e (6.3) non sono verificate la serie (6.2) e l'integrale (6.4) non convergono e si dice che la v.a. X non possiede valore d'attesa. In pratica l'attesa non è altro che la somma dei valori assunti da X moltiplicati per le rispettive probabilità (si ricordi che nel caso continuo $f_X(x) dx$ può essere inteso come la probabilità che X assuma valori in $[x, x + dx]$), e quindi non è altro che la media dei valori assunti da X se si suppone che le frequenze relative coincidano con le rispettive probabilità. Nel seguito, quando non diversamente specificato, le definizioni e le proprietà enunciate con il simbolo \mathbf{E} saranno valide sia per il caso discreto che per quello continuo.

Teorema 6.1. Se X, X_1, \dots, X_n sono v.a. e a, b due numeri arbitrari, allora

$$\mathbf{E}(aX + b) = a\mathbf{E}(X) + b, \quad (6.5)$$

$$\mathbf{E}(X_1 + \dots + X_n) = \mathbf{E}(X_1) + \dots + \mathbf{E}(X_n). \quad (6.6)$$

Definizione 6.2. Chiameremo **varianza** della v.a. X con attesa μ la quantità

$$\sigma^2 = \mathbf{Var}(X) = \mathbf{E}[(X - \mu)^2]; \quad (6.7)$$

si chiama invece **deviazione standard** la quantità $\sigma = \sqrt{\mathbf{Var}(X)}$. Chiameremo poi **covarianza** di due v.a. X e Y con attese rispettivamente μ e ν la quantità

$$\mathbf{Cov}(X, Y) = \mathbf{E}[(X - \mu)(Y - \nu)], \quad (6.8)$$

e **coefficiente di correlazione** la quantità

$$\rho_{XY} = \frac{\mathbf{Cov}(X, Y)}{\sqrt{\mathbf{Var}(X)}\sqrt{\mathbf{Var}(Y)}}. \quad (6.9)$$

Se $\mathbf{Cov}(X, Y) = 0$, o equivalentemente $\rho_{X,Y} = 0$, diremo che le v.a. X e Y sono **non correlate**.

Sarà utile dare anche la definizione (6.7) di varianza in termini espliciti: nel caso di v.a. discrete a valori interi k avremo

$$\sigma^2 = \sum_k (k - \mu)^2 p_k; \quad (6.10)$$

mentre nel caso di v.a. continue si ha

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f_X(x) dx. \quad (6.11)$$

Si vede subito dalle definizioni (6.7) e (6.8) che

$$\mathbf{Var}(X) = \mathbf{Cov}(X, X), \quad (6.12)$$

per cui la covarianza non è altro che una generalizzazione del concetto di varianza al caso di due v.a. Come nella corrispondente discussione svolta nella parte di Statistica Descrittiva si può mostrare (e gli argomenti sono gli stessi) che la varianza di una v.a. rappresenta la dispersione dei valori di X attorno al suo valore d'attesa: una varianza grande indica che X tende a prendere valori anche molto lontani da μ ; viceversa nel caso di varianza piccola i valori di X sono piuttosto concentrati attorno a μ . Il coefficiente di correlazione, d'altra parte, non è altro che una covarianza *ridotta*, nel senso che, mentre $\mathbf{Cov}(X, Y)$ può assumere tutti i valori reali positivi e negativi, si può dimostrare che $-1 \leq \rho_{XY} \leq +1$. Per il resto i significati di coefficiente di correlazione e covarianza sono molto simili.

Teorema 6.2. *Se due v.a. X e Y sono indipendenti allora esse sono anche non correlate.*

Viceversa è facile far vedere con degli esempi che esistono v.a. non correlate che *non sono indipendenti*. In altri termini indipendenza e non correlazione non sono concetti equivalenti: l'indipendenza implica la non correlazione, ma in generale il viceversa non è vero. Ciononostante per comodità la non correlazione viene spesso utilizzata come una *indipendenza debole*, nel senso che v.a. non correlate sono considerate *quasi indipendenti*. Se $\mathbf{Cov}(X, Y) > 0$ si parla di **correlazione positiva** e questo indica che Y tende ad assumere valori grandi (rispettivamente: piccoli) quando X assume valori grandi (rispettivamente: piccoli). Viceversa se $\mathbf{Cov}(X, Y) < 0$ si parla di **correlazione negativa** e in tal caso Y tende ad assumere valori grandi (rispettivamente: piccoli) quando X assume valori piccoli (rispettivamente: grandi). A questo proposito si veda anche la corrispondente discussione svolta nella parte di Statistica Descrittiva.

Teorema 6.3. *Date due v.a. X e Y si ha*

$$\mathbf{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y); \quad (6.13)$$

In particolare si ha anche che

$$\sigma^2 = \mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2. \quad (6.14)$$

Dimostrazione: Partendo dalla definizione (6.8), ricordando che $\mu = \mathbf{E}(X)$ e $\nu = \mathbf{E}(Y)$ e usando i risultati del Teorema 6.1 si ricava che

$$\begin{aligned} \mathbf{Cov}(X, Y) &= \mathbf{E}[(X - \mu)(Y - \nu)] = \mathbf{E}(XY - \nu X - \mu Y + \mu\nu) \\ &= \mathbf{E}(XY) - \nu\mathbf{E}(X) - \mu\mathbf{E}(Y) + \mu\nu = \mathbf{E}(XY) - 2\mu\nu + \mu\nu \\ &= \mathbf{E}(XY) - \mu\nu = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y). \end{aligned}$$

La relazione (6.14) segue direttamente da (6.13) tramite (6.12). \square

Si noti che i valori d'attesa di prodotti di v.a. come quelli delle formule (6.8) e (6.13) sono definiti tramite le leggi congiunte di X e Y : se queste v.a. sono discrete con valori interi k e ℓ , e legge congiunta $p(k, \ell)$, tali valori d'attesa sono del tipo

$$\mathbf{E}(XY) = \sum_{k, \ell} k\ell p(k, \ell). \quad (6.15)$$

Non riporteremo invece le formule per il caso di v.a. continue perché esse richiedono l'uso di integrali multipli. Inoltre, data l'utilità di (6.14) nel calcolo della varianza, sarà utile dare in termini espliciti anche questa formula: con le solite notazioni nel caso discreto avremo

$$\sigma^2 = \sum_k k^2 p_k - \left(\sum_k k p_k \right)^2 = \sum_k k^2 p_k - \mu^2, \quad (6.16)$$

mentre nel caso continuo avremo

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 f_X(x) dx - \left(\int_{-\infty}^{+\infty} x f_X(x) dx \right)^2 = \int_{-\infty}^{+\infty} x^2 f_X(x) dx - \mu^2. \quad (6.17)$$

Il Teorema 6.1 afferma in sostanza che il calcolo dell'attesa $\mathbf{E}(X)$ di una v.a. è un'operazione lineare; non si può dire invece la stessa cosa per la varianza che è un'operazione in cui compaiono v.a. elevate al quadrato. Il seguente Teorema precisa alcune importanti proprietà della varianza.

Teorema 6.4. *Se X è una v.a. e a e b due numeri si ha*

$$\mathbf{Var}(aX + b) = a^2 \mathbf{Var}(X); \quad (6.18)$$

se X e Y sono due v.a. si ha

$$\mathbf{Var}(X + Y) = \mathbf{Var}(X) + \mathbf{Var}(Y) + 2\mathbf{Cov}(X, Y), \quad (6.19)$$

e quindi, solo se X e Y sono non correlate potremo scrivere

$$\mathbf{Var}(X + Y) = \mathbf{Var}(X) + \mathbf{Var}(Y). \quad (6.20)$$

Dimostrazione: Posto $\mu = \mathbf{E}(X)$ da (6.5) si ha $\mathbf{E}(aX + b) = a\mu + b$; dalla definizione (6.7) e dal Teorema 6.1 si ha quindi immediatamente la (6.18)

$$\mathbf{Var}(aX + b) = \mathbf{E}[(aX - a\mu)^2] = \mathbf{E}[a^2(X - \mu)^2] = a^2 \mathbf{E}[(X - \mu)^2] = a^2 \mathbf{Var}(X).$$

Per dimostrare la relazione (6.19), posto come al solito $\mu = \mathbf{E}(X)$ e $\nu = \mathbf{E}(Y)$, da (6.6) si ha $\mathbf{E}(X + Y) = \mu + \nu$, e quindi dalle definizioni (6.7) e (6.8) applicando il Teorema 6.1 si ha

$$\begin{aligned} \mathbf{Var}(X + Y) &= \mathbf{E} [((X + Y) - (\mu + \nu))^2] = \mathbf{E} [((X - \mu) + (Y - \nu))^2] \\ &= \mathbf{E}[(X - \mu)^2] + \mathbf{E}[(Y - \nu)^2] + 2\mathbf{E}[(X - \mu)(Y - \nu)] \\ &= \mathbf{Var}(X) + \mathbf{Var}(Y) + 2\mathbf{Cov}(X, Y). \end{aligned}$$

Se però X e Y sono non correlate si ha $\mathbf{Cov}(X, Y) = 0$ e quindi si ottiene (6.20). \square

Con qualche complicazione di calcolo si può dimostrare anche una generalizzazione della relazione (6.19): date n v.a. X_1, \dots, X_n si ha

$$\begin{aligned} \mathbf{Var}(X_1 + \dots + X_n) &= \sum_{i,j=1}^n \mathbf{Cov}(X_i, X_j) \\ &= \mathbf{Var}(X_1) + \dots + \mathbf{Var}(X_n) + \sum_{i \neq j} \mathbf{Cov}(X_i, X_j); \quad (6.21) \end{aligned}$$

se poi le X_i sono anche non correlate si ha

$$\mathbf{Var}(X_1 + \dots + X_n) = \mathbf{Var}(X_1) + \dots + \mathbf{Var}(X_n). \quad (6.22)$$

Teorema 6.5. Se X è una v.a. con attesa μ e varianza σ^2 , allora la v.a.

$$Y = \frac{X - \mu}{\sigma}$$

avrà attesa 0 e varianza 1 e prenderà il nome di **v.a. standardizzata**

Dimostrazione: Dalla relazione (6.5) si ha

$$\mathbf{E}(Y) = \frac{1}{\sigma} [\mathbf{E}(X) - \mu] = 0;$$

tenendo conto poi di (6.18) e del fatto che $\mathbf{E}(X - \mu) = 0$ si ha

$$\mathbf{Var}(Y) = \frac{\mathbf{Var}(X - \mu)}{\sigma^2} = \frac{\mathbf{E}[(X - \mu)^2]}{\sigma^2} = \frac{\mathbf{Var}(X)}{\sigma^2} = 1$$

il che conclude la dimostrazione. \square

Ad esempio, con questo tipo di terminologia, si noterà che nel Teorema 5.9 la v.a. Y non è altro che la v.a. \bar{X} standardizzata.

Definizione 6.3. Chiameremo rispettivamente **momento di ordine k** e **momento centrato di ordine k** di una v.a. X le quantità

$$m_k = \mathbf{E}(X^k), \quad \mu_k = \mathbf{E}[(X - \mathbf{E}(X))^k].$$

Ovviamente $m_1 = \mathbf{E}(X)$, e $\mu_2 = \mathbf{Var}(X)$. Chiameremo inoltre **asimmetria** e **curtosi** di X rispettivamente le quantità

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}, \quad \gamma_2 = \frac{\mu_4}{\mu_2^2}.$$

6.2 Esempi di attese e varianze

Osserviamo innanzitutto che il valore dell'attesa e della varianza dipendono soltanto dalla legge di una v.a. e pertanto che v.a. diverse hanno la stessa attesa e la stessa varianza se sono identicamente distribuite. Nel seguito esamineremo alcuni semplici esempi di calcolo di queste quantità.

Inizieremo con le v.a. discrete: qualunque v.a. distribuita secondo una legge di Bernoulli $B(1, p)$ assume solo i due valori 0 e 1 con probabilità rispettivamente $1 - p$ e p . Pertanto dalla definizione si ha

$$\begin{aligned} \mathbf{E}(X) &= 0 \cdot (1 - p) + 1 \cdot p = p \\ \mathbf{E}(X^2) &= 0^2 \cdot (1 - p) + 1^2 \cdot p = p \end{aligned}$$

e quindi da (6.14) si ha

$$\mathbf{E}(X) = p, \quad \mathbf{Var}(X) = p(1 - p). \quad (6.23)$$

D'altra parte per il Teorema 5.3 ogni v.a. Binomiale $B(n, p)$ è la somma $X_1 + \dots + X_n$ di n v.a. indipendenti e tutte identicamente distribuite secondo la legge di Bernoulli $B(1, p)$; quindi da (6.6), (6.22) e (6.23) si ottiene per una **legge Binomiale** $B(n, p)$

$$\boxed{\mathbf{E}(X) = np, \quad \mathbf{Var}(X) = np(1 - p).} \quad (6.24)$$

Si noti che (6.23) è un caso particolare di (6.24) per $n = 1$. Per una **legge di Poisson** $\mathcal{P}(\lambda)$ si ha invece

$$\boxed{\mathbf{E}(X) = \lambda, \quad \mathbf{Var}(X) = \lambda.} \quad (6.25)$$

Proveremo soltanto la prima relazione, dato che la dimostrazione della seconda ripercorre gli stessi passaggi con qualche piccola complicazione nei calcoli: ricordando che una v.a. di Poisson assume tutti i valori interi $k = 0, 1, \dots$, da (6.2) e (5.8), e omettendo il primo termine della serie perché nullo, si ha

$$\mathbf{E}(X) = \sum_{k=0}^{\infty} k p_k = \sum_{k=1}^{\infty} k p_k = \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!};$$

riscaldando ora l'indice di somma con la posizione $j = k - 1$ e ricordando il risultato notevole

$$\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda},$$

si ottiene infine la (6.25):

$$\mathbf{E}(X) = e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^{j+1}}{j!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda.$$

Per le v.a. continue ci limiteremo a ricordare solo alcuni risultati rilevanti. Innanzitutto per una **legge Uniforme** $\mathcal{U}([a, b])$ si prova che

$$\boxed{\mathbf{E}(X) = \frac{a+b}{2}, \quad \mathbf{Var}(X) = \frac{(b-a)^2}{12}.} \quad (6.26)$$

Per una **legge Normale** $\mathcal{N}(\mu, \sigma^2)$ si prova poi che

$$\boxed{\mathbf{E}(X) = \mu, \quad \mathbf{Var}(X) = \sigma^2.} \quad (6.27)$$

Questo attribuisce un significato probabilistico preciso ai due parametri μ e σ che finora avevano avuto solo un ruolo puramente analitico nella descrizione del grafico della fd di $\mathcal{N}(\mu, \sigma^2)$. Si osservi che in particolare per una **legge Normale standard** $\mathcal{N}(0, 1)$ si avrà

$$\mathbf{E}(X) = 0, \quad \mathbf{Var}(X) = \mathbf{E}(X^2) = 1 \quad (6.28)$$

come per tutte le v.a. standardizzate. Per una **legge Chi quadro** $\chi^2(n)$ si prova infine che

$$\boxed{\mathbf{E}(X) = n, \quad \mathbf{Var}(X) = 2n.} \quad (6.29)$$

In particolare la prima di (6.29) si dimostra facilmente ricordando che in base al Teorema 5.6 una v.a. X con legge $\chi^2(n)$ coincide con la somma $X_1^2 + \dots + X_n^2$ se le X_k sono v.a. indipendenti e tutte Normali standard $\mathcal{N}(0, 1)$. Segue allora facilmente da (6.6) e da (6.28) che

$$\mathbf{E}(X) = \mathbf{E}(X_1^2) + \dots + \mathbf{E}(X_n^2) = \mathbf{Var}(X_1) + \dots + \mathbf{Var}(X_n) = n.$$

Capitolo 7

Teoremi limite

7.1 Legge dei Grandi Numeri

Come vedremo meglio nella parte di Statistica Inferenziale, lo scopo della statistica è quello di estrarre delle informazioni sulla distribuzione di una v.a. X a partire da un certo numero di osservazioni empiriche (misure). Qui anticiperemo qualche idea per spiegare il ruolo di alcuni importanti risultati del Calcolo delle Probabilità nell'analisi dei dati sperimentali.

Esempio 7.1. Supponiamo di avere una popolazione di individui di due tipi A e B : maschi e femmine di una specie biologica; divisione in due partiti in un gruppo di persone; palline di due colori in un'urna e così via. Supponiamo inoltre che la proporzione del gruppo A sia p e quella del gruppo B sia $1 - p$ (ovviamente con $0 < p < 1$). Se p è sconosciuta e la popolazione è piccola e tutta disponibile per un'osservazione, il valore di p potrà essere ottenuto semplicemente contando tutti gli individui di tipo A e B . In generale però succede che la popolazione sia molto grande (come nel caso dei cittadini di uno stato) o anche semplicemente non tutta disponibile per un'osservazione (come nel caso degli individui di una specie biologica). Per attribuire un valore attendibile, ossia per *stimare*, la proporzione p si procederà allora come nei sondaggi pre-elettorali: si estrae un campione casuale di n individui e li si esamina contando il numero N_A di quelli di tipo A . È piuttosto intuitivo ritenere che $\bar{p} = N_A/n$ rappresenti una stima accettabile di p , e che tale stima sia tanto più attendibile quanto più grande è il numero n . Si noti che la differenza fra p e \bar{p} non sta solo nel fatto che i due valori numerici in generale non coincidono, ma anche e soprattutto nel fatto che p è un ben determinato (ancorché sconosciuto) numero, mentre \bar{p} è una v.a. Infatti ogni volta che ripetiamo l'estrazione casuale degli n individui da esaminare otterremo normalmente un diverso valore di \bar{p} . Da un punto di vista più formale la situazione può essere descritta nel modo seguente: si considerano n v.a. indipendenti X_1, \dots, X_n che rappresentano il nostro *campione casuale* nel senso che

$$X_k = \begin{cases} 1 & \text{se l'individuo } k\text{-mo è di tipo } A, \\ 0 & \text{altrimenti,} \end{cases} \quad k = 1, 2, \dots, n.$$

Nelle condizioni del nostro problema le X_k sono quindi tutte v.a. di Bernoulli $B(1, p)$ dato

che $\mathbf{P}\{X_k = 1\} = p$. In altri termini le X_k sono n v.a. indipendenti tutte distribuite come una generica v.a. X di Bernoulli $B(1, p)$. Inoltre è chiaro che per noi $N_A = X_1 + \dots + X_n$, per cui la nostra stima di p sarà data dalla v.a.

$$\bar{p} = \frac{N_A}{n} = \frac{X_1 + \dots + X_n}{n} = \bar{X}, \quad (7.1)$$

cioè dalla media aritmetica dei valori delle X_k . Si noterà ora dal Teorema 5.3 la v.a. N_A sarà una Binomiale $B(n, p)$. Siccome inoltre da (6.23) si ha $\mathbf{E}(X) = p$, il nostro problema assume la seguente forma: *data la v.a. di Bernoulli X con legge $B(1, p)$ vogliamo stimarne il valore d'attesa $\mathbf{E}(X) = p$* . La stima (7.1), la media aritmetica di un campione casuale di valori di X , rappresenta la nostra soluzione intuitiva, e sarà argomento di questa sezione fornire qualche strumento matematico per motivare meglio questa scelta.

Dall'esempio che precede si ricava innanzitutto l'idea che in statistica esistono due tipi di quantità: quelle *teoriche*, e quelle *empiriche* associate alle prime con lo scopo di stimarle. Così in particolare il valore d'attesa (o media) $\mu = \mathbf{E}(X)$ di una v.a. è una quantità teorica (essa dipende dal modello matematico del nostro esperimento e si calcola dalla legge di X), mentre la **media aritmetica**

$$\boxed{\bar{X} = \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_{k=1}^n X_k.} \quad (7.2)$$

calcolata da un campione casuale di v.a. distribuite come X è la corrispondente quantità empirica (essa cioè si calcola dalle misure sperimentali effettuate sul nostro sistema), e noi abbiamo visto che \bar{X} viene usata per stimare $\mathbf{E}(X)$. Ricorderemo inoltre che mentre le quantità teoriche sono tipicamente dei numeri (come μ), le quantità empiriche usate per stimarle sono delle v.a. (come \bar{X}): questo serve a rendere conto della inevitabile variabilità delle misure sperimentali. Naturalmente le medie, teoriche ed empiriche, non sono le uniche quantità rilevanti in statistica: ad esempio sarà anche importante stimare delle varianze. In questo caso, sempre seguendo lo stesso criterio intuitivo di sostituire il calcolo empirico di medie aritmetiche al calcolo teorico dei valori d'attesa, per stimare una varianza $\sigma^2 = \mathbf{Var}(X)$ si potrà utilizzare ad esempio la v.a. **varianza campionaria** (vedi Teorema 1.5)

$$\boxed{S_c^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2 = \overline{X^2} - \bar{X}^2 = \frac{1}{n} \sum_{k=1}^n X_k^2 - \bar{X}^2.} \quad (7.3)$$

Vedremo nel seguito che sarà necessario introdurre qualche correzione della espressione (7.3), ma per il momento la useremo senza ulteriori discussioni. Si noterà infine che sia \bar{X} che S_c^2 coincidono numericamente con le analoghe quantità introdotte nella parte di Statistica Descrittiva con una importante differenza: mentre nella statistica descrittiva medie e varianze erano calcolate su tutta la popolazione, qui esse sono calcolate su un campione casuale estratto dalla popolazione teorica

(cioè dalla v.a. X) che in generale non è tutta disponibile per le osservazioni. Per questo motivo ora le quantità (7.2) e (7.3) sono delle v.a.

Abbiamo osservato nella discussione dell'Esempio 7.1 che, almeno intuitivamente, l'affidabilità delle stime effettuate aumentava con l'aumentare del numero n delle osservazioni effettuate. Il significato preciso di questa affermazione è il contenuto della *Legge dei Grandi Numeri* che ora enunceremo, ancorché in una forma semplificata. Siccome si tratta di un *teorema limite* per $n \rightarrow \infty$ sarà ovviamente necessario pensare di potere, almeno ipoteticamente, eseguire un numero infinito di misure: cioè dovremo supporre che le X_k formino una intera successione con $k \in \mathbf{N}$. Inoltre ora anche \bar{X} e S_c^2 varieranno con n , cioè formeranno delle successioni, anche se, per non appesantire la notazione, noi eviteremo in questi casi di indicare esplicitamente l'indice n . Infine, siccome si tratta di successioni di v.a. e non di numeri, sarà necessario precisare – anche se in maniera semplificata – in che senso parleremo di convergenza.

Definizione 7.1. *Data una successione di v.a. Z_n con $n \in \mathbf{N}$, diremo che per $n \rightarrow \infty$ essa converge verso il numero a , e scriveremo*

$$Z_n \xrightarrow{n} a, \quad (7.4)$$

quando

$$\lim_n \mathbf{E}(Z_n) = a, \quad \lim_n \mathbf{Var}(Z_n) = 0. \quad (7.5)$$

Questa definizione deriva il suo significato da quello di varianza: se la successione delle varianze di Z_n converge verso 0, questo vuol dire che la distribuzione di Z_n è sempre più concentrata attorno al valore d'attesa $\mathbf{E}(Z_n)$; ma a sua volta la successione $\mathbf{E}(Z_n)$ si avvicina sempre più al numero a . Al limite per $n \rightarrow \infty$ potremo quindi dire che le Z_n diventano v.a. che assumono invariabilmente il valore a .

Teorema 7.1. Legge dei Grandi Numeri (LGN): *Se X_k con $k \in \mathbf{N}$ è una successione di v.a. indipendenti, tutte con la stessa attesa μ e la stessa varianza σ^2 , avremo*

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{n} \mu, \quad (7.6)$$

$$S_c^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2 \xrightarrow{n} \sigma^2. \quad (7.7)$$

Dimostrazione: Per dimostrare (7.6) osserviamo innanzitutto che in base a (6.5) e (6.6) si ha

$$\mathbf{E}(\bar{X}) = \frac{\mathbf{E}(X_1) + \dots + \mathbf{E}(X_n)}{n} = \frac{n \cdot \mu}{n} = \mu;$$

inoltre ricordando che le X_k sono indipendenti (e quindi non correlate) da (6.18) e (6.22) si ha anche

$$\mathbf{Var}(\bar{X}) = \frac{\mathbf{Var}(X_1) + \dots + \mathbf{Var}(X_n)}{n^2} = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n} \xrightarrow{n} 0,$$

e quindi $\bar{X} \xrightarrow{n} \mu$ nel senso della definizione 7.1. Per dimostrare (7.7) cominciamo invece con l'osservare che

$$\begin{aligned} S_c^2 &= \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2 = \frac{1}{n} \sum_{k=1}^n [(X_k - \mu) + (\mu - \bar{X})]^2 \\ &= \frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2 + \frac{2}{n} \sum_{k=1}^n (X_k - \mu)(\mu - \bar{X}) + \frac{1}{n} \sum_{k=1}^n (\mu - \bar{X})^2 \\ &= \frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2 + \frac{2}{n} (\mu - \bar{X}) \sum_{k=1}^n (X_k - \mu) + \frac{n \cdot (\mu - \bar{X})^2}{n} \\ &= \frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2 + \frac{2}{n} (\mu - \bar{X})(n\bar{X} - n\mu) + (\mu - \bar{X})^2 \\ &= \frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2 + 2(\mu - \bar{X})(\bar{X} - \mu) + (\mu - \bar{X})^2 \\ &= \frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2 - (\mu - \bar{X})^2 = \frac{1}{n} \sum_{k=1}^n Z_k - (\mu - \bar{X})^2 \\ &= \bar{Z} - (\mu - \bar{X})^2 \end{aligned}$$

dove abbiamo posto per comodità $Z_k = (X_k - \mu)^2$. Siccome dalle definizioni si ha $\mathbf{E}(Z_k) = \mathbf{Var}(X_k) = \sigma^2$, applicando la (7.6) alla successione delle Z_k risulta $\bar{Z} \xrightarrow{n} \sigma^2$; e d'altra parte, sempre per (7.6), si ha anche che $(\mu - \bar{X})^2 \xrightarrow{n} 0$. In conclusione quindi potremo dire che $S_c^2 \xrightarrow{n} \sigma^2$, cioè la (7.7). \square

7.2 Teorema Limite Centrale

La LGN fornisce delle preziose indicazioni sulla stima dei parametri statistici, ma non permette di valutare quantitativamente l'attendibilità di queste stime. Sarebbe molto utile invece avere degli strumenti che ci permettessero di dire, ad esempio, con quale probabilità il *valore stimato*, che è sempre una v.a., coincide con – o almeno cade ad una certa distanza da – il *valore vero* del parametro in esame. In pratica abbiamo bisogno di poter dire che errore si commette approssimando un parametro con una stima. È peraltro evidente che per fare questo bisognerebbe avere o delle informazioni sulla distribuzione della v.a. usata per stimare il parametro, o almeno una buona approssimazione di tale legge. Sarà utile allora ricordare un altro

importante risultato noto come *Teorema Limite Centrale* che stabilisce una certa forma di universalità delle distribuzioni normali: se una v.a. X è la somma di un gran numero di piccole v.a. allora essa è approssimativamente distribuita secondo una legge normale. Questa proprietà è particolarmente importante in molti settori della statistica

Esempio 7.2. Supponiamo di dover misurare una certa quantità il cui valore vero sconosciuto è μ . A causa degli inevitabili errori sperimentali, però, il risultato della misura non è in generale coincidente con μ , anzi sarà diverso ogni volta che ripeteremo la misura: si tratta cioè di una vera e propria v.a. X . Per stimare μ useremo allora una campione casuale di n misure, ma per poter ricavare delle informazioni sull'attendibilità dalla stima dobbiamo sapere quale è la legge di X . Supponendo per semplicità stiamo che non vi siano errori sistematici, potremmo allora pensare che il valore di X sia composto di due parti: una deterministica costituita dal valore vero μ cercato, e una aleatoria Y dovuta agli errori sperimentali, in modo che complessivamente $X = \mu + Y$. Siccome Y può essere pensata come la somma di innumerevoli piccoli disturbi dovuti alle condizioni sperimentali, e senza una direzione privilegiata, il Teorema Limite Centrale ci autorizza a ritenere che Y segue una legge normale con media nulla $\mathcal{N}(0, \sigma^2)$. In base al Teorema 5.5 la v.a. X risulta quindi distribuita secondo la legge normale $\mathcal{N}(\mu, \sigma^2)$. Abbiamo così ricavato delle utili informazioni sulla forma della legge di X , anche se restano da stimare i due parametri μ e σ^2 .

Il Teorema Limite Centrale, che ora enunceremo senza dimostrazione, prende la forma di una affermazione sui limiti di successioni di FDC: questo permette anche di darne una versione unificata per v.a. discrete e continue. Il limite è qui inteso nel senso ordinario del *limite punto a punto* di una successione di funzioni e non ha bisogno di particolari nuove definizioni. Ricorderemo infine che come in (5.16) con il simbolo $\Phi(t)$ indichiamo la FDC della legge normale standard $\mathcal{N}(0, 1)$.

Teorema 7.2. Teorema Limite Centrale (TLC): *Data una successione X_k con $k \in \mathbf{N}$ di v.a. indipendenti con la stessa media μ e la stessa varianza $\sigma^2 < +\infty$, posto $S_n = X_1 + \dots + X_n = n\bar{X}$ e*

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$$

e detta $F_n(t) = \mathbf{P}\{S_n^* \leq t\}$ la successione delle FDC di S_n^* , risulta

$$\lim_n F_n(t) = \Phi(t)$$

dove $\Phi(t)$ è la FDC della legge normale standard $\mathcal{N}(0, 1)$.

In pratica, per n abbastanza grande, questo Teorema ci autorizza ad approssimare con la legge normale standard $\mathcal{N}(0, 1)$ la legge delle somme *standardizzate* S_n^* . Si vede facilmente infatti da (6.6) che $\mathbf{E}(S_n) = n\mu$; inoltre dall'indipendenza delle X_k

e da (6.22) si ha $\mathbf{Var}(S_n) = n\sigma^2$: segue allora facilmente da (6.5) che $\mathbf{E}(S_n^*) = 0$ e $\mathbf{Var}(S_n^*) = 1$, cioè che S_n^* è una v.a. standardizzata. Si noti la generalità di questo risultato: in esso l'unica ipotesi che si fa sulle v.a. X_k è che esse abbiano media e varianza finite. Per il resto la loro legge può essere del tutto arbitraria: un fatto molto utile soprattutto se tale legge non ci è nota. Così, se le X_k sono generiche v.a. indipendenti, tutte con media μ varianza σ^2 , con le notazioni del TLC e per n abbastanza grande avremo l'approssimazione

$$\mathbf{P}\{S_n \leq x\} = \mathbf{P}\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{x - n\mu}{\sigma\sqrt{n}}\right\} = \mathbf{P}\left\{S_n^* \leq \frac{x - n\mu}{\sigma\sqrt{n}}\right\} \simeq \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right).$$

Quando si usano relazioni di questo tipo si parla anche di *approssimazione normale*: la loro importanza sta nel fatto che, sotto le condizioni del TLC, essa permette di calcolare i valori della FDC di somme di generiche v.a. indipendenti mediante le tavole dei valori della FDC normale standard. Si noti inoltre, pur essendo $\mathcal{N}(0, 1)$ la legge di una v.a. continua, il teorema resta vero anche se le X_k sono v.a. discrete; in questo caso però, come vedremo nell'esempio seguente, sarà bene usare dei piccoli accorgimenti per migliorare le approssimazioni.

