Approfondimento
Statistical Data Analysis course

A.A. 2023-2024 / Prof. A.Pompili / alexis.pompili@ba.infn.it

Content of this in-depth part: - Significance of an observed signal. Wilks' theorem and Profile Likelihood (ratio). Upper limits.
- p-value and search for a new signal. Statistical significance of a new signal.

SIGNIFICANCE of an observed physical SIGNAL

A simple type of goodness-of-fit to claim a discovery - example - I

A simple type of goodness-of-fit is often carried out to judge …
whether a discrepancy between data and expectation is enough significant to merit a claim for a new discovery:

Let us assume we are in a situation in which we may/might see evidence for a special type of signal event;
- suppose the # of the **signal candidates** are n_s can be treated as a Poisson variable with mean ν_s ;
- in addition to the signal candidates suppose to find also a certain # of **background events** n_b that can be also treated as Poisson variable;
- the total # of candidates found $n = n_s + n_b$ is therefore a Poissonian variable with mean $\nu = \nu_s + \nu_b$ (remember the "reproductive" property of Poisson distribution ?). Thus, the probability to observe n events is:

$$f(n; \nu_s, \nu_b) = \frac{(\nu_s + \nu_b)^n}{n!} e^{-(\nu_s + \nu_b)}$$

Suppose we carried out the experiment and found n_{obs} candidates.
In order to quantify our degree of confidence in the discovery of a new effect/signal (namely $\nu_s \neq 0$) …
… we can compute how likely it is to find n_{obs} candidates or more (namely $n > n_{obs}$) from background fluctuation alone!
In other words, we have to calculate the **p-value** :

$$P(n \geq n_{obs}) = \sum_{n=n_{obs}}^{\infty} f(n; \nu_s = 0, \nu_b) = 1 - \sum_{n=0}^{n_{obs}-1} f(n; \nu_s = 0, \nu_b) = 1 - \sum_{n=0}^{n_{obs}-1} \frac{(\nu_b)^n}{n!} e^{-(\nu_b)}$$

NOTE: this is **NOT** the probability of the (null) hypothesis $\nu_s = 0$!
It's rather the probability - under the assumption $\nu_s = 0$ - of obtaining as many candidates/events as observed or more **!**
Dispite this subtlety in its interpretation the **p-value** is a useful number to consider when deciding if a new effect/signal is found.
Numerical example:
if we expect $\nu_b = 0.5$ and we observe $n_{obs} = 5$ the **p-value** is $= 1 - e^{-(0.5)} \sum_{n=0}^{4} \frac{(0.5)^n}{n!} = 1.7 \cdot 10^{-4} = 0.017\%$

A simple type of goodness-of-fit to claim a discovery - example - II

Further NOTE:

standard deviation of a Poisson variable/observable

If you consider the $n_{obs} \pm \sqrt{n_{obs}}$ as an estimate for $\nu = \nu_s + \nu_b$, or better, after subtracting the background $\nu_b = 0.5$, you consider 4.5 ± 2.2 as an estimate for ν_s , this would be misleading since it's only about 2 standard deviations from 0, thus giving the wrong impression that ν_s is not very incompatible with zero ("wrong" because of the **p-value**)!

This is a problem of misinterpretation.

Indeed here we are interested in the probability that a Poisson variable of mean ν_b will fluctuate upwardto n_{obs} or higher, and not in the probability that a variable with mean n_{obs} will fluctuate downward to ν_b or lower.

Moreover, ν_b has been wrongly assumed without error. It is instead important to quantify the systematic uncertainty in the background when evaluating the significance of a new effect/signal.

To illustrate this, consider that just with $\nu_b = 0.8$, the **p-value** would be $\cong 0.14\%$, namely higher by about an order of magnitude.

Wilks' Theorem - I

When a large # of measurements is available the Wilks' theorem allows to find … an **approximate asymptotic expression for a test statistic based on a *likelihood ratio*** $\lambda(\vec{x}) = \frac{L(\vec{x}|H_1)}{L(\vec{x}|H_0)}$ (namely of the kind inspired by the Nyman-Pearson Lemma).

Let us assume that the two hypotheses H_0 and H_1 can be defined in terms of a set of parameters $\vec{\theta} = (\theta_1, \dots, \theta_m)$ that appear in the definition of of the likelihood function; now…

- the condition that H_1 is trues can be expressed as … $\vec{\theta} \in \Theta_1$
- the condition that H_0 is trues can be expressed as … $\vec{\theta} \in \Theta_0$

Let us assume that $\Theta_0 \subseteq \Theta_1$ or, in other words, that the **hypotheses are nested**.

