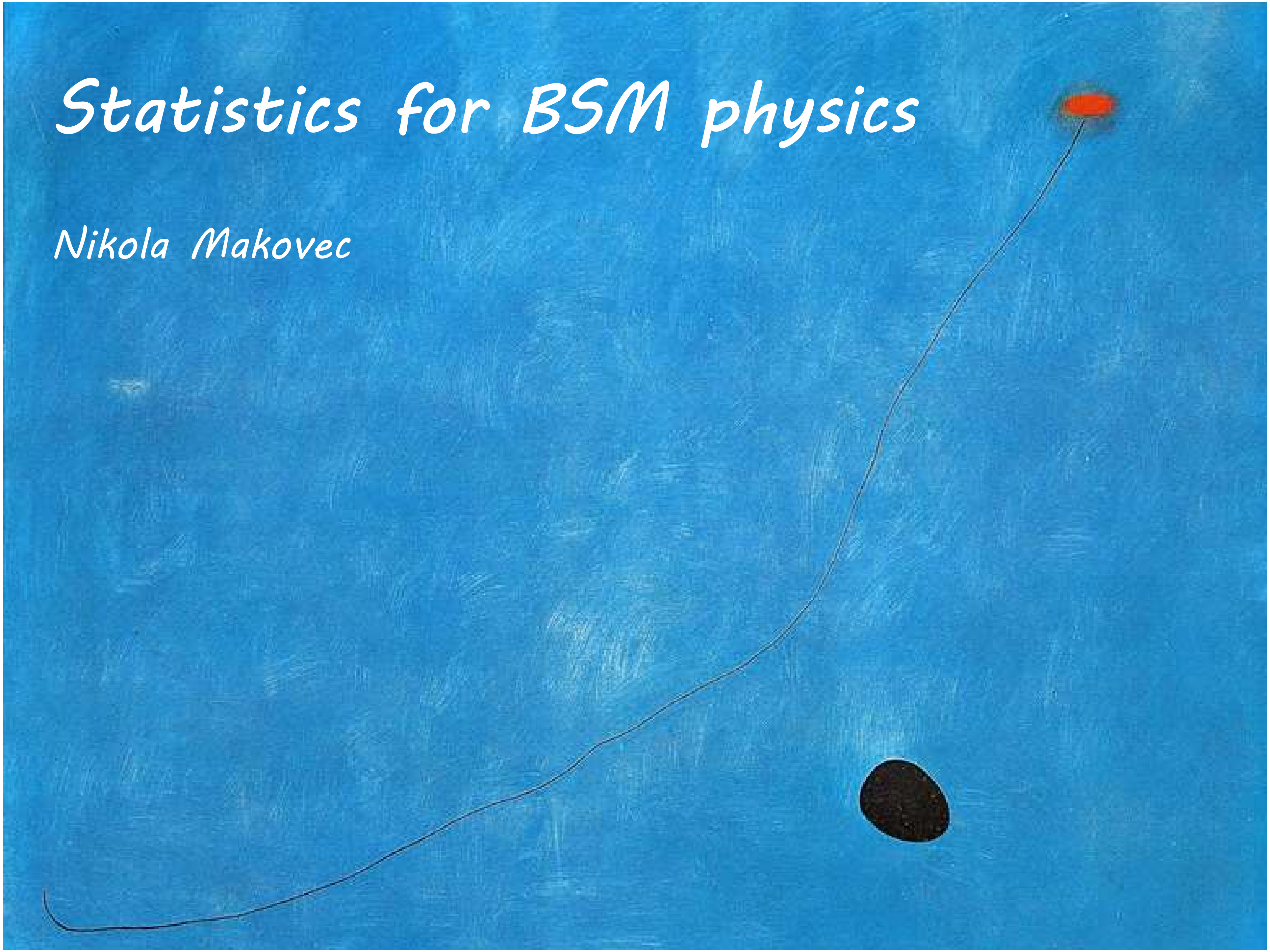
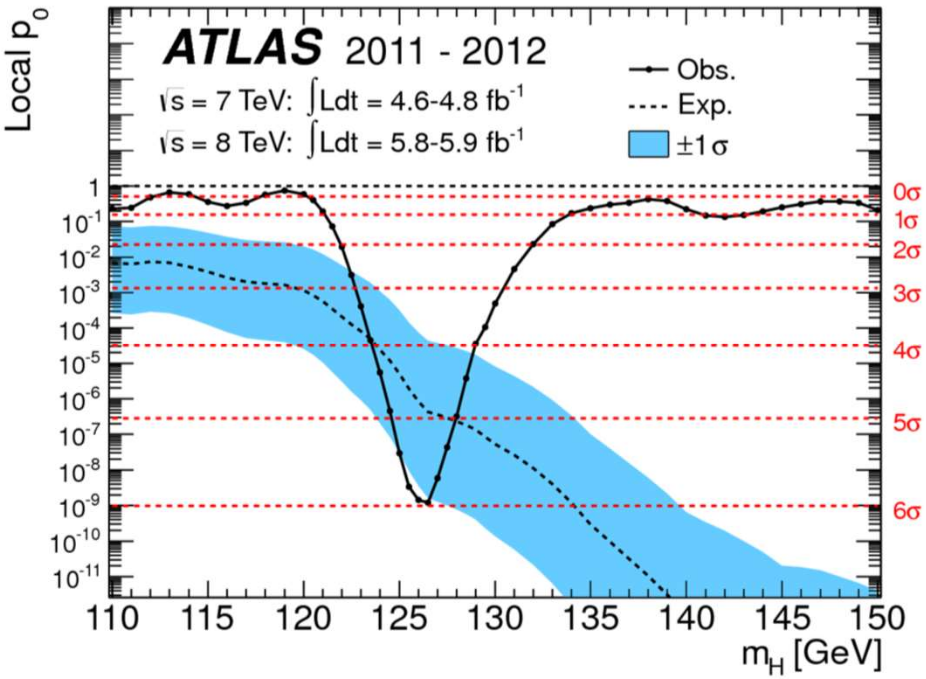
Statistics for BSM physics