Esempio 7.3. Torniamo a considerare il modello esaminato nell'Esempio 7.1 supponendo per semplicità che le proporzioni degli individui di tipo A e B siano uguali, in modo che $p = \frac{1}{2}$. Estraiamo un campione casuale di $n = 100$ individui, e poniamoci il problema di calcolare la probabilità che in esso ve ne siano più di 60 di tipo A . Con le notazioni dell'Esempio 7.1 sappiamo che tale numero $N_A = X_1 + \dots + X_{100}$ è una v.a. Binomiale $B(100, \frac{1}{2})$, per cui la probabilità richiesta si scrive come

$$\mathbf{P}\{N_A > 60\} = \sum_{k=61}^{100} \binom{100}{k} \frac{1}{2^{100}}.$$

Il valore numerico di questa probabilità non è però facilmente ricavabile, se non con l'aiuto di qualche macchina calcolatrice. L'approssimazione normale ci permette invece di ottenere una approssimazione con il semplice uso delle Tavole numeriche della FDC Normale standard Φ nell'Appendice D.1. Infatti osserviamo che la v.a. N_A non è altro che la somma S_n del Teorema 7.2 con $n = 100$; detta allora $F_{100}(x)$ la FDC della N_A standardizzata S_{100}^* , e ricordando che le X_k sono tutte Bernoulli $B(1, \frac{1}{2})$, per cui da (6.23) si ha

$$\begin{aligned} \mu &= \mathbf{E}(X_k) = p = \frac{1}{2} \\ \sigma^2 &= \mathbf{Var}(X_k) = p(1-p) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \end{aligned}$$

potremo scrivere

$$\begin{aligned} \mathbf{P}\{N_A > 60\} &= 1 - \mathbf{P}\{N_A \leq 60\} = 1 - \mathbf{P}\left\{\frac{N_A - 100\mu}{\sigma\sqrt{100}} \leq \frac{60 - 100\mu}{\sigma\sqrt{100}}\right\} \\ &= 1 - F_{100}\left(\frac{60 - 100\mu}{\sigma\sqrt{100}}\right) = 1 - F_{100}(2) \simeq 1 - \Phi(2) = 0.02275 \end{aligned}$$

avendo ricavato dalle Tavole che $\Phi(2) = 0.97725$. Il calcolo esatto di $\mathbf{P}\{N_A > 60\}$ con la distribuzione Binomiale non è però impossibile e fornisce il valore 0,01760 che è un po' diverso dal valore trovato con l'approssimazione normale. Questa discrepanza è dovuta al fatto che stiamo applicando il TLC a v.a. discrete come le Binomiali: infatti bisogna osservare che N_A assume solo valori interi, ad esempio ... , 59, 60, 61, ..., e che quindi il valore di $\mathbf{P}\{N_A \leq x\}$ resta sempre lo stesso per le x comprese fra due interi consecutivi. In particolare

$$F_{100} \left(\frac{x - 100\mu}{\sigma\sqrt{100}} \right)$$

resta costante per $60 \leq x < 61$. Viceversa, come si vede dalle Tavole dell'Appendice D.1, nel medesimo intervallo il valore approssimante

$$\Phi \left(\frac{x - 100\mu}{\sigma\sqrt{100}} \right)$$

varia con continuità da $\Phi(2) = 0.97725$ a $\Phi(2.2) = 0.98610$. È intuitivo, allora, che la scelta migliore per il valore di x non è l'estremo inferiore 60 dell'intervallo $[60, 61)$, ma il suo punto di mezzo $x = 60.5$. Infatti eseguendo di nuovo il calcolo si ha ora

$$\begin{aligned} \mathbf{P}\{N_A > 60\} &= 1 - \mathbf{P}\{N_A \leq 60\} = 1 - \mathbf{P}\{N_A \leq 60.5\} \\ &= 1 - \mathbf{P} \left\{ \frac{N_A - 100\mu}{\sigma\sqrt{100}} \leq \frac{60.5 - 100\mu}{\sigma\sqrt{100}} \right\} = 1 - F_{100} \left(\frac{60.5 - 100\mu}{\sigma\sqrt{100}} \right) \\ &= 1 - F_{100}(2.1) \simeq 1 - \Phi(2.1) = 0.01786 \end{aligned}$$

che costituisce una migliore approssimazione del valore corretto 0.01760. Come regola generale, quindi, se S_n è una v.a. a valori interi, si ottiene una migliore approssimazione normale se invece di calcolare $\mathbf{P}\{S_n \leq k\}$ si calcola $\mathbf{P}\{S_n \leq k + \frac{1}{2}\}$, dove k è un intero.

Resta infine da fare qualche osservazione sul numero minimo n per il quale l'approssimazione normale del Teorema 7.2 si possa considerare applicabile alla somma S_n . A questo proposito va subito detto che non ci sono risultati generali, ma solo delle regole empiriche, piuttosto variabili secondo le fonti consultate. In particolare l'approssimazione normale si considera applicabile quando n vale almeno 30–50. Ancora una volta però va adoperata qualche cautela nel caso in cui S_n è una v.a. discreta. Se, come nell'Esempio 7.3, S_n è Binomiale $B(n, p)$ si può mostrare con degli esempi che per valori prossimi a 0 di p oppure di $1 - p$ l'approssimazione normale non è molto buona anche per valori di n considerati grandi secondo il nostro precedente criterio. In questi casi il criterio deve tenere conto anche del valore di p e in generale si considera applicabile l'approssimazione normale quando sono verificate *ambidue* le condizioni $np \geq 5$, e $n(1 - p) \geq 5$. Quindi per soddisfare questo criterio, se ad esempio $p = 0.05$ si deve avere $n \geq 100$; se invece $p = 0.01$ si deve avere $n \geq 500$.

Parte III
Statistica inferenziale

Capitolo 8

Stima di parametri

8.1 Stima puntuale

La LGN e il TLC sono gli strumenti principali con i quali affronteremo i problemi di stima ai quali abbiamo accennato nel Capitolo 7. Ricorderemo che lo scopo è quello di utilizzare i risultati X_1, \dots, X_n di n misure di una data quantità aleatoria X per ottenere dei numeri che approssimino in maniera affidabile il valore di qualche parametro sconosciuto della distribuzione di X (una proporzione p come nell'Esempio 7.1 o un valore d'attesa μ come nell'Esempio 7.2). Si tratta quindi di un tipico argomento di statistica inferenziale visto che si vogliono ricavare informazioni generali su una v.a. X a partire da un particolare insieme di misure. Prima di dare veste generale al problema della stima, sarà però utile discutere ancora qualche esempio in continuità con gli Esempi 7.1, 7.2 e 7.3 del Capitolo 7.

Esempio 8.1. (Stima di una media) Si voglia determinare il peso medio degli individui adulti di sesso maschile di una determinata specie di animali. In questo caso descriveremo il peso degli animali con una v.a. X e il nostro problema sarà quello di determinare $\mu = \mathbf{E}(X)$. Non abbiamo a priori informazioni sulla legge di X anche se potremmo ragionevolmente fare delle ipotesi (almeno come approssimazione) su tale legge. La nostra richiesta, però, è solo quella di stimare μ e noi seguiremo una procedura simile a quella dell'Esempio 7.1: estrarremo un campione casuale di n individui, ne misureremo il peso e prenderemo la media aritmetica di tali misure come stima di μ . Da un punto di vista formale ciò vuol dire prendere n v.a. indipendenti X_1, \dots, X_n che rappresentano il peso degli animali del campione casuale – quindi tali v.a. sono tutte distribuite con la stessa legge (sconosciuta) di X e hanno lo stesso valore d'attesa μ – e calcolarne la media aritmetica \bar{X} come in (7.2). Ovviamente \bar{X} è una v.a. il cui valore cambia al variare del campione aleatorio estratto: noi prenderemo il valore ricavato dal campione a nostra disposizione come una stima della media μ fidandoci del fatto che, in base alla LGN Teorema 7.1, se il campione è sufficientemente grande tale stima sarà ragionevolmente vicina al valore vero μ . Si noti come la stima di una proporzione dell'Esempio 7.1 non sia che un caso particolare di stima di una media.

Esempio 8.2. (Stima di un parametro) In altri casi sulla base di qualche ragionamento è possibile ricavare la forma qualitativa della distribuzione di una data v.a. X , i cui parametri sono però delle quantità incognite da stimare con una misura. Così, riprendendo la discussione dell'Esempio 5.7, potremmo dire che il numero aleatorio X di telefonate che arrivano ad un centralino telefonico in un generico intervallo di tempo T è una v.a. con legge di Poisson $\mathcal{P}(\lambda)$. Il parametro λ però è sconosciuto: esso ovviamente dipende dal particolare centralino studiato e dal particolare periodo della giornata considerato, e il nostro problema è ora quello di stimarlo. Ricorderemo a questo proposito che in base all'equazione (6.25) il parametro λ è anche il valore d'attesa $\mathbf{E}(X)$ della nostra v.a. di Poisson. Pertanto la maniera più naturale per stimare λ consisterà nel misurare in n giorni diversi il numero di telefonate pervenute al dato centralino in un ben determinato periodo di tempo, e nel calcolare poi la media aritmetica di queste misure. Ancora una volta avremo un campione casuale composto di n v.a. indipendenti X_1, \dots, X_n tutte con legge $\mathcal{P}(\lambda)$, a partire dal quale calcoliamo la v.a. \bar{X} (7.2) fidandoci del fatto che in base alla LGN, se n è abbastanza grande, il valore numerico osservato non sarà molto diverso dal valore di λ .

Esempio 8.3. (Stima di una distribuzione discreta) Sia X una v.a. discreta della quale si voglia stimare la distribuzione teorica

$$p_k = \mathbf{P}\{X = k\}, \quad k = 0, 1, 2, \dots$$

Misureremo allora n volte la v.a. X ottenendo un campione X_j con $j = 1, 2, \dots, n$ di n v.a. indipendenti e tutte distribuite come X , nel senso che, comunque scelto j , avremo $\mathbf{P}\{X_j = k\} = p_k$ per ogni valore di k . Fissato un valore di k , le n v.a.

$$Y_j(k) = \begin{cases} 1 & \text{se } X_j = k \\ 0 & \text{altrimenti} \end{cases} \quad j = 1, 2, \dots, n$$

saranno allora tutte indipendenti e di Bernoulli $B(1, p_k)$, sicché da (6.23) risulterà anche $\mathbf{E}[Y_j(k)] = p_k$. Posto allora

$$N_k = \sum_{j=1}^n Y_j(k) = \text{numero delle } X_j \text{ che valgono } k$$

la (7.6) della LGN Teorema 7.1 ci autorizza ad affermare che al limite per $n \rightarrow \infty$ si ha

$$\bar{p}_k = \frac{N_k}{n} = \frac{1}{n} \sum_{j=1}^n Y_j(k) = \bar{Y}(k) \xrightarrow{n} p_k,$$

e quindi per grandi n potremo stimare i valori teorici p_k con i valori empirici $\bar{p}_k = N_k/n$.

Esempio 8.4. (Stima di una distribuzione continua) Estenderemo ora le idee esposte nell'Esempio 8.3 anche al caso in cui X sia una v.a. continua con fd $f(x)$. Anche in questo caso misureremo n volte X ottenendo un campione X_1, \dots, X_n di v.a. indipendenti e tutte distribuite come X ; questa volta però non ha molto senso domandarsi quante volte i valori

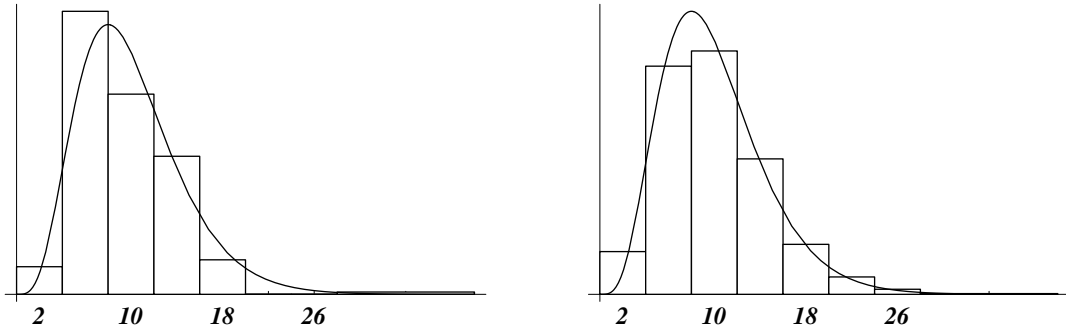


Figura 8.1: Approssimazione della fd di una legge $\chi^2(10)$ con istogrammi ottenuti da dati simulati. L'approssimazione migliora passando da $n = 100$ (primo grafico) a $n = 1000$ campioni (secondo grafico).

delle X_j coincidono con uno dei possibili valori x di X . Si dividerà invece l'intervallo dei valori di X in un numero finito di intervalli J_1, \dots, J_m e si costruirà l'*istogramma* delle misure, nel senso che si tratterà su ciascun intervallo J_k un rettangolo la cui area sarà proporzionale al numero delle X_j il cui valore cade in J_k . La LGN, Teorema 7.1, ci consente allora di affermare che il profilo di tale istogramma approssima l'andamento della fd $f(x)$ della v.a. X . Infatti, fissato un intervallo J_k , si definiscono le v.a.

$$Y_j(k) = \begin{cases} 1 & \text{se } X_j \in J_k \\ 0 & \text{altrimenti} \end{cases} \quad j = 1, 2, \dots, n$$

che seguono tutte una legge di Bernoulli $B(1, p_k)$ dove

$$p_k = \mathbf{P}\{Y_j(k) = 1\} = \mathbf{P}\{X_j \in J_k\} = \int_{J_k} f(x) dx.$$

In particolare avremo allora che $\mathbf{E}[Y_j(k)] = p_k$. Posto allora

$$N_k = \sum_{j=1}^n Y_j(k) = \text{numero delle } X_j \text{ che cadono in } J_k$$

la (7.6) del Teorema 7.1 ci autorizza ad affermare che al limite per $n \rightarrow \infty$ si ha

$$\bar{p}_k = \frac{N_k}{n} = \frac{1}{n} \sum_{j=1}^n Y_j(k) = \bar{Y}(k) \xrightarrow{n} p_k = \int_{J_k} f(x) dx,$$

e quindi per grandi n avremo

$$\bar{p}_k \simeq \int_{J_k} f(x) dx.$$

Ora, nel nostro istogramma le aree dei rettangoli sono proporzionali a N_k e quindi anche a $\bar{p}_k = N_k/n$; pertanto, indicata con $|J_k|$ la lunghezza dell'intervallo J_k , l'altezza del rettangolo k -mo è

$$\frac{\bar{p}_k}{|J_k|} \simeq \frac{1}{|J_k|} \int_{J_k} f(x) dx.$$

Ma se l'intervallo J_k è abbastanza piccolo si ha anche approssimativamente

$$\frac{\bar{p}_k}{|J_k|} \simeq \frac{1}{|J_k|} \int_{J_k} f(x) dx \simeq \frac{1}{|J_k|} \cdot f(x_k)|J_k| = f(x_k),$$

dove x_k è un opportuno punto dell'intervallo J_k . In conclusione quindi le altezze dell'istogramma seguiranno l'andamento della fd $f(x)$. Un esempio concreto di approssimazione di una fd continua è mostrato nella Figura 8.1: la curva continua rappresenta la fd della legge $\chi^2(10)$, mentre gli istogrammi sono stati ottenuti simulando dei campioni di v.a. che seguono tale legge. L'altezza delle barre $\bar{p}_k/|J_k|$ approssima i valori della fd, ma si noterà dai grafici che l'approssimazione migliora passando da un campione di $n = 100$ elementi (grafico a sinistra) a uno di $n = 1000$ elementi (grafico a destra).

Esempio 8.5. Vediamo ora con un esempio concreto come la LGN, cioè il Teorema 7.1, può essere usata nell'analisi dei dati empirici. Supponiamo di voler studiare la distribuzione della v.a. X che rappresenta il numero di figli maschi nelle famiglie di 12 figli. Ovviamente X è una v.a. discreta che assume solo i 13 valori interi $k = 0, 1, \dots, 12$. Per determinare la legge di X possiamo iniziare costruendo un modello sulla base delle ipotesi più elementari possibili:

1. in ogni famiglia gli esiti di parti differenti sono indipendenti;
2. in ogni parto nasce un maschio o una femmina con eguale probabilità $\frac{1}{2}$.

Definiamo allora le 12 v.a. X_1, \dots, X_{12}

$$X_j = \begin{cases} 1 & \text{se al parto } j\text{-mo nasce un figlio maschio,} \\ 0 & \text{se al parto } j\text{-mo nasce una figlia femmina,} \end{cases} \quad j = 1, 2, \dots, 12$$

ciascuna con legge di Bernoulli $B(1, \frac{1}{2})$: esse rappresentano gli esiti dei 12 parti e ovviamente $X = X_1 + \dots + X_{12}$. Siccome inoltre le X_j sono indipendenti per ipotesi, la X è Binomiale $B(12, \frac{1}{2})$ in base al Teorema 5.3. Questo ovviamente fissa le p_k teoriche:

$$p_k = \binom{12}{k} \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{12-k} = \binom{12}{k} \frac{1}{2^{12}}.$$

Confrontiamo ora questo modello con alcuni risultati sperimentali. Dall'esame dei dati anagrafici di $n = 6115$ famiglie con 12 figli si ricavano i numeri N_k (frequenze assolute) di famiglie con k figli maschi; si costruiscono quindi, come nell'Esempio 8.3, le stime $\bar{p}_k = N_k/n$ (frequenze relative) e le si confronta con i valori teorici p_k : in base alla LGN infatti le \bar{p}_k dovrebbero essere abbastanza prossime ai valori delle probabilità di avere k figli maschi. I risultati sono riportati nella Tabella 8.1 e nei grafici di Figura 8.2. Sebbene i dati teorici riproducano qualitativamente l'andamento dei dati empirici (si guardi il comportamento delle barre nere e bianche nella prima parte della Figura 8.2), ad un'analisi più accurata non sfugge che l'accordo non è particolarmente buono. Per mettere in risalto le differenze fra la distribuzione teorica e quella empirica possiamo calcolare i rapporti p_k/\bar{p}_k per vedere di quanto essi si discostano dal valore ottimale 1. I valori di questo rapporto sono riportati sempre nella Tabella 8.1 e nella seconda parte della Figura 8.2. Si

k	N_k	p_k	\bar{p}_k	p_k/\bar{p}_k
0	3	0.00024	0.00049	0.49764
1	24	0.00293	0.00392	0.74646
2	104	0.01611	0.01701	0.94743
3	286	0.05371	0.04677	1.14840
4	670	0.12085	0.10957	1.10298
5	1033	0.19336	0.16893	1.14462
6	1343	0.22559	0.21962	1.02715
7	1112	0.19336	0.18185	1.06330
8	829	0.12085	0.13557	0.89143
9	478	0.05371	0.07817	0.68712
10	181	0.01611	0.02960	0.54438
11	45	0.00293	0.00736	0.39811
12	7	0.00024	0.00114	0.21327

Tabella 8.1: Frequenze assolute N_k di famiglie con k figli maschi su $n = 6115$ famiglie con 12 figli. Le corrispondenti frequenze relative $\bar{p}_k = N_k/n$ sono usate per stimare la distribuzione della v.a. X numero di figli maschi nelle famiglie di 12 figli, e sono quindi confrontate con i valori teorici delle p_k di una legge Binomiale $B(12, \frac{1}{2})$ associata ad X in base ad alcune semplici ipotesi.

può notare così ad esempio che i dati teorici sono sistematicamente più grandi di quelli empirici per valori centrali di k , e soprattutto più piccoli per i valori estremi. Per migliorare l'accordo con i dati sperimentali potremmo allora modificare le ipotesi 1 e 2 che sono alla base del nostro modello teorico. Siccome dall'ipotesi 1 dipende la forma Binomiale della legge di X , una sua modifica produrrebbe sicuramente dei cambiamenti più profondi nella natura del nostro modello teorico. Partiremo allora dalla discussione, sicuramente più elementare, dell'ipotesi 2: se essa non fosse vera la X sarebbe sempre distribuita secondo una legge Binomiale $B(12, p)$, ma con un parametro p diverso da $\frac{1}{2}$. Dobbiamo allora trovare un modo per assegnare il valore di p in assenza dell'ipotesi 2. La strada più semplice potrebbe essere allora quella di determinare una stima \bar{p} di p a partire dai dati empirici a nostra disposizione. Se X è Binomiale $B(12, p)$ da (6.24) la sua attesa sarà $\mathbf{E}(X) = 12p$. D'altra parte, come negli Esempi iniziali di questa Sezione, noi possiamo, in base alla LGN, stimare il valore di $\mathbf{E}(X)$ con una media aritmetica dei valori sperimentali di X : siccome su $n = 6115$ osservazioni ogni valore k di X viene assunto N_k volte, nella formula (7.2) della media aritmetica ogni valore k si ripete N_k volte per cui avremo

$$\bar{X} = \frac{1}{n} \sum_{k=0}^{12} kN_k = \sum_{k=0}^{12} k\bar{p}_k.$$

La LGN ci dice ora che \bar{X} coincide con l'attesa di una Binomiale $B(12, \bar{p})$ con parametro \bar{p} ottenuto imponendo la relazione $12\bar{p} = \bar{X}$; quindi per p otterremo la stima

$$\bar{p} = \frac{1}{12n} \sum_{k=0}^{12} kN_k = 0.51922$$

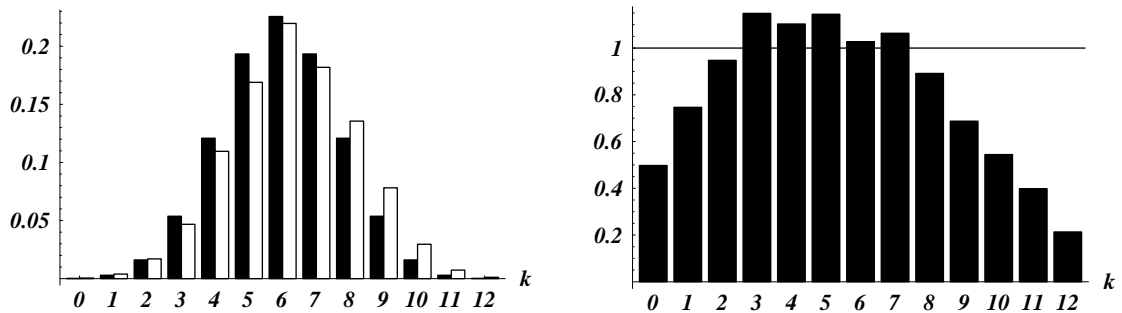


Figura 8.2: Confronto fra le frequenze teoriche p_k di $B(12, \frac{1}{2})$ (barre nere), e le frequenze empiriche \bar{p}_k (barre bianche). Il secondo grafico riporta i valori dei rapporti p_k/\bar{p}_k .

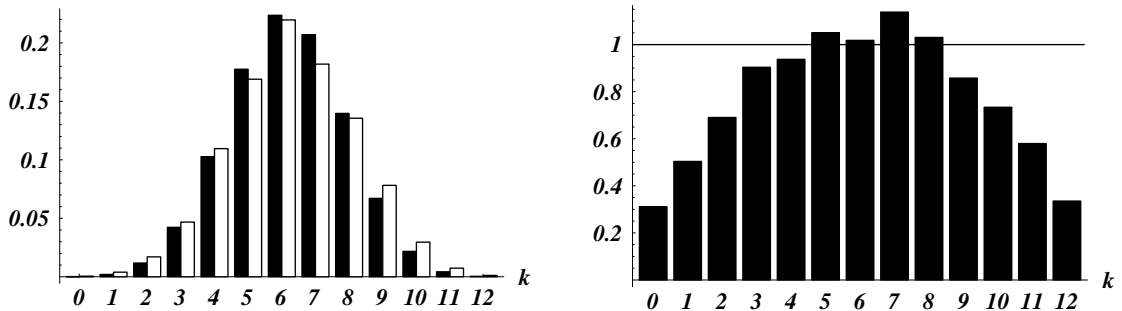


Figura 8.3: Confronto fra le frequenze teoriche p_k di $B(12, \bar{p})$ con $\bar{p} = 0.51922$ stimato a partire dai dati sperimentali (barre nere), e le frequenze empiriche \bar{p}_k (barre bianche). Il secondo grafico riporta i valori dei rapporti p_k/\bar{p}_k .

leggermente diversa dal valore $\frac{1}{2}$ del nostro modello iniziale. Possiamo allora ripetere il confronto fra dati sperimentali e teorici nel nuovo modello Binomiale con $\bar{p} = 0.51922$. Senza riportare i dati numerici in una nuova Tabella, eseguiremo il confronto sui nuovi grafici di Figura 8.3. Come si può notare i grafici sono ora leggermente diversi: ad esempio il grafico dei rapporti p_k/\bar{p}_k è un po' più simmetrico. Non si può però dire che l'accordo con i dati empirici sia sostanzialmente migliorato, visto che le previsioni teoriche presentano gli stessi problemi del modello precedente: una sistematica sottovalutazione dei dati sperimentali sulle code (valori estremi di k) della distribuzione. In conclusione questa discussione sembra suggerire che le difficoltà del nostro modello non sono nel valore del parametro p , ma nel carattere Binomiale della distribuzione di X . In questo caso dovremmo rivedere l'ipotesi 1 di indipendenza che è alla base del modello Binomiale. Da un punto di vista sostanziale questo è un fatto importante: la possibile non indipendenza degli esiti (maschio o femmina) dei diversi parti all'interno della stessa famiglia indicherebbe infatti che ci sono famiglie con una tendenza ad avere figli maschi, e famiglie con una tendenza ad avere figlie femmine. Discuteremo brevemente questo punto come ulteriore applicazione della LGN. Riprendendo le 12 v.a. X_1, \dots, X_{12} definite inizialmente supporremo ora che

ciascuna segua una legge di Bernoulli $B(1, \bar{p})$ dove $\bar{p} = 0.51922$ è il valore da noi stimato dai dati sperimentali. Ora, se le X_j fossero indipendenti la X sarebbe Binomiale $B(12, \bar{p})$ in base al Teorema 5.3, per cui da (6.24) dovrebbe risultare

$$\mathbf{Var}(X) = \mathbf{Var}(X_1) + \dots + \mathbf{Var}(X_{12}) = 12\bar{p}(1 - \bar{p}) = 2.996;$$

se invece le X_j non fossero indipendenti X non sarebbe più Binomiale, e noi non potremmo più usare la (6.24). Ricordiamo infatti che in questo caso la varianza della somma $X_1 + \dots + X_{12}$ non è più la semplice somma delle varianze delle X_j , ma deve essere invece usata l'equazione (6.21). La LGN, però, ci mette in condizione di stimare la varianza di X direttamente dai dati sperimentali tramite la (7.7) e senza fare ipotesi sull'indipendenza delle X_j . Infatti da (7.6) e (7.7), sempre ricordando che nelle $n = 6115$ misure ogni valore k di X si ripete N_k volte, abbiamo

$$\bar{X} = \frac{1}{n} \sum_{k=0}^{12} kN_k = 6.231; \quad S_c^2 = \frac{1}{n} \sum_{k=0}^{12} (k - \bar{X})^2 N_k = 3.489.$$

Come si vede il valore $S_c^2 = 3.489$ così stimato appare abbastanza diverso dal valore 2.996 ottenuto sommando le varianze delle X_j supposte indipendenti, e questo sembra suggerire ancora una volta una dipendenza reciproca delle X_j . Anzi, tenendo conto di questi valori e della relazione (6.21), si ha

$$\begin{aligned} \sum_{i \neq j} \mathbf{Cov}(X_i, X_j) &= \mathbf{Var}(X_1 + \dots + X_{12}) - [\mathbf{Var}(X_1) + \dots + \mathbf{Var}(X_{12})] \\ &= 3.489 - 2.996 = 0.493; \end{aligned}$$

cioè i dati sperimentali sembrano indicare che le covarianze degli esiti X_j di parti diversi siano complessivamente positive, ossia che le X_j siano positivamente correlate. Dato il significato del concetto di covarianza come discusso nella Sezione 6.1, questo suggerirebbe che vi sono famiglie con la tendenza a generare figli maschi, e famiglie con la tendenza a generare figlie femmine. L'esito dei parto potrebbe, cioè, non essere una faccenda puramente lasciata al caso: se in una famiglia si osservano nascite maschili (rispettivamente: femminili), la probabilità che anche le nascite successive siano maschili (rispettivamente: femminili) aumenta.

In tutti gli esempi precedenti siamo in presenza di una situazione generale che potremmo descrivere nel modo seguente: il nostro problema è quello di studiare una v.a. X la cui legge dipende da uno o più parametri numerici sconosciuti che noi chiameremo genericamente θ . Ad esempio se la legge di X è di Poisson $\mathcal{P}(\lambda)$ come nell'Esempio 8.2 il parametro tipicamente è $\theta = \lambda$; se invece la legge è normale come nell'Esempio 7.2 $\mathcal{N}(\mu, \sigma^2)$ il parametro sconosciuto θ in genere è la coppia di parametri $\theta = (\mu, \sigma)$ con attesa e varianza ambedue incognite. Il nostro scopo è allora quello di stimare il valore di θ o anche, più in generale, il valore di una qualche funzione $h(\theta)$: ad esempio nel caso di Poisson con $\theta = \lambda$ potremmo essere interessati a stimare $h(\lambda) = 1/\lambda$; nel caso normale con $\theta = (\mu, \sigma)$ potremmo essere interessati a stimare $h(\mu, \sigma) = \mu$, oppure $h(\mu, \sigma) = \mu^2$ e così via. È importante a questo punto introdurre le seguenti Definizioni:

Definizione 8.1. Diremo che n v.a. X_1, \dots, X_n costituiscono un **campione casuale** della v.a. X se esse sono indipendenti e tutte distribuite con la stessa legge di X ; si chiama poi **statistica** qualunque v.a. $T = t(X_1, \dots, X_n)$.

Definizione 8.2. Dato un campione X_1, \dots, X_n di v.a. con legge dipendente dal parametro θ , chiameremo **stimatore** di $h(\theta)$ una statistica T che sia una funzione $t(X_1, \dots, X_n)$ del campione dato. Diremo inoltre che T è uno **stimatore non distorto** se $\mathbf{E}_\theta(T) = h(\theta)$ (dove \mathbf{E}_θ indica che l'attesa è calcolata supponendo che il parametro sconosciuto abbia valore θ), e che esso è uno **stimatore consistente** se esso converge ad $h(\theta)$ per $n \rightarrow \infty$.

Definizione 8.3. Quando la stima di $h(\theta)$ viene effettuata con un solo valore dello stimatore T si parla di **stima puntuale** di $h(\theta)$.

In linea di principio, quindi, uno stimatore è una qualsiasi v.a. – ottenuta come funzione delle v.a. del campione – che viene usata per stimare $h(\theta)$. Ovviamente, però, uno stimatore T è buono se i suoi valori sono vicini al valore vero $h(\theta)$. Questa richiesta spiega l'introduzione dei concetti di stimatore non distorto e di stimatore consistente: essi infatti, per definizione, garantiscono in qualche modo che T prenda valori prossimi ad $h(\theta)$. Osserveremo inoltre che in tutti gli esempi trattati finora abbiamo eseguito delle stime puntuali: nell'Esempio 7.1 abbiamo stimato il parametro p di una $B(1, p)$, nell'Esempio 8.1 abbiamo stimato l'attesa μ di una v.a. X , nell'Esempio 8.2 abbiamo stimato il parametro λ di una $\mathcal{P}(\lambda)$ e così via, usando in ogni caso un solo valore dello stimatore che era sempre la media aritmetica \bar{X} . Questa scelta ci era stata suggerita dalla LGN Teorema 7.1 che garantiva la consistenza dello stimatore \bar{X} . Per lo stesso motivo avevamo proposto la varianza campionaria (7.3) come stimatore della varianza, ma avevamo anche osservato che essa, pur essendo consistente come si vede da (7.7) del Teorema 7.1, richiede una piccola correzione. Vedremo successivamente altri metodi per determinare la forma più opportuna di uno stimatore.

Nel seguito adotteremo le seguenti notazioni: X_1, \dots, X_n sarà un campione di n v.a. indipendenti, tutte con la stessa legge dipendente da uno o più parametri θ , con attesa μ e varianza σ^2 , mentre \bar{X} e S_c^2 rappresenteranno la media aritmetica (7.2) e la varianza campionaria (7.3) del campione. Ovviamente μ e σ sono delle $h(\theta)$ cioè dipendono da θ , ma questa dipendenza non verrà resa esplicita per non appesantire la notazione. Per la stessa ragione eviteremo di indicare l'indice sottoscritto θ nei simboli di attesa e varianza: scriveremo cioè sempre \mathbf{E} invece di \mathbf{E}_θ . Introdurremo infine la **varianza corretta**

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2 \quad (8.1)$$

per la quale ovviamente valgono le relazioni

$$S^2 = \frac{n}{n-1} S_c^2 = \frac{n}{n-1} (\overline{X^2} - \bar{X}^2). \quad (8.2)$$

Teorema 8.1. \bar{X} , S_c^2 e S^2 sono stimatori consistenti rispettivamente di μ e σ^2 ; \bar{X} e S^2 , ma non S_c^2 , sono anche stimatori non distorti di μ e σ^2 .