Following an opposite convention (with H_0 at the numerator) w.r.t. the ratio in Neyman-Pearson Lemma)

Given a data sample of **independent measurements** $(\vec{x}_1, \dots, \vec{x}_N)$ the theorem ensures that, **assuming some regularity conditions of the likelihood function, the following quantity …** …

$$-2ln \frac{\sup_{\vec{\theta} \in \Theta_0} \prod_{i=1}^{N} L(\vec{x}_i; \vec{\theta})}{\sup_{\vec{\theta} \in \Theta_1} \prod_{i=1}^{N} L(\vec{x}_i; \vec{\theta})}$$

has a distribution that can be approximated, for $N \to \infty$ and if H_0 is true, with a χ^2 distribution having a n.d.o.f. = difference between the dimentionalities of the sets Θ_1 and Θ_0 .

Note: the **sup** expresses the maximization of the product of the likelihoods for the N independent measurements (for a set of variables) when a certain hypothesis is true

To understand better the theorem we can consider the example in the next slide.

Wilks' Theorem - II / example

Let us assume that μ is the **only** *parameter-of-interest*, whereas the remaining parameters $\vec{\theta} = (\theta_1, \dots, \theta_m)$ are nuisance ones.
For instance, μ could be a ***signal strength***, namely the ratio of a signal cross section to its theoretical value (say in the SM theory).

- H_0 hypothesis : $\mu = \mu_0$ (say the value foreseen by the current theory model)
- H_1 hypothesis : $\mu \geq 0$ (i.e. it may have any possible positive (or null) value)

The Wilks' theorem ensures that the quantity…

$$-2ln \frac{\sup\limits_{\vec{\theta}} \prod_{i=1}^{N} L(\vec{x}_i; \mu, \vec{\theta})}{\sup\limits_{\mu,\vec{\theta}} \prod_{i=1}^{N} L(\vec{x}_i; \mu, \vec{\theta})}$$

… is asymptotically distributed as a χ^2 with 1 d.o.f.

Wilks' Theorem - II / example

Let us assume that μ is the **only** *parameter-of-interest*, whereas the remaining parameters $\vec{\theta} = (\theta_1, \dots, \theta_m)$ are nuisance ones.

For instance, μ could be a ***signal strength***, namely the ratio of a signal cross section to its theoretical value (say in the SM theory).

H_0 hypothesis : $\mu = \mu_0$ (say the value foreseen by the current theory model)

H_1 hypothesis : $\mu \geq 0$ (i.e. it may have any possible positive (or null) value)

The Wilks' theorem ensures that the quantity…

$$-2ln \frac{\sup_{\vec{\theta}} \prod_{i=1}^{N} L\left(\vec{x}_i; \mu, \vec{\theta}\right)}{\sup_{\mu, \vec{\theta}} \prod_{i=1}^{N} L\left(\vec{x}_i; \mu, \vec{\theta}\right)}$$

… is asymptotically distributed as a χ^2 with 1 d.o.f.

Likelihood function evaluated when the parameters assume the values ($\mu = \hat{\mu}$, $\vec{\theta} = \hat{\vec{\theta}}$) that maximize it!

$$\prod_{i=1}^{N} L\left(\vec{x}_i;\ \hat{\mu}, \hat{\vec{\theta}}\right)$$

Wilks' Theorem - II / example

Let us assume that μ is the **only** *parameter-of-interest*, whereas the remaining parameters $\vec{\theta} = (\theta_1, \ldots, \theta_m)$ are nuisance ones.
For instance, μ could be a ***signal strength***, namely the ratio of a signal cross section to its theoretical value (say in the SM theory).

- H_0 hypothesis : $\mu = \mu_0$ (say the value foreseen by the current theory model)
- H_1 hypothesis : $\mu \geq 0$ (i.e. it may have any possible positive (or null) value)

The Wilks' theorem ensures that the quantity…

$$-2ln \frac{\sup\limits_{\vec{\theta}} \prod_{i=1}^{N} L(\vec{x}_i; \mu, \vec{\theta})}{\sup\limits_{\mu,\vec{\theta}} \prod_{i=1}^{N} L(\vec{x}_i; \mu, \vec{\theta})}$$

… is asymptotically distributed as a χ^2 with 1 d.o.f.