Nikola Makovec

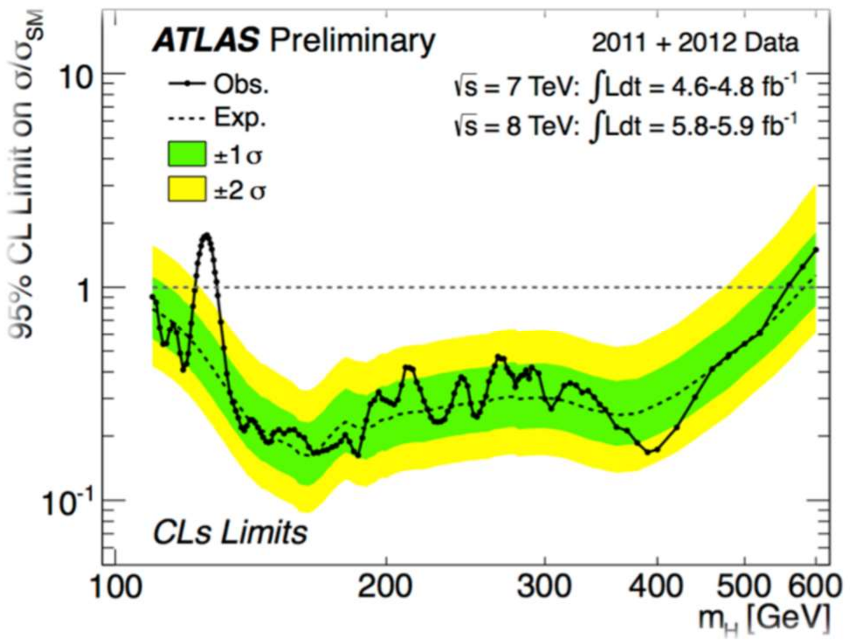


Statistics in particle physics

Discovery plot



Exclusion plot



The goal of this lecture is to understand how these plots are made and how to interpret them

Outline and references

Outline:

1. Probability density function
2. Parameters estimation with the method of maximum likelihood
3. Modelling the data
4. Hypothesis tests
    1. Discovery
    2. Exclusion