Dimostrazione: La consistenza di \bar{X} e S_c^2 discende direttamente dalla LGN Teorema 7.1; quella di S^2 deriva poi da quella di S_c^2 tenendo conto della relazione (8.2). Per dimostrare che X è uno stimatore non distorto di μ basterà osservare che

$$\mathbf{E}(\bar{X}) = \mathbf{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\mathbf{E}(X_1) + \dots + \mathbf{E}(X_n)}{n} = \frac{n \cdot \mu}{n} = \mu. \quad (8.3)$$

Viceversa S_c^2 non è uno stimatore non distorto di σ^2 . Infatti, siccome le X_k hanno tutte la stessa legge, da

$$\sigma^2 = \mathbf{Var}(X_k) = \mathbf{E}(X_k^2) - \mathbf{E}(X_k)^2 = \mathbf{E}(X_k^2) - \mu^2$$

si ottiene $\mathbf{E}(X_k^2) = \sigma^2 + \mu^2$, e analogamente tenendo conto di (8.3) si ha

$$\begin{aligned} \mathbf{E}(\bar{X}^2) &= \mathbf{Var}(\bar{X}) + \mathbf{E}(\bar{X})^2 = \mathbf{Var}\left(\frac{1}{n} \sum_{k=1}^n X_k\right) + \mu^2 \\ &= \frac{1}{n^2} \sum_{k=1}^n \mathbf{Var}(X_k) + \mu^2 = \frac{n \cdot \sigma^2}{n^2} + \mu^2 = \frac{\sigma^2}{n} + \mu^2. \end{aligned}$$

Ne segue allora che

$$\begin{aligned} \mathbf{E}(S_c^2) &= \mathbf{E}\left(\frac{1}{n} \sum_{k=1}^n X_k^2\right) - \mathbf{E}(\bar{X}^2) = \frac{1}{n} \sum_{k=1}^n \mathbf{E}(X_k^2) - \mathbf{E}(\bar{X}^2) \\ &= \frac{n \cdot (\sigma^2 + \mu^2)}{n} - \left(\frac{\sigma^2}{n} + \mu^2\right) = \frac{n-1}{n} \sigma^2 \end{aligned}$$

cioè che S_c^2 è uno stimatore distorto per σ^2 . Da questa stessa relazione e da (8.2) segue però immediatamente che invece la varianza corretta S^2 è uno stimatore non distorto della varianza σ^2 . \square

Dal Teorema 8.1 discende che la media aritmetica (7.2) \bar{X} è lo stimatore più naturale per l'attesa μ di un v.a. X ; inoltre sia la varianza campionaria (7.3) S_c^2 che la varianza corretta (8.1) sono stimatori consistenti della varianza σ^2 , ma solo S^2 è anche uno stimatore non distorto: per questo motivo in generale nei problemi di stima si preferisce usare la varianza corretta.

Esempio 8.6. $n = 20$ misure di una v.a. X danno i risultati raccolti nella Tabella 8.2. La stima puntuale dell'attesa μ di X è fornita dalla media aritmetica (7.2)

$$\bar{X} = \frac{1}{20} \sum_{k=1}^{20} X_k = 4.86$$

6.03	5.95	7.26	5.27	5.44	3.84	3.94	3.62	3.30	5.36
4.18	3.80	5.42	4.39	4.92	4.93	3.89	5.14	5.70	4.89

Tabella 8.2: Campione di $n = 20$ misure di una v.a. X .

mentre per stimare la varianza σ^2 è possibile calcolare sia la varianza campionaria (7.3) S_c^2 che la varianza corretta (8.2) S^2

$$S_c^2 = \overline{X^2} - \bar{X}^2 = 0.94, \quad S^2 = \frac{n}{n-1} (\overline{X^2} - \bar{X}^2) = 0.99.$$

Come si vede le due stime puntuali della varianza sono leggermente diverse e in generale si preferisce quella non distorta S^2 . È evidente comunque che la differenza fra le due stime puntuali diventa sempre più piccola al crescere di n .

8.2 Stima per intervalli

La stima puntuale di $h(\theta)$ resta una risposta piuttosto grossolana al problema di determinare una ragionevole approssimazione del valore vero incognito. In particolare è evidente che il valore stimato non sarà mai uguale al valore vero $h(\theta)$, e che la teoria della stima puntuale non permette di valutare neanche probabilisticamente l'entità della differenza fra i due valori. A questa necessità risponde invece la teoria della stima per intervalli: in pratica si rinuncia a stimare $h(\theta)$ con un solo valore di uno stimatore aleatorio, e si preferisce determinare – sempre a partire dai valori del campione di misure – i due estremi aleatori di un intero intervallo prefissando in maniera opportuna il valore della probabilità dell'evento “l'intervallo contiene $h(\theta)$ ”. La differenza principale sta nel fatto che, mentre la probabilità di ottenere il valore vero con una stima puntuale è sempre nulla, la probabilità che un intervallo con estremi aleatori contenga $h(\theta)$ è diversa da zero e in generale può anche essere calcolata. Nel seguito supporremo sempre di avere a disposizione un campione casuale X_1, \dots, X_n .

Definizione 8.4. Diremo che le due v.a. $T_1 = t_1(X_1, \dots, X_n)$ e $T_2 = t_2(X_1, \dots, X_n)$ sono gli estremi di un **intervallo di fiducia** $[T_1, T_2]$ **di livello** α (con $0 < \alpha < 1$) per $h(\theta)$ quando

$$\mathbf{P}_\theta\{T_1 \leq h(\theta) \leq T_2\} = 1 - \alpha$$

dove il simbolo \mathbf{P}_θ indica che la probabilità è calcolata supponendo che il valore del parametro incognito sia proprio θ .

Per un α fissato l'intervallo di fiducia non è unico. In particolare ci sono molti modi in cui si può ripartire la probabilità α che l'intervallo *non* contenga $h(\theta)$. In genere, però, si preferisce scegliere T_1 e T_2 in modo che

$$\mathbf{P}_\theta\{h(\theta) < T_1\} = \mathbf{P}_\theta\{T_2 < h(\theta)\} = \frac{\alpha}{2},$$

cioè si suddivide α in due parti eguali, in modo che le probabilità che i due estremi siano entrambi troppo grandi o troppo piccoli valgano $\alpha/2$. Con questa precisazione in generale (ma non sempre) gli estremi dell'intervallo di fiducia assumono la forma simmetrica $T \pm \Delta$, dove il valore centrale T è un opportuno stimatore di $h(\theta)$ e 2Δ è l'ampiezza aleatoria dell'intervallo $[T - \Delta, T + \Delta]$. Naturalmente, per un dato campione, l'ampiezza dell'intervallo di fiducia dipende dalla scelta del valore di α . Tipicamente si scelgono valori piccoli di α (ad esempio 0.05 oppure 0.01), in modo che la probabilità $1 - \alpha$ che l'intervallo contenga il valore vero sia corrispondentemente grande (ad esempio 0.95 oppure 0.99). È abbastanza intuitivo quindi che al diminuire di α , cioè richiedendo che la probabilità $1 - \alpha$ che l'intervallo contenga il valore vero aumenti, l'intervallo di fiducia debba allargarsi.

8.2.1 Intervallo di fiducia per l'attesa μ

Supporremo inizialmente di avere un *campione gaussiano* X_1, \dots, X_n con legge normale $N(\mu, \sigma^2)$, e di voler stimare il valore di μ . Abbiamo già osservato che la media aritmetica \bar{X} è uno stimatore non distorto e consistente per μ . Ora, se **il valore della varianza σ^2 è noto**, dal Teorema 5.9 abbiamo che

$$Y = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$$

è distribuita secondo la legge normale standard $\mathcal{N}(0, 1)$. Ricordando allora che in base a (5.21) si ha

$$\mathbf{P}\{|Y| \leq \varphi_{1-\frac{\alpha}{2}}\} = 1 - \alpha,$$

potremo scrivere che

$$\begin{aligned} 1 - \alpha &= \mathbf{P}\left\{\left|\sqrt{n} \frac{\bar{X} - \mu}{\sigma}\right| \leq \varphi_{1-\frac{\alpha}{2}}\right\} = \mathbf{P}\left\{|\bar{X} - \mu| \leq \frac{\sigma}{\sqrt{n}} \varphi_{1-\frac{\alpha}{2}}\right\} \\ &= \mathbf{P}\left\{\bar{X} - \frac{\sigma}{\sqrt{n}} \varphi_{1-\frac{\alpha}{2}} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} \varphi_{1-\frac{\alpha}{2}}\right\} \end{aligned}$$

e quindi l'intervallo di fiducia di livello α assume la forma simmetrica

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n}} \varphi_{1-\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} \varphi_{1-\frac{\alpha}{2}}\right]$$

che può anche essere riscritta come

$$\boxed{\bar{X} \pm \frac{\sigma}{\sqrt{n}} \varphi_{1-\frac{\alpha}{2}}}. \quad (8.4)$$

Se invece, come più spesso accade, **il valore della varianza σ^2 non è noto** bisogna prima stimare il valore di σ^2 usando lo stimatore corretto S^2 , e poi bisogna ricordare che sempre in base a (5.21) la v.a.

$$T = \sqrt{n} \frac{\bar{X} - \mu}{S}$$

è distribuita secondo la legge di Student $t(n - 1)$. Pertanto, seguendo gli stessi passaggi del caso precedente e ricordando la relazione (5.23) si ha che l'intervallo di fiducia di livello α assume la nuova forma simmetrica

$$\left[\bar{X} - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1), \bar{X} + \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1) \right]$$

che può anche essere riscritta come

$$\boxed{\bar{X} \pm \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1)}. \quad (8.5)$$

In pratica gli intervalli di fiducia per la media hanno sempre il centro in \bar{X} e un'ampiezza Δ che varia secondo i casi. Innanzitutto essa dipende dal campione dato nel senso che Δ cresce con σ (o con la sua stima S) e diminuisce con \sqrt{n} . Inoltre Δ dipende da un opportuno quantile della legge (normale o di Student) che meglio descrive la v.a. standardizzata. Il fatto che tali quantili crescano quando α prende valori sempre più vicini a zero, indica chiaramente che per avere una maggiore probabilità di contenere μ bisogna allargare l'intervallo di fiducia.

Abbiamo già osservato che le formule (8.4) e (8.5) forniscono gli intervalli di fiducia per μ solo se il campione è gaussiano. Queste formule restano comunque approssimativamente valide anche nel caso generale di *campioni non gaussiani* a causa dell'approssimazione normale discussa nella Sezione 7.2, purché n sia abbastanza grande. In pratica esse possono sempre essere applicate se $n \geq 20$. Va infine notato che le Tavole D.2 si arrestano a $n = 120$; va però ricordato che la differenza fra i quantili normali e i quantili di Student diminuisce all'aumentare di n : per $n \geq 120$ essa diviene irrilevante e i quantili $\varphi_{1-\frac{\alpha}{2}}$ possono essere senz'altro usati al posto dei quantili $t_{1-\frac{\alpha}{2}}(n-1)$ nella (8.5) che diviene

$$\boxed{\bar{X} \pm \frac{S}{\sqrt{n}} \varphi_{1-\frac{\alpha}{2}}}. \quad (8.6)$$

Esempio 8.7. Supponiamo di effettuare $n = 100$ misure di una quantità fisica affetta da un errore sperimentale casuale W con legge $\mathcal{N}(0, \sigma^2)$ e $\sigma^2 = 4$. Se il valore vero (sconosciuto) della quantità fisica è μ , i risultati delle misure sono v.a. del tipo $X = \mu + W$, e avremo che anche X è Normale $\mathcal{N}(\mu, \sigma^2)$ sicché il campione di misure è gaussiano. Supponendo che la media delle 100 misure sia $\bar{X} = 50$, vogliamo ora determinare un intervallo di fiducia di livello $\alpha = 0.01$ per μ . Siccome in questo caso la varianza $\sigma^2 = 4$ è nota, possiamo semplicemente applicare (8.4) e otteniamo

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}} \varphi_{1-\frac{\alpha}{2}} = 50 \pm \frac{2}{\sqrt{100}} 2.58 = 50 \pm 0.52$$

cioè con il 99% di probabilità l'intervallo $[49.48, 50.52]$ contiene il valore vero della nostra quantità fisica.

Esempio 8.8. Torniamo al problema – introdotto nell'Esempio 7.1 – della stima della proporzione p di individui di tipo A in una popolazione composta di individui di tipo A e B : supponiamo di aver estratto un campione di $n = 100$ misure delle quali 57 sono di tipo A , e determiniamo un intervallo di fiducia di livello $\alpha = 0.05$ per p . Formalmente siamo in presenza di un campione X_1, \dots, X_{100} di $n = 100$ v.a. tutte indipendenti e Bernoulli $B(1, p)$ dal quale ricaviamo immediatamente una stima puntuale di p come

$$\bar{X} = \frac{1}{100} \sum_{j=1}^{100} X_j = \frac{57}{100} = 0.570$$

Il valore $n = 100$ può essere ritenuto abbastanza grande per poter supporre che la media standardizzata delle X_j sia Normale standard. Siccome però non abbiamo informazioni sul valore della varianza (che per una Binomiale implicherebbe anche la conoscenza di p) dobbiamo preventivamente stimarne il valore mediante la varianza corretta (8.2): ricordando che per una Bernoulli si ha sempre $X_j^2 = X_j$ avremo che

$$S^2 = \frac{n}{n-1} (\overline{X^2} - \bar{X}^2) = \frac{100}{99} \left(\frac{1}{100} \sum_{j=1}^{100} X_j^2 - 0.570^2 \right) = \frac{100}{99} (0.570 - 0.570^2) = 0.248;$$

e usando l'equazione (8.5) otterremo l'intervallo

$$\bar{X} \pm \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1) = 0.570 \pm \frac{\sqrt{0.248}}{\sqrt{100}} t_{0.975}(99) = 0.570 \pm \frac{0.498}{10} 1.984 = 0.570 \pm 0.099$$

In realtà il quantile $t_{0.975}(99)$ necessario per il calcolo dell'intervallo non è presente nelle Tavole D.2: per aggirare questo problema abbiamo usato il quantile più vicino $t_{0.975}(100)$ avendo peraltro osservato che la variabilità di questi valori fra 90 e 100 gradi di libertà è limitata a poche unità sulla terza cifra decimale. Infine si noti che, essendo $n = 100$ abbastanza grande, il valore del quantile di Student $t_{0.975}(99) \simeq 1.984$ non differisce eccessivamente dal corrispondente quantile della Normale standard $\varphi_{0.975} = 1.960$, e se avessimo calcolato l'intervallo di fiducia con la formula (8.6) avremmo trovato

$$\bar{X} \pm \frac{S}{\sqrt{n}} \varphi_{1-\frac{\alpha}{2}} = 0.570 \pm \frac{0.498}{10} 1.960 = 0.570 \pm 0.098$$

ancora un intervallo non molto diverso dagli altri due. In conclusione possiamo dire che un intervallo di fiducia che contiene il valore vero di p con il 95% di probabilità sarà approssimativamente del tipo $[0.47, 0.67]$.

8.2.2 Intervallo di fiducia per la varianza σ^2

Supponiamo ora nuovamente di avere un *campione gaussiano* X_1, \dots, X_n con legge normale $N(\mu, \sigma^2)$, e di voler stimare la varianza σ^2 . Già sappiamo che la varianza corretta S^2 è uno stimatore non distorto e consistente per σ^2 , e inoltre dal Teorema 5.9 abbiamo che

$$Z = (n-1) \frac{S^2}{\sigma^2}$$

segue una legge $\chi^2(n-1)$. Tenendo conto di (5.24) si ha allora

$$\begin{aligned} 1 - \alpha &= \mathbf{P}\{\chi_{\frac{\alpha}{2}}^2(n-1) \leq Z \leq \chi_{1-\frac{\alpha}{2}}^2(n-1)\} \\ &= \mathbf{P}\left\{\chi_{\frac{\alpha}{2}}^2(n-1) \leq (n-1) \frac{S^2}{\sigma^2} \leq \chi_{1-\frac{\alpha}{2}}^2(n-1)\right\} \\ &= \mathbf{P}\left\{\frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)}\right\} \end{aligned}$$

e quindi l'intervallo di fiducia di livello α assume la forma

$$\left[\frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)} \right] \quad (8.7)$$

Esempio 8.9. Proviamo a fornire un intervallo di fiducia di livello $\alpha = 0.05$ per la stima della varianza dell'Esempio 8.8. Da (8.7) e dai calcoli sviluppati in precedenza abbiamo che l'estremo sinistro dell'intervallo è ora

$$\frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)} = \frac{99 \times 0.248}{\chi_{0.975}^2(99)}$$

ma siccome le Tavole D.3 si arrestano a 35 gradi di libertà, dovremo usare una formula approssimata richiamata in fondo alle stesse Tavole:

$$\chi_{\alpha}^2(n) \simeq \frac{1}{2}(\varphi_{\alpha} + \sqrt{2n-1})^2, \quad n > 35$$

Avremo allora che

$$\chi_{0.975}^2(99) \simeq \frac{1}{2}(\varphi_{0.975} + \sqrt{199})^2 = \frac{1}{2}(1.960 + \sqrt{199})^2 = 129.07$$

per cui in definitiva il valore dell'estremo sinistro è

$$\frac{99 \times 0.248}{129.07} = 0.19$$

Per l'estremo destro si ha analogamente

$$\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)} = \frac{99 \times 0.248}{\chi_{0.025}^2(99)}$$

e per il quantile richiesto (ricordando anche che $\varphi_{\alpha} = -\varphi_{1-\alpha}$)

$$\chi_{0.025}^2(99) \simeq \frac{1}{2}(\varphi_{0.025} + \sqrt{199})^2 = \frac{1}{2}(-\varphi_{0.975} + \sqrt{199})^2 = \frac{1}{2}(-1.960 + \sqrt{199})^2 = 73.77$$

Pertanto l'estremo destro è

$$\frac{99 \times 0.248}{73.77} = 0.33$$

per cui complessivamente il richiesto intervallo di fiducia è $[0.19, 0.33]$.

8.3 Stima di Massima Verosimiglianza

Non sempre la forma dello stimatore può essere indovinata in maniera naturale, come nel caso della media e della varianza. Sarà utile quindi avere un criterio generale per determinare un opportuno stimatore di un parametro θ .

Definizione 8.5. Dato il campione X_1, \dots, X_n , se le X_k sono v.a. discrete con valori x_k chiameremo **funzione di verosimiglianza** la funzione di θ

$$L(\theta) = \mathbf{P}_\theta\{X_1 = x_1\} \cdot \dots \cdot \mathbf{P}_\theta\{X_n = x_n\} = p_\theta(x_1) \cdot \dots \cdot p_\theta(x_n); \quad (8.8)$$

se invece le X_k sono v.a. continue con fd f_θ chiameremo **funzione di verosimiglianza** la funzione di θ

$$L(\theta) = f_\theta(x_1) \cdot \dots \cdot f_\theta(x_n). \quad (8.9)$$

Si noti che nel caso discreto, tenendo conto di (5.27) (ricordiamo che le X_k di un campione sono indipendenti e identicamente distribuite), $L(\theta)$ non è altro che la probabilità congiunta (vedi Sezione 5.5) che le X_k assumano i rispettivi valori x_k , supponendo che il parametro sconosciuto abbia il valore θ ; analogamente, nel caso continuo $L(\theta)$ è la densità congiunta delle X_k indipendenti e identicamente distribuite (vedi equazione (5.30)), supponendo che il parametro da stimare abbia il valore θ . Se i valori x_1, \dots, x_n del campione sono assegnati, il valore di $L(\theta)$ dipenderà solo da θ , e sarà in generale possibile determinare il valore $\hat{\theta}$ di θ per il quale $L(\theta)$ è massima. Nel caso di v.a. discrete, $\hat{\theta}$ è quindi il valore per il quale i valori x_1, \dots, x_n hanno la massima probabilità di essere osservati; nel caso continuo invece $\hat{\theta}$ è il valore per il quale x_1, \dots, x_n sono la moda della distribuzione del campione. Ovviamente $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ dipenderà dai valori di x_1, \dots, x_n anche se in genere, per semplificare la notazione, noi eviteremo di indicarlo. Potremo allora definire la v.a. $\hat{\Theta} = \hat{\theta}(X_1, \dots, X_n)$ che, sulla base delle precedenti osservazioni, sarà adottato come un buon stimatore di θ :

Definizione 8.6. Chiameremo **stimatore di massima verosimiglianza (MV)** la v.a. $\hat{\Theta} = \hat{\theta}(X_1, \dots, X_n)$ dove $\hat{\theta}$ è il valore che rende massima la funzione di verosimiglianza $L(\theta)$.

Il massimo della $L(\theta)$ è in genere determinato con i tradizionali metodi analitici (annullamento della derivata); siccome però $L(\theta)$ si presenta come un *prodotto* di n funzioni, e siccome non è sempre agevole derivare tali prodotti, spesso si preferisce determinare il massimo della funzione $\log L(\theta)$ che si presenta invece come una *somma* di funzioni di θ . Le due procedure sono equivalenti dato che la funzione $\log x$ è monotona.

Esempio 8.10. Supponiamo di avere qualche buona ragione per ritenere che la v.a. X sia una Binomiale $B(m, p)$, con m noto ma con parametro $p \in [0, 1]$ sconosciuto, e sia x_1, \dots, x_n un campione di misure di X (le x_j sono qui tutte numeri interi compresi fra 0 e m) dal quale vogliamo estrarre una stima di p . L'Esempio 8.5 (X era il numero di figli

maschi in una famiglia di $m = 12$ figli) costituisce un caso particolare di questo problema con $m = 12$: nella seconda parte di quell'Esempio, infatti, avevamo abbandonato l'ipotesi semplificativa $p = 1/2$, e per stimare il valore di p dal campione dato avevamo adottato una procedura basata sulla LGN che ci aveva condotto al seguente risultato:

$$\bar{p} = \frac{1}{12n} \sum_{k=0}^{12} kN_k = \frac{1}{m} \sum_{j=0}^n x_j = \frac{\bar{x}}{m} \quad (8.10)$$

Mostriamo ora che lo stimatore di MV del parametro $\theta = p$ per questo problema coincide proprio con (8.10). Infatti nel nostro caso (con x intero fra 0 e m)

$$\mathbf{P}\{X = x\} = \binom{m}{x} p^x (1-p)^{m-x}$$

per cui avremo

$$\begin{aligned} L(p) &= \prod_{j=1}^n \binom{m}{x_j} p^{x_j} (1-p)^{m-x_j}, \\ \log L(p) &= \sum_{j=1}^n \log \binom{m}{x_j} + \sum_{j=1}^n x_j \log p + \sum_{j=1}^n (m-x_j) \log(1-p). \end{aligned}$$

Annullando la derivata rispetto a p si ha allora

$$\frac{d}{dp} \log L(p) = \frac{1}{p} \sum_{j=1}^n x_j - \frac{1}{1-p} \sum_{j=1}^n (m-x_j) = 0$$

da cui otteniamo

$$(1-p) \sum_{j=1}^n x_j = p \sum_{j=1}^n (m-x_j) = nmp - p \sum_{j=1}^n x_j$$

ovvero, semplificando,

$$\sum_{j=1}^n x_j = nmp.$$

in definitiva quindi la stima di MV è

$$\hat{p} = \frac{1}{m} \sum_{j=1}^n x_j = \frac{\bar{x}}{m} \quad (8.11)$$

che coincide con (8.10). Si noti che per $m = 1$ la X è una v.a. di Bernoulli e il nostro problema si riconduce alla stima di una proporzione trattata nell'Esempio 7.1, e il risultato (7.1) coincide con la (8.11) per $m = 1$.

Esempio 8.11. Supponiamo ora di avere un campione x_1, \dots, x_n estratto da una v.a. X con legge di Poisson $\mathcal{P}(\lambda)$ e parametro λ incognito. In questo caso le x_j sono numeri interi senza limitazione superiore, e siccome

$$\mathbf{P}\{X = x\} = e^{-\lambda} \frac{\lambda^x}{x!}$$

per la funzione di verosimiglianza si ha

$$L(\lambda) = \prod_{j=1}^n e^{-\lambda} \frac{\lambda^{x_j}}{x_j!},$$

$$\log L(\lambda) = -n\lambda + \log \lambda \sum_{j=1}^n x_j - \sum_{j=1}^n \log(x_j!).$$

Annullando la derivata rispetto a λ si ha allora

$$\frac{d}{d\lambda} \log L(\lambda) = -n + \frac{1}{\lambda} \sum_{j=1}^n x_j = 0$$

da cui otteniamo come stima di MV

$$\hat{\lambda} = \frac{1}{n} \sum_{j=1}^n x_j = \bar{x}$$

che ancora una volta coincide con la media aritmetica del campione. Questo risultato non è sorprendente se si riflette al fatto che in base a (6.25) il parametro λ gioca anche il ruolo di valore d'attesa di una v.a. di Poisson, e che in base al Teorema 8.1 la media è un buon stimatore del valore d'attesa.

Esempio 8.12. Sia ora x_1, \dots, x_n è un campione estratto da una v.a. X con legge Normale $\mathcal{N}(\mu, \sigma^2)$, e sia $\theta = (\mu, \sigma^2)$ la coppia di parametri da stimare. Questa volta le x_j assumono valori reali e continui distribuiti secondo la fd normale (5.13) per cui la funzione di verosimiglianza sarà

$$L(\mu, \sigma) = \prod_{j=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_j - \mu)^2/2\sigma^2},$$

$$\log L(\mu, \sigma) = -n \log(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2.$$

Siccome questa volta dobbiamo stimare due parametri, bisognerà risolvere un sistema di due equazioni ottenute annullando ambedue le derivate rispetto a μ e σ . Annullando la derivata rispetto a μ si ha prima di tutto

$$\frac{d}{d\mu} \log L(\mu, \sigma) = \frac{1}{\sigma^2} \sum_{j=1}^n (x_j - \mu) = 0$$

e quindi si ottiene lo stimatore di MV per μ

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^n x_j = \bar{x}$$

che come negli altri casi è la media aritmetica del campione. Annullando poi la derivata rispetto a σ

$$\frac{d}{d\sigma} \log L(\mu, \sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{j=1}^n (x_j - \mu)^2 = 0$$

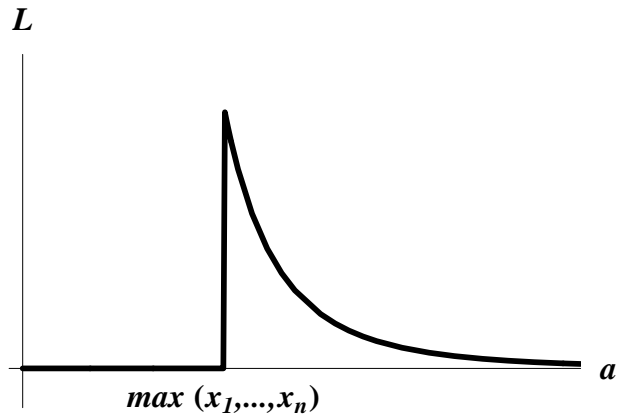


Figura 8.4: Funzione di verosimiglianza $L(a)$ dell'esempio 8.13.

e sostituendo la soluzione $\hat{\mu} = \bar{x}$, si ha poi lo stimatore di MV per σ^2

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \hat{\mu})^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2.$$

Si noter  che lo stimatore di MV della varianza ottenuto nell'Esempio precedente coincide con la varianza campionaria S_c^2 (7.3), e non con la varianza corretta S^2 (8.1). Pertanto, come si deduce dal Teorema 8.1, in questo caso lo stimatore di MV   distorto. Questo fenomeno non   eccezionale: sebbene gli stimatori di MV godano di molte buone propriet  anche asintotiche, non   garantito che essi siano anche non distorti. Si tratta comunque di un difetto relativamente poco importante perch , come si pu  vedere dall'equazione (8.2), nella maggior parte dei casi uno stimatore distorto pu  essere facilmente corretto moltiplicandolo per un opportuno fattore. Negli esempi precedenti, comunque, gli stimatori di MV non sono risultati molto diversi da quelli che potevano essere intuiti anche per altra via. Mostriamo ora solo un caso in cui la forma dello stimatore di MV   diversa da quella degli stimatori pi  comuni.

Esempio 8.13. Sia x_1, \dots, x_n un campione di una v.a. X uniforme $\mathcal{U}(0, a)$ (vedi Esempio 5.8), con a parametro incognito da stimare. Ora le x_j sono numeri reali e positivi e, dato un arbitrario valore di a , pi  piccoli di a . Siccome la fd di X  

$$f_X(x) = \begin{cases} 1/a & \text{se } 0 \leq x \leq a, \\ 0 & \text{altrimenti.} \end{cases}$$

la funzione di verosimiglianza sar  diversa da zero solo se a risulta maggiore di tutte le x_j , ovvero se $a > \max x_1, \dots, x_n$, e avr  quindi la forma

$$L(a) = \begin{cases} 1/a^n & \text{se } a > \max x_1, \dots, x_n \\ 0 & \text{altrimenti} \end{cases}$$

riportata nella Figura 8.4. In questo caso si vede subito dal grafico che per un dato campione x_1, \dots, x_n la funzione di verosimiglianza $L(a)$ assume il valore massimo proprio in

$$\hat{a} = \max(x_1, \dots, x_n)$$

formula che permette di definire lo stimatore di MV come $\hat{A} = \max(X_1, \dots, X_n)$. Come si potrà notare questo risultato è meno ovvio di quelli discussi in precedenza, anche se è abbastanza intuitivo che si proponga di stimare l'estremo superiore dell'intervallo $[0, a]$ proprio con il più grande degli x_j . Si può però anche dimostrare che lo stimatore \hat{A} è distorto nel senso che

$$\mathbf{E}(\hat{A}) = \frac{n}{n+1} a < a.$$

Questa distorsione può comunque essere facilmente eliminata introducendo lo stimatore corretto

$$\hat{A}_c = \frac{n+1}{n} \max(X_1, \dots, X_n).$$

Come si vede facilmente, infine, questa correzione diviene del tutto irrilevante per grandi valori di n .

Capitolo 9

Test di Ipotesi

9.1 Ipotesi ed errori

L'esito della procedura di stima di un parametro è un numero, o un intervallo di numeri; viceversa l'esito di un test statistico è una *decisione* fra ipotesi alternative. Discutiamo brevemente qualche esempio per chiarire meglio questa affermazione.

Esempio 9.1. In un campione di $n = 11\,712$ bambini nati in un paese tra il 1968 e il 1973 ci sono state $m = 5\,934$ nascite maschili; la proporzione empirica di maschi è quindi

$$\bar{p} = \frac{5\,934}{11\,712} = 0.507;$$

possiamo affermare che c'è stata una prevalenza di nascite maschili? O dobbiamo solo attribuire al caso il fatto che \bar{p} non sia esattamente $1/2$ come sarebbe naturale attendersi? Abbiamo quindi bisogno di una procedura affidabile che ci consenta di accettare una affermazione oppure l'altra, e di stimare la probabilità di sbagliare.

Analogamente, riprendendo la discussione dell'Esempio 8.5, possiamo affermare che gli esiti delle nascite successive nelle famiglie di 12 figli siano indipendenti fra loro? O dobbiamo supporre che ci siano famiglie con la tendenza a generare figli maschi, e famiglie con la tendenza a generare figlie femmine? Detto in altri termini: con quale criterio e con quale affidabilità possiamo decidere di accettare una delle due ipotesi alternative che abbiamo appena formulato?

Una ditta farmaceutica ha prodotto un nuovo farmaco per la cura di una determinata malattia: con quale procedura e con quale affidabilità possiamo pervenire a stabilire se il farmaco è realmente efficace? In che maniera dovremmo rilevare i dati empirici, e come possiamo usarli per giungere alla decisione richiesta?

In tutti gli esempi precedenti lo scopo dell'analisi statistica è quello di pervenire, con un preciso livello di affidabilità, alla accettazione di una fra due possibili ipotesi. Nel seguito avremo a che fare prevalentemente – ma non obbligatoriamente – con ipotesi che riguardano i valori di qualche parametro della distribuzione delle v.a. ritenute rilevanti per il nostro problema: così nel primo caso dell'Esempio 9.1 il parametro rilevante è la probabilità p che in un parto si produca una nascita maschile. Possiamo

allora definire la v.a. X di Bernoulli $B(1, p)$ che prende valori 1 e 0 con probabilità p e $1 - p$ secondo che in un parto si verifica rispettivamente una nascita maschile o una femminile. I nostri dati empirici sono allora le n v.a. di Bernoulli indipendenti X_j con $j = 1, \dots, n$ (e con $n = 11\,712$) che rappresentano gli esiti dei parti registrati, e in base alla LGN potremo stimare p mediante la media \bar{X} il cui valore è la proporzione empirica \bar{p} : noi vogliamo usare questi dati sperimentali per decidere se il valore di p è o no uguale a $1/2$.