Likelihood function evaluated when $\mu = \mu_0$ and the nuisance parameters are fit and assume the values $\vec{\theta} = \hat{\hat{\vec{\theta}}}$ that maximize it for a fixed $\mu = \mu_0$!

$$\prod_{i=1}^{N} L(\vec{x}_i; \mu_0, \hat{\hat{\vec{\theta}}}(\mu_0))$$

Likelihood function evaluated when the parameters assume the values ($\mu = \hat{\mu}$, $\vec{\theta} = \hat{\vec{\theta}}$) that maximize it!

$$\prod_{i=1}^{N} L(\vec{x}_i; \hat{\mu}, \hat{\vec{\theta}})$$

Wilks' Theorem - II / example

- Let us assume that μ is the **only** *parameter-of-interest*, whereas the remaining parameters $\vec{\theta} = (\theta_1, \dots, \theta_m)$ are nuisance ones. For instance, μ could be a ***signal strength***, namely the ratio of a signal cross section to its theoretical value (say in the SM theory).

H_0 hypothesis : $\mu = \mu_0$ (say the value foreseen by the current theory model)

H_1 hypothesis : $\mu \geq 0$ (i.e. it may have any possible positive (or null) value)

The Wilks' theorem ensures that the quantity…

$$-2ln \frac{\sup_{\vec{\theta}} \prod_{i=1}^{N} L(\vec{x}_i; \mu, \vec{\theta})}{\sup_{\mu, \vec{\theta}} \prod_{i=1}^{N} L(\vec{x}_i; \mu, \vec{\theta})}$$

… is asymptotically distributed as a χ^2 with 1 d.o.f.

Likelihood function evaluated when $\mu = \mu_0$ and the nuisance parameters are fit and assume the values $\vec{\theta} = \hat{\hat{\vec{\theta}}}$ that maximize it for a fixed $\mu = \mu_0$!

$$\prod_{i=1}^{N} L(\vec{x}_i; \mu_0, \hat{\hat{\vec{\theta}}}(\mu_0))$$

Likelihood function evaluated when the parameters assume the values ($\mu = \hat{\mu}$, $\vec{\theta} = \hat{\vec{\theta}}$) that maximize it!

$$\prod_{i=1}^{N} L(\vec{x}_i; \hat{\mu}, \hat{\vec{\theta}})$$

Note: it's **not** effectively a ratio since the denominator is a real number

The test statistic (for a generic value μ) $t(\mu) = -2ln\lambda(\mu) = -2ln \frac{L(\vec{x}; \mu, \hat{\hat{\vec{\theta}}}(\mu))}{L(\vec{x}; \hat{\mu}, \hat{\vec{\theta}})}$ **… is called Profile Likelihood (ratio) … that has important application in Upper Limits calculations.**

Wilks' Theorem & Profile Likelihood (ratio)

- A minimum of $t(\mu) = -2\ln\lambda(\mu)$ at $\mu = \hat{\mu}$ indicates the possible presence of a signal having a signal strength equal to $\hat{\mu}$.

  Therefore, **this test statistics is suitable for searches of a new signal** (as will be clear later).

  Indeed, a scan of $t(\mu)$ as function of μ reveals a minimum at the value $\mu = \hat{\mu}$ and the minimum value of $t(\mu)$, namely $t(\hat{\mu})$ is 0 by contruction. As discussed elsewhere, an **uncertainty interval of $t(\mu)$** can be determined from the excursion of $t(\mu)$ around the minimum $\hat{\mu}$.

  To recap: **the Profile Likelihood is introduced in order to satisfy the conditions required by Wilk's theorem according to which, if μ corresponds to the true value, then $t(\mu)$ follows a χ^2 distribution with 1 d.o.f.**

  - Usually, **the addition of nuisance parameters broadens the shape of the profile likelihood as a function of the POI μ**, comparing with the case where nuisance parameters are not added. Consequently, the uncertainty on μ increases when nuisance parameters (typically modelling the sources of systematic) are included in the test statistic (i.e. in the likelihood). This will be clearer later.

- As will be discussed later extensively, **the test statistic $t_\mu \equiv t(\mu)$ can be used to compute p-values corresponding to the various hypotheses on μ in order to determine a statistical significance or an upper limit (different variations can deal various analysis cases).** We will argue that **those p-values can be computed in general by generating sufficiently large Monte Carlo pseudo-experiments but in many cases asymptotic approximations allow a much faster evaluation.**

Wilks' theorem : an example application - I

Again, let us assume that μ is the **only** *parameter-of-interest* (a ***signal strength***) whereas $\vec{\theta} = (\theta_1, \ldots, \theta_m)$ are the nuisance parameters.