References:

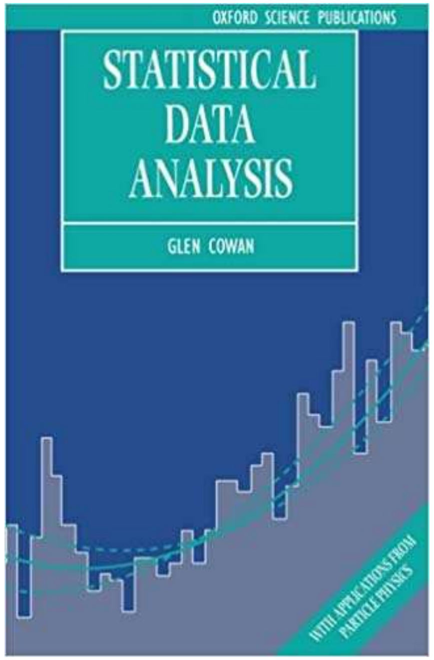
- Statistical data analysis, G. Cowan (Oxford University Press)
    - A reference book covering the basic of statistics for HEP

- Statistics for searches at the LHC, G. Cowan
    - https://arxiv.org/abs/1307.2487
    - Include material not covered in his book (eg: CLs, Profile-Likelihood,…)

- Introduction to Statistical Methods for High Energy Physics, G. Cowan
    - https://indico.cern.ch/event/134153/
    - Summer Student Lecture Programme Course

- Foundations of statistics, A. Hoecker
    - https://indico.cern.ch/event/713464/
    - Summer Student Lecture Programme Course

- Statistical analysis methods in HEP, N. Berger
    - https://indico.lal.in2p3.fr/event/4738/
    - LAL Winter Lecture



Probability density function

Probability distribution

A *random variable* represents the outcome of a repeatable experiment whose result is uncertain.

Probabilistic treatment of possible outcomes

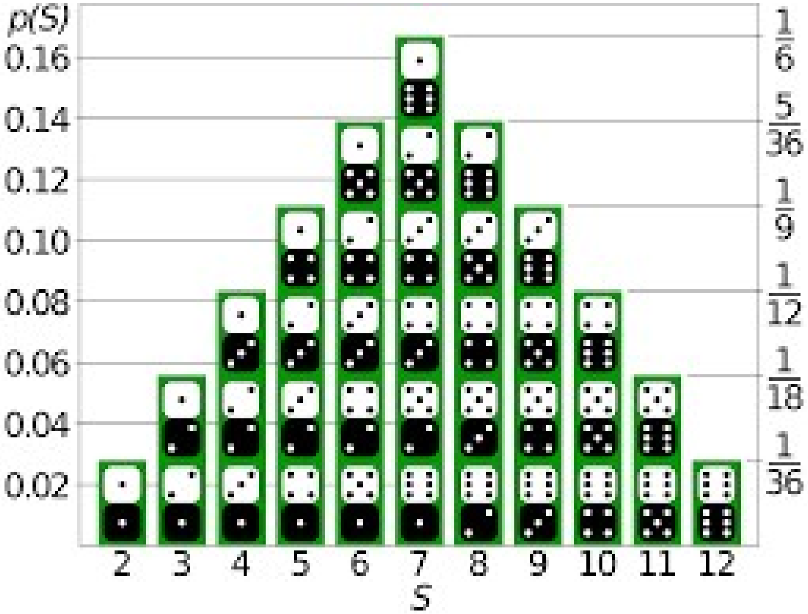
→ Probability distribution for discrete variables

Properties:

$$P_i \geq 0$$

$$\sum_i P_i = 1$$

Example: two dices roll probability



Probability density function (pdf)