La procedura più intuitiva in questo caso sarà quella di confrontare il valore empirico \bar{p} con il valore ipotetico $1/2$ esaminando il valore assunto da $|\bar{p} - \frac{1}{2}|$: se questa differenza è troppo grande noi rifiuteremo l'ipotesi $p = \frac{1}{2}$. Per rendere precisa e quantitativa questa procedura dovremo però trovare il modo di fissare un numero $\delta > 0$ che ci consenta di rifiutare l'ipotesi $p = \frac{1}{2}$ quando $|\bar{p} - \frac{1}{2}| > \delta$: in pratica dobbiamo essere in grado di fissare una soglia numerica δ al di là della quale noi riterremo che la differenza $|\bar{p} - \frac{1}{2}|$ sia troppo grande perché si possa accettare l'ipotesi $p = \frac{1}{2}$.

Ogni volta che sarà possibile, dunque, le ipotesi alternative saranno formulate in termini dei valori del parametro θ di qualche distribuzione: si divide l'insieme di tutti i possibili valori di θ in due parti Θ_0 e Θ_1 : un'ipotesi sarà $\theta \in \Theta_0$, mentre l'ipotesi alternativa sarà $\theta \in \Theta_1$. Nell'esempio che stiamo discutendo il parametro p assume valori in $(0, 1)$, e i due sottoinsiemi che definiscono le nostre due ipotesi sono

$$\Theta_0 = \left\{ \frac{1}{2} \right\}, \quad \Theta_1 = \left(0, \frac{1}{2} \right) \cup \left(\frac{1}{2}, 1 \right).$$

Naturalmente va precisato subito che decisioni assunte in questo modo sulla base di test statistici non possono che essere solo *probabilmente* corrette. In altri termini ci sarà sempre la possibilità di commettere degli errori ed è importante essere in grado di valutare la probabilità di questi errori. Vedremo peraltro nella discussione seguente che in realtà il valore della soglia δ e il valore della probabilità di errore sono legati fra loro.

Proseguendo nella discussione dell'esempio delle nascite maschili potremo dunque dire che siamo in presenza di due ipotesi alternative per il parametro $p \in (0, 1)$:

$$H_0 : p = \frac{1}{2}, \quad H_1 : p \neq \frac{1}{2}$$

Potremmo allora provare a fissare il valore della soglia δ di accettazione delle due ipotesi scegliendo tentativamente $\delta = 0.01$. Questo significa che dai dati empirici calcoleremo innanzitutto \bar{p} ; poi, se $|\bar{p} - \frac{1}{2}| < 0.01$ accetteremo H_0 , mentre se $|\bar{p} - \frac{1}{2}| > 0.01$ accetteremo H_1 . Ovviamente, però, può accadere che il test così eseguito ci conduca a decisioni errate in uno dei due seguenti modi: può succedere che p sia realmente uguale a $1/2$, ma che casualmente i dati empirici producano un valore di $|\bar{p} - \frac{1}{2}|$ più grande di 0.01 inducendoci a concludere che p è diverso da $1/2$; viceversa può succedere che p sia realmente diverso da $1/2$, ma che casualmente i dati empirici producano un valore di $|\bar{p} - \frac{1}{2}|$ più piccolo di 0.01 inducendoci a concludere che p è invece proprio uguale a $1/2$.

Gli errori possibili sono quindi di due tipi principali: si può affermare che H_0 è falsa quando essa è vera, o affermare che H_0 è vera quando essa è falsa. È importante essere in grado di controllare la probabilità di commettere questi errori; per il momento noi ci limiteremo solo ad osservare che, nel nostro esempio, tali probabilità dipendono ovviamente dal valore della soglia δ nel modo seguente: se il valore di δ *aumenta* il test sarà *meno severo* per l'ipotesi H_0 nel senso che sarà più difficile rifiutarla. Conseguentemente diminuisce la probabilità di commettere un errore del primo tipo (affermare che H_0 è falsa quando essa è vera). È altrettanto evidente però che in questo caso aumenterà la probabilità di commettere l'altro tipo di errore (affermare che H_0 è vera quando essa è falsa). Il valore di δ quindi deve essere scelto in maniera molto prudente, evitando di prenderlo troppo grande o troppo piccolo per non fare aumentare troppo la probabilità di uno dei due tipi di errore. D'altra parte queste osservazioni mostrano che il valore di δ deve essere scelto in conseguenza del rischio di errore che intendiamo correre prendendo la nostra decisione sulle due ipotesi H_0 e H_1 . Riassumeremo ora il contenuto di questa discussione preliminare in alcune Definizioni.

La formulazione del test parte dalla individuazione delle due ipotesi alternative che vogliamo esaminare: H_0 detta **ipotesi nulla**, e H_1 detta **ipotesi alternativa**. In generale il ruolo di queste due ipotesi non è simmetrico: tipicamente si sceglie come ipotesi nulla l'ipotesi più conservativa e prudente, ricordando però che lo sperimentatore è normalmente interessato a verificare se l'ipotesi H_0 può essere rifiutata. Così, se si sta sperimentando un nuovo farmaco, H_0 è in genere l'ipotesi "il farmaco è inefficace", ma lo sperimentatore è interessato a verificare piuttosto se può essere sostenuta l'ipotesi alternativa H_1 : "il farmaco è efficace".

Definizione 9.1. *Chiameremo **errore di prima specie** quello che si commette rifiutando H_0 quando H_0 è vera; viceversa chiameremo **errore di seconda specie** quello che si commette accettando H_0 quando H_0 è falsa.*

I due tipi di errore non sono considerati in genere sullo stesso piano per due ragioni principali: innanzitutto gli errori di prima specie (fare affermazioni imprudenti e false) sono considerati più gravi di quelli di seconda (perdere l'occasione di mettere in evidenza qualcosa di nuovo). Nel caso della sperimentazione di un farmaco si giudica più grave mettere in circolazione un farmaco inutile (o addirittura dannoso), che perdere l'occasione di produrre un farmaco efficace. In secondo luogo l'errore di prima specie si commette sotto l'ipotesi che H_0 sia vera: siccome in generale H_0 è un'ipotesi più precisa dell'alternativa H_1 , supporre che H_0 sia vera permette nella maggior parte dei casi di valutare la probabilità dell'errore. Viceversa, siccome l'errore di seconda specie si commette sotto l'ipotesi che sia vera H_1 (che spesso è definita solo dall'essere il contrario di H_0) in generale è piuttosto difficile poter calcolare la probabilità di questo secondo tipo di errore. Nel seguito supporremo che l'ipotesi H_0 sia caratterizzata dal fatto che un certo parametro θ cada in un sottoinsieme Θ_0 dell'insieme Θ dei suoi valori, e che l'ipotesi H_1 corrisponda invece al fatto che θ appartenga al sottoinsieme Θ_1 complementare di Θ_0 .

Definizione 9.2. Chiameremo **regione critica** l'evento D che conduce al rifiuto dell'ipotesi nulla H_0 ; chiameremo poi **livello** del test la quantità

$$\alpha = \sup_{\theta \in \Theta_0} \mathbf{P}_\theta(D),$$

e **potenza** del test la funzione

$$\pi(\theta) = \mathbf{P}_\theta(D), \quad \theta \in \Theta_1;$$

chiameremo infine **significatività** del test il più piccolo valore α_s del livello che, per un campione dato, conduce al rifiuto di H_0 .

In generale la regione critica D sarà un evento del tipo $\{T \geq t\}$ definito in termini di una qualche statistica $T = t(X_1, \dots, X_n)$ la cui legge è completamente nota se l'ipotesi H_0 è vera, cioè se $\theta \in \Theta_0$; inoltre \mathbf{P}_θ indica come al solito la probabilità calcolata sotto l'ipotesi che il parametro sconosciuto abbia proprio il valore θ . Conseguentemente il livello α di un test è il massimo della probabilità di commettere un errore di prima specie. Viceversa la potenza del test è la probabilità di rifiutare H_0 al variare di θ in Θ_1 , cioè quando H_0 è falsa: in pratica si tratta della probabilità di mettere in evidenza la correttezza di H_1 quando questa è vera, e quindi la probabilità di errori di seconda specie vale $1 - \pi(\theta)$ con $\theta \in \Theta_1$. In generale il valore del livello è una scelta operata inizialmente dello sperimentatore che decide il rischio di errore di prima specie che vuole correre: i valori più usati sono $\alpha = 0.05$ e $\alpha = 0.01$, ma anche $\alpha = 0.10$. La forma della regione critica D e la potenza sono una conseguenza di questa scelta. Invece la significatività α_s dipende principalmente dai dati empirici: i dati confermeranno l'ipotesi H_0 per tutti i livelli $\alpha \leq \alpha_s$; e quindi se α_s è piccola l'ipotesi H_0 risulta confermata dai dati per piccoli valori del livello α , cioè con piccoli rischi di errori di prima specie. Ovviamente un test è buono se il livello α e la significatività α_s sono piccoli, e la potenza $\pi(\theta)$ è grande.

Esempio 9.2. Siamo ora in condizione di discutere più compiutamente il problema della proporzione di figli maschi introdotto nell'Esempio 9.1. A questo scopo introduciamo $n = 11\,712$ v.a. indipendenti X_j , con $j = 1, \dots, n$, tutte di Bernoulli $B(1, p)$ con $X_j = 1$ se il parto j -mo ha dato luogo ad una nascita maschile, e $X_j = 0$ viceversa. Il parametro p prende valori in $(0, 1)$ e noi vogliamo decidere quale delle due seguenti ipotesi accettare:

$$H_0 : p = \frac{1}{2}, \quad H_1 : p \neq \frac{1}{2}$$

ricordando che il valore osservato delle nascite maschili è 5934. Definiremo allora le v.a.

$$S_n = \sum_{j=1}^n X_j, \quad \bar{X} = \frac{S_n}{n} = \frac{1}{n} \sum_{j=1}^n X_j$$

i cui valori empirici sono rispettivamente

$$s = 5\,934, \quad \bar{p} = \frac{5\,934}{11\,712} = 0.507$$

e osserveremo che in base al Teorema 5.3 S_n è Binomiale $B(n, p)$, per cui da (6.24), (6.5) e (6.18) avremo

$$\begin{aligned}\mathbf{E}(S_n) &= np, & \mathbf{Var}(S_n) &= np(1-p), \\ \mathbf{E}(\bar{X}) &= p, & \mathbf{Var}(\bar{X}) &= \frac{p(1-p)}{n}.\end{aligned}$$

Il Teorema 7.2 (TLC), con $\mu = p$ e $\sigma^2 = p(1-p)$ per le nostre v.a. di Bernoulli X_j , ci dice inoltre che, siccome $n = 11\,712$ può essere considerato grande, detta Z una v.a. normale standard $\mathcal{N}(0, 1)$, avremo con buona approssimazione

$$\sqrt{n} \frac{\bar{X} - p}{\sqrt{p(1-p)}} \simeq Z$$

nel senso che la legge della v.a. al primo membro è ben approssimata dalla legge $\mathcal{N}(0, 1)$; nello stesso senso potremo allora scrivere

$$\bar{X} \simeq p + \sqrt{\frac{p(1-p)}{n}} Z$$

dove, tenendo conto del Teorema 5.5, la legge della v.a. al secondo membro è normale $\mathcal{N}(p, p(1-p)/n)$. A questo punto, preso $\delta > 0$ e ricordando che con il simbolo Φ indichiamo la FD della Normale standard (5.16), siamo in grado di eseguire il calcolo della seguente probabilità che giocherà un ruolo importante nella discussione:

$$\begin{aligned}\mathbf{P} \left\{ \left| \bar{X} - \frac{1}{2} \right| \geq \delta \right\} &= \mathbf{P} \left\{ \bar{X} \leq \frac{1}{2} - \delta \right\} + \mathbf{P} \left\{ \bar{X} \geq \frac{1}{2} + \delta \right\} \\ &\simeq \mathbf{P} \left\{ p + \sqrt{\frac{p(1-p)}{n}} Z \leq \frac{1}{2} - \delta \right\} \\ &\quad + \mathbf{P} \left\{ p + \sqrt{\frac{p(1-p)}{n}} Z \geq \frac{1}{2} + \delta \right\} \\ &= \mathbf{P} \left\{ Z \leq \left(\frac{1}{2} - p - \delta \right) \sqrt{\frac{n}{p(1-p)}} \right\} \\ &\quad + \mathbf{P} \left\{ Z \geq \left(\frac{1}{2} - p + \delta \right) \sqrt{\frac{n}{p(1-p)}} \right\} \\ &= \mathbf{P} \left\{ Z \leq \left(\frac{1}{2} - p - \delta \right) \sqrt{\frac{n}{p(1-p)}} \right\} \\ &\quad + 1 - \mathbf{P} \left\{ Z \leq \left(\frac{1}{2} - p + \delta \right) \sqrt{\frac{n}{p(1-p)}} \right\} \\ &= \Phi \left(\frac{1 - 2p - 2\delta}{2} \sqrt{\frac{n}{p(1-p)}} \right) \\ &\quad + 1 - \Phi \left(\frac{1 - 2p + 2\delta}{2} \sqrt{\frac{n}{p(1-p)}} \right) \quad (9.1)\end{aligned}$$

Notiamo che in particolare, se si accetta l'ipotesi H_0 che $p = \frac{1}{2}$, e ricordando le proprietà di simmetria (5.17) di Φ , la (9.1) si semplifica in

$$\mathbf{P} \left\{ \left| \bar{X} - \frac{1}{2} \right| \geq \delta \right\} = \Phi(-2\delta\sqrt{n}) + 1 - \Phi(2\delta\sqrt{n}) = 2[1 - \Phi(2\delta\sqrt{n})] \quad (9.2)$$

Fissiamo ora, come suggerito in precedenza, la regione critica D scegliendo tentativamente $\delta = 0.01$; avremo allora

$$D = \left\{ \left| \bar{X} - \frac{1}{2} \right| \geq 0.010 \right\},$$

e siccome il valore empirico è $|\bar{p} - \frac{1}{2}| = |0.507 - 0.500| = 0.007 < 0.010$ l'esito del test è che accetteremo l'ipotesi H_0 . Dobbiamo però ora calcolare il livello α del test eseguito in questo modo: supponendo H_0 vera, ponendo $n = 11\,712$ e $\delta = 0.01$ in (9.2) e facendo uso delle Tavole D.1 avremo allora

$$\alpha = \mathbf{P} \left\{ \left| \bar{X} - \frac{1}{2} \right| \geq 0.01 \right\} = 2[1 - \Phi(2.164)] = 0.030.$$

Dunque la scelta di $\delta = 0.01$ produce un test di livello 0.03 che ci induce ad accettare H_0 . Tipicamente, però, si preferisce fissare inizialmente il livello α del test deducendo poi il valore di δ e la regione critica D . Ad esempio supponiamo di voler eseguire un test di livello $\alpha = 0.05$: per determinare δ osserveremo allora che, supponendo vera H_0 , da (9.2) si ha

$$\alpha = \mathbf{P} \left\{ \left| \bar{X} - \frac{1}{2} \right| \geq \delta \right\} = 2[1 - \Phi(2\delta\sqrt{n})]$$

ovvero

$$\Phi(2\delta\sqrt{n}) = 1 - \frac{\alpha}{2}$$

e quindi ricordando la definizione (5.12) di quantile

$$2\delta\sqrt{n} = \varphi_{1-\frac{\alpha}{2}}$$

dove come al solito φ_α indica il quantile di ordine α della Normale standard. In conclusione, ponendo $\alpha = 0.05$ e $n = 11\,712$, e usando le Tavole D.1 per calcolare $\varphi_{0.975}$, avremo

$$\delta = \frac{1}{2\sqrt{n}} \varphi_{1-\frac{\alpha}{2}} = 0.009$$

La regione critica di livello $\alpha = 0.05$ è allora

$$D = \left\{ \left| \bar{X} - \frac{1}{2} \right| \geq 0.009 \right\},$$

e siccome $|\bar{p} - \frac{1}{2}| = |0.507 - 0.500| = 0.007 < 0.009$ l'esito del test è ancora favorevole all'ipotesi H_0 . Noteremo però che scegliendo un α più grande (0.05 invece di 0.03) il test per H_0 è diventato più severo, nel senso che δ è diminuito (passando da 0.010 a 0.009), la regione critica D si è allargata e quindi è divenuto più probabile rifiutare H_0 . È evidente allora che mantenendo sempre gli stessi dati empirici ($\bar{p} = 0.507$) e aumentando progressivamente il livello si arriverà ad un valore α_s di α (detto, come sappiamo, *significatività*) tale che l'ipotesi H_0 non sia più in grado di superare il test. Dato che per noi $|\bar{p} - \frac{1}{2}| = 0.007$, è chiaro che α_s sarà il valore di α che produce per δ il valore $\bar{\delta} = 0.007$, e quindi, sempre da (9.2) e dalle Tavole D.1, la significatività del nostro test è

$$\alpha_s = \mathbf{P} \left\{ \left| \bar{X} - \frac{1}{2} \right| \geq \bar{\delta} \right\} = 2[1 - \Phi(2\bar{\delta}\sqrt{n})] = 0.130$$

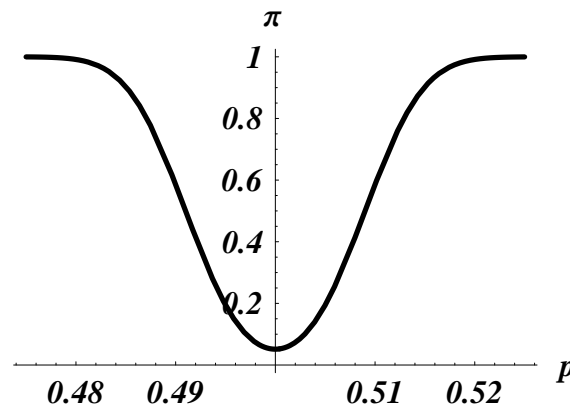


Figura 9.1: Funzione potenza $\pi(p)$ per il test di livello $\alpha = 0.05$ dell'esempio 9.2.

cioè l'ipotesi H_0 verrà rifiutata da tutti i test con livello maggiore o uguale di 0.13. Infine possiamo anche calcolare la funzione potenza di un test di livello $\alpha = 0.05$, cioè con $\delta = 0.009$ e regione critica $D = \{|\bar{X} - \frac{1}{2}| \geq 0.009\}$. In questo caso dobbiamo supporre vera l'ipotesi H_1 (cioè $p \neq \frac{1}{2}$) e usare (9.1) per calcolare la potenza:

$$\begin{aligned} \pi(p) &= \mathbf{P} \left\{ \left| \bar{X} - \frac{1}{2} \right| \geq \delta \right\} \\ &= \Phi \left(\frac{1 - 2p - 2\delta}{2} \sqrt{\frac{n}{p(1-p)}} \right) + 1 - \Phi \left(\frac{1 - 2p + 2\delta}{2} \sqrt{\frac{n}{p(1-p)}} \right) \\ &= \Phi \left(\frac{108.2(0.491 - p)}{\sqrt{p(1-p)}} \right) + 1 - \Phi \left(\frac{108.2(0.509 - p)}{\sqrt{p(1-p)}} \right) \end{aligned}$$

Il grafico di questa funzione è riportato nella Figura 9.1. Da questa si vede che la potenza del test di livello 0.05 è molto buona non appena p differisce da $\frac{1}{2}$ di più di 0.02: in questo caso infatti il test rileva che $p \neq \frac{1}{2}$ praticamente con probabilità uguale a 1. Viceversa se la differenza fra p e $\frac{1}{2}$ è minore di 0.02 la potenza diminuisce molto pur mantenendosi sempre superiore ad un minimo che vale 0.05: così, se p differisce molto poco da $\frac{1}{2}$, il test di livello $\alpha = 0.05$ rivela la differenza solo con una probabilità di circa 0.05. Ma la funzione potenza ci permette di fare affermazioni anche nei casi intermedi: ad esempio, se p differisce da $\frac{1}{2}$ di 0.1 (cioè se $p = 0.51$ oppure $p = 0.49$) la probabilità che il nostro test di livello $\alpha = 0.05$ lo rilevi è 0.59, e corrispondentemente la probabilità di errori del secondo tipo (accettare H_0 quando essa è falsa) è 0.41.

L'Esempio appena discusso ci ha dato l'occasione di illustrare concretamente i concetti generali introdotti nella Definizione 9.2; nelle Sezioni che seguono ci concentreremo invece sull'esame di alcuni particolari tipi di test di uso piuttosto frequente e relativi sia al valore di alcuni parametri (medie e varianze) che al tipo di distribuzione delle v.a. osservate.

9.2 Test sulla media

Il primo tipo di test riguarda il valore μ – sconosciuto – dell’attesa di una v.a. dalla quale abbiamo estratto un campione X_1, \dots, X_n : tipicamente si vuol conoscere se μ può essere considerata uguale ad un ben determinato valore μ_0 , oppure no. In questo caso le ipotesi da valutare sono

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0 \quad (9.3)$$

e si parla di **test bilaterale** per mettere in evidenza il fatto che nell’ipotesi alternativa μ può essere sia più grande che più piccola di μ_0 . A volte però potremmo essere interessati a mettere in evidenza solo che μ è più grande (o più piccolo) di μ_0 : in questo caso le ipotesi sono del tipo

$$H_0 : \mu \leq \mu_0, \quad H_1 : \mu > \mu_0 \quad (9.4)$$

$$H_0 : \mu \geq \mu_0, \quad H_1 : \mu < \mu_0 \quad (9.5)$$

e si parla rispettivamente di **test unilaterale destro e sinistro**.

Esaminiamo innanzitutto il caso di un test bilaterale (9.3): per determinare la regione critica dobbiamo scegliere un opportuno stimatore da osservare, e per i test sull’attesa μ è naturale prendere in considerazione la media aritmetica del campione $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ che come sappiamo è un buon stimatore del valore d’attesa. Il test consisterà nell’esaminare se il valore osservato \bar{X} può significativamente essere considerato diverso da valore μ_0 , per cui la regione critica prenderà la forma dell’evento $\{|\bar{X} - \mu_0| > \delta\}$, nel senso che se la differenza $|\bar{X} - \mu_0|$ fra valore osservato e valore ipotetico assume valori troppo grandi superando una soglia $\delta > 0$, allora rifiuteremo l’ipotesi nulla H_0 e accetteremo l’ipotesi alternativa H_1 . Per rendere quantitativa questa procedura dobbiamo calcolare un opportuno valore numerico per la soglia δ fissando a priori il valore del livello α del test e richiedendo poi, in base alla Definizione 9.2, che sia verificata la relazione

$$\mathbf{P}\{|\bar{X} - \mu_0| > \delta\} = \alpha, \quad \text{supponendo } H_0 : \mu = \mu_0 \quad (9.6)$$

Questa equazione ci consentirà di determinare il valore della soglia δ per un test bilaterale di livello α prefissato.

Nel caso di un test unilaterale destro (9.4) la regione critica assume invece la forma $\{\bar{X} - \mu_0 > \delta\}$, nel senso che ora si esamina la differenza $\bar{X} - \mu_0$ senza valore assoluto perché siamo interessati a mettere in evidenza che μ è più grande di (e non solo diverso da) μ_0 . Se tale differenza assume valori troppo grandi superando una soglia $\delta > 0$, allora rifiuteremo l’ipotesi nulla H_0 e accetteremo l’ipotesi alternativa H_1 . Anche in questo caso per determinare la soglia δ si fissa il livello α del test e si applica la Definizione 9.2

$$\sup \mathbf{P}\{\bar{X} - \mu_0 > \delta\} = \alpha, \quad \text{supponendo } H_0 : \mu \leq \mu_0 \quad (9.7)$$

Per i test unilaterali converrà inoltre osservare che, supponendo vera l'ipotesi H_0 secondo la quale $\mu_0 - \mu \geq 0$, si ha per la regione critica

$$\begin{aligned} \{\bar{X} - \mu_0 > \delta\} &= \{(\bar{X} - \mu) + (\mu - \mu_0) > \delta\} = \{\bar{X} - \mu > \delta + \mu_0 - \mu\} \\ &\subseteq \{\bar{X} - \mu > \delta\} \end{aligned}$$

e quindi anche

$$\mathbf{P}\{\bar{X} - \mu_0 > \delta\} \leq \mathbf{P}\{\bar{X} - \mu > \delta\}, \quad \mu \leq \mu_0$$

Pertanto, ricordando anche che la legge della v.a. centrata $\bar{X} - \mu$ è indipendente da μ , da (9.7) si ha

$$\mathbf{P}\{\bar{X} - \mu > \delta\} = \alpha, \quad \text{supponendo } H_0 : \mu \leq \mu_0 \quad (9.8)$$

In maniera analoga per un test unilaterale sinistro (9.5) la regione critica è della forma $\{\bar{X} - \mu_0 < -\delta\}$, e la soglia si calcola scegliendo il livello α e richiedendo che

$$\mathbf{P}\{\bar{X} - \mu < -\delta\} = \alpha, \quad \text{supponendo } H_0 : \mu \geq \mu_0 \quad (9.9)$$

A questo punto per procedere è necessario avere delle informazioni sulla legge di \bar{X} , cioè sulla legge delle osservazioni X_1, \dots, X_n . Noi supporremo sempre nel seguito che le X_k siano v.a. normali $\mathcal{N}(\mu, \sigma^2)$ (o possano essere ritenute tali in base al TLC 7.2) in modo da poter usare i risultati del Teorema 5.9, ricordando peraltro che – sempre per il TLC – tali risultati sono comunque approssimativamente veri se il numero n delle osservazioni è abbastanza grande. Si noti che la legge $\mathcal{N}(\mu, \sigma^2)$ delle v.a. X_k non è completamente conosciuta: non solo non ci è noto il valore di μ , ma in generale potrebbe non esserci noto neanche il valore di σ^2 . Per questo motivo dovremo ora distinguere due tipi di test.

9.2.1 Test di Gauss

Se la legge delle X_k è $\mathcal{N}(\mu, \sigma^2)$ il Teorema 5.9 afferma che la v.a.

$$U = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$$

è una Normale standard $\mathcal{N}(0, 1)$. Basandoci su questo risultato riprendiamo la discussione dei test nel caso in cui **la varianza σ^2 è nota**. Per un test bilaterale delle ipotesi (9.3) la relazione (9.6) può ovviamente essere riscritta come

$$\mathbf{P} \left\{ \left| \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} \right| > \frac{\delta \sqrt{n}}{\sigma} \right\} = \mathbf{P} \left\{ |U_0| > \frac{\delta \sqrt{n}}{\sigma} \right\} = \alpha, \quad \text{supponendo } H_0 : \mu = \mu_0$$

dove la v.a.

$$\boxed{U_0 = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma}} \quad (9.10)$$

è una $\mathcal{N}(0, 1)$; ricordando allora il risultato (5.21) si ha facilmente

$$\frac{\delta\sqrt{n}}{\sigma} = \varphi_{1-\frac{\alpha}{2}}.$$

Dato il livello α , il valore $\varphi_{1-\frac{\alpha}{2}}$ si trova facilmente consultando le Tavole dell'Appendice D.1, e la regione critica $\{|\bar{X} - \mu_0| > \delta\}$ di livello α si scriverà equivalentemente come

$$\{|U_0| > \varphi_{1-\frac{\alpha}{2}}\}.$$

In conclusione un test bilaterale di livello α delle ipotesi (9.3) si esegue in questo modo: *si calcola il valore empirico della v.a. U_0 (9.10) e lo si confronta con il quantile $\varphi_{1-\frac{\alpha}{2}}$ della Normale standard ricavato dalle Tavole D.1: se risulta $|U_0| > \varphi_{1-\frac{\alpha}{2}}$, allora si rifiuta l'ipotesi H_0 e si accetta H_1 ; se invece risulta $|U_0| \leq \varphi_{1-\frac{\alpha}{2}}$ si accetta l'ipotesi H_0 .* Siccome la significatività α_s è il più piccolo valore del livello α per il quale i dati empirici sono in regime critica, per calcolare α_s bisogna imporre $\varphi_{1-\frac{\alpha_s}{2}} = |U_0|$ dove U_0 è il valore empirico della statistica (9.10). Dalla definizione di quantile si ha allora che $1 - \frac{\alpha_s}{2} = \Phi(|U_0|)$ dove Φ è la FD Normale standard, ovvero

$$\alpha_s = 2[1 - \Phi(|U_0|)].$$

Analogamente si costruisce il test unilaterale destro di livello α per le ipotesi (9.4): la relazione (9.8) si riscrive come

$$\mathbf{P}\left\{\sqrt{n}\frac{\bar{X} - \mu}{\sigma} > \frac{\delta\sqrt{n}}{\sigma}\right\} = \mathbf{P}\left\{U > \frac{\delta\sqrt{n}}{\sigma}\right\} = \alpha, \quad \text{supponendo } H_0 : \mu \leq \mu_0$$

dove U è $\mathcal{N}(0, 1)$; siccome

$$\mathbf{P}\left\{U \leq \frac{\delta\sqrt{n}}{\sigma}\right\} = 1 - \mathbf{P}\left\{U > \frac{\delta\sqrt{n}}{\sigma}\right\} = 1 - \alpha,$$

per semplice definizione di quantile avremo allora questa volta che

$$\frac{\delta\sqrt{n}}{\sigma} = \varphi_{1-\alpha},$$

e quindi per (9.8) la regione critica è ora

$$\{U_0 > \varphi_{1-\alpha}\}$$

dove U_0 è sempre data da (9.10). Il test unilaterale destro si esegue allora in questo modo: *si calcola il valore empirico della U_0 e lo si confronta con il quantile $\varphi_{1-\alpha}$ della Normale standard ricavato dalle Tavole D.1: se risulta $U_0 > \varphi_{1-\alpha}$, allora si rifiuta l'ipotesi H_0 e si accetta H_1 ; se invece risulta $U_0 \leq \varphi_{1-\alpha}$ si accetta l'ipotesi H_0 .* La significatività del test unilaterale destro si ottiene poi imponendo $\varphi_{1-\alpha_s} = U_0$, cioè $1 - \alpha_s = \Phi(U_0)$ e quindi

$$\alpha_s = 1 - \Phi(U_0).$$

166.6	169.3	168.2	176.4	168.6	170.1	167.7	168.1	164.3	171.1
172.5	165.7	166.1	171.3	176.5	168.8	169.7	168.1	167.1	172.8
173.5	168.9	169.7	167.7	173.0	159.4	168.8	163.7	174.4	174.0
164.4	171.1	168.1	171.4	174.6	168.7	169.4	165.7	159.5	164.1
166.0	168.1	169.0	172.6	172.2	170.4	173.4	181.5	165.5	167.9
168.9									

Tabella 9.1: Altezze in *cm* di un campione di $n = 51$ reclute.

Il test unilaterale sinistro si costruisce in maniera identica con l'unica differenza che la regione critica sarà ora

$$\{U_0 < -\varphi_{1-\alpha}\}$$

La procedura per il test sarà quindi: *si calcola il valore empirico della U_0 : se risulta $U_0 < -\varphi_{1-\alpha}$, allora si rifiuta l'ipotesi H_0 e si accetta H_1 ; se invece risulta $U_0 \geq -\varphi_{1-\alpha}$ si accetta l'ipotesi H_0* . La significatività del test unilaterale sinistro si ottiene allora da $\varphi_{1-\alpha_s} = -U_0$, cioè $1 - \alpha_s = \Phi(-U_0)$ e quindi

$$\alpha_s = 1 - \Phi(-U_0).$$

Esempio 9.3. È noto che l'altezza X delle persone di un determinato paese è una v.a. che segue una legge Normale $\mathcal{N}(\mu, \sigma^2)$: supponiamo ora di sapere, in base ai dati di un censimento del 1950, che per gli individui di sesso maschile si ha $\mu_0 = 168$ *cm* e $\sigma^2 = 19$ *cm*². Nel 1965 viene esaminato un campione di $n = 51$ reclute alla visita di leva e si riportano le loro altezze nella Tabella 9.1. Si constata a questo punto che la media del campione è $\bar{X} = 169.3$ *cm*. Supponendo di poter considerare la varianza $\sigma^2 = 19$ *cm*² come ancora attendibile e quindi nota, si vuol sapere se al livello $\alpha = 0.05$ possiamo dire che la media delle altezze è aumentata? Il test richiesto è dunque unilaterale destro con ipotesi del tipo (9.4): dai dati a nostra disposizione abbiamo che

$$U_0 = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n} = \frac{169.3 - 168}{\sqrt{19}} \sqrt{51} = 2.13$$

Siccome dalle Tavole D.1 risulta $\varphi_{0.95} = 1.65 < 2.13 = U_0$, si vede subito che i dati sono nella regione critica per cui il test unilaterale destro di livello $\alpha = 0.05$ ci conduce a rifiutare H_0 e ad accettare l'ipotesi che l'altezza sia aumentata. Dato il valore di $U_0 = 2.13$ possiamo anche calcolare la significatività del test che è $\alpha_s = 1 - \Phi(2.13) = 0.017$, un valore non particolarmente buono che lascia qualche dubbio sulla sicurezza dell'esito del test. Infatti se avessimo svolto i calcoli con un livello $\alpha = 0.01$ ci saremmo accorti che $\varphi_{0.99} = 2.33 > 2.13 = U_0$, per cui in questo caso il risultato del test sarebbe stato quello di accettare l'ipotesi H_0 secondo la quale l'altezza media è rimasta di 168 *cm*. Il problema dovrebbe essere risolto in maniera un po' diversa se inizialmente si chiedesse di verificare che l'altezza media è *cambiata* (non *aumentata*). In questo caso il test dovrebbe essere bilaterale: il calcolo di U_0 resta invariato, ma il suo valore deve ora essere confrontato con $\varphi_{1-\frac{\alpha}{2}}$. Al livello $\alpha = 0.05$ abbiamo allora che $\varphi_{0.975} = 1.96 < 2.13 = U_0$ per cui anche il test bilaterale conferma il rifiuto di H_0 : l'altezza è cambiata. La significatività è però peggiorata: $\alpha_s = 2[1 - \Phi(2.13)] = 0.033$, e quindi l'esito del test è ancora meno sicuro.