Previously the likelihood function was considered for a set of **independent measurements** $(\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_N)$ with parameters $(\mu, \vec{\theta})$: →

$$L(\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_N; \mu, \vec{\theta}) = \prod_{i=1}^{N} f(\vec{x}_i; \mu, \vec{\theta})$$

In general, the # of events N can also be used as information and we need to consider the **extended** likelihood function: →
(Note that in the poissonian term the expected # of events ν may also depend on the parameters).

$$L(\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_N; \mu, \vec{\theta}) = \frac{e^{-\nu(\mu,\vec{\theta})}\nu(\mu,\vec{\theta})^N}{N!} \cdot \prod_{i=1}^{N} f(\vec{x}_i; \mu, \vec{\theta})$$

The two hypotheses H_0 and H_1 are represented as two possible sets of values Θ_1 and Θ_0 of the parameters $(\mu, \vec{\theta})$.
Typically, H_1 represents the presence of both signal and background (i.e. $\nu = \mu s + b$) while…
… H_0 represents the presence of only background events in our data samples (i.e. $\nu = b$, namely $\mu = 0$).
This means that hypothesis H_0 is nested in H_1 since $\nu = b$ is $\nu = \mu s + b$ with $\mu = 0$!

Note that the multiplicative parameter μ , called ***signal strength***, is typical of many data analyses performed at the LHC; it was introduced assuming that the expected signal yield from theory is s and all possible values of the expected signal are obtained by varying μ (after assuming that $\mu = 1$ corresponds to the theory prediction).

Wilks' theorem : an example application - II

- The PDF $f(\vec{x};\mu,\vec{\theta})$ - for a generic index i so we can drop the index - can be expressed as the superposition of two components:
  - one PDF for the signal : $f_s(\vec{x};\mu,\vec{\theta})$ [it typically represents a resonance peak]
  - one PDF for the background : $f_b(\vec{x};\mu,\vec{\theta})$

… to be weighted by the expected signal and background fractions : $f(\vec{x};\mu,\vec{\theta}) = \left(\frac{\mu s}{\mu s+b}\right) f_s(\vec{x};\mu,\vec{\theta}) + \left(\frac{b}{\mu s+b}\right) f_b(\vec{x};\mu,\vec{\theta})$

Note that in general s and b depend also on the unknown parameters, namely $s = s(\vec{\theta})$ and $b = b(\vec{\theta})$.
An example to understand this: in a search for the Higgs boson the theoretical cross section may depend on the Higgs boson's mass.

In this case the extended likelihood can be written as:

$$L_{s+b}(\vec{\boldsymbol{x}}_1,\dots,\vec{\boldsymbol{x}}_N;\mu,\vec{\theta}) = \frac{e^{-(\mu s(\vec{\theta})+b(\vec{\theta}))}\cancel{(\mu s(\vec{\theta})+b(\vec{\theta}))^N}}{N!}\cdot\prod_{i=1}^{N}\frac{1}{\cancel{\mu s(\vec{\theta})+b(\vec{\theta})}}\left[\mu s(\vec{\theta})f_s(\vec{x}_i;\mu,\vec{\theta}) + b(\vec{\theta})f_b(\vec{x}_i;\mu,\vec{\theta})\right]$$

$$= \frac{e^{-(\mu s(\vec{\theta})+b(\vec{\theta}))}}{N!}\cdot\prod_{i=1}^{N}\left[\mu s(\vec{\theta})f_s(\vec{x}_i;\mu,\vec{\theta}) + b(\vec{\theta})f_b(\vec{x}_i;\mu,\vec{\theta})\right]$$

Under the background-only (null) hypothesis (H_0) : $\mu = 0$ ⇨ $L_b(\vec{\boldsymbol{x}}_1,\dots,\vec{\boldsymbol{x}}_N;\vec{\theta}) = \frac{e^{-b(\vec{\theta})}}{N!}\cdot\prod_{i=1}^{N} b(\vec{\theta})f_b(\vec{x}_i;\mu,\vec{\theta})$