A random variable can also be a continuous variable

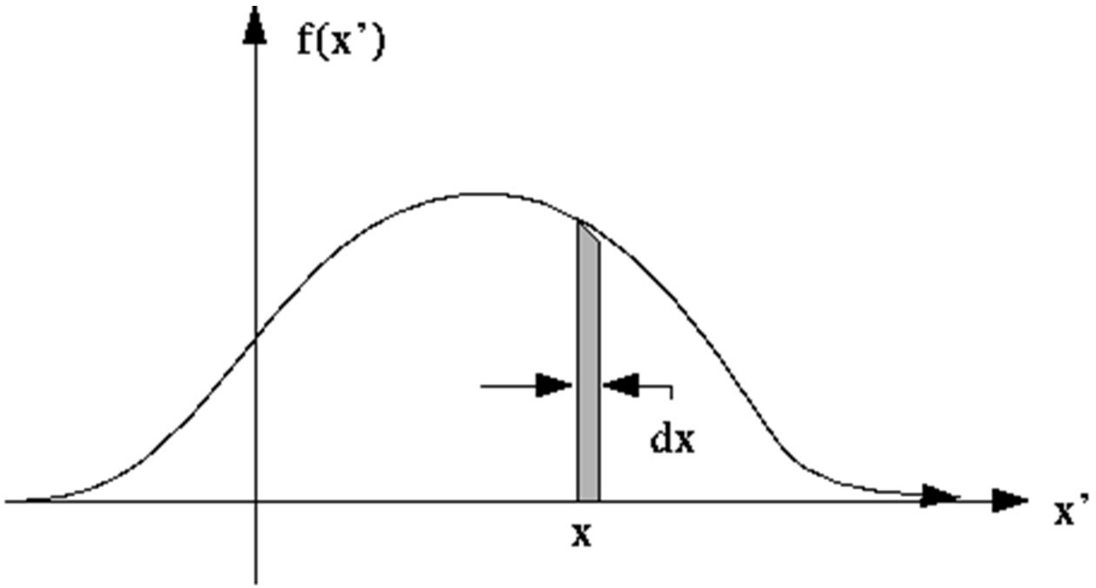
→ Probability distribution function: p(x)

p(x)dx gives the probability that x is observed in [x, x + dx]

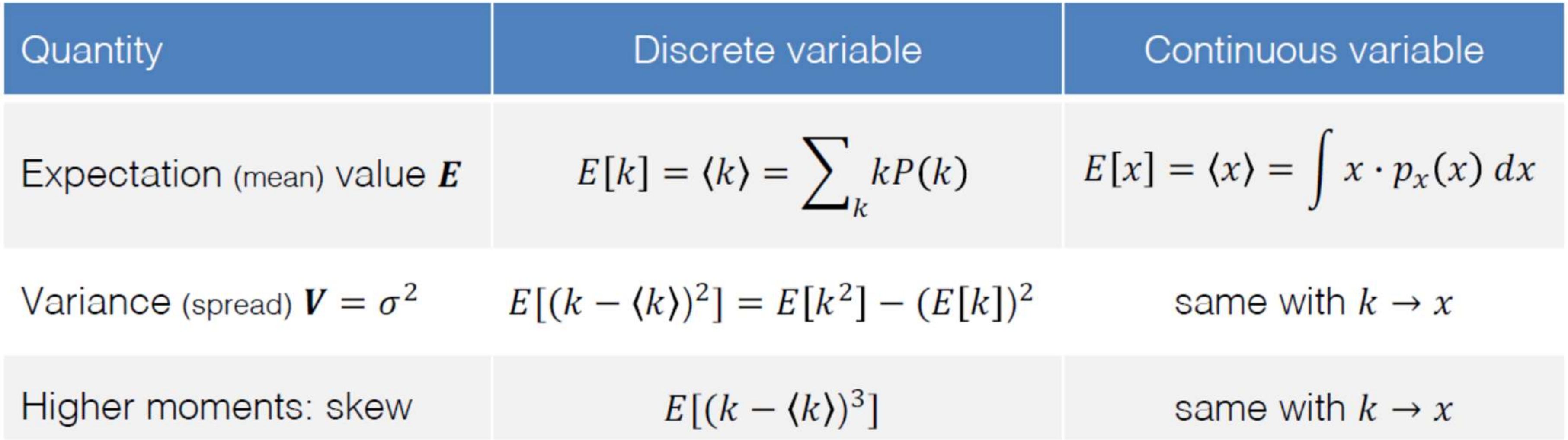
Properties:

$$p(x) \geq 0$$

$$\int p(x)dx = 1$$



Properties



Quantity	Discrete variable	Continuous variable
Expectation (mean) value $\boldsymbol{E}$	$E[k] = \langle k \rangle = \sum_k kP(k)$	$E[x] = \langle x \rangle = \int x \cdot p_x(x)\, dx$
Variance (spread) $\boldsymbol{V} = \sigma^2$	$E[(k - \langle k \rangle)^2] = E[k^2] - (E[k])^2$	same with $k \rightarrow x$
Higher moments: skew	$E[(k - \langle k \rangle)^3]$	same with $k \rightarrow x$

The variance represents the width of the PDF about the mean

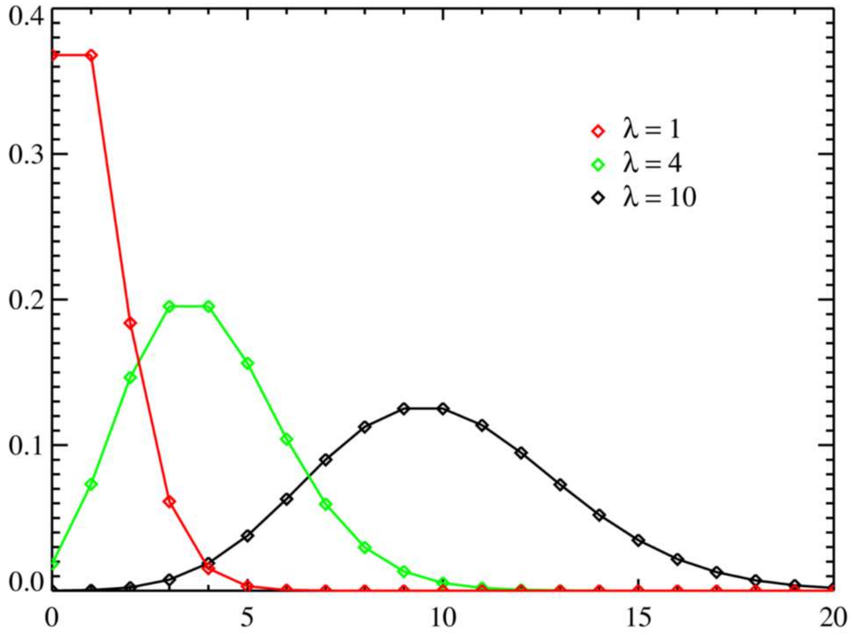
Convenient to express this in terms of the standard deviation $\sigma=\sqrt{V}$

Higher moment (like skew) can be defined and are not very useful in practice

Poisson distribution

n is a discrete random variable

$$P(n,\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$



Properties:

- E[x]=λ
- V[x]= λ
- $P(n,\alpha).P(n,\beta) = P(n,\alpha+\beta)$

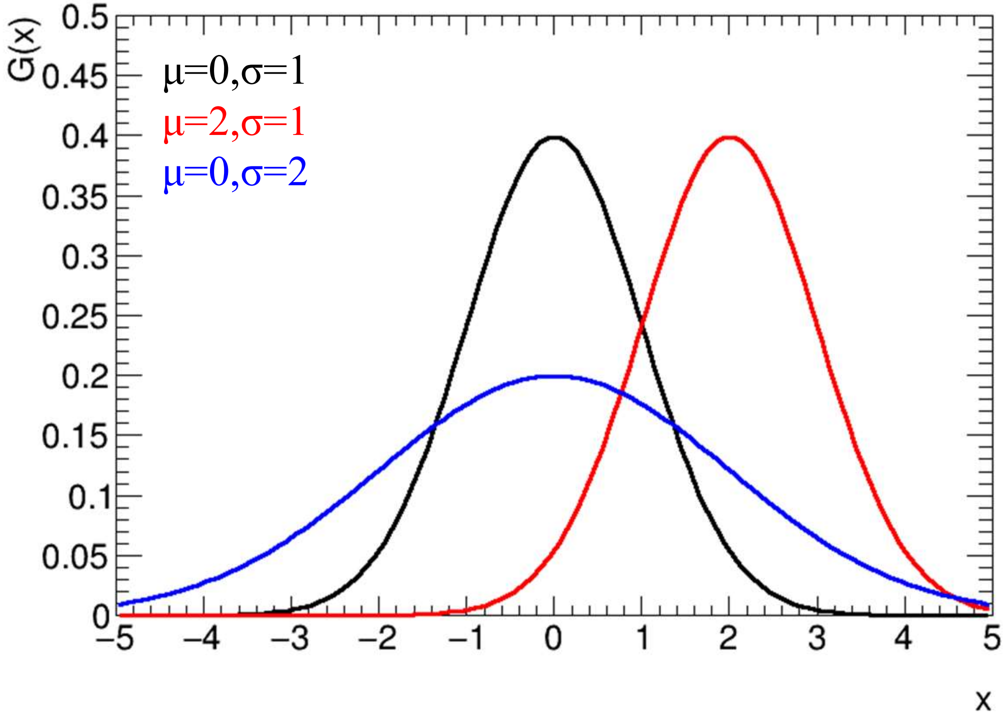
An example of a Poisson random variable is the number of events of a certain type observed in a particle scattering experiment with a given integrated luminosity *L* in the limit that the total number of events is very large and the probability for an individual decay within the time period is very small.

The Poisson distribution approaches the Gaussian distribution for large λ.

Gaussian (aka Normal) distribution

x is a continous random variable

$$G(x;\mu,\sigma) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{\text{Normalisation factor}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