9.2.2 Test di Student

Sempre supponendo che le X_k siano tutte $\mathcal{N}(\mu, \sigma^2)$ esaminiamo ora il caso in cui **la varianza σ^2 non è nota**. La differenza pratica con il caso precedente risiede nel fatto che ora il valore empirico di U_0 in (9.10) non può più essere calcolato compiutamente perché ci manca il valore di σ . Per superare questo problema dovremo allora stimare σ^2 tramite la varianza corretta S^2 (8.1) e ricordare che il Teorema 5.9 afferma che la v.a.

$$T = \sqrt{n} \frac{\bar{X} - \mu}{S}$$

è una Student con $n - 1$ gradi di libertà $t(n - 1)$. Per un test bilaterale delle ipotesi (9.3) la relazione (9.6) può allora essere riscritta come

$$\mathbf{P} \left\{ \left| \sqrt{n} \frac{\bar{X} - \mu_0}{S} \right| > \frac{\delta \sqrt{n}}{S} \right\} = \mathbf{P} \left\{ |T_0| > \frac{\delta \sqrt{n}}{S} \right\} = \alpha, \quad \text{supponendo } H_0 : \mu = \mu_0$$

dove la v.a.

$$\boxed{T_0 = \sqrt{n} \frac{\bar{X} - \mu_0}{S}} \tag{9.11}$$

è una $t(n - 1)$; ricordando allora il risultato (5.23) si ha facilmente

$$\frac{\delta \sqrt{n}}{S} = t_{1-\frac{\alpha}{2}}(n - 1).$$

Dato il livello α , il valore $t_{1-\frac{\alpha}{2}}(n - 1)$ si trova facilmente consultando le Tavole dell'Appendice D.2, e la regione critica $\{|\bar{X} - \mu_0| > \delta\}$ di livello α si scriverà equivalentemente come

$$\{|T_0| > t_{1-\frac{\alpha}{2}}(n - 1)\}.$$

In conclusione un test bilaterale di livello α delle ipotesi (9.3) si esegue ora secondo questa procedura: *si calcola il valore empirico della v.a. T_0 (9.11) e lo si confronta con il quantile $t_{1-\frac{\alpha}{2}}(n - 1)$ della Student ricavato dalle Tavole D.2: se risulta $|T_0| > t_{1-\frac{\alpha}{2}}(n - 1)$, allora si rifiuta l'ipotesi H_0 e si accetta H_1 ; se invece risulta $|T_0| \leq t_{1-\frac{\alpha}{2}}(n - 1)$ si accetta l'ipotesi H_0 . Inoltre la significatività del test bilaterale è*

$$\alpha_s = 2 [1 - F_{n-1}(|T_0|)]$$

dove $F_{n-1}(x)$ è la FD della legge di Student $t(n - 1)$ e T_0 è il valore empirico (9.11).

Analogamente si costruiscono i test unilaterali destro e sinistro di livello α per le ipotesi (9.4) e (9.5): le regioni critiche sono ora rispettivamente

$$\{T_0 > t_{1-\alpha}(n - 1)\}, \quad \{T_0 < -t_{1-\alpha}(n - 1)\}$$

dove T_0 è sempre data da (9.11), e i test si eseguono con la seguente procedura: *si calcola il valore empirico della T_0 e lo si confronta con il quantile $t_{1-\alpha}(n - 1)$ della*

Student ricavato dalle Tavole D.2: se risulta $T_0 > t_{1-\alpha}(n-1)$ (rispettivamente: $T_0 < -t_{1-\alpha}(n-1)$), allora si rifiuta l'ipotesi H_0 e si accetta H_1 ; se invece risulta $T_0 \leq t_{1-\alpha}(n-1)$ (rispettivamente: $T_0 \geq -t_{1-\alpha}(n-1)$) si accetta l'ipotesi H_0 . Le significatività dei test unilaterali destro e sinistro sono poi rispettivamente

$$\alpha_s = 1 - F_{n-1}(T_0), \quad \alpha_s = 1 - F_{n-1}(-T_0).$$

Si noti che i quantili della distribuzione di Student con $n-1$ gradi di libertà dipendono ovviamente da n , e nelle Tavole D.2 i valori dei quantili sono riportati solo per $n \leq 120$. Sembrerebbe dunque impossibile eseguire un test di Student per campioni con $n > 120$. Per capire come si affronta questa difficoltà bisogna ricordare che per n molto grande la distribuzione di Student tende a coincidere con la distribuzione Normale standard. Conseguentemente, per un dato valore del livello α , se $n > 120$ i quantili calcolati dalle Tavole D.1 della Normale standard e quelli ricavati dalle Tavole D.2 della Student sono praticamente coincidenti. In questo caso il test per la media può essere effettuato calcolando il valore empirico T_0 da (9.11), ma usando i quantili della Normale standard. In pratica per $n > 120$ la procedura per il test bilaterale di Student si modifica nel modo seguente: si calcola il valore empirico della v.a. T_0 (9.11) e lo si confronta con il quantile $\varphi_{1-\frac{\alpha}{2}}$ della Normale standard ricavato dalle Tavole D.1: se $|T_0| > \varphi_{1-\frac{\alpha}{2}}$, si rifiuta l'ipotesi H_0 ; se invece risulta $|T_0| \leq \varphi_{1-\frac{\alpha}{2}}$ si accetta l'ipotesi H_0 . Analogamente per i test unilaterali destro e sinistro si calcola T_0 e si determina il quantile $\varphi_{1-\alpha}$ della Normale standard dalle Tavole D.1: se risulta $T_0 > \varphi_{1-\alpha}$ (rispettivamente: $T_0 < -\varphi_{1-\alpha}$), si rifiuta l'ipotesi H_0 ; se invece $T_0 \leq \varphi_{1-\alpha}$ (rispettivamente: $T_0 \geq -\varphi_{1-\alpha}$) si accetta l'ipotesi H_0 .

Esempio 9.4. Riprendendo l'Esempio 9.3 supponiamo ora di non poter considerare come attendibile il valore 19 per la varianza σ^2 della nostra v.a. In questo caso dobbiamo innanzitutto stimare la varianza corretta che risulta essere $S^2 = 16.5$ e calcolare poi la statistica di Student

$$T_0 = \frac{\bar{X} - \mu_0}{S} \sqrt{n} = \frac{169.3 - 168}{\sqrt{16.5}} \sqrt{51} = 2.29$$

Siccome dalle Tavole D.2 risulta $t_{0.95}(50) = 1.68 < 2.29 = T_0$, anche questa volta i dati sono nella regione critica per cui il test unilaterale destro di livello $\alpha = 0.05$ implica il rifiuto di H_0 . Per il valore $T_0 = 2.29$ possiamo anche calcolare la significatività del test che è $\alpha_s = 0.013$, un valore che lascia di nuovo qualche dubbio sulla affidabilità del test. Infatti se avessimo svolto i calcoli con un livello $\alpha = 0.01$ ci saremmo accorti che $t_{0.99}(50) = 2.40 > 2.29 = T_0$, per cui in questo caso il risultato del test sarebbe stato quello di accettare l'ipotesi H_0 secondo la quale l'altezza media è rimasta di 168 cm.

9.3 Test per il confronto delle medie

Un altro tipo di test riguarda il confronto fra le medie μ_X e μ_Y di due v.a. X e Y . Dovremo qui distinguere due casi secondo che i due campioni siano *accoppiati* o *indipendenti*. Chiameremo **campione accoppiato** un campione $(X_1, Y_1), \dots, (X_n, Y_n)$

di misure di due v.a. quando è importante conservare memoria della connessione fra X_k e Y_k : ad esempio, se vogliamo studiare l'effetto di un farmaco possiamo somministrarlo ad un numero n di pazienti misurando su ciascuno di essi qualche parametro rilevante prima (X) e dopo (Y) la somministrazione per mettere in evidenza eventuali differenze fra i loro comportamenti medi (μ_X e μ_Y). In questo caso ovviamente è importante non perdere l'accoppiamento: X_k e Y_k sono le misure eseguite sul medesimo paziente, il k -mo. Parleremo invece di **campioni indipendenti** quando non vi è nessuna relazione rilevante fra gli elementi X_1, \dots, X_n del primo campione e gli elementi Y_1, \dots, Y_m del secondo, nel senso che gli elementi *corrispondenti* X_k e Y_k non sono misurati sullo stesso individuo e la loro collocazione al posto k -mo è priva di significato statistico. Peraltro due campioni indipendenti possono avere anche cardinalità diverse n ed m , situazione evidentemente non consentita nel caso di un campione accoppiato. Un esempio di campioni indipendenti si ha quando si studia un farmaco confrontando il suo effetto con quello di un *placebo*: in questo caso si somministrano il farmaco e il *placebo* a due gruppi distinti di pazienti (anche di numero diverso) e si misura qualche parametro rilevante sui due gruppi (X e Y) per confrontare gli effetti medi μ_X e μ_Y .

Come al solito il confronto delle medie viene formalizzato tramite delle ipotesi: avremo un test bilaterale se le ipotesi sono

$$H_0 : \mu_X = \mu_Y, \quad H_1 : \mu_X \neq \mu_Y \quad (9.12)$$

invece il test è unilaterale quando le ipotesi sono del tipo

$$H_0 : \mu_X \leq \mu_Y, \quad H_1 : \mu_X > \mu_Y \quad (9.13)$$

In ambedue i casi, comunque, il confronto deve essere fatto seguendo procedure diverse per campioni accoppiati e campioni indipendenti come vedremo nella discussione seguente. Si noti che per brevità abbiamo indicato solo le ipotesi del test unilaterale destro senza però perdere molto in generalità: il test unilaterale sinistro è infatti descritto dalle medesime ipotesi *scambiando il ruolo di X e Y* .

9.3.1 Campioni accoppiati

Nel caso di campioni accoppiati si costruisce innanzitutto il campione delle differenze $Z_k = X_k - Y_k$ con $k = 1, \dots, n$, e poi si esegue un test sull'unico campione Z_1, \dots, Z_n per verificare se la sua media μ_Z può essere considerata nulla o diversa da zero. Le ipotesi per i test bilaterali (9.12) e unilaterali (9.13) saranno allora rispettivamente

$$H_0 : \mu_Z = 0, \quad H_1 : \mu_Z \neq 0 \quad (9.14)$$

$$H_0 : \mu_Z \leq 0, \quad H_1 : \mu_Z > 0 \quad (9.15)$$

e quindi si ricade sostanzialmente nei test studiati nella Sezione 9.2 con il particolare valore $\mu_0 = 0$. Se le v.a. Z_k possono essere considerate (almeno approssimativamen-

te) $\mathcal{N}(\mu, \sigma^2)$ e **se la varianza σ^2 è nota**, allora posto

$$\bar{Z} = \frac{1}{n} \sum_{k=1}^n Z_k$$

nell'ipotesi H_0 la v.a.

$$\boxed{U_0 = \sqrt{n} \frac{\bar{Z}}{\sigma}} \quad (9.16)$$

è Normale standard $\mathcal{N}(0, 1)$. Viceversa **se la varianza σ^2 non è nota** si introduce la varianza corretta

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (Z_k - \bar{Z})^2$$

e si osserva che nell'ipotesi H_0 la v.a.

$$\boxed{T_0 = \sqrt{n} \frac{\bar{Z}}{S}} \quad (9.17)$$

è Student con $n-1$ gradi di libertà $t(n-1)$. A questo punto le procedure per i test sono le stesse della Sezione 9.2: per un dato livello α , se la varianza è nota le regioni critiche dei test bilaterale e unilaterale sono rispettivamente

$$\{|U_0| > \varphi_{1-\frac{\alpha}{2}}\}, \quad \{U_0 > \varphi_{1-\alpha}\}$$

con significatività

$$\alpha_s = 2[1 - \Phi(|U_0|)], \quad \alpha_s = 1 - \Phi(U_0).$$

Se invece la varianza non è nota le regioni critiche dei test bilaterale e unilaterale sono rispettivamente

$$\{|T_0| > t_{1-\frac{\alpha}{2}}(n-1)\}, \quad \{T_0 > t_{1-\alpha}(n-1)\},$$

con significatività

$$\alpha_s = 2[1 - F_{n-1}(|T_0|)], \quad \alpha_s = 1 - F_{n-1}(T_0),$$

dove F_{n-1} è la FD della legge di Student $t(n-1)$. Così ad esempio per un test bilaterale di livello α con varianza σ^2 non nota si esegue un test di Student: *si calcola il valore empirico della v.a. T_0 (9.17) e lo si confronta con il quantile $t_{1-\frac{\alpha}{2}}(n-1)$ della Student ricavato dalle Tavole D.2; se risulta $|T_0| > t_{1-\frac{\alpha}{2}}(n-1)$, si rifiuta H_0 ; se invece risulta $|T_0| \leq t_{1-\frac{\alpha}{2}}(n-1)$ si accetta l'ipotesi H_0 . Gli altri casi si trattano in modo analogo.*

X	Y	Z	X	Y	Z	X	Y	Z
80	85	5	70	82	12	78	70	-8
80	84	4	65	73	8	75	77	2
82	87	5	83	89	6	76	76	0
75	81	6	74	85	11	78	82	4
80	79	-1	81	86	5	77	83	6
74	85	11	68	72	4	75	80	5
80	87	7	69	74	5	72	80	8
72	78	6	71	77	6	71	81	10
91	86	-5	70	75	5	75	76	1
88	80	-8	73	81	8	78	77	-1

Tabella 9.2: Pulsazioni di $n = 30$ pazienti prima (X) e dopo (Y) l'assunzione di un farmaco.

Esempio 9.5. Si sperimenta un farmaco su un campione di $n = 30$ pazienti rilevando il numero delle pulsazioni al minuto prima (X) e dopo (Y) la somministrazione: i dati sono riportati nella Tabella 9.2. Possiamo dire in base a questi valori, e ad un livello $\alpha = 0.05$, che la frequenza delle pulsazioni è aumentata? Osserviamo innanzitutto che si tratta effettivamente di un campione accoppiato: infatti è importante conservare la relazione fra le coppie di valori X e Y relative ad ogni singolo paziente. Ci viene richiesto di eseguire un test unilaterale destro con ipotesi (9.15), e senza conoscenza preventiva della varianza di Z : passiamo quindi a calcolare media, varianza corretta e statistica di Student (9.17) per il campione delle differenze $Z = Y - X$ riportato in Tabella:

$$\bar{Z} = 4.23, \quad S = 5.01, \quad T_0 = \frac{\bar{Z}}{S} \sqrt{n} = 4.63$$

A questo punto dovremo paragonare il valore di T_0 con l'opportuno quantile di Student: $t_{0.95}(29) = 1.70 < 4.63 = T_0$ per cui al livello $\alpha = 0.05$ il campione è in regione critica, e quindi possiamo dire che il farmaco ha prodotto un aumento della frequenza delle pulsazioni. La significatività del test questa volta è piuttosto buona: $\alpha_s = 1 - F_{n-1}(4.63)$ dove $F_{n-1}(x)$ è la FD della legge $t(n-1)$; questo valore non è presente sulle Tavole D.2 ma può essere calcolato in altro modo ed è $\alpha_s = 0.00003$. Possiamo poi calcolare la potenza del test di livello $\alpha = 0.05$ supponendo ad esempio – per semplificare la procedura – che l'ipotesi alternativa si riduca a $\mu_Z = 2$ (cioè calcoliamo la funzione potenza nell'unico valore $\mu_Z = 2$ del parametro). Abbiamo già visto che la regione critica del nostro test è

$$D = \{T_0 > t_{0.95}(29) = 1.70\}$$

e noi dobbiamo ora calcolare la potenza come $\mathbf{P}(D)$ nell'ipotesi $H_1 : \mu_Z = 2$. Se H_1 è vera la v.a.

$$T_1 = \frac{\bar{Z} - \mu_Z}{S} \sqrt{n} = \frac{\bar{Z} - 2}{5.01} \sqrt{30}$$

segue una legge di Student $t(29)$, per cui la potenza è

$$\begin{aligned}\pi(2) &= \mathbf{P}(D) = \mathbf{P}\{T_0 > 1.70\} = \mathbf{P}\left\{\frac{\bar{Z}}{S}\sqrt{n} > 1.70\right\} = \mathbf{P}\left\{\frac{\bar{Z} - \mu_Z}{S}\sqrt{n} > 1.70 - \frac{\mu_Z}{S}\sqrt{n}\right\} \\ &= \mathbf{P}\left\{\frac{\bar{Z} - 2}{5.01}\sqrt{30} > 1.70 - \frac{2}{5.01}\sqrt{30}\right\} = \mathbf{P}\{T_1 > -0.49\} = 0.686\end{aligned}$$

Anche il valore di questa probabilità non è desumibile dalle Tavole D.2, e deve essere ottenuto usando opportuni programmi di calcolo. Il valore della potenza $\pi(2) = 0.686$ indica dunque che se la media μ_Z è effettivamente diversa da 0 e vale 2, il test lo rivela con probabilità del 68.6%. Pertanto se $\mu_Z = 2$ la probabilità di errore di seconda specie vale $1 - 0.686 = 0.314$: un valore non del tutto trascurabile. La maniera più semplice di migliorare la potenza di un test di livello α è quella di aumentare il numero n delle osservazioni. Mostriamo ora come è possibile stimare il valore di n che, a parità degli altri dati, consente di avere una potenza del 95%. In pratica noi vogliamo determinare n in modo tale che per un dato α , e per $\mu_Z = 2$ risulti

$$\mathbf{P}\{T_0 > t_{1-\alpha}(n-1)\} = \mathbf{P}\left\{\frac{\bar{Z}}{S}\sqrt{n} > t_{1-\alpha}(n-1)\right\} = 0.95 \quad (9.18)$$

Siccome la statistica

$$T_1 = \frac{\bar{Z} - \mu_Z}{S}\sqrt{n}$$

segue una legge di Student $t(n-1)$ la (9.18) si riscrive come

$$\mathbf{P}\left\{\frac{\bar{Z} - \mu_Z}{S}\sqrt{n} > t_{1-\alpha}(n-1) - \frac{\mu_Z\sqrt{n}}{S}\right\} = 1 - \mathbf{P}\left\{T_1 \leq t_{1-\alpha}(n-1) - \frac{\mu_Z\sqrt{n}}{S}\right\} = 0.95$$

ovvero

$$\mathbf{P}\left\{T_1 \leq t_{1-\alpha}(n-1) - \frac{\mu_Z\sqrt{n}}{S}\right\} = \mathbf{P}\left\{T_1 \leq t_{0.95}(n-1) - \frac{\mu_Z\sqrt{n}}{S}\right\} = 0.05$$

Quest'ultima relazione è una equazione nell'incognita n che però non può essere banalmente risolta andando a consultare le Tavole D.2 perché l'incognita n compare anche nella legge $t(n-1)$ di T_1 . Per trovare una soluzione approssimata, allora, si ricorda che per n abbastanza grande la legge $t(n)$ è ben approssimata dalla legge Normale standard $\mathcal{N}(0, 1)$: potremo in pratica supporre che T_1 abbia legge $\mathcal{N}(0, 1)$, e quindi avremo

$$\varphi_{0.05} = -\varphi_{0.95} \simeq t_{0.95}(n-1) - \frac{\mu_Z\sqrt{n}}{S} \simeq \varphi_{0.95} - \frac{\mu_Z\sqrt{n}}{S}$$

ovvero in conclusione

$$\sqrt{n} \simeq \frac{2S\varphi_{0.95}}{\mu_Z}$$

Sostituendo ora il valore empirico $S = 5.01$, il valore tabulato $\varphi_{0.95} = 1.64$ e il valore $\mu_Z = 2$ della nostra ipotesi alternativa si ottiene

$$n \simeq (5.01 \times 1.64)^2 \simeq 68$$

Gli elementi del campione devono quindi essere più del doppio di quelli del nostro campione iniziale per garantire una potenza adeguata ad un test di livello 0.05.

9.3.2 Campioni indipendenti

Nel caso di campioni indipendenti con numerosità rispettive n ed m non ha alcun senso studiare il campione delle differenze. In questo caso supporremo che (almeno approssimativamente) le X_j e le Y_k siano rispettivamente $\mathcal{N}(\mu_X, \sigma_X^2)$ e $\mathcal{N}(\mu_Y, \sigma_Y^2)$, e partiremo calcolando separatamente le medie dei due campioni

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j, \quad \bar{Y} = \frac{1}{m} \sum_{k=1}^m Y_k.$$

Se le varianze σ_X^2 e σ_Y^2 sono note si può dimostrare che la v.a.

$$U = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

è Normale standard $\mathcal{N}(0, 1)$. In tal caso si introduce la quantità empirica

$$U_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \quad (9.19)$$

e in analogia con quanto ottenuto nella Sezione 9.2 si definiscono le procedure per i test. Nel caso bilaterale con ipotesi (9.12) la regione critica è

$$\{|U_0| > \varphi_{1-\frac{\alpha}{2}}\}$$

per cui si calcola il valore empirico della U_0 (9.19) e lo si confronta con il quantile $\varphi_{1-\frac{\alpha}{2}}$ della Normale standard: se $|U_0| > \varphi_{1-\frac{\alpha}{2}}$, si rifiuta H_0 ; se invece risulta $|U_0| \leq \varphi_{1-\frac{\alpha}{2}}$ si accetta H_0 . Allo stesso modo nel caso unilaterale con ipotesi (9.13) la regione critica è

$$\{U_0 > \varphi_{1-\alpha}\}$$

per cui si calcola il valore empirico della U_0 e lo si confronta con il quantile $\varphi_{1-\alpha}$ della Normale standard: se $U_0 > \varphi_{1-\alpha}$, si rifiuta H_0 ; se invece risulta $U_0 \leq \varphi_{1-\alpha}$ si accetta H_0 .

Viceversa se le varianze σ_X^2 e σ_Y^2 non sono note si introduce la **varianza combinata**

$$V^2 = \frac{1}{n+m-2} \left[\sum_{j=1}^n (X_j - \bar{X})^2 + \sum_{k=1}^m (Y_k - \bar{Y})^2 \right] = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} \quad (9.20)$$

Si può dimostrare che, se $\sigma_X^2 = \sigma_Y^2$ (o almeno sono approssimativamente uguali), V^2 è un buon stimatore della varianza comune, e

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{V \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

X	9.2	8.3	10.3	11.0	12.0	8.6	9.3	10.3	9.7	9.0
Y	10.9	11.3	10.9	10.2	9.3	10.4	10.5	11.3	10.8	12.0
	10.6	10.4	12.2	10.9	10.4					

Tabella 9.3: Produttività mensili di una ditta prima (Y) e dopo (X) aver introdotto cambiamenti nel processo produttivo.

è una Student con $n + m - 2$ gradi di libertà $t(n + m - 2)$. In tal caso si introduce la quantità empirica

$$T_0 = \frac{\bar{X} - \bar{Y}}{V \sqrt{\frac{1}{n} + \frac{1}{m}}} \quad (9.21)$$

e le regioni critiche dei test bilaterale (9.12) e unilaterale (9.13) di livello α sono rispettivamente

$$\{|T_0| > t_{1-\frac{\alpha}{2}}(n + m - 2)\}, \quad \{T_0 > t_{1-\alpha}(n + m - 2)\},$$

con conseguente definizione delle procedure per i test.

Esempio 9.6. Una ditta introduce dei cambiamenti nel proprio processo produttivo e vuol sapere se questo ha modificato la produttività. Sono disponibili i dati delle produzioni mensili (in una opportuna unità di misura): nella Tabella 9.6 sono riportati $n = 10$ valori successivi (X), e $m = 15$ precedenti (Y) al cambiamento del processo produttivo. I campioni sono indipendenti dato che non si attribuisce particolare significato all'accoppiamento dei loro valori (che peraltro sono in numero differente). Sia inoltre noto che ambedue le v.a. X e Y sono normali $\mathcal{N}(\mu_X, \sigma_X^2)$ e $\mathcal{N}(\mu_Y, \sigma_Y^2)$. Si vuol sapere se, sulla base dei dati empirici per i quali

$$\bar{X} = 9.8, \quad \bar{Y} = 10.8$$

e al livello $\alpha = 0.02$, possiamo affermare che $\mu_X = \mu_Y$. Supponiamo prima di tutto di sapere che la deviazione standard della produttività non è stata modificata dai cambiamenti e vale $\sigma_X = \sigma_Y = 1.1$. In questo caso potremo eseguire un test di Gauss con la statistica (9.19) il cui valore è $U_0 = -2.23$: dalle Tavole abbiamo allora che $\varphi_{0.99} = 2.33 > 2.23 = |U_0|$, per cui il test suggerisce di accettare l'ipotesi $H_0 : \mu_X = \mu_Y$. Viceversa, se supponiamo di non conoscere nulla sulle deviazioni standard di X e Y dovremo applicare un test di Student con la statistica (9.21). Sarà necessario allora stimare dai dati empirici le varianze corrette e la varianza combinata

$$S_X^2 = 1.30, \quad S_Y^2 = 0.51, \quad V^2 = 0.82$$

e il valore della statistica $T_0 = -2.71$. Dalle Tavole ricaviamo allora che $t_{0.99}(23) = 2.50 < 2.71 = |T_0|$: cioè il test di Student ci suggerisce di rifiutare l'ipotesi $H_0 : \mu_X = \mu_Y$. I risultati discordanti dei due tipi di test sono ovviamente un segno di incertezza: un esame della significatività dei due test conferma peraltro questa situazione. Infatti si ha $\alpha_s = 2[1 - \Phi(|U_0|)] = 0.03$ nel caso di Gauss e $\alpha_s = 2[1 - F_{23}(|T_0|)] = 0.01$ nel caso di Student: due valori piuttosto elevati che sollevano qualche dubbio sulla attendibilità dei risultati.

9.4 Test di Fisher sulla varianza

Anche la stima della varianza è un aspetto importante dell'analisi statistica che abbiamo esaminato nella Sezione 8.7: qui esamineremo il modo di verificare le ipotesi relative alla eguaglianza delle varianze di due campioni. Supporremo allora di avere due campioni X_1, \dots, X_n e Y_1, \dots, Y_m estratti da due v.a. X e Y normali $\mathcal{N}(\mu_X, \sigma_X^2)$, $\mathcal{N}(\mu_Y, \sigma_Y^2)$: le ipotesi che vogliamo sottoporre a test – bilaterale e unilaterale – saranno allora rispettivamente

$$\begin{aligned} H_0 : \sigma_X^2 &= \sigma_Y^2 & H_1 : \sigma_X^2 &\neq \sigma_Y^2 \\ H_0 : \sigma_X^2 &\leq \sigma_Y^2 & H_1 : \sigma_X^2 &> \sigma_Y^2 \end{aligned}$$

Si noti che per brevità abbiamo indicato solo le ipotesi del test unilaterale destro senza però perdere molto in generalità: il test unilaterale sinistro è infatti descritto dalle medesime ipotesi *scambiando il ruolo di X e Y* . Sappiamo dal Teorema 5.9 che le v.a.

$$\frac{(n-1)S_X^2}{\sigma_X^2} = \frac{1}{\sigma_X^2} \sum_{j=1}^n (X_j - \bar{X})^2, \quad \frac{(m-1)S_Y^2}{\sigma_Y^2} = \frac{1}{\sigma_Y^2} \sum_{k=1}^m (Y_k - \bar{Y})^2;$$

seguono rispettivamente le leggi $\chi^2(n)$ e $\chi^2(m)$, e quindi in base al Teorema 5.8 potremo dire che la v.a.

$$F = \frac{\frac{(n-1)S_X^2}{\sigma_X^2} / (n-1)}{\frac{(m-1)S_Y^2}{\sigma_Y^2} / (m-1)}$$

segue la legge di Fisher $F(n-1, m-1)$. Pertanto, supponendo che $\sigma_X^2 = \sigma_Y^2$, anche la v.a.

$$\boxed{F_0 = \frac{S_X^2}{S_Y^2}} \tag{9.22}$$

segue la legge di Fisher $F(n-1, m-1)$. Ora è chiaro che le due varianze σ_X^2 e σ_Y^2 saranno giudicate uguali quando F_0 assume valori vicini a 1, mentre potremo dire che le due varianze sono diverse se F_0 è abbastanza diversa da 1. Pertanto, tenendo conto della asimmetria della fd della legge di Fisher, le regioni critiche di livello α per i test bilaterale e unilaterale saranno rispettivamente

$$\begin{aligned} \{F_0 < f_{\frac{\alpha}{2}}(n-1, m-1)\} \cup \{F_0 > f_{1-\frac{\alpha}{2}}(n-1, m-1)\} \\ \{F_0 > f_{1-\alpha}(n-1, m-1)\} \end{aligned}$$

dove le $f_\alpha(n, m)$ indicano i quantili delle corrispondenti leggi di Fisher che possono essere ricavati dalle Tavole D.4

Esempio 9.7. Riprendendo la discussione dell'Esempio 9.6 possiamo porci il problema di verificare se, al livello $\alpha = 0.05$, le varianze σ_X^2 e σ_Y^2 delle v.a. X e Y possono essere considerate uguali. Bisogna allora eseguire un test bilaterale di Fisher: dai calcoli

dell'Esempio 9.6 ricaviamo prima di tutto il valore della statistica di Fisher $F_0 = 2.55$. Calcoliamo poi i quantili della legge di Fisher che definiscono la regione critica: il valore $f_{0.975}(9, 14) = 3.21$ si trova immediatamente dalle Tavole D.4; l'altro quantile che invece non si trova immediatamente sulle Tavole va calcolato da

$$f_{0.025}(9, 14) = \frac{1}{f_{0.975}(14, 9)} = 0.26$$

A questo punto possiamo eseguire il test: siccome $f_{0.025}(9, 14) = 0.26 \leq F_0 = 2.55 \leq 3.21 = f_{0.975}(9, 14)$, al livello $\alpha = 0.05$ accetteremo l'ipotesi nulla $H_0 : \sigma_X^2 = \sigma_Y^2$.