Wilks' theorem : an example application - III

At this point we can write down the likelihood ratio $\lambda(\vec{x}) = \frac{L(\vec{x}|H_1)}{L(\vec{x}|H_0)}$ for the specific considered case:

$$\lambda(\vec{\boldsymbol{x}}_1, \dots, \vec{\boldsymbol{x}}_N; \mu, \vec{\theta}) = \frac{L_{s+b}(\vec{\boldsymbol{x}}_1, \dots, \vec{\boldsymbol{x}}_N; \mu, \vec{\theta})}{L_b(\vec{\boldsymbol{x}}_1, \dots, \vec{\boldsymbol{x}}_N; \vec{\theta})} = \frac{\frac{e^{-(\mu s(\vec{\theta})+b(\vec{\theta}))}}{N!} \cdot \prod_{i=1}^N \left[\mu s(\vec{\theta}) f_s(\vec{x}_i; \mu, \vec{\theta}) + b(\vec{\theta}) f_b(\vec{x}_i; \mu, \vec{\theta})\right]}{\frac{e^{-b(\vec{\theta})}}{N!} \cdot \prod_{i=1}^N b(\vec{\theta}) f_b(\vec{x}_i; \mu, \vec{\theta})}$$

$$= \frac{e^{-(\mu s(\vec{\theta})+b(\vec{\theta}))}}{e^{-b(\vec{\theta})}} \cdot \prod_{i=1}^N \frac{\left[\mu s(\vec{\theta}) f_s(\vec{x}_i; \mu, \vec{\theta}) + b(\vec{\theta}) f_b(\vec{x}_i; \mu, \vec{\theta})\right]}{b(\vec{\theta}) f_b(\vec{x}_i; \mu, \vec{\theta})} = e^{-(\mu s(\vec{\theta}))} \cdot \prod_{i=1}^N \left[\frac{\mu s(\vec{\theta}) f_s(\vec{x}_i; \mu, \vec{\theta})}{b(\vec{\theta}) f_b(\vec{x}_i; \mu, \vec{\theta})} + 1\right]$$

… and thus the negative logarithm of the likelihood ratio is (applying as usual the logarithm's properties):

$$-\ln \lambda(\vec{\boldsymbol{x}}_1, \dots, \vec{\boldsymbol{x}}_N; \mu, \vec{\theta}) = -\ln e^{-(\mu s(\vec{\theta}))} - \ln \prod_{i=1}^N \left[\frac{\mu s(\vec{\theta}) f_s(\vec{x}_i; \mu, \vec{\theta})}{b(\vec{\theta}) f_b(\vec{x}_i; \mu, \vec{\theta})} + 1\right] = +\mu s(\vec{\theta}) - \sum_{i=1}^N \ln \left[\frac{\mu s(\vec{\theta}) f_s(\vec{x}_i; \mu, \vec{\theta})}{b(\vec{\theta}) f_b(\vec{x}_i; \mu, \vec{\theta})} + 1\right]$$

This equation can be used to determine Upper Limits in searches for new signals (L.Lista's book pagg. 222-223 -CLs method)!
Despite the fact that this neg-log-likelihood ratio is written with H_0 at the denominator and H_1 at the numerator, that is the inverse convention w.r.t. that used for the Wilks' theorem (but identical to the ratio defined in the framework of the Nyman-Pearson Lemma)
… Wilk's theorem can apply also in this case with the only change of an extra "-" sign in the definition of the test statistic
(a "-" in front of the logarithm of a ratio just makes the inversion of the ratio).

Wilks' theorem : (simple counting experiment example) - IV

In the case of a **simple counting experiment** … the likelihood function **only** accounts for the Poissonian probability term which only depends on the # of observed events N and the dependence on the parameters only appears in the expected signal and background yields:

$$\lambda(N;\mu,\vec{\theta}) = \frac{L_{s+b}(N;\mu,\vec{\theta})}{L_b(N;\vec{\theta})} = \frac{e^{-(\mu s(\vec{\theta})+b(\vec{\theta}))}}{e^{-b(\vec{\theta})}} \cdot \prod_{i=1}^{N} \frac{[\mu s(\vec{\theta}) + b(\vec{\theta})]}{b(\vec{\theta})} = e^{-(\mu s(\vec{\theta}))} \cdot \prod_{i=1}^{N} \left[\frac{\mu s(\vec{\theta})}{b(\vec{\theta})} + 1\right] = e^{-(\mu s(\vec{\theta}))} \cdot \left[\frac{\mu s(\vec{\theta})}{b(\vec{\theta})} + 1\right]^N$$

$$\Rightarrow \quad -\ln\lambda(N;\mu,\vec{\theta}) = -\ln e^{-(\mu s(\vec{\theta}))} - \ln\left[\frac{\mu s(\vec{\theta})}{b(\vec{\theta})} + 1\right]^N = +\mu s(\vec{\theta}) - N\ln\left[\frac{\mu s(\vec{\theta})}{b(\vec{\theta})} + 1\right]$$

… which is a simplified version of the previous expressions with the terms f_s and f_b dropped.