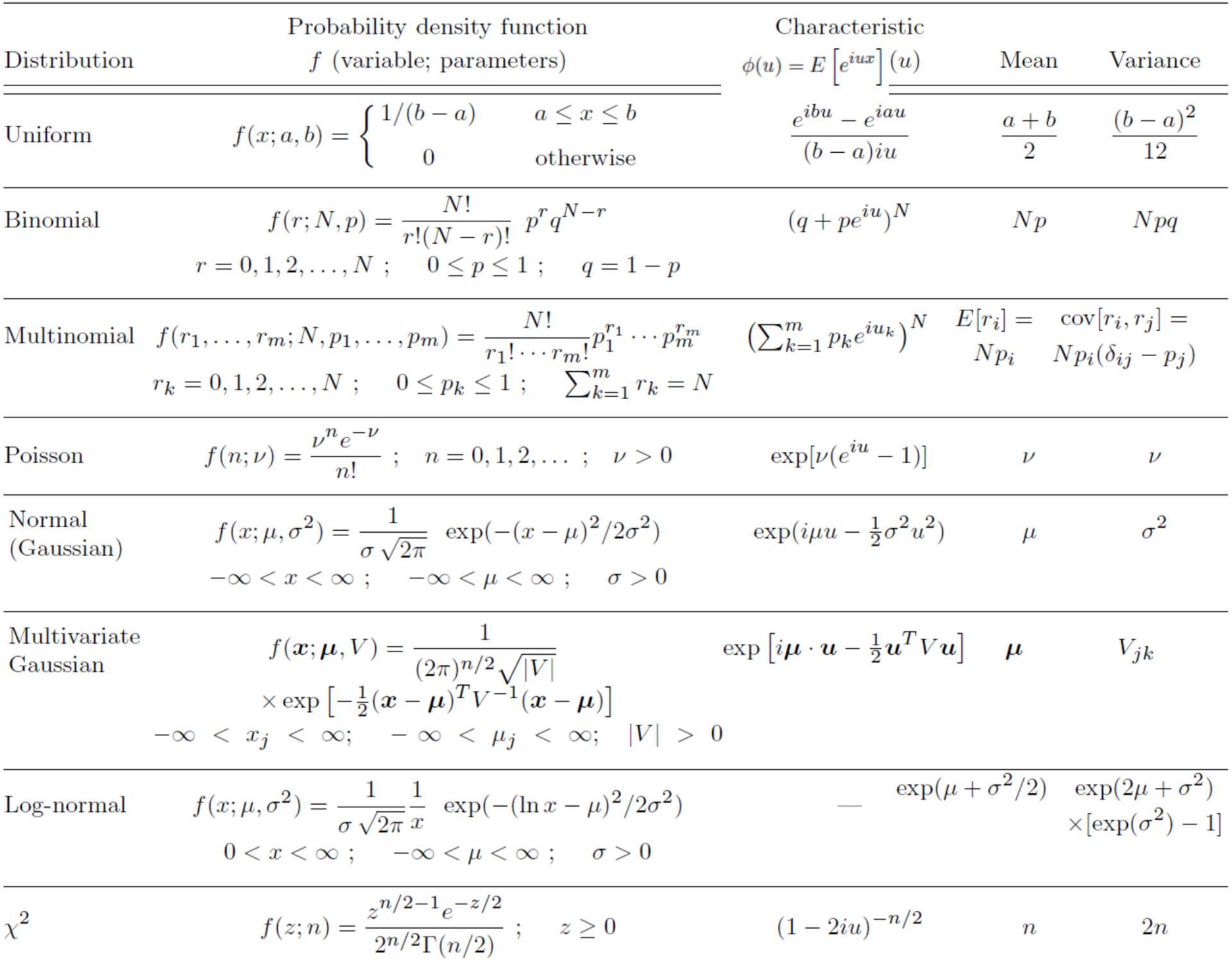


Properties:

- E[x]=μ
- V[x]=σ^2

Thanks to the Central Limit Theorem Limit (CLT), the Gaussian pdf plays an important role in statistics

Common probability density functions



Distribution	Probability density function f (variable; parameters)	Characteristic $\phi(u) = E\left[e^{iux}\right](u)$	Mean	Variance
Uniform	$f(x;a,b) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{e^{ibu} - e^{iau}}{(b-a)iu}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Binomial	$f(r;N,p) = \dfrac{N!}{r!(N-r)!}\ p^r q^{N-r}$ $r = 0,1,2,\ldots,N\ ;\quad 0 \le p \le 1\ ;\quad q = 1-p$	$(q + pe^{iu})^N$	Np	Npq
Multinomial	$f(r_1,\ldots,r_m;N,p_1,\ldots,p_m) = \dfrac{N!}{r_1!\cdots r_m!} p_1^{r_1}\cdots p_m^{r_m}$ $r_k = 0,1,2,\ldots,N\ ;\quad 0 \le p_k \le 1\ ;\quad \sum_{k=1}^m r_k = N$	$\left(\sum_{k=1}^m p_k e^{iu_k}\right)^N$	$E[r_i] =$ Np_i	$\mathrm{cov}[r_i,r_j] =$ $Np_i(\delta_{ij} - p_j)$
Poisson	$f(n;\nu) = \dfrac{\nu^n e^{-\nu}}{n!}\ ;\quad n = 0,1,2,\ldots\ ;\quad \nu > 0$	$\exp[\nu(e^{iu} - 1)]$	ν	ν
Normal (Gaussian)	$f(x;\mu,\sigma^2) = \dfrac{1}{\sigma\sqrt{2\pi}}\ \exp(-(x-\mu)^2/2\sigma^2)$ $-\infty < x < \infty\ ;\quad -\infty < \mu < \infty\ ;\quad \sigma > 0$	$\exp(i\mu u - \frac{1}{2}\sigma^2 u^2)$	μ	σ^2
Multivariate Gaussian	$f(\boldsymbol{x};\boldsymbol{\mu},V) = \dfrac{1}{(2\pi)^{n/2}\sqrt{\lvert V\rvert}}$ $\times \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T V^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$ $-\infty < x_j < \infty;\quad -\infty < \mu_j < \infty;\quad \lvert V\rvert > 0$	$\exp\left[i\boldsymbol{\mu}\cdot\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^T V\boldsymbol{u}\right]$	$\boldsymbol{\mu}$	V_{jk}
Log-normal	$f(x;\mu,\sigma^2) = \dfrac{1}{\sigma\sqrt{2\pi}}\dfrac{1}{x}\ \exp(-(\ln x - \mu)^2/2\sigma^2)$ $0 < x < \infty\ ;\quad -\infty < \mu < \infty\ ;