9.5 Test del χ^2 di adattamento

Prenderemo ora in esame il problema di decidere se un dato campione di n misure possa essere considerato estratto da una v.a. X che segue una determinata legge. L'oggetto della nostra indagine, quindi, non è più il valore di qualche parametro θ della legge di X , ma l'eventuale **adattamento** dei dati sperimentali ad una ipotetica legge teorica di X . Per semplicità supporremo che X assuma solo un numero finito m di possibili valori e che le probabilità *teoriche* di tali valori siano p_1, \dots, p_m . Le ipotesi del nostro test sono quindi ora

$$\begin{aligned} H_0 & : \text{ il campione segue la legge delle } p_1, \dots, p_m, \\ H_1 & : \text{ il campione segue un'altra legge.} \end{aligned}$$

Siccome nella discussione che segue i valori numerici di X non giocano nessun ruolo, tutta la procedura può essere facilmente adattata anche al caso di variabili qualitative. Inoltre il test può essere eseguito anche se X assume infiniti valori, o è continua, scegliendo qualche opportuno metodo per raggruppare i dati in un numero finito m di classi. Supponiamo allora che il valore j -mo di X venga sperimentalmente trovato N_j volte con $j = 1, \dots, m$, e che quindi $\bar{p}_j = N_j/n$ siano le corrispondenti frequenze relative (si noti che N_j e \bar{p}_j sono v.a., mentre p_j sono numeri): è evidente che il test sarà eseguito con un confronto fra le p_j teoriche e le \bar{p}_j empiriche. Si introduce a questo scopo la statistica

$$D_0 = n \sum_{j=1}^m \frac{(\bar{p}_j - p_j)^2}{p_j} = \sum_{j=1}^m \frac{(N_j - np_j)^2}{np_j} \quad (9.23)$$

e si prova che, se l'ipotesi H_0 è verificata e se n è abbastanza grande, allora D_0 segue approssimativamente una legge Chi quadro $\chi^2(m-1)$ con $m-1$ gradi di libertà. Viceversa, se H_0 non è verificata le differenze $\bar{p}_j - p_j$ non sono trascurabili e D_0 tende a prendere valori piuttosto grandi. La regione critica di un test di livello α è allora

$$\{D_0 > \chi_{1-\alpha}^2(m-1)\}$$

dove $\chi_{1-\alpha}^2(m-1)$ sono i quantili della legge del Chi quadro che possono essere ricavati dalle Tavole D.3. La procedura del test è quindi la seguente: *a partire dalle frequenze osservate si calcola il valore di D_0 in (9.23) e lo si confronta con il quantile $\chi_{1-\alpha}^2(m-1)$ (dove m è il numero dei possibili valori di X): se $D_0 \leq \chi_{1-\alpha}^2(m-1)$ si accetta H_0 ; se invece $D_0 > \chi_{1-\alpha}^2(m-1)$ l'ipotesi H_0 viene rifiutata.*

L'uso di questo test richiede infine alcune precisazioni. Prima di tutto non sono noti risultati rigorosi circa la grandezza di n necessaria per rendere applicabile questo test, ci sono solo regole empiriche la più nota delle quali è la seguente: deve sempre risultare $np_j \geq 5$ per tutti gli indici $j = 1, \dots, m$. Se questo non avviene l'espedito più comune per aggirare la difficoltà è quello di unificare le classi con le frequenze N_j più piccole in modo da formarne altre più grandi che rispettino la condizione $np_j \geq 5$. In secondo luogo va osservato che a volte la distribuzione teorica p_j non è completamente nota. Ad esempio ci si potrebbe porre il problema di un adattamento dei dati ad una legge Binomiale $B(n, p)$ per la quale p sia sconosciuto. In questo caso prima di applicare il test bisognerà stimare i parametri necessari a partire dai dati X_1, \dots, X_n , e la procedura del test deve essere modificata nel modo seguente: *se la conoscenza completa della distribuzione teorica richiede la determinazione di $q < m-1$ parametri, bisogna prima stimare questi parametri mediante gli opportuni stimatori di MV, poi si deve calcolare D_0 di (9.23) e infine si applica il test usando come regione critica di livello α l'evento*

$$\{D_0 > \chi_{1-\alpha}^2(m-q-1)\}.$$

In pratica si devono sottrarre tanti gradi di libertà quanti sono i parametri stimati.

Esempio 9.8. Riprendiamo la discussione dell'Esempio 8.5 sul numero X di figli maschi nelle famiglie con 12 figli. Si ricorderà che le ipotesi iniziali più semplici ci avevano condotto ad un modello nel quale la v.a. X è Binomiale $B(12, \frac{1}{2})$: vogliamo sottoporre questa ipotesi ad un test del χ^2 ricordando che ora

$$p_k = \binom{12}{k} \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{12-k} = \binom{12}{k} \frac{1}{2^{12}}, \quad k = 0, 1, \dots, 12$$

Bisogna però subito dire che per questa distribuzione le condizioni di applicabilità del test non sono rispettate. Infatti per le classi estreme si ha $p_0 = p_{12} = 2^{-12}$ e quindi

$$np_0 = np_{12} = 6115 \times 2^{-12} = 1.49 < 5$$

Per risolvere questo problema dovremo allora unificare le classi $\{X = 0\}$ e $\{X = 12\}$ rispettivamente con le contigue $\{X = 1\}$ e $\{X = 11\}$ in modo da avere

$$\begin{aligned} n(p_0 + p_1) &= 6115 \times (2^{-12} + 12 \times 2^{-12}) = 6115 \times 13 \times 2^{-12} = 19.47 > 5 \\ n(p_{11} + p_{12}) &= 6115 \times (12 \times 2^{-12} + 2^{-12}) = 6115 \times 13 \times 2^{-12} = 19.47 > 5 \end{aligned}$$

In conclusione dovremo eseguire un test del χ^2 non sulla originaria distribuzione Binomiale delle 13 p_k con $k = 0, 1, \dots, 12$, ma sulla distribuzione Binomiale *raggruppata* delle 11 q_j

così definite:

$$q_j = \begin{cases} p_0 + p_1 & \text{per } j = 1, \\ p_j & \text{per } j = 2, \dots, 10, \\ p_{11} + p_{12} & \text{per } j = 11. \end{cases}$$

Tenendo allora conto dei valori empirici delle frequenze N_0, \dots, N_{12} riportati nella Tabella 8.1, si introducono le frequenze per le classi raggruppate

$$M_j = \begin{cases} N_0 + N_1 & \text{per } j = 1, \\ N_j & \text{per } j = 2, \dots, 10, \\ N_{11} + N_{12} & \text{per } j = 11. \end{cases}$$

e con $m = 11$ si calcola la statistica (9.23) che vale

$$D_0 = \sum_{j=1}^m \frac{(M_j - nq_j)^2}{nq_j} = 242.05$$

Siccome il corrispondente quantile del χ^2 è $\chi_{1-\alpha}^2(m-1) = \chi_{0.95}^2(10) = 18.31$, il test decisamente rigetta l'ipotesi che la distribuzione empirica si adatti alla distribuzione Binomiale $B(12, \frac{1}{2})$. Come seconda possibilità possiamo allora provare a verificare l'ipotesi – avanzata sempre nell'Esempio 8.5 – secondo la quale la distribuzione di X è Binomiale $B(12, \bar{p})$, con \bar{p} stima di MV di p ottenuta a partire dai dati empirici. Il valore di \bar{p} è già stato calcolato nell'Esempio 8.5 mediante lo stimatore di MV che coincide con la media dei valori empirici di X . In realtà per essere rigorosi dovremmo rideterminare lo stimatore perché ora abbiamo modificato il modello raggruppando le classi estreme. Siccome però la correzione sarebbe piccola e non modificherebbe il risultato del test, noi assumeremo che la vecchia stima $\bar{p} = 0.519$ possa essere tranquillamente usata come stima di MV del parametro p . A questo punto dovremo seguire la stessa procedura già seguita in precedenza con due modifiche: bisogna prima di tutto ricalcolare le p_k e le q_j tenendo conto del fatto che ora p vale 0.519, e non $\frac{1}{2}$; in secondo luogo bisogna calcolare di nuovo il valore della statistica (9.23) con i nuovi parametri $D_0 = 105.79$, e confrontarlo con il valore del quantile $\chi_{1-\alpha}^2(m-1-1) = \chi_{0.95}^2(9) = 16.92$ dato che siamo stati obbligati a stimare un parametro per definire la distribuzione teorica. Sebbene in questo secondo calcolo il valore di D_0 si sia più che dimezzato, il test del χ^2 chiaramente indica ancora che la distribuzione empirica non può essere considerata ben adattata ad una distribuzione Binomiale. Va infine notato che l'esito del test non sarebbe modificato se scegliessimo valori diversi e ragionevolmente piccoli per il livello α

Esempio 9.9. Supponiamo di voler verificare empiricamente il Teorema 5.4 di Poisson secondo il quale, in particolari condizioni, una legge Binomiale può essere ben approssimata da una legge di Poisson. Per far questo simuliamo un campione di $n = 2000$ numeri estratti da una v.a. X Binomiale $B(200, 0.01)$ e confrontiamo le frequenze empiriche dei suoi valori con le frequenze teoriche della legge di Poisson $\mathcal{P}(2)$. I valori delle frequenze assolute del campione sono riportati nella Tabella 9.4. Prima di applicare il test bisogna però osservare che per una legge di Poisson le p_k sono in numero infinito, mentre il test del χ^2 è stato

k	0	1	2	3	4	5	6	7	8
N_k	280	545	544	355	186	55	25	9	1

Tabella 9.4: Frequenze di un campione simulato di 2 000 numeri estratti secondo la legge Binomiale $B(200, 0.01)$.

classi	0	1	2	3	4	5	6	≥ 7
M_j	280	545	544	355	186	55	25	10

Tabella 9.5: Rispetto alla Tabella 9.4 le frequenze dei valori 7 e 8 sono state unificate in un'unica classe ≥ 7 .

costruito solo per un numero finito m di valori; inoltre per la legge $\mathcal{P}(2)$ da noi considerata si ha

$$2000 \times p_7 = 6.87, \quad 2000 \times p_8 = 1.72, \quad \dots$$

per cui praticamente la condizione di applicabilità del test non è verificata per tutti i valori $k \geq 7$. Ambedue i problemi possono essere superati raggruppando tutti i valori $k \geq 7$ in un'unica classe. In questo modo non solo avremo un numero $m = 8$ finito di classi di valori con le frequenze riportate nella Tabella 9.5, ma anche la condizione di applicabilità del test sarà rispettata perché ora

$$2000 \times (p_7 + p_8 + \dots) = 9.07$$

A questo punto con $m = 8$ possiamo calcolare la statistica (9.23)

$$D_0 = \sum_{j=1}^m \frac{(M_j - nq_j)^2}{nq_j} = 4.85$$

ed eseguire un test di livello $\alpha = 0.05$: siccome risulta $\chi_{0.95}^2(7) = 14.07 > 4.85 = D_0$ il test suggerisce di accettare l'ipotesi H_0 secondo la quale il campione segue una legge di Poisson $\mathcal{P}(2)$.

Esempio 9.10. Proviamo a controllare empiricamente le conclusioni del TLC 7.2: con le notazioni di quel Teorema, la legge di S_{30}^* , somma standardizzata di $k = 30$ v.a. indipendenti e uniformi $\mathcal{U}([0, 1])$ (con media $\mu = \frac{1}{2}$ e varianza $\sigma^2 = \frac{1}{12}$, vedi (6.26)), è ben approssimata dalla legge Normale standard $\mathcal{N}(0, 1)$. Viene simulato un campione di $n = 1000$ valori di S_{30}^* : siccome si tratta di una v.a. continua, l'applicazione del test del χ^2 richiede preventivamente la definizione degli intervalli rispetto ai quali vanno calcolate le frequenze. Naturalmente la scelta va operata tenendo conto delle condizioni di applicabilità del test. Osserviamo a questo scopo che per una v.a. Z Normale standard $\mathcal{N}(0, 1)$ si ha dalle Tavole D.1

$$\begin{aligned} n \mathbf{P}\{Z \leq -2\} &= 1000 \times \Phi(-2) = 1000 \times [1 - \Phi(2)] = 22.75 > 5 \\ n \mathbf{P}\{Z > 2\} &= n[1 - \mathbf{P}\{Z \leq 2\}] = 1000 \times [1 - \Phi(2)] = 22.75 > 5 \end{aligned}$$

similmente si provano relazioni analoghe per tutti gli intervalli di ampiezza 0.5 presi fra -2 e 2; così ad esempio

$$n \mathbf{P}\{1.5 \leq Z \leq 2.0\} = 1000 \times [\Phi(2.0) - \Phi(1.5)] = 44.06 > 5$$

classi	≤ -2.0	$[-2.0,-1.5]$	$[-1.5,-1.0]$	$[-1.0,-0.5]$	$[-0.5,0.0]$
N_j	27	53	100	147	191
classi	$[0.0,0.5]$	$[0.5,1.0]$	$[1.0,1.5]$	$[1.5,2.0]$	≥ 2.0
N_j	183	143	79	38	39

Tabella 9.6: Frequenze di un campione di $n = 1000$ numeri ottenuti come somma standardizzata di $k = 30$ v.a. Uniformi $\mathcal{U}([0, 1])$ indipendenti.

Definiremo allora $m = 10$ classi utilizzando gli 8 intervalli finiti delimitati dai punti $-2.0, -1.5, \dots, 2.0$, e i due intervalli infiniti $(-\infty, -2.0]$ e $[2.0, +\infty)$. Le frequenze del nostro campione simulato in queste classi sono riportate nella Tabella 9.6. A questo punto, ricordando che per noi, con Z Normale standard, si ha

$$p_j = \begin{cases} \mathbf{P}\{Z \leq -2\} = 1 - \Phi(2) & \text{per } j = 1, \\ \mathbf{P}\left\{-3 + \frac{j}{2} \leq Z \leq -3 + \frac{j+1}{2}\right\} = \Phi\left(-3 + \frac{j+1}{2}\right) - \Phi\left(-3 + \frac{j}{2}\right) & \text{per } j = 2, \dots, 9, \\ \mathbf{P}\{Z > 2\} = 1 - \Phi(2) & \text{per } j = 10. \end{cases}$$

siamo in grado, usando le Tavole D.1, di calcolare la statistica (9.23)

$$D_0 = \sum_{j=1}^m \frac{(N_j - np_j)^2}{np_j} = 18.32$$

e di eseguire un test di livello $\alpha = 0.05$: siccome risulta $\chi_{0.95}^2(9) = 16.92 < 18.32 = D_0$ il test suggerisce di rifiutare l'ipotesi H_0 secondo la quale il campione delle somme standardizzate segue una legge Normale standard. Le conclusioni di questo test non sono però particolarmente solide: infatti, se avessimo scelto un livello $\alpha = 0.01$, dalla relazione $\chi_{0.99}^2(9) = 21.67 > 18.32 = D_0$ avremmo ricavato per H_0 la conclusione opposta.

9.6 Test del χ^2 di indipendenza

Una versione particolare del test del χ^2 di adattamento viene usata come test di indipendenza di due campioni accoppiati X_1, \dots, X_n e Y_1, \dots, Y_n di due v.a. X e Y . Supponendo che queste due v.a. possano assumere solo un numero finito di valori (anche di tipo qualitativo), rispettivamente u_1, \dots, u_r e v_1, \dots, v_s , e riprendendo una notazione già introdotta nella Sezione 2.1 indicheremo con N_{jk} il numero di volte in cui nel campione compare la coppia di valori (u_j, v_k) , e con $N_{j\cdot}$ e $N_{\cdot,k}$ il numero di volte in cui si presentano separatamente rispettivamente i valori u_j e v_k . Le probabilità teoriche congiunte p_{jk} , e marginali p_j e q_k con le quali si presentano i valori delle v.a. non sono note, ma possono essere stimate tramite i loro stimatori MV, cioè tramite le frequenze empiriche relative congiunte e marginali

$$\bar{p}_{jk} = \frac{N_{jk}}{n}, \quad \bar{p}_j = \frac{N_{j\cdot}}{n}, \quad \bar{q}_k = \frac{N_{\cdot,k}}{n}; \quad \begin{cases} j = 1, \dots, r \\ k = 1, \dots, s \end{cases}$$

		FIS			marginali
		<i>alti</i>	<i>medi</i>	<i>bassi</i>	
MAT	<i>alti</i>	56 <i>30.80</i>	71 <i>72.66</i>	12 <i>35.54</i>	139
	<i>medi</i>	47 <i>54.95</i>	163 <i>129.64</i>	38 <i>63.41</i>	248
	<i>bassi</i>	14 <i>31.24</i>	42 <i>73.70</i>	85 <i>36.05</i>	141
marginali		117	276	135	528

Tabella 9.7: Frequenze dei voti in fisica e matematica di $n = 528$ studenti.

Inoltre ricorderemo da (5.27) che l'indipendenza di X e Y equivale alla richiesta che fra le probabilità teoriche sussista la relazione $p_{jk} = p_j q_k$; pertanto è intuitivo pensare di realizzare un test di indipendenza confrontando le frequenze congiunte empiriche \bar{p}_{jk} con i prodotti $\bar{p}_j \bar{q}_k$. A questo scopo si introduce la statistica

$$D_0 = n \sum_{j,k} \frac{(\bar{p}_{jk} - \bar{p}_j \bar{q}_k)^2}{\bar{p}_j \bar{q}_k} = \sum_{j,k} \frac{(N_{jk} - n \bar{p}_j \bar{q}_k)^2}{n \bar{p}_j \bar{q}_k} \quad (9.24)$$

e si dimostra che, se le frequenze relative marginali non sono troppo piccole, essa segue una legge $\chi^2[(r-1)(s-1)]$ con $(r-1)(s-1)$ gradi di libertà. Pertanto in questo caso la regione critica di livello α è

$$\{D_0 > \chi_{1-\alpha}^2[(r-1)(s-1)]\}.$$

Si noti che dalle definizioni date risulta che le quantità che in D_0 sono paragonate con le N_{jk} possono anche essere calcolate come

$$n \bar{p}_j \bar{q}_k = \frac{N_{j \cdot} N_{\cdot k}}{n} \quad (9.25)$$

espressione che risulta di uso più facile nelle tabelle di contingenza. Il test di indipendenza può essere applicato anche raggruppando i possibili valori di X e Y in un piccolo numero di classi: in particolare questa osservazione rende possibile l'applicazione del test anche al caso di v.a. con un numero infinito, anche continuo, di valori.

Esempio 9.11. Nella Tabella di contingenza 9.7 sono riportate le frequenze dei voti (raggruppati in tre classi: *alti*, *medi* o *bassi*) riportati in esami di Fisica e Matematica da $n = 528$ studenti. Si vuole esaminare l'ipotesi H_0 secondo la quale i risultati degli esami

		ETÀ					marginali
		18-30	31-40	41-50	51-60	61-70	
INCIDENTI	0	748 <i>770.13</i>	821 <i>818.37</i>	786 <i>772.81</i>	720 <i>720.99</i>	672 <i>664.70</i>	3747
	1	74 <i>61.87</i>	60 <i>65.74</i>	51 <i>62.08</i>	66 <i>57.92</i>	50 <i>53.40</i>	301
	2	31 <i>22.40</i>	25 <i>23.81</i>	22 <i>22.48</i>	16 <i>20.97</i>	15 <i>19.34</i>	109
	> 2	9 <i>7.60</i>	10 <i>8.08</i>	6 <i>7.63</i>	5 <i>7.12</i>	7 <i>6.56</i>	37
marginali		862	916	865	807	744	4194

Tabella 9.8: Numero di incidenti per classi di età di $n = 4194$ guidatori.

in Fisica e Matematica sono indipendenti. Innanzitutto è conveniente completare subito la tabella delle frequenze congiunte N_{jk} con le frequenze marginali $N_{j\cdot}$ e $N_{\cdot k}$ ottenute sommando lungo le righe e lungo le colonne, e con il valore del numero totale delle misure n . Nella stessa tabella, poi, sotto ogni valore delle frequenze congiunte è stato riportato il valore delle quantità (9.25) che servono per calcolare la statistica (9.24), il valore della quale è

$$D_0 = \sum_{j,k} \frac{(N_{jk} - n\bar{p}_j\bar{q}_k)^2}{n\bar{p}_j\bar{q}_k} = 145.78$$

Siccome i numeri delle classi sono $r = s = 3$ il test del χ^2 di livello $\alpha = 0.05$ consiste semplicemente nell'osservare che $\chi_{0.95}^2(4) = 9.49 < 145.78 = D_0$ per cui la conclusione è che al livello $\alpha = 0.05$ si rifiuta l'ipotesi H_0 di indipendenza fra i voti di Fisica e Matematica. La conclusione non cambia se si passa al livello $\alpha = 0.01$ dato che $\chi_{0.99}^2(4) = 13.28 < 145.78 = D_0$.

Esempio 9.12. Nella Tabella 9.8 sono raccolti i risultati di un'inchiesta volta a stabilire se l'età di un guidatore (divisa in $r = 5$ classi) influenzi il numero degli incidenti automobilistici (diviso in $s = 4$ classi). Nella tabella sono anche riportati i valori delle quantità (9.25) che servono per calcolare la statistica (9.24) per la quale si ha $D_0 = 14.40$. Per un test di livello $\alpha = 0.05$ si ha $\chi_{0.95}^2(12) = 21.03 > 14.40 = D_0$, e per un test di livello $\alpha = 0.01$ si ha $\chi_{0.99}^2(12) = 26.22 > 14.40 = D_0$. Pertanto in ambedue i casi il test suggerisce di accettare l'ipotesi H_0 di indipendenza fra età e numero di incidenti. Va però detto che qualche dubbio sulla validità del test deve essere avanzato a cauda del fatto che, come si vede dai dati, nella tabella di contingenza sono presenti frequenze relative piuttosto piccole.

Parte IV
Appendici

Appendice A

Esercizi

A.1 Esercizi di Statistica Descrittiva

Esercizio A.1.1. Sia dato il seguente campione composto di $n = 19$ numeri interi:

5, 4, 5, 4, 4, 1, 5, 6, 5, 4, 2, 3, 7, 5, 3, 3, 4, 3, 4.

Calcolare: la media, la varianza, lo scarto quadratico, il range, la mediana, i due quartili, la media armonica e la media geometrica. Costruire la tabella di frequenza. Disegnare il diagramma a barre del campione e il suo *boxplot*.

Esercizio A.1.2. Nella fascia oraria fra le 12:00 e le 13:00 di ogni giorno un centralino telefonico riceve un numero aleatorio X di chiamate. Il valore di X è stato registrato in 23 giorni diversi ottenendo i seguenti risultati:

10, 14, 11, 10, 11, 10, 8, 6, 12, 7, 16, 10, 11, 6, 8, 4, 9, 12, 7, 6, 14, 4, 11.

Costruire la tabella di frequenza e disegnare il diagramma a barre del campione. Calcolare la media, la varianza, lo scarto quadratico, la mediana, i due quartili, la media armonica e la media geometrica e disegnare il *boxplot*.

Esercizio A.1.3. Il numero di particelle α emesso da un campione radioattivo in ogni periodo di 10 secondi è una v.a. X : supponendo che in 31 misurazioni (di 10 secondi l'una) le frequenze N_k dei valori k di X siano state:

$$\begin{array}{cccccccc} k = & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ N_k = & 3 & 7 & 10 & 8 & 1 & 1 & 0 & 1 \end{array}$$

calcolare la media, la varianza, la mediana e i quartili del campione, e disegnarne l'istogramma delle frequenze assolute e il *boxplot*.

Esercizio A.1.4. $n = 100$ giocatori di roulette partono con un capitale di 5\$ ciascuno e alla fine del gioco hanno perduto tutto. Si registrano i valori massimi del capitale raggiunto da ogni giocatore durante il gioco ottenendo la seguente tabella:

25	9	5	5	5	9	6	5	15	45	35	6	5	6	24	21	16	5	8	7
7	5	5	35	13	9	5	18	6	10	19	16	21	8	13	5	9	10	10	6
23	8	5	10	15	7	5	5	24	9	11	34	12	11	17	11	16	5	15	5
12	6	5	5	7	6	17	20	7	8	8	6	10	11	6	7	5	12	11	18
6	21	6	5	24	7	16	21	23	15	11	8	6	8	14	11	6	9	6	10

Costruire la tabella di frequenza e disegnare il diagramma a barre. Calcolare media, varianza, moda, mediana e quartili, e disegnare il *boxplot*.

Esercizio A.1.5. $n = 53$ misure di una quantità aleatoria forniscono i seguenti risultati riportati in ordine crescente:

1.70 1.93 2.42 2.52 2.59 2.66 2.72 2.76 2.88 3.01
 3.05 3.12 3.12 3.15 3.15 3.17 3.32 3.36 3.40 3.54
 3.63 3.71 3.71 3.72 3.81 3.95 4.01 4.01 4.04 4.04
 4.07 4.07 4.15 4.17 4.42 4.43 4.46 4.52 4.56 4.78
 4.83 5.13 5.15 5.15 5.30 5.33 5.33 5.39 5.44 5.61
 5.66 5.83 6.90

- Determinare *range*, mediana e quartili, e disegnare il *boxplot*.
- Calcolare la media.
- Costruire la tabella di frequenza raggruppando i dati nei seguenti 8 intervalli di ampiezze differenti:

[0.0, 2.0] (2.0, 3.0] (3.0, 3.5] (3.5, 4.0]
 (4.0, 4.5] (4.5, 5.0] (5.0, 6.0] (6.0, 8.0]

disegnare l'istogramma e calcolare la moda.

- Calcolare media e varianza approssimate con i dati raggruppati.

Esercizio A.1.6. $n = 40$ misure di una quantità aleatoria forniscono i seguenti risultati:

0,12 0,13 0,18 0,18 0,21 0,25 0,30 0,35 0,46 0,54
 0,87 0,92 1,10 1,19 1,43 1,45 1,47 1,67 1,79 1,84
 1,89 1,90 1,91 1,91 1,97 1,98 2,09 2,26 2,35 2,70
 2,75 3,39 3,58 3,62 3,89 4,20 5,50 6,43 6,96 8,48

- Determinare *range*, mediana e quartili, e disegnare il *boxplot*.
- Calcolare la media e la moda del campione.
- Costruire la tabella di frequenza raggruppando i dati negli intervalli

[0, 1], (1, 2], (2, 3], ..., (8, 9].

e disegnare l'istogramma.

- Calcolare media e varianza approssimate con i dati raggruppati.

Esercizio A.1.7. $n = 40$ misure di una quantità aleatoria forniscono i seguenti risultati:

0,04 0,12 0,23 0,37 0,47 0,59 0,64 0,76 0,80 0,97
 1,00 1,01 1,08 1,11 1,22 1,33 1,53 1,61 1,63 2,00
 2,03 2,19 2,25 2,36 2,77 2,96 3,05 3,10 3,34 3,79
 3,85 4,56 5,27 5,79 5,82 6,41 7,88 7,99 8,16 10,00

- Determinare *range*, mediana e quartili, e disegnare il *boxplot*.
- Calcolare la media e la moda del campione.
- Costruire la tabella di frequenza raggruppando i dati negli intervalli

$[0, 1], (1, 2], (2, 3], \dots, (9, 10]$.

e disegnare l'istogramma.

- Calcolare media e varianza approssimate con i dati raggruppati.

Esercizio A.1.8. $n = 40$ misure di velocità del vento in una stazione meteorologica forniscono i seguenti risultati:

0.12 0.23 0.24 0.50 0.50 0.56 0.77 1.01 1.03 1.10
 1.40 1.45 1.64 1.68 1.72 1.72 1.81 1.83 1.84 2.14
 2.28 2.31 2.34 2.43 2.55 2.91 3.00 3.12 3.53 3.59
 3.94 4.17 4.70 4.73 5.02 5.07 6.73 6.97 7.09 9.74

- Costruire la tabella di frequenza raggruppando i dati in 5 classi con estremi degli intervalli 0, 2, 4, 6, 8, 10, e disegnare l'istogramma.
- Calcolare media e varianza per dati raggruppati. Determinare primo secondo e terzo quartile e disegnare il *boxplot*.

Esercizio A.1.9. Un'azienda vuol condurre un'indagine sulla propria clientela misurando i consumi di un determinato prodotto in 90 famiglie. Si ottenengono i seguenti risultati:

1.69 1.71 1.72 1.92 1.99 2.01 2.04 2.06 2.09 2.50
 2.52 2.53 2.55 2.66 2.81 2.82 2.93 2.98 3.04 3.36
 3.48 3.54 3.64 3.71 3.75 3.82 3.85 3.91 4.01 4.11
 4.23 4.23 4.23 4.36 4.38 4.54 4.55 4.59 4.76 4.83
 5.10 5.19 5.29 5.29 5.61 5.69 5.71 5.78 5.91 5.92
 5.93 6.08 6.22 6.38 6.70 6.85 6.89 7.00 7.06 7.09
 7.23 7.42 7.45 7.55 7.66 7.80 7.81 7.94 7.96 7.99
 8.77 8.88 9.10 9.20 9.38 9.41 9.75 10.02 10.07 10.41
 11.28 11.38 11.82 11.86 12.37 12.53 13.54 14.22 15.80 18.37

- Costruire la tabella di frequenza raggruppando i dati in 9 classi con estremi degli intervalli 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 e disegnare l'istogramma.
- Calcolare media e varianza per dati raggruppati. Determinare primo secondo e terzo quartile e disegnare il *boxplot*.

Esercizio A.1.10. La seguente tabella contiene i pesi in grammi di 40 prodotti:

21.3 21.6 21.8 21.8 22.1 22.2 22.2 22.2 22.4 22.4
 22.5 22.5 22.5 22.6 22.6 22.8 22.8 22.8 22.9 23.0
 23.0 23.0 23.1 23.2 23.2 23.4 23.5 23.5 23.5 23.6
 23.8 23.8 23.9 23.9 24.0 24.0 24.3 24.6 24.6 24.9

- Costruire la tabella di frequenza raggruppando i dati in 8 classi di ampiezza 0.5 considerando come primo intervallo $[20.95, 21.45]$, e disegnare l'istogramma.
- Calcolare media e varianza. Determinare primo secondo e terzo quartile e disegnare il *boxplot*.

Esercizio A.1.11. $n = 28$ misure di una quantità aleatoria forniscono i seguenti risultati:

14.00 5.99 26.35 35.95 15.95 24.95 19.95 32.95 59.00 9.95
 69.95 61.35 14.95 12.95 16.95 10.95 57.35 29.95 5.95 41.95
 66.95 19.85 11.95 15.95 50.25 74.65 68.00 69.95

- Costruire la tabella di frequenza raggruppando i dati in 7 classi di ampiezza 10 considerando come primo intervallo $[5.455, 15.455]$, e disegnare l'istogramma.
- Calcolare media e varianza per dati raggruppati. Determinare primo secondo e terzo quartile e disegnare il *boxplot*.

Esercizio A.1.12. $n = 20$ misure di due caratteri X e Y forniscono i seguenti risultati

X	Y	X	Y	X	Y	X	Y
0.02	-0.13	0.36	-0.34	0.58	0.62	0.82	2.62
0.03	0.82	0.37	0.38	0.63	0.05	0.84	1.44
0.16	-0.96	0.52	-0.37	0.66	-0.13	0.87	0.43
0.23	-0.57	0.54	0.26	0.67	0.22	0.89	0.61
0.27	1.59	0.57	0.84	0.72	1.59	0.96	-0.38

Calcolare le medie e le varianze di X e Y , la covarianza, il coefficiente di correlazione e i parametri della retta di regressione. Rappresentare i dati e la retta di regressione nel piano X, Y .

Esercizio A.1.13. $n = 28$ misure di due caratteri X e Y forniscono i seguenti risultati

X	Y	X	Y	X	Y	X	Y
0.04	1.41	0.28	0.53	0.45	1.05	0.65	0.26
0.07	-0.31	0.29	0.39	0.45	1.23	0.65	1.08
0.18	-1.18	0.30	-0.89	0.51	0.40	0.83	1.88
0.18	-1.01	0.30	1.20	0.55	1.04	0.88	0.48
0.21	-0.15	0.39	-0.52	0.57	-0.91	0.88	1.42
0.22	-0.31	0.42	1.79	0.58	-1.33	0.89	0.36
0.22	0.32	0.45	0.74	0.62	1.73	0.97	-0.37

Calcolare le medie e le varianze di X e Y , la covarianza, il coefficiente di correlazione e i parametri della retta di regressione. Rappresentare i dati e la retta di regressione nel piano X, Y .

A.2 Esercizi di Probabilità

Esercizio A.2.1. Per incoraggiare la carriera tennistica di Mario suo padre gli promette un premio se egli riesce a vincere almeno due partite di seguito in una serie di tre partite giocate alternativamente con suo padre (P) e con il campione della loro associazione (C). Mario può scegliere di iniziare la serie con suo padre (PCP) o con il campione (CPC). Sapendo che il campione gioca meglio del padre, quale successione di partite gli converrà scegliere?

Esercizio A.2.2. Un cassetto contiene calzini rossi e bianchi e si sa che, se si estraggono a caso due calzini, la probabilità che siano ambedue rossi è $\frac{1}{2}$. Qual è il più piccolo numero di calzini (bianchi e rossi) che rende possibile questa situazione? Qual è la risposta alla domanda precedente se è noto che il numero dei calzini bianchi è pari?

Esercizio A.2.3. Un gruppo di candidati viene sottoposto ad un test con un questionario di $n = 8$ domande a risposta multipla. Ogni domanda ha 4 possibili risposte. In test viene superato rispondendo correttamente ad almeno 6 domande. Un candidato completamente impreparato risponde in maniera del tutto casuale:

- in media a quante domande risponderà correttamente?
- con quale probabilità supererà il test?

Esercizio A.2.4. Una partita di 100 oggetti ne contiene 5 difettosi. La partita viene sottoposta ad una verifica alla fine della quale essa può essere accettata o rifiutata. Per quale delle seguenti procedure di verifica è maggiore la probabilità di respingere la partita:

- si scelgono 6 oggetti e si respinge la partita se se ne trova almeno uno difettoso;
- si scelgono 20 oggetti e si respinge la partita se se ne trova non più di uno difettoso.

Esercizio A.2.5. Un venditore deve fare 10 telefonate ogni giorno per convincere dei clienti ad acquistare un prodotto; dall'esperienza precedente si sa che il cliente acquista il prodotto nel 15% dei casi. Supponendo che gli esiti delle telefonate siano indipendenti, calcolare:

- il numero medio μ di prodotti venduti al giorno e lo scarto quadratico σ ;
- la probabilità di vendere meno di tre prodotti al giorno.

Se la percentuale dei successi fosse solo del 5%, con quale probabilità 10 venditori venderebbero esattamente 4 prodotti in un giorno? (Suggerimento: usare l'approssimazione di Poisson).

Esercizio A.2.6. Un telegrafo trasmette *punti* e *linee*, ed è noto che la frequenza dei *punti* al momento della trasmissione è $\frac{5}{8}$ mentre quella delle *linee* è $\frac{3}{8}$. Disturbi di trasmissione modificano $\frac{2}{5}$ dei *punti* in *linee* ed $\frac{1}{3}$ delle *linee* in *punti*. Supponendo di ricevere un messaggio, calcolare la probabilità che il segnale ricevuto sia proprio quello trasmesso, nel caso in cui si riceve *punto* ed in quello in cui si riceve *linea*.

Esercizio A.2.7. Quando le condizioni meteorologiche sono favorevoli (H) e il pilota può vedere la pista, un aereo atterra felicemente (A) con probabilità $\mathbf{P}(A|H) = p$. Se invece le condizioni meteorologiche impediscono di vedere la pista (\overline{H}) il pilota deve eseguire un atterraggio strumentale, e l'affidabilità degli strumenti – cioè la probabilità di funzionare correttamente (T) in condizioni di tempo sfavorevole – è $\mathbf{P}(T|\overline{H}) = q$: se gli strumenti funzionano correttamente la probabilità di un atterraggio felice resta invariata $\mathbf{P}(A|\overline{H} \cap T) = p$; se invece gli strumenti danno problemi la probabilità di un atterraggio felice è $\mathbf{P}(A|\overline{H} \cap \overline{T}) = p^* < p$.

- Sapendo che le condizioni meteorologiche sono favorevoli con probabilità $\mathbf{P}(H) = s$ determinare la probabilità totale $\mathbf{P}(A)$ di un atterraggio felice.
- Supponendo che un aereo sia atterrato felicemente determinare la probabilità $\mathbf{P}(\overline{H}|A)$ che il pilota sia stato costretto dalle sfavorevoli condizioni meteorologiche ad un atterraggio strumentale.

Esercizio A.2.8. Sia X una variabile aleatoria normale con media $\mu = -2$ e varianza $\sigma^2 = 4$: facendo uso delle tavole calcolare la probabilità $\mathbf{P}(-3 \leq X \leq -1)$.

Esercizio A.2.9. Sia X una v.a. Binomiale $B(n, p)$ con $n = 1000$ e $p = 0.003$: usando l'approssimazione di Poisson si calcoli la probabilità $\mathbf{P}\{2 \leq X \leq 4\}$.

Esercizio A.2.10. Supponiamo di estrarre $n = 40$ valori da una v.a. normale con media $\mu = 3$ e varianza $\sigma^2 = 6$: quale è il numero medio di valori atteso nell'intervallo $[3, 4]$?

Esercizio A.2.11. Il numero di particelle α emesso da un campione radioattivo in ogni intervallo di 10 secondi è una v.a. X che segue una distribuzione di Poisson di parametro $\lambda = 2$: calcolare $\mathbf{P}(X > 7)$.

Esercizio A.2.12. Nella fascia oraria fra le 12:00 e le 13:00 di ogni giorno un centralino telefonico riceve un numero aleatorio X di chiamate con una media $\lambda = 10$. Supponendo che X sia distribuita secondo Poisson, calcolare la probabilità $\mathbf{P}(X \leq 4)$ di ricevere non più di 4 telefonate.

Esercizio A.2.13. Data una v.a. X normale $\mathcal{N}(0, \sigma^2)$, si fissi in maniera arbitraria un intervallo $[a, b]$ con $0 < a < b$. Facendo uso delle funzioni di distribuzione cumulativa e di densità normali standard

$$\Phi(x) = \int_{-\infty}^x \varphi(y) dy, \quad \varphi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

- determinare (in funzione di a e b) il valore di σ che rende massima la probabilità $\mathbf{P}(a \leq X \leq b)$;
- facendo uso delle Tavole, calcolare esplicitamente σ e $\mathbf{P}(a \leq X \leq b)$ nel caso in cui $a = 1$ e $b = 2$.

Esercizio A.2.14. Una v.a. X ha legge con densità $f_A(x)$ se si verifica l'evento A , e legge con densità $f_B(x)$ se si verifica l'evento B ; inoltre $\mathbf{P}(A) = p$, e $\mathbf{P}(B) = q$, con $p + q = 1$.

- Determinare la densità della v.a. X (*suggerimento*: calcolare prima la funzione di distribuzione $F(x)$, e poi calcolare $f(x)$ per derivazione).
- Supponendo che le leggi con densità f_A ed f_B abbiano rispettivamente attese m_A ed m_B , e varianze σ_A^2 e σ_B^2 , calcolare l'attesa m e la varianza σ^2 di X (*suggerimento*: usare la relazione $\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$).

Esercizio A.2.15. Un segnale arriva nell'intervallo di tempo $[0, \tau]$ con probabilità p . Se esso arriva in tale intervallo l'istante di arrivo è distribuito uniformemente in $[0, \tau]$. Preso un istante t con $0 \leq t \leq \tau$, e posto

$$\begin{aligned} A &= \text{il segnale arriva in } [0, \tau] \\ B &= \text{il segnale arriva in } [0, t] \\ C &= \text{il segnale arriva in } [t, \tau] \end{aligned}$$

calcolare la probabilità che il segnale arrivi in $[t, \tau]$ supponendo che esso non sia arrivato in $[0, t]$.

Esercizio A.2.16. Due v.a. X ed Y sono legate dalla relazione $Y = 2 - 3X$. Sapendo che $\mathbf{E}X = -1$ e $\sigma_X^2 = 4$, calcolare l'attesa e la varianza di Y , la covarianza e il coefficiente di correlazione delle due v.a.

Esercizio A.2.17. Una coda di $n = 60$ persone attende di ritirare del denaro ad uno sportello bancario: la quantità di denaro prelevata da ciascuno è una v.a. con media $\mu = 50$ Euro, ed deviazione standard $\sigma = 20$ Euro. I prelevamenti sono indipendenti. Usando l'approssimazione normale, determinare l'ammontare del denaro che deve essere inizialmente in cassa per soddisfare le richieste di tutti con una probabilità di 0.95.

Esercizio A.2.18. Una fabbrica produce delle partite di $n = 10\,000$ pezzi e la probabilità che uno di tali pezzi sia difettoso è $p = 0.05$. Le cause dei difetti sono indipendenti per i diversi pezzi. I pezzi difettosi sono accumulati in un recipiente: usando l'approssimazione normale, determinare il numero di pezzi per i quali tale recipiente deve essere progettato in modo che esso non risulti insufficiente con una probabilità di 0.99.

A.3 Esercizi di Statistica Inferenziale

Esercizio A.3.1. Le misure di una quantità X sono soggette ad errori casuali: X è quindi una v.a. normale; $n = 15$ valori delle misure sono:

2.05	1.42	6.18	6.69	4.47
5.36	3.47	6.74	5.19	0.91
3.22	9.50	5.85	3.41	5.66

Determinare un intervallo di fiducia di livello $\alpha = 0.05$ per l'attesa.

Esercizio A.3.2. $n = 10$ misure di una v.a. X , gaussiana con attesa e varianza sconosciute, danno i seguenti risultati:

1.01	2.25	1.60	1.75	1.49
1.45	2.51	1.87	3.95	2.10

Determinare un intervallo di fiducia di livello $\alpha = 0.02$ per l'attesa.

Esercizio A.3.3. $n = 18$ misure di una v.a. gaussiana X forniscono i seguenti risultati:

4.09	4.56	5.01	5.49	4.82	5.56	3.95	4.04	2.63
3.78	3.58	4.52	4.86	3.65	4.44	4.62	3.97	3.63

Supponendo che la varianza $\sigma_X^2 = 1$ sia conosciuta:

- determinare l'intervallo di fiducia di livello $\alpha = 0.05$ per l'attesa μ_X ;
- eseguire un test di Student bilaterale di livello $\alpha = 0.05$ per verificare l'ipotesi $H_0 : \mu_X = 4$ contro l'ipotesi $H_1 : \mu_X \neq 4$.

Esercizio A.3.4. $n = 20$ misure di una quantità aleatoria X forniscono i seguenti risultati:

2.23 4.09 3.97 5.57 3.09 3.00 2.85 2.12 3.26 2.11
3.10 1.82 2.82 1.99 3.25 1.53 2.99 1.03 3.86 2.45

Supponendo di sapere che X è normale con media μ e varianza σ^2 non note, eseguire un test di livello $\alpha = 0.05$ per decidere tra le due ipotesi

$$H_0 : \mu = 3, \quad H_1 : \mu \neq 3.$$

Esercizio A.3.5. Le misure di pressione di un campione di $n = 200$ pneumatici di automobile hanno una media $\bar{x} = 33.57$ e una varianza $s^2 = 1.723$. Decidere tra le due ipotesi

$$H_0 : \mu = 34, \quad H_1 : \mu \neq 34$$

con un test di livello $\alpha = 0.01$.

Esercizio A.3.6. $n = 20$ misure dell'energia cinetica delle particelle di un gas forniscono – nelle opportune unità di misura – i seguenti risultati:

2.58 0.25 3.96 4.89 3.80 1.42 0.96 7.99 2.47 4.32
2.19 0.66 1.37 6.22 1.41 2.56 1.06 1.40 0.45 1.40

Supponendo che l'attesa μ e la varianza σ^2 siano sconosciute:

- determinare l'intervallo di fiducia di livello $\alpha = 0.05$ per l'attesa μ ;
- eseguire un test di Student bilaterale di livello $\alpha = 0.05$ per verificare l'ipotesi $H_0 : \mu = 3$ contro l'ipotesi $H_1 : \mu \neq 3$.

Esercizio A.3.7. $n = 10$ misure di una v.a. normale X con media e varianza sconosciute danno i seguenti risultati:

2.47 1.79 0.01 2.94 4.10
2.13 4.51 0.72 -2.99 0.83

- determinare l'intervallo di fiducia di livello $\alpha = 0.05$ per la media;
- eseguire un test di Student unilaterale di livello $\alpha = 0.05$ per verificare l'ipotesi $H_0 : \mu \leq 0$ contro l'ipotesi $H_1 : \mu > 0$.

Esercizio A.3.8. Sia dato il seguente campione di una v.a. normale X con attesa μ e varianza σ^2 sconosciute:

$$X = \begin{array}{cccccc} 2.02 & -0.87 & -1.68 & -1.39 & -0.05 \\ -2.69 & 3.14 & -2.46 & -0.05 & 1.83 \end{array}$$

- calcolare le stime \bar{X} della media e S^2 della varianza;
- determinare gli intervalli di fiducia di livello $\alpha = 0.05$ per la media e la varianza del campione;
- con un test bilaterale di livello $\alpha = 0.05$ si controlli l'ipotesi $\mu = 0$, contro l'ipotesi $\mu \neq 0$.

Esercizio A.3.9. Si vuol verificare se in un determinato impianto il livello medio del rumore non superi 90 *dB*: si effettuano 9 misure durante una giornata e si trovano (in *dB*) i seguenti valori:

$$95 \quad 98 \quad 92 \quad 84 \quad 105 \quad 92 \quad 110 \quad 86 \quad 98$$

Supponendo che il livello di rumore sia una v.a. X normale con media e varianza sconosciute, valutare con un test unilaterale di livello $\alpha = 0.05$ le due ipotesi $H_0 : \mu \leq 90$ e $H_1 : \mu > 90$.

Esercizio A.3.10. Una fabbrica produce automezzi che consumano in media 10 *lt* di carburante ogni 100 *Km* ad una data velocità. Per verificare gli effetti di un dispositivo introdotto per ridurre il consumo si misurano i *lt* di carburante consumati da un campione di 10 automezzi ottenendo i seguenti risultati:

$$9.0 \quad 12.0 \quad 11.0 \quad 7.5 \quad 10.2 \quad 9.8 \quad 13.0 \quad 12.0 \quad 12.5 \quad 10.4$$

Supponendo che il consumo sia una v.a. normale con media e varianza sconosciute, valutare con un test bilaterale di livello $\alpha = 0.05$ le due ipotesi $H_0 : \mu = 10$, e $H_1 : \mu \neq 10$.

Esercizio A.3.11. Siano X e Y la quantità di nicotina depositata rispettivamente da sigarette con e senza filtro: da un campione di $n = 9$ misure di X si ha $\bar{x} = 0.7$ e $s_X^2 = 0.03$, mentre da un campione di $m = 11$ misure di Y si ha $\bar{y} = 1.36$ e $s_Y^2 = 0.22$. Stabilire con un test unilaterale di livello $\alpha = 0.01$ se la media μ_Y è significativamente più grande della media μ_X .

Esercizio A.3.12. Per controllare l'efficacia di un nuovo medicinale si paragonano i risultati X ed Y di un certo tipo di analisi clinica eseguita su due campioni indipendenti rispettivamente di $n = 10$ ed $m = 12$ pazienti: ai 10 pazienti del primo gruppo è stato somministrato il nuovo farmaco; ai 12 del secondo gruppo è stato

somministrato solo un *placebo*. Se il farmaco è efficace la media di X deve essere più grande della media di Y . I risultati delle analisi sono i seguenti:

$$\begin{array}{rcccccc} X & = & 9.51 & 8.39 & 8.62 & 9.48 & 8.85 \\ & & 9.29 & 8.43 & 9.57 & 9.30 & 9.21 \\ Y & = & 8.10 & 8.58 & 9.05 & 7.28 & 7.64 & 5.83 \\ & & 8.61 & 7.10 & 6.44 & 7.43 & 8.63 & 7.94 \end{array}$$

Stabilire con un test unilaterale di livello $\alpha = 0.05$ se la media μ_X è significativamente più grande della media μ_Y .

Esercizio A.3.13. La differenza fra entrate e uscite (in migliaia di Euro) di una ditta è una v.a. normale con varianza 1. Vengono introdotte alcune modifiche nel sistema di vendita dei prodotti: per controllare se la situazione finanziaria è migliorata si confrontano i bilanci X di $n = 10$ mesi successivi all'introduzione di tali modifiche con quelli Y di $m = 12$ mesi precedenti ottenendo i seguenti risultati

$$\begin{array}{rcccccc} X & = & 0.61 & 0.90 & 2.76 & 1.31 & 3.33 \\ & & 2.08 & 1.42 & -0.67 & 2.22 & 3.28 \\ Y & = & 0.80 & 2.28 & 1.11 & -1.26 & 0.70 & 1.26 \\ & & 0.42 & 2.24 & 1.58 & -0.21 & -0.26 & -2.02 \end{array}$$

Supponendo che le modifiche abbiano lasciato immutata la varianza, stabilire con un test di livello $\alpha = 0.05$ se la media μ_X è aumentata rispetto alla media μ_Y .

Esercizio A.3.14. Per controllare se una nuova procedura di fabbricazione ha modificato la qualità dei prodotti di una azienda si paragonano le misure X ed Y di una data caratteristica dei prodotti prima e dopo l'introduzione della nuova procedura. Si ottengono così due campioni indipendenti rispettivamente di $n = 10$ ed $m = 12$ valori:

$$\begin{array}{rcccccc} X & = & 9.41 & 9.71 & 10.32 & 9.05 & 8.63 \\ & & 9.12 & 8.65 & 8.91 & 10.36 & 8.80 \\ Y & = & 8.10 & 7.58 & 8.06 & 8.43 & 8.63 & 8.69 \\ & & 7.61 & 8.39 & 10.57 & 9.11 & 8.60 & 8.62 \end{array}$$

Stabilire con un test bilaterale di livello $\alpha = 0.05$ se la media μ_X è significativamente diversa dalla media μ_Y .

Esercizio A.3.15. Siano dati due campioni gaussiani indipendenti X_1, \dots, X_n e Y_1, \dots, Y_m , composti rispettivamente di $n = 21$ e $m = 31$ valori. Supponiamo che le varianze empiriche calcolate a partire dai dati siano rispettivamente

$$S_X^2 = 2, \quad S_Y^2 = 1;$$

- determinare gli intervalli di fiducia di livello $\alpha = 0.05$ per le varianze σ_X^2 e σ_Y^2 ;
- eseguire un test bilaterale di Fisher di livello $\alpha = 0.05$ per le due ipotesi

$$H_0 = \sigma_X^2 = \sigma_Y^2 \quad H_1 = \sigma_X^2 \neq \sigma_Y^2$$

Esercizio A.3.16. Due serie indipendenti di $n = 10$ ed $m = 9$ misure rispettivamente di due quantità aleatorie X ed Y danno i seguenti risultati

$$\begin{array}{cccccccccc} X = & -1.34 & 1.32 & -0.96 & 0.29 & -1.41 & 0.23 & -0.56 & -0.32 & 0.66 & 1.27 \\ Y = & 4.83 & -2.52 & -1.79 & -2.85 & 1.45 & 1.09 & 1.87 & 2.03 & -2.60 & \end{array}$$

- Calcolare le due medie \bar{X}, \bar{Y} , e le due varianze empiriche S_X^2, S_Y^2 .
- Determinare gli intervalli di fiducia di livello $\alpha = 0.05$ per le due varianze sconosciute σ_X^2 e σ_Y^2 .
- Eseguire un test di Fisher unilaterale di livello $\alpha = 0.05$ per verificare l'ipotesi $H_0 : \sigma_Y^2 \leq \sigma_X^2$ contro l'ipotesi $H_1 : \sigma_Y^2 > \sigma_X^2$.

Esercizio A.3.17. Siano dati i seguenti due campioni indipendenti:

$$X = \begin{array}{cccccc} 3.16 & 4.37 & 6.67 & 2.49 & 2.06 \\ 4.10 & -0.88 & 3.84 & 1.45 & \end{array}$$

$$Y = \begin{array}{ccccc} 3.51 & 6.94 & 5.46 & 5.27 & 7.48 \\ 5.76 & 2.21 & 7.29 & 7.80 & 6.06 \end{array}$$

- Si calcolino le medie \bar{X}, \bar{Y} e le varianze S_X^2, S_Y^2 dei due campioni.
- Si determini l'intervallo di fiducia di livello $\alpha = 0.05$ per la varianza del campione delle X .
- Con un test unilaterale destro di livello $\alpha = 0.05$ si controlli l'ipotesi $\sigma_X^2 \leq \sigma_Y^2$, contro l'ipotesi $\sigma_X^2 > \sigma_Y^2$.
- Con un test bilaterale di livello $\alpha = 0.05$ si controlli l'ipotesi $H_0 : \sigma_X^2 = \sigma_Y^2$ contro l'ipotesi $H_1 : \sigma_X^2 \neq \sigma_Y^2$.

Esercizio A.3.18. Per studiare l'attività di un centralino telefonico si rileva il numero X di telefonate che arrivano tra le 11.00 e le 12.00 in $n = 100$ giorni lavorativi tipici. Le frequenze N_j con cui si ritrovano i diversi valori $j = 0, 1, \dots$ di X sono le seguenti:

$$\begin{array}{cccccccccc} j = & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \geq 9 \\ N_j = & 1 & 4 & 15 & 18 & 22 & 17 & 10 & 8 & 3 & 2 \end{array}$$

Verificare con un test del χ^2 di livello $\alpha = 0,05$ se questi dati sono compatibili con l'ipotesi H_0 secondo la quale la v.a. X segue una distribuzione di Poisson con parametro $\lambda = 4$.

Esercizio A.3.19. Si misura 500 volte una v.a. X discreta che assume i cinque valori $j = 0, 1, 2, 3, 4$, e si ottengono le seguenti frequenze dei risultati:

$$\begin{array}{cccccc} j = & 0 & 1 & 2 & 3 & 4 \\ N_j = & 4 & 51 & 163 & 173 & 109 \end{array}$$

Controllare se questi dati sono compatibili con l'ipotesi che X sia Binomiale $B(4, \frac{2}{3})$ usando un test del χ^2 di livello $\alpha = 0.05$.

Esercizio A.3.20. Si ripete per $n = 200$ volte il lancio di 5 monetine, si conta in ogni ripetizione il numero di teste (con valori $j = 0, 1, 2, 3, 4, 5$), e si ottengono le seguenti frequenze

$$\begin{array}{cccccc} j = & 0 & 1 & 2 & 3 & 4 & 5 \\ N_j = & 8 & 37 & 62 & 56 & 34 & 3 \end{array}$$

Eeguire un test del χ^2 di livello $\alpha = 0.05$ per verificare se questi dati sono compatibili con l'ipotesi H_0 che il numero di teste si distribuisce in maniera Binomiale con $p = \frac{1}{2}$, cioè secondo $B(5, \frac{1}{2})$.

Esercizio A.3.21. $n = 200$ misure di una v.a. discreta X che prende valori $j = 0, \dots, 4$ presentano le seguenti frequenze

$$\begin{array}{cccccc} j = & 0 & 1 & 2 & 3 & 4 \\ N_j = & 30 & 72 & 70 & 24 & 4 \end{array}$$

Stabilire con un test del χ^2 di livello $\alpha = 0.05$ se queste misure sono compatibili con l'ipotesi che la nostra v.a. sia Binomiale $B(4, 0.4)$.

Esercizio A.3.22. Si lanciano 5 dadi $n = 2000$ volte, e si conta il numero X dei sei (X ha valori $j = 0, 1, 2, 3, 4, 5$) ottenendo le seguenti frequenze

$$\begin{array}{cccccc} j = & 0 & 1 & 2 & 3 & 4 & 5 \\ N_j = & 822 & 804 & 300 & 67 & 7 & 0 \end{array}$$

Eeguire un test del χ^2 di livello $\alpha = 0.05$ per verificare se questi dati sono compatibili con l'ipotesi H_0 che X si distribuisce in maniera Binomiale con $p = \frac{1}{6}$, cioè secondo la legge $B(5, \frac{1}{6})$.

Esercizio A.3.23. Un apparecchio contiene 4 componenti elettronici identici. Dopo aver funzionato per un certo tempo ogni componente ha una probabilità p (non nota) di essere ancora funzionante. Un campione di $n = 50000$ apparecchi viene

esaminato, e per ciascuno di essi si conta il numero ($j = 0, 1, 2, 3, 4$) di componenti ancora funzionanti ottenendo i seguenti risultati

$j =$	0	1	2	3	4
$N_j =$	6	201	2400	14644	32749

Calcolare la stima \bar{p} (di MV) per il parametro p , e stabilire con un test del χ^2 di Pearson di livello $\alpha = 0.05$ se i risultati sono compatibili con l'ipotesi H_0 secondo la quale la v.a. "numero di componenti ancora funzionanti" è Binomiale $B(4, \bar{p})$.

Esercizio A.3.24. $n = 40$ misure di una quantità aleatoria X danno i seguenti risultati:

0.44	0.37	0.87	1.73	-0.41	2.84	1.40	-0.17	0.29	1.59
0.39	2.39	1.68	-0.05	1.01	1.17	0.62	2.83	0.73	0.91
0.31	-0.92	2.28	0.74	1.02	0.70	2.06	2.56	0.94	2.56
-0.34	1.40	1.42	-0.09	2.17	1.83	1.80	-0.14	1.40	0.91

Usando un test del χ^2 di livello $\alpha = 0,05$, e con i seguenti $k = 4$ eventi

$$X \leq 0, \quad 0 < X \leq 1, \quad 1 < X \leq 2, \quad 2 < X,$$

stabilire con quale delle due seguenti ipotesi questi dati sono compatibili:

- X è normale $\mathcal{N}(1, 1)$;
- X è uniforme nell'intervallo $[-1, 3]$.

Esercizio A.3.25. $n = 20$ misure di una quantità aleatoria X danno i seguenti risultati:

-3.601	1.064	3.370	1.535	1.017
1.933	3.100	-0.569	1.141	1.815
2.267	0.195	-0.506	-0.167	-2.936
-0.211	-0.659	-0.375	0.024	2.765

Controllare se questi dati sono compatibili con l'ipotesi che X sia $\mathcal{N}(1, 4)$ ($\sigma = 2$) usando un test del χ^2 di livello $\alpha = 0.05$, e con i seguenti $k = 3$ eventi

$$X \leq 0, \quad 0 < X \leq 2, \quad 2 < X.$$

Supponendo poi che X sia $\mathcal{N}(\mu, 4)$, determinare un intervallo di fiducia di livello $\alpha = 0.05$ per μ .

Esercizio A.3.26. Le misure della larghezza del cranio (in mm) effettuate su un campione di $n = 84$ scheletri etruschi hanno una media $\bar{x} = 143.8$ e una deviazione standard $s = 6.0$. La tabella delle frequenze assolute negli intervalli indicati è

<i>larghezza (mm)</i>	<i>frequenza</i>
< 135	5
135–140	10
140–145	33
145–150	24
> 150	12

Eeguire un test del χ^2 di livello $\alpha = 0.05$ per verificare l'ipotesi che il campione provenga da una distribuzione normale con la media e la varianza stimate.

Esercizio A.3.27. Si misurano la lunghezza X e il peso Y di $n = 600$ pezzi prodotti da una fabbrica per controllare se sono: *troppo lunghi*, *giusti*, o *troppo corti* in X ; *troppo pesanti*, *giusti*, *troppo leggeri* in Y . I risultati della verifica sono riassunti nella seguente tabella:

	<i>corti</i>	<i>giusti</i>	<i>lunghi</i>	TOT
<i>leggeri</i>	6	48	8	
<i>giusti</i>	52	402	36	
<i>pesanti</i>	6	38	4	
TOT				

Stabilire con un test del χ^2 di livello $\alpha = 0.05$ se queste deviazioni di X e Y dai valori ammissibili sono indipendenti.

Esercizio A.3.28. Si sceglie un campione di individui che usano quotidianamente un automezzo privato per raggiungere il posto di lavoro: ogni soggetto è classificato in base alla potenza della propria auto e alla distanza in Km che percorre ogni giorno ottenendo i dati della seguente tabella:

	0–10 Km	10–20 Km	> 20 Km	TOT
<i>molto potente</i>	6	27	19	
<i>potente</i>	8	36	17	
<i>normale</i>	21	45	33	
<i>piccola</i>	14	18	6	
TOT				

Stabilire con un test del χ^2 di livello $\alpha = 0.05$ se questi dati mostrano una dipendenza tra potenza dell'auto e distanza percorsa.

Esercizio A.3.29. Viene effettuata un'indagine per sapere quale mezzo di comunicazione è considerato più affidabile: ad ogni individuo viene richiesta età, sesso, livello di studio e mezzo di comunicazione ritenuto più affidabile. I risultati sono riassunti nelle seguenti tre tabelle

	<i>giornale</i>	<i>televisione</i>	<i>radio</i>	TOT
<i>< 35 anni</i>	30	68	10	
<i>35–54 anni</i>	61	78	21	
<i>> 54 anni</i>	98	43	21	
TOT				

	<i>giornale</i>	<i>televisione</i>	<i>radio</i>	TOT
<i>maschio</i>	92	108	20	
<i>femmina</i>	97	81	32	
TOT				

	<i>giornale</i>	<i>televisione</i>	<i>radio</i>	TOT
<i>media inferiore</i>	45	22	6	
<i>media superiore</i>	95	115	33	
<i>università</i>	49	52	13	
TOT				

Stabilire con un test del χ^2 di livello $\alpha = 0.05$ se questi dati mostrano una dipendenza tra giudizio sull'affidabilità dei mezzi di comunicazione e, rispettivamente, età, sesso e livello di studio.

Esercizio A.3.30. Un'azienda vuole verificare l'affidabilità di tre diverse configurazioni (A , B e C) di una macchina industriale esaminando i guasti a cui essa è soggetta: sapendo che ci sono quattro possibili tipi di guasti (1, 2, 3 e 4) e che i dati sono quelli della seguente tabella

	1	2	3	4	TOT
A	20	44	17	9	
B	4	17	7	12	
C	10	31	14	5	
TOT					

stabilire con un test del χ^2 di livello $\alpha = 0.05$ se questi dati mostrano una dipendenza tra configurazioni e tipi di guasti.

Appendice B

Schemi

B.1 Formulario di Statistica Inferenziale

STIMA E TEST PER L'ATTESA DI UNA V.A. X

Sia X_1, \dots, X_n una campione di una v.a. X con attesa μ e varianza σ^2 . Adotteremo le notazioni:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2;$$

φ_α e $t_\alpha(n)$ saranno i quantili di ordine α rispettivamente delle leggi $\mathcal{N}(0, 1)$ e $t(n)$; infine poniamo:

$$U_0 = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}, \quad T_0 = \frac{\bar{X} - \mu_0}{S} \sqrt{n}$$

Intervalli di fiducia di livello α per μ :

1. varianza σ^2 di X nota: $\bar{X} \pm \varphi_{1-\frac{\alpha}{2}} \sigma / \sqrt{n}$
2. varianza σ^2 di X non nota, e $n \geq 120$: $\bar{X} \pm \varphi_{1-\frac{\alpha}{2}} S / \sqrt{n}$
3. varianza σ^2 di X non nota, e $n < 120$: $\bar{X} \pm t_{1-\frac{\alpha}{2}}(n-1) S / \sqrt{n}$

Test bilaterale per le seguenti ipotesi sui valori dell'attesa μ :

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

Regioni critiche di livello α :

1. σ^2 nota: $\{|U_0| > \varphi_{1-\frac{\alpha}{2}}\}$
2. σ^2 non nota, e $n \geq 120$: $\{|T_0| > \varphi_{1-\frac{\alpha}{2}}\}$
3. σ^2 non nota, e $n < 120$: $\{|T_0| > t_{1-\frac{\alpha}{2}}(n-1)\}$

Test unilaterali per le seguenti ipotesi sui valori dell'attesa μ :

$$\begin{aligned} H_0 : \mu \leq \mu_0, & \quad H_1 : \mu > \mu_0 \\ H_0 : \mu \geq \mu_0, & \quad H_1 : \mu < \mu_0 \end{aligned}$$

Regioni critiche di livello α : rispettivamente

1. σ^2 nota: $\{U_0 > \varphi_{1-\alpha}\}$ e $\{U_0 < -\varphi_{1-\alpha}\}$
2. σ^2 non nota, e $n \geq 120$: $\{T_0 > \varphi_{1-\alpha}\}$ e $\{T_0 < -\varphi_{1-\alpha}\}$
3. σ^2 non nota, e $n < 120$: $\{T_0 > t_{1-\alpha}(n-1)\}$ e $\{T_0 < -t_{1-\alpha}(n-1)\}$

CONFRONTO FRA LE ATTESE DI DUE V.A. X E Y

Ipotesi per i **test bilaterale e unilaterale** sull'eguaglianza delle attese:

$$\begin{aligned} H_0 : \mu_X = \mu_Y, & & H_1 : \mu_X \neq \mu_Y \\ H_0 : \mu_X \leq \mu_Y, & & H_1 : \mu_X > \mu_Y \end{aligned}$$

Campioni accoppiati X_1, \dots, X_n e Y_1, \dots, Y_n ; σ^2 varianza di $Z = X - Y$

$$\begin{aligned} Z_i = X_i - Y_i, & & \bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i, & & S^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2 \\ U_0 = \frac{\bar{Z}}{\sigma} \sqrt{n}, & & T_0 = \frac{\bar{Z}}{S} \sqrt{n} \end{aligned}$$

Se σ^2 è nota: regioni critiche bilaterale e unilaterale di livello α rispettivamente

$$\{|U_0| > \varphi_{1-\frac{\alpha}{2}}\}, \quad \{U_0 > \varphi_{1-\alpha}\}$$

Se σ^2 non è nota: regioni critiche bilaterale e unilaterale di livello α rispettivamente

$$\{|T_0| > t_{1-\frac{\alpha}{2}}(n-1)\}, \quad \{T_0 > t_{1-\alpha}(n-1)\}$$

Campioni indipendenti X_1, \dots, X_n e Y_1, \dots, Y_m di X e Y con attese μ_X e μ_Y e varianze σ_X^2 e σ_Y^2 . Notazioni:

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i, & \bar{Y} &= \frac{1}{m} \sum_{i=1}^m Y_i \\ V^2 &= \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} \\ U_0 &= \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} & T_0 &= \frac{\bar{X} - \bar{Y}}{V \sqrt{\frac{1}{n} + \frac{1}{m}}} \end{aligned}$$

Se σ_X^2 e σ_Y^2 sono note: regioni critiche bilaterale e unilaterale di livello α rispettivamente

$$\{|U_0| > \varphi_{1-\frac{\alpha}{2}}\} \quad \{U_0 > \varphi_{1-\alpha}\}$$

Se σ_X^2 e σ_Y^2 non sono note: regioni critiche bilaterale e unilaterale di livello α rispettivamente

$$\{|T_0| > t_{1-\frac{\alpha}{2}}(n+m-2)\} \quad \{T_0 > t_{1-\alpha}(n+m-2)\}$$

STIME E TEST PER LE VARIANZE

X_1, \dots, X_n campione di X con attesa μ e varianza σ^2 sconosciute. Notazioni:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

È necessario supporre che $X \sim \mathcal{N}(\mu, \sigma^2)$, o che n sia abbastanza grande da poter usare il TLC. Infine $\chi_\alpha^2(n)$ sono i quantili di ordine α della legge $\chi^2(n)$.

Intervallo di fiducia di livello α per σ^2 :

$$\left[\frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)} \right]$$

Per $n \geq 35$ i quantili $\chi_\alpha^2(n)$ non tabulati si possono calcolare dai quantili normali:

$$\chi_\alpha^2(n) \approx \frac{1}{2} (\varphi_\alpha + \sqrt{2n-1})^2$$

Confronto fra le varianze di due campioni indipendenti X_1, \dots, X_n e Y_1, \dots, Y_m .
Notazioni:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{Y} = \frac{1}{m} \sum_{j=1}^m Y_j$$

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S_Y^2 = \frac{1}{m-1} \sum_{j=1}^m (Y_j - \bar{Y})^2, \quad F_0 = \frac{S_X^2}{S_Y^2}$$

$f_\alpha(n, m)$ quantili di ordine α della legge di Fisher $F(n, m)$; per l'uso delle tavole ricordare che

$$f_\alpha(n, m) = \frac{1}{f_{1-\alpha}(m, n)}$$

Test bilaterale (di Fisher) sull'eguaglianza delle varianze:

$$H_0 : \sigma_X^2 = \sigma_Y^2, \quad H_1 : \sigma_X^2 \neq \sigma_Y^2$$

Regione critica di livello α :

$$\{F_0 < f_{\frac{\alpha}{2}}(n-1, m-1)\} \cup \{F_0 > f_{1-\frac{\alpha}{2}}(n-1, m-1)\}$$

Test unilaterale (di Fisher) sull'eguaglianza delle varianze:

$$H_0 : \sigma_X^2 \leq \sigma_Y^2, \quad H_1 : \sigma_X^2 > \sigma_Y^2$$

Regione critica di livello α :

$$\{F_0 > f_{1-\alpha}(n-1, m-1)\}$$

TEST DI ADATTAMENTO E DI INDIPENDENZA

Test del χ^2 per il confronto fra frequenze empiriche e distribuzioni teoriche: v.a. con k valori, campione di numerosità n , frequenze assolute N_1, \dots, N_k , distribuzione teorica p_1, \dots, p_k . Notazione:

$$\bar{p}_j = \frac{N_j}{n}, \quad D_0 = n \sum_{j=1}^k \frac{(\bar{p}_j - p_j)^2}{p_j} = \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j}$$

Ipotesi: n abbastanza grande da avere $np_j \geq 5$ per $j = 1, \dots, k$; $\chi_\alpha^2(n)$ sono i quantili di ordine α della legge $\chi^2(n)$.

Regione critica di livello α :

$$\{D_0 > \chi_{1-\alpha}^2(k-1)\}$$

Se la conoscenza della distribuzione teorica richiede la stima (di MV) di $q < k - 1$ parametri la regione critica diviene

$$\{D_0 > \chi_{1-\alpha}^2(k-1-q)\}$$

Test del chi quadro per l'indipendenza: X prende valori u_1, \dots, u_r , Y prende valori v_1, \dots, v_s . Valori del campione accoppiato (u_j, v_k) con $j = 1, \dots, r$ e $k = 1, \dots, s$. Notazioni

N_{jk} frequenza congiunta assoluta della classe (u_j, v_k)

$N_{j\cdot}$ frequenza marginale assoluta della classe u_j

$N_{\cdot,k}$ frequenza marginale assoluta della classe v_k

$$\bar{p}_j = \frac{N_{j\cdot}}{n}, \quad \bar{q}_k = \frac{N_{\cdot,k}}{n}, \quad D_0 = \sum_{j,k} \frac{(N_{jk} - n\bar{p}_j\bar{q}_k)^2}{n\bar{p}_j\bar{q}_k}$$

Ricordare anche che

$$n\bar{p}_j\bar{q}_k = \frac{N_{j\cdot} N_{\cdot,k}}{n}$$

Regione critica di livello α :

$$\{D_0 > \chi_{1-\alpha}^2[(r-1)(s-1)]\}$$

Appendice C

Notazioni

C.1 Notazioni vettoriali

In uno spazio reale p -dimensionale \mathbf{R}^p i punti $\mathbf{x} = (x_1, \dots, x_p)$ possono essere considerati come **vettori** applicati nell'origine dello spazio e con componenti x_k . Per questo motivo parleremo indifferentemente di *vettori* e di *punti* in \mathbf{R}^p . Chiameremo **vettore nullo** il vettore $\mathbf{0}$ le cui componenti sono tutte uguali a 0: ovviamente la sua rappresentazione coincide con l'origine di \mathbf{R}^p . I concetti geometrici che si introducono non sono altro che la naturale generalizzazione dei concetti usati nel ben noto caso $p = 3$ dello spazio tridimensionale naturale.

Chiameremo **modulo** del vettore \mathbf{x} il numero positivo

$$|\mathbf{x}| = \sqrt{\sum_{k=1}^p x_k^2}.$$

Si introducono poi tre operazioni:

- la somma $\mathbf{z} = \mathbf{x} + \mathbf{y}$ di due vettori è il vettore le cui componenti $z_k = x_k + y_k$ sono le somme delle corrispondenti componenti dei due addendi;
- il prodotto $\mathbf{z} = a\mathbf{x}$ di un numero reale a per un vettore è il vettore le cui componenti $z_k = ax_k$ sono il prodotto per a delle corrispondenti componenti di \mathbf{x} ;
- il **prodotto scalare** $\mathbf{x} \cdot \mathbf{y}$ di due vettori è il numero reale dato da

$$\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^p x_k y_k,$$

e si può anche definire l'angolo θ fra i vettori \mathbf{x} e \mathbf{y} in modo che $\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}||\mathbf{y}| \cos \theta$.

È facile vedere che il prodotto scalare è simmetrico, cioè $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$, e gode della seguente proprietà: dati due numeri reali a e b e tre vettori $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{R}^p$ si ha

$$\mathbf{z} \cdot (a\mathbf{x} + b\mathbf{y}) = a\mathbf{z} \cdot \mathbf{x} + b\mathbf{z} \cdot \mathbf{y}. \quad (\text{C.1})$$

Naturalmente si ha anche dalle definizioni che $|\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$. Si dice che due vettori \mathbf{x} e \mathbf{y} sono **ortogonali** se $\mathbf{x} \cdot \mathbf{y} = 0$. Un ruolo importante è giocato dai vettori \mathbf{v} di modulo $|\mathbf{v}| = 1$ detti anche **versori**: in particolare i punti della forma $a\mathbf{v}$ (cioè i vettori di direzione \mathbf{v} e modulo $|a|$) descrivono, al variare di a , una retta passante per l'origine; anzi si può provare che i punti di ogni retta passante per l'origine hanno questa forma. La **proiezione** di un punto \mathbf{x} sulla retta individuata da \mathbf{v} si costruisce determinando il punto di questa retta che si trova alla minore distanza dal vertice del vettore \mathbf{x} . Possiamo anche dire che la proiezione è quindi un punto del tipo $a\mathbf{v}$,

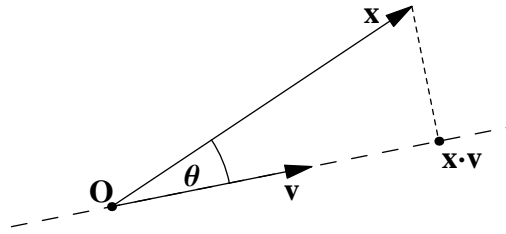


Figura C.1: Proiezione di \mathbf{x} lungo la direzione di \mathbf{v} : il modulo del vettore proiettato (disegnato solo come un punto per non appesantire il grafico) è il prodotto scalare $\mathbf{x} \cdot \mathbf{v}$.

e che quindi, per un dato \mathbf{v} , tutto quel che bisogna calcolare è il modulo a della proiezione. Si dimostra che il modulo del vettore proiettato è $a = \mathbf{x} \cdot \mathbf{v}$, e pertanto che la proiezione di un punto \mathbf{x} su una retta \mathbf{v} si ottiene eseguendo semplicemente il prodotto scalare $\mathbf{x} \cdot \mathbf{v}$: il significato geometrico di questa operazione è illustrato nella Figura C.1.

Nello spazio \mathbf{R}^p le **matrici** $p \times p$ indicate con la notazione $\mathbb{A} = \|a_{k\ell}\|$ permettono di definire delle trasformazioni dei vettori \mathbf{x} mediante il prodotto *righe per colonne*: più precisamente con la notazione $\mathbb{A}\mathbf{x}$ indicheremo il vettore con componenti

$$(\mathbb{A}\mathbf{x})_k = \sum_{\ell=1}^p a_{k\ell}x_\ell.$$

Chiameremo **matrice trasposta** la matrice $\mathbb{A}^T = \|a_{\ell k}\|$ che si ottiene da $\mathbb{A} = \|a_{k\ell}\|$ scambiando le righe con le colonne, e diremo che \mathbb{A} è una **matrice simmetrica** quando coincide con la sua trasposta, cioè se $a_{k\ell} = a_{\ell k}$. Data una matrice \mathbb{A} si consideri ora l'equazione

$$\mathbb{A}\mathbf{x} = \lambda\mathbf{x}; \tag{C.2}$$

chiameremo **autovalore** di \mathbb{A} ogni numero λ tale che esista un vettore $\mathbf{x} \neq \mathbf{0}$ soluzione di (C.2); in tal caso \mathbf{x} si chiama **autovettore** di \mathbb{A} associato all'autovalore λ . Se \mathbf{x} è autovettore di \mathbb{A} associato a λ , si prova facilmente che anche tutti i vettori $a\mathbf{x}$, con a numero reale arbitrario, sono autovettori associati allo stesso autovalore λ . In generale una matrice \mathbb{A} in \mathbf{R}^p ha p autovalori, anche non tutti distinti, e si dimostra che quando \mathbb{A} è simmetrica tali autovalori sono reali. Converrà in questo caso riordinare gli autovalori in ordine decrescente

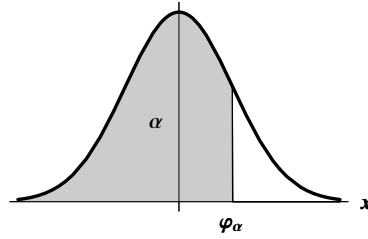
$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

indicando con $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ i corrispondenti autovettori che potranno sempre essere scelti in modo da essere **ortonormali**, cioè ortogonali fra loro e tutti di modulo 1. Avremo quindi per gli autovettori la relazione

$$\mathbf{v}_k \cdot \mathbf{v}_\ell = \begin{cases} 0 & \text{se } k \neq \ell, \\ |\mathbf{v}_k|^2 = 1 & \text{se } k = \ell. \end{cases}$$

Appendice D

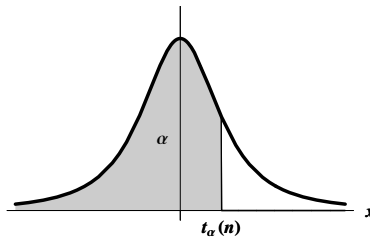
Tavole Numeriche



D.1 Legge Normale standard $\mathcal{N}(0, 1)$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861

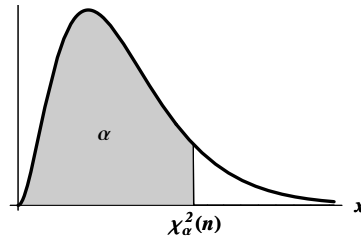
$$-\varphi_\alpha = \varphi_{1-\alpha}$$



D.2 Legge di Student $t(n)$

n	0.950	0.975	0.990	0.995	n	0.950	0.975	0.990	0.995
1	6.31375	12.70620	31.82050	63.65670	31	1.69552	2.03951	2.45282	2.74404
2	2.91999	4.30265	6.96456	9.92484	32	1.69389	2.03693	2.44868	2.73848
3	2.35336	3.18245	4.54070	5.84091	33	1.69236	2.03452	2.44479	2.73328
4	2.13185	2.77645	3.74695	4.60409	34	1.69092	2.03224	2.44115	2.72839
5	2.01505	2.57058	3.36493	4.03214	35	1.68957	2.03011	2.43772	2.72381
6	1.94318	2.44691	3.14267	3.70743	36	1.68830	2.02809	2.43449	2.71948
7	1.89458	2.36462	2.99795	3.49948	37	1.68709	2.02619	2.43145	2.71541
8	1.85955	2.30600	2.89646	3.35539	38	1.68595	2.02439	2.42857	2.71156
9	1.83311	2.26216	2.82144	3.24984	39	1.68488	2.02269	2.42584	2.70791
10	1.81246	2.22814	2.76377	3.16927	40	1.68385	2.02108	2.42326	2.70446
11	1.79588	2.20099	2.71808	3.10581	41	1.68288	2.01954	2.42080	2.70118
12	1.78229	2.17881	2.68100	3.05454	42	1.68195	2.01808	2.41847	2.69807
13	1.77093	2.16037	2.65031	3.01228	43	1.68107	2.01669	2.41625	2.69510
14	1.76131	2.14479	2.62449	2.97684	44	1.68023	2.01537	2.41413	2.69228
15	1.75305	2.13145	2.60248	2.94671	45	1.67943	2.01410	2.41212	2.68959
16	1.74588	2.11991	2.58349	2.92078	46	1.67866	2.01290	2.41019	2.68701
17	1.73961	2.10982	2.56693	2.89823	47	1.67793	2.01174	2.40835	2.68456
18	1.73406	2.10092	2.55238	2.87844	48	1.67722	2.01063	2.40658	2.68220
19	1.72913	2.09302	2.53948	2.86093	49	1.67655	2.00958	2.40489	2.67995
20	1.72472	2.08596	2.52798	2.84534	50	1.67591	2.00856	2.40327	2.67779
21	1.72074	2.07961	2.51765	2.83136	55	1.67303	2.00404	2.39608	2.66822
22	1.71714	2.07387	2.50832	2.81876	60	1.67065	2.00030	2.39012	2.66028
23	1.71387	2.06866	2.49987	2.80734	65	1.66864	1.99714	2.38510	2.65360
24	1.71088	2.06390	2.49216	2.79694	70	1.66691	1.99444	2.38081	2.64790
25	1.70814	2.05954	2.48511	2.78744	75	1.66543	1.99210	2.37710	2.64298
26	1.70562	2.05553	2.47863	2.77871	80	1.66412	1.99006	2.37387	2.63869
27	1.70329	2.05183	2.47266	2.77068	90	1.66196	1.98667	2.36850	2.63157
28	1.70113	2.04841	2.46714	2.76326	100	1.66023	1.98397	2.36422	2.62589
29	1.69913	2.04523	2.46202	2.75639	110	1.65882	1.98177	2.36073	2.62126
30	1.69726	2.04227	2.45726	2.75000	120	1.65765	1.97993	2.35782	2.61742

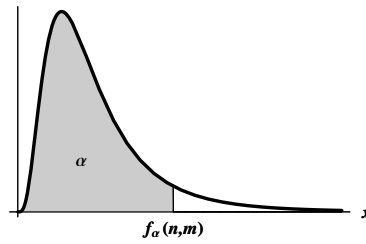
$$-t_{\alpha}(n) = t_{1-\alpha}(n)$$



D.3 Legge del Chi Quadro $\chi^2(n)$

n	0.005	0.010	0.025	0.050	0.950	0.975	0.990	0.995
1	0.00004	0.00016	0.00098	0.00393	3.84146	5.02389	6.63490	7.87944
2	0.01003	0.02010	0.05064	0.10259	5.99146	7.37776	9.21034	10.59663
3	0.07172	0.11483	0.21580	0.35185	7.81473	9.34840	11.34487	12.83816
4	0.20699	0.29711	0.48442	0.71072	9.48773	11.14329	13.27670	14.86026
5	0.41174	0.55430	0.83121	1.14548	11.07050	12.83250	15.08627	16.74960
6	0.67573	0.87209	1.23734	1.63538	12.59159	14.44938	16.81189	18.54758
7	0.98926	1.23904	1.68987	2.16735	14.06714	16.01276	18.47531	20.27774
8	1.34441	1.64650	2.17973	2.73264	15.50731	17.53455	20.09024	21.95495
9	1.73493	2.08790	2.70039	3.32511	16.91898	19.02277	21.66599	23.58935
10	2.15586	2.55821	3.24697	3.94030	18.30704	20.48318	23.20925	25.18818
11	2.60322	3.05348	3.81575	4.57481	19.67514	21.92005	24.72497	26.75685
12	3.07382	3.57057	4.40379	5.22603	21.02607	23.33666	26.21697	28.29952
13	3.56503	4.10692	5.00875	5.89186	22.36203	24.73560	27.68825	29.81947
14	4.07467	4.66043	5.62873	6.57063	23.68479	26.11895	29.14124	31.31935
15	4.60092	5.22935	6.26214	7.26094	24.99579	27.48839	30.57791	32.80132
16	5.14221	5.81221	6.90766	7.96165	26.29623	28.84535	31.99993	34.26719
17	5.69722	6.40776	7.56419	8.67176	27.58711	30.19101	33.40866	35.71847
18	6.26480	7.01491	8.23075	9.39046	28.86930	31.52638	34.80531	37.15645
19	6.84397	7.63273	8.90652	10.11701	30.14353	32.85233	36.19087	38.58226
20	7.43384	8.26040	9.59078	10.85081	31.41043	34.16961	37.56623	39.99685
21	8.03365	8.97200	10.28290	11.59131	32.67057	35.47888	38.93217	41.40106
22	8.64272	9.54249	10.98232	12.33801	33.92444	36.78071	40.28936	42.79565
23	9.26042	10.19572	11.68855	13.09051	35.17246	38.07563	41.63840	44.18128
24	9.88623	10.85636	12.40115	13.84843	36.41503	39.36408	42.97982	45.55851
25	10.51965	11.52398	13.11972	14.61141	37.65248	40.64647	44.31410	46.92789
26	11.16024	12.19815	13.84390	15.37916	38.88514	41.92317	45.64168	48.28988
27	11.80759	12.87850	14.57338	16.15140	40.11327	43.19451	46.96294	49.64492
28	12.46134	13.56471	15.30786	16.92788	41.33714	44.46079	48.27824	50.99338
29	13.12115	14.25645	16.04707	17.70837	42.55697	45.72229	49.58788	52.33562
30	13.78672	14.95346	16.79077	18.49266	43.77297	46.97924	50.89218	53.67196
31	14.45777	15.65546	17.53874	19.28057	44.98534	48.23189	52.19139	55.00270
32	15.13403	16.36222	18.29076	20.07191	46.19426	49.48044	53.48577	56.32811
33	15.81527	17.07351	19.04666	20.86653	47.39988	50.72508	54.77554	57.64845
34	16.50127	17.78915	19.80625	21.66428	48.60237	51.96600	56.06091	58.96393
35	17.19182	18.50893	20.56938	22.46502	49.80185	53.20335	57.34207	60.27477

$$\chi^2_\alpha(n) \simeq \frac{1}{2}(\varphi_\alpha + \sqrt{2n-1})^2, \quad n > 35$$



D.4 Legge di Fisher $F(n, m)$

$$\alpha = 0.950$$

	n	1	2	3	4	5	6	7	8	9	10	15	20	30	60	∞
m																
3		10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.62	8.57	8.53
4		7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.75	5.69	5.63
5		6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.50	4.43	4.37
6		5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.81	3.74	3.67
7		5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.38	3.30	3.23
8		5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.08	3.01	2.93
9		5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.86	2.79	2.71
10		4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.70	2.62	2.54
11		4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.57	2.49	2.40
12		4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.47	2.38	2.30
13		4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.38	2.30	2.21
14		4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.31	2.22	2.13
15		4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.25	2.16	2.07
16		4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.19	2.11	2.01
17		4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.15	2.06	1.96
18		4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.11	2.02	1.92
19		4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.07	1.98	1.88
20		4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.04	1.95	1.84
21		4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.18	2.10	2.01	1.92	1.81
22		4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	1.98	1.89	1.78
23		4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.13	2.05	1.96	1.86	1.76
24		4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	2.03	1.94	1.84	1.73
25		4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.09	2.01	1.92	1.82	1.71
26		4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.99	1.90	1.80	1.69
27		4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.06	1.97	1.88	1.79	1.67
28		4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.96	1.87	1.77	1.65
29		4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.03	1.94	1.85	1.75	1.64
30		4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.84	1.74	1.62
31		4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15	2.00	1.92	1.83	1.73	1.61
32		4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	1.99	1.91	1.82	1.71	1.59
33		4.14	3.28	2.89	2.66	2.50	2.39	2.30	2.23	2.18	2.13	1.98	1.90	1.81	1.70	1.58
34		4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	1.97	1.89	1.80	1.69	1.57
35		4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	1.96	1.88	1.79	1.68	1.56
36		4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	1.95	1.87	1.78	1.67	1.55
37		4.11	3.25	2.86	2.63	2.47	2.36	2.27	2.20	2.14	2.10	1.95	1.86	1.77	1.66	1.54
38		4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	1.94	1.85	1.76	1.65	1.53
39		4.09	3.24	2.85	2.61	2.46	2.34	2.26	2.19	2.13	2.08	1.93	1.85	1.75	1.65	1.52
40		4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.74	1.64	1.51
60		4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.65	1.53	1.39
120		3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.75	1.66	1.55	1.43	1.25
∞		3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.67	1.57	1.46	1.32	1.00

$$f_{\alpha}(n, m) = \frac{1}{f_{1-\alpha}(m, n)}$$

$$\alpha = 0.975$$

	n	1	2	3	4	5	6	7	8	9	10	15	20	30	60	∞
m																
4		12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.66	8.56	8.46	8.36	8.26
5		10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.43	6.33	6.23	6.12	6.02
6		8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.27	5.17	5.07	4.96	4.85
7		8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.57	4.47	4.36	4.25	4.14
8		7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.10	4.00	3.89	3.78	3.67
9		7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.77	3.67	3.56	3.45	3.33
10		6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.52	3.42	3.31	3.20	3.08
11		6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.33	3.23	3.12	3.00	2.88
12		6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.18	3.07	2.96	2.85	2.72
13		6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.05	2.95	2.84	2.72	2.60
14		6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	2.95	2.84	2.73	2.61	2.49
15		6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.86	2.76	2.64	2.52	2.40
16		6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.79	2.68	2.57	2.45	2.32
17		6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.72	2.62	2.50	2.38	2.25
18		5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.67	2.56	2.44	2.32	2.19
19		5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.62	2.51	2.39	2.27	2.13
20		5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.57	2.46	2.35	2.22	2.09
21		5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.53	2.42	2.31	2.18	2.04
22		5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.50	2.39	2.27	2.14	2.00
23		5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.47	2.36	2.24	2.11	1.97
24		5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.44	2.33	2.21	2.08	1.94
25		5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.41	2.30	2.18	2.05	1.91
26		5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.39	2.28	2.16	2.03	1.88
27		5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.36	2.25	2.13	2.00	1.85
28		5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.34	2.23	2.11	1.98	1.83
29		5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.32	2.21	2.09	1.96	1.81
30		5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.31	2.20	2.07	1.94	1.79
31		5.55	4.16	3.57	3.23	3.01	2.85	2.73	2.64	2.56	2.50	2.29	2.18	2.06	1.92	1.77
32		5.53	4.15	3.56	3.22	3.00	2.84	2.71	2.62	2.54	2.48	2.28	2.16	2.04	1.91	1.75
33		5.51	4.13	3.54	3.20	2.98	2.82	2.70	2.61	2.53	2.47	2.26	2.15	2.03	1.89	1.73
34		5.50	4.12	3.53	3.19	2.97	2.81	2.69	2.59	2.52	2.45	2.25	2.13	2.01	1.88	1.72
35		5.48	4.11	3.52	3.18	2.96	2.80	2.68	2.58	2.50	2.44	2.23	2.12	2.00	1.86	1.70
36		5.47	4.09	3.50	3.17	2.94	2.78	2.66	2.57	2.49	2.43	2.22	2.11	1.99	1.85	1.69
37		5.46	4.08	3.49	3.16	2.93	2.77	2.65	2.56	2.48	2.42	2.21	2.10	1.97	1.84	1.67
38		5.45	4.07	3.48	3.15	2.92	2.76	2.64	2.55	2.47	2.41	2.20	2.09	1.96	1.82	1.66
39		5.43	4.06	3.47	3.14	2.91	2.75	2.63	2.54	2.46	2.40	2.19	2.08	1.95	1.81	1.65
40		5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.18	2.07	1.94	1.80	1.64
60		5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.06	1.94	1.82	1.67	1.48
120		5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	1.94	1.82	1.69	1.53	1.31
∞		5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.83	1.71	1.57	1.39	1.00

D.5 Valori di $e^{-\lambda}$

λ	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	1.0000	0.9900	0.9802	0.9704	0.9608	0.9512	0.9418	0.9324	0.9231	0.9139
0.1	0.9048	0.8958	0.8869	0.8781	0.8694	0.8607	0.8521	0.8437	0.8353	0.8270
0.2	0.8187	0.8106	0.8025	0.7945	0.7866	0.7788	0.7711	0.7634	0.7558	0.7483
0.3	0.7408	0.7334	0.7261	0.7189	0.7118	0.7047	0.6977	0.6907	0.6839	0.6771
0.4	0.6703	0.6637	0.6570	0.6505	0.6440	0.6376	0.6313	0.6250	0.6188	0.6126
0.5	0.6065	0.6005	0.5945	0.5886	0.5827	0.5769	0.5712	0.5655	0.5599	0.5543
0.6	0.5488	0.5434	0.5379	0.5326	0.5273	0.5220	0.5169	0.5117	0.5066	0.5016
0.7	0.4966	0.4916	0.4868	0.4819	0.4771	0.4724	0.4677	0.4630	0.4584	0.4538
0.8	0.4493	0.4449	0.4404	0.4360	0.4317	0.4274	0.4232	0.4190	0.4148	0.4107
0.9	0.4066	0.4025	0.3985	0.3946	0.3906	0.3867	0.3829	0.3791	0.3753	0.3716
1.0	0.367879									
2.0	0.135335									
3.0	0.049787									
4.0	0.018316									
5.0	0.006738									
6.0	0.002479									
7.0	0.000912									
8.0	0.000335									
9.0	0.000123									
10.0	0.000045									

$$\lambda = [\lambda] + r, \quad [\lambda] = 0, 1, 2, \dots, \quad 0 < r < 1$$

$$e^{-\lambda} = e^{-[\lambda]}e^{-r}$$

Indice analitico

- adattamento, 129
- additività, 44
- algebra, 42
 - generata, 42
- approssimazione normale, 84
- asimmetria, 18, 75
- autovalore, 166
- autovettore, 166
- autovettori ortonormali, 166
- baricentro, 29
- boxplot, 17
- campione, 4
 - casuale, 79, 96
 - gaussiano, 99, 101
 - ordinato, 15
 - standardizzato, 14
- campioni
 - accoppiati, 121
 - indipendenti, 122
- carattere, 3
- classificazione, 34
- cluster, 34
- componente principale, 31
- convergenza
 - di successioni di v.a., 81
- correlazione
 - matrice di, 29
 - coefficiente di, 25, 72
 - negativa, 25, 73
 - positiva, 25, 73
- covarianza, 25, 72
 - matrice di, 29
- curtosi, 18, 75
- dati
 - multidimensionali, 4
 - qualitativi, 3
 - quantitativi, 3
 - raggruppati, 12, 14
- decile, 15
- decomposizione, 42
- deviazione standard, 12, 72
- diagramma a barre, 6
- differenza interquartile, 17
- direzione principale, 31
- dispersione, 30
 - totale, 29
- distribuzione, 52
 - congiunta, 54
 - marginale, 54
- equiprobabilità, 39
- errore, 111
 - di prima specie, 111
 - di seconda specie, 111
- errore quadratico medio, 14, 26
- eventi
 - disgiunti, 41
 - incompatibili, 41
 - indipendenti, 47
- evento, 41
 - elementare, 39, 41
- famiglia di leggi, 55
- fedeltà, 32
- formula
 - della probabilità totale, 46
 - di Bayes, 47
- frequenza
 - assoluta, 4
 - cumulata, 5

- relativa, 5
- frequenze
 - congiunte, 23
 - marginali, 23
- funzione
 - di densità, 59
 - di distribuzione, 53
 - congiunta, 54
 - marginali, 54
 - di v.a., 54
- gradi di libertà, 63
- indicatore, 52
- indice
 - di centralità, 9
 - di dispersione, 9
 - robusto, 16
- intervallo di fiducia, 98
- ipotesi, 109
 - nulla, 111
- istogramma, 6, 91
- legge, 52
 - Bernoulli, 55
 - Binomiale, 56
 - Chi quadro, 63
 - congiunta, 54
 - Fisher, 63
 - Gauss, 61
 - marginale, 54
 - Multinomiale, 68
 - Normale, 61
 - Normale standard, 62
 - Poisson, 57
 - Student, 63
 - Uniforme, 60
- Legge dei Grandi Numeri, 81
- livello, 98, 112
- matrice, 166
 - simmetrica, 166
 - trasposta, 166
- media, 10, 71
 - aritmetica, 80
 - armonica, 20
 - geometrica, 20
 - pesata, 12
 - quadratica, 20
- mediana, 15, 65
- moda, 9, 60
- modalità, 4
- modulo, 165
- momento, 18, 75
 - centrato, 18, 75
- non correlazione, 25
- normalizzazione, 5
- percentile, 15
- piano principale, 31
- popolazione, 3
- potenza, 112
- probabilità, 39, 44
 - a posteriori, 47
 - a priori, 47
 - condizionata, 45
 - congiunta, 45
 - definizione classica, 40
 - spazio di, 44
- prodotto scalare, 165
- proiezione, 165
- quantile, 15, 65
- quartile, 15, 65
- range, 17
- regione critica, 112
- retta di regressione, 26
- risultato, 39
- scarto quadratico, 12
- significatività, 112
- somma di v.a., 54
- spazio
 - degli eventi elementari, 40
 - dei campioni, 40
- statistica, 96
- stima

- di un parametro, 90
- di una distribuzione continua, 90
- di una distribuzione discreta, 90
- di una media, 89
- di una proporzione, 79
- puntuale, 96
- stimatore, 96
 - consistente, 96
 - di massima verosimiglianza, 103
 - non distorto, 96
- tabella
 - di contingenza, 23
 - di frequenza, 6
- Teorema
 - di Poisson, 57
 - Limite Centrale, 83
- valore
 - centrale, 5
 - d'attesa, 71
- variabile aleatoria, 51
 - continua, 59
 - discreta, 54
 - standardizzata, 75
- variabili aleatorie
 - identicamente distribuite, 52
 - indipendenti, 53
 - non correlate, 72
- varianza, 12, 72
 - campionaria, 80
 - combinata, 126
 - corretta, 96
- variazione
 - coefficiente di, 12
- verosimiglianza
 - funzione di, 103
- versore, 165
- vettore, 165
 - nullo, 165
- vettori
 - ortogonali, 165