The same considerations about the application of Wilks' theorem hold.

Introduction to the search for New Signals - I

The goal of many experiments is to search for new physical phenomena.
If an experiment provides a convincing measurement of a new signal the result should be published and claimed as discovery, otherwise, it can be nonetheless interesting to quote an ***upper limit*** to the yield of the possible new signal.

Given an observed data sample, claiming the discovery of a new signal requires determining that the sample is sufficiently ***inconsistent*** with the hypothesis that only background is present in the data (null hypothesis H_0).
A test statistic can be used to measure this inconsistency of the observation in the hypothesis of the presence of background only.

To claim a discovery one needs to quote a **p-value** or alternatively a **statistical significance** given as an equivalent number of standard deviations !

$p - \text{value}$ ⟺ **Probability** that
the considered test statistic **t** assumes
a value **greater or equal to the observed one**
in the case of pure background fluctuation

[large values of the test statistic correspond to a more signal-like sample]

In the case of an ***event counting experiment*** (in which the number of observed events is adopted as test statistic, the p-value can be determined as the probability to count a number of events *equal to* or *greater than* the observed one assuming the presence of no signal and the expected background level (see example next slide).

Introduction to the search for New Signals - II

- From L.Lista's book (pagg. 206-7):

Example 10.25 ***p*-Value for a Poissonian Counting**
Figure 10.1 shows a Poisson distribution corresponding to an expected number of (background-only) events equal to 4.5. In case the observed number of events is 8, the *p*-value is equal to the probability to observe 8 or more events, i.e. it is given by:

$$p = P(n \geq 8) = \sum_{n=8}^{\infty} \mathrm{Pois}(n;\, 4.5) = 1 - e^{-4.5} \sum_{n=0}^{7} \frac{4.5^n}{n!}\,.$$

Performing the computation explicitly, a *p*-value of 0.087 can be determined.

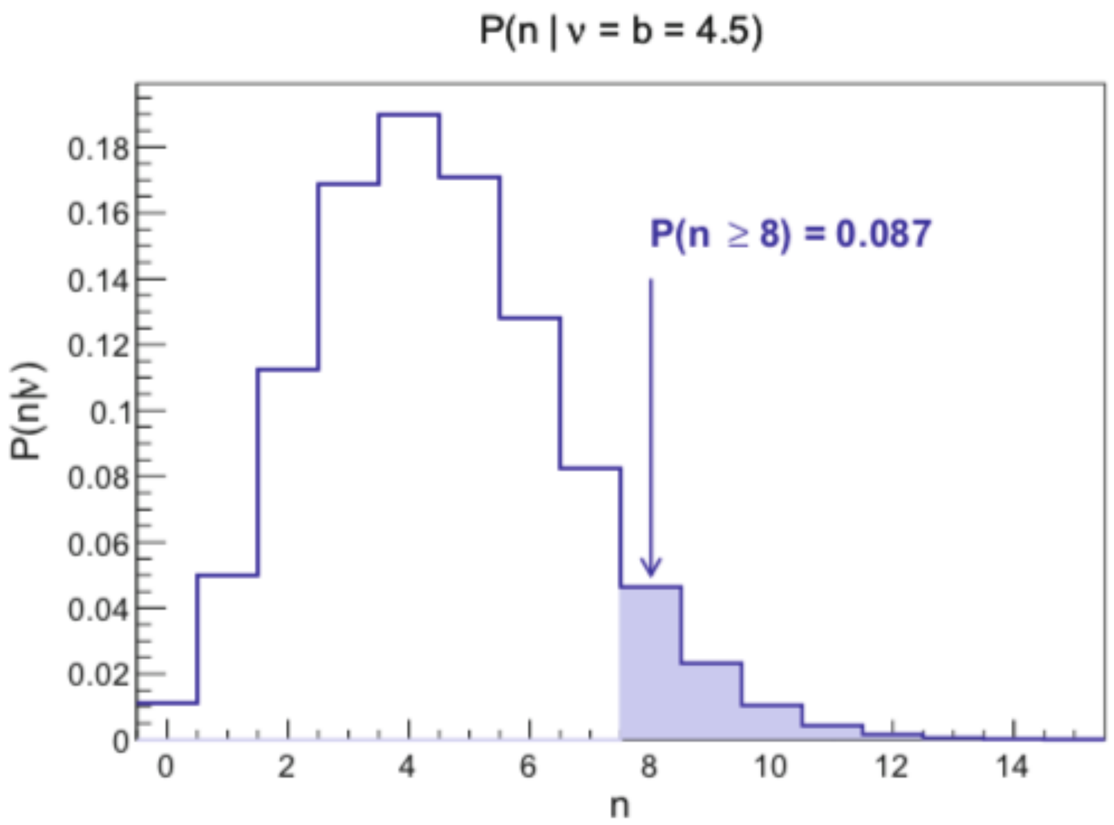

Fig. 10.1 Poisson distribution for a null signal and an expected background of $b = 4.5$. The probability corresponding to $n \geq 8$ (*light blue area*) is 0.087, and is equal to the p-value, assuming the event counting as test statistic

Introduction to the search for New Signals - III

- Instead of quoting a p-value, it's often preferred to report the **equivalent number of standard deviations that correspond to an area equal to the p-value under the right-most tail of a normal distribution**.
  Thus one quotes a Zσ significance corresponding to a given p-value using the following transformation:

$$p = \int_Z^{\infty} \frac{e^{-x^2/2}dx}{\sqrt{2\pi}} = 1 - \Phi(Z) = \Phi(-Z) = \frac{1}{2}\left[1 - erf\left(\frac{Z}{\sqrt{2}}\right)\right]$$

This table provides the correspondence between Zσ & p-value :

$Z(\sigma)$	p
1.00	1.59×10^{-1}
1.28	1.00×10^{-1}
1.64	5.00×10^{-2}
2.00	2.28×10^{-2}
2.32	1.00×10^{-2}
3.00	1.35×10^{-3}
3.09	1.00×10^{-3}
3.71	1.00×10^{-4}
4.00	3.17×10^{-5}
5.00	2.87×10^{-7}
6.00	9.87×10^{-10}

Typical convention

Evidence (>3σ)

Observation (>5σ)

Significance for Poissonian counting experiment

In a counting experiment the # of observed events is the only considered information.
The **selected event sample** contains - in general - a mixture of n events due to both signal and background process;
the **expected total number of events** is $s + b$ where s and b are the expected # of signal and background events respectively.

Assuming the expected background is known (from theory or from a control data sample with negligible uncertainty) the main unknown parameter of the problem is s and the likelihood function is:

$$L(n; s, b) = \frac{(s+b)^n}{n!} e^{-(s+b)}$$

The # of observed events n must be compared with the expected number of background events b in the null hypothesis ($s = 0$)

If b is sufficiently large, the distribution can be approximated with a Gaussian with average b and standard deviation = $\sqrt{b}$).
An excess in data, quantified as $s = n - b$ should be compared with the expected standard deviation $\sqrt{b}$ and the statistical significance can be approximately evauated with a well-popular expression:

$$Z = \frac{s}{\sqrt{b}}$$

In case the expected background is affected by a non-negligible uncertainty the previous expression must be modified: $Z = \frac{s}{\sqrt{b + \sigma_b^2}}$

Cowan suggests a better approximation valid even in the case $b \ll 1$:

$$Z = \sqrt{2\left[(s+b)\ln\left(1+\frac{s}{b}\right) - s\right]} \xrightarrow{s \ll b} Z = \frac{s}{\sqrt{b}}$$

Significance with Likelihood ratio - I

- As already pointed out, **a test statisic suitable for searches for a new signal is the likelihood ratio**: $\lambda(\vec{x}) = \frac{L(\vec{x}|H_1)}{L(\vec{x}|H_0)}$

For instance, as discussed before, **a likelihood ratio of the form** $\lambda(\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_N; \mu, \vec{\theta}) = \frac{L_{s+b}(\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_N; \mu, \vec{\theta})}{L_b(\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_N; \vec{\theta})}$

Of course, a minimum of the test statistic $-2\ln \boldsymbol{\lambda}(\boldsymbol{\mu})$ …

[I write here compactly $\lambda(\vec{x}_1, \ldots, \vec{x}_N; \mu) \equiv \lambda(\mu)$**,** having dropped the dependence on nuisance parameters **(*)**]

… at $\boldsymbol{\mu} = \widehat{\boldsymbol{\mu}}$ indicates the possible presence of a signal having a signal strength equal to $\widehat{\boldsymbol{\mu}}$.

Important note: The advantage of the (negative-log) likelihood ratio as test statistic is that H_0 , assumed in the denominator, can be taken as a special case of the H_1 , assumed in the nominator, with $\boldsymbol{\mu} = \mathbf{0}$.
This represents a case of nested hypothesis and, assuming the likelihood function is sufficiently regular to satisfy the Wilks' theorem requisites, the theorem holds!
Again, note that the convention is the opposite of the Wilks theorem (numerator and denominator hypotheses are exchanged and an extra "-" sign is involved. Thus, the test statistic must correctly expressed as $+\mathbf{2ln}\ \boldsymbol{\lambda}(\boldsymbol{\mu})$.

According to Wilks' theorem, the distribution of $2\ln \lambda(\widehat{\mu})$ can be approximated by a χ^2 distribution with 1 degree of freedom.
In particular, an approximate estimate of the significance level Z is given by : $\mathbf{Z} \cong \sqrt{2 \ln \boldsymbol{\lambda}(\widehat{\boldsymbol{\mu}})}$

(*) this significance is called " *local* " in the sense that it corresponds to a fixed set of values for the nuisance parameter(s) $\vec{\theta}$!

Significance with Likelihood ratio - II

- In case one or more parameters are estimated from data … **the local significance at fixed values of the measured parameters can be affected by the *look-elsewhere-effect*** as we will discuss in the ***annex slides***.

- An accurate estimate of the statistical significance corresponding to the test statistic $-2\ln\lambda$ can be achieved by generating a large number of **Monte Carlo pseudo-experiments** assuming the presence of no signal ($\mu = 0$), which gives a good approximation of the expected distribution of $-2\ln\lambda$ which is not known when the Wilks' theorem does not apply/hold.

  In order to determine large significance values ($\geq 5\sigma$) with sufficient precision, very large samples of these " *MC toys* " are needed, as we will discuss later.

- A convenient statistics that accounts for nuisance parameters (all the parameters are treated as nuisance with the exception of μ treated as the only parameter-of-interest) is the **Profile Likelihood (ratio)**, introduced earlier.
  A scan of the test statistic $t_\mu(\mu) = -2\,ln\,\lambda\,(\mu)$ **as a function of μ reveals a minimum at the value $\mu = \hat{\mu}$.**
  The minimum value $t_\mu(\hat{\mu}) = 0$ **by construction!** An **uncertainty interval for μ can be obtained** with the method discussed in an earlier lesson (connection between MINOS and Profile Likelihhod); the interval extremes happen at $t_\mu = 1$.
  To be clear, let me stress here that **the Profile Likelihood is introduced in order to satisfy the conditions required by the Wilks' theorem,** according to which **if μ corresponds to the true value then t_μ follows a χ^2 distribution with 1 d.o.f.!**

Profile likelihood as test statistic for *Observation*

- **In order to enforce the condition $\mu \geq 0$, since the signal yield cannot have negative values, the test statistic $t_\mu(\mu) = -2\ ln\ \lambda(\mu)$ can be modified as follows:**

$$\tilde{t}_\mu = -2\log\tilde{\lambda}(\mu) = \begin{cases} -2\log\frac{L(\vec{x}\,|\,\mu,\hat{\hat{\theta}}(\mu))}{L(\vec{x}\,|\,\hat{\mu},\hat{\theta})} & \hat{\mu} \geq 0\,, \\ -2\log\frac{L(\vec{x}\,|\,\mu,\hat{\hat{\theta}}(\mu))}{L(\vec{x}\,|\,0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0\,. \end{cases}$$

In practise, the estimate of μ is replaced with zero if the best fit value $\hat{\mu}$ is negative, which may occur in case of a downward fluctuation in data.

- **In order to assess the presence of a new signal, the hypothesis of positive signal strength μ is tested against the hypothesis $\mu = 0$. This is done with the test statistic $t_\mu(\mu) = -2\ ln\ \lambda(\mu)$ evaluated for $\mu = 0$. However, the test statistic $t_0 = -2\ ln\ \lambda(0)$ may reject the hypothesis of null signal ($\mu = 0$) in case of a downward fluctuation in data. Therefore, a modification of t_0 has been proposed that is only sensitive to an excess in data that produces a positive value of $\hat{\mu}$:**

$$q_0 = \begin{cases} -2\log\lambda(0) & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0\,. \end{cases}$$

The p-value corresponding to the test statistic q_0 can be also evaluated with MC pseudo-experiments, see *annex slides*.

For completeness have a reading to the ***annex slides***
(my talk at the conference Charm 2020 given in may 2021)
and the related Proceedings.
Links are on the web page of this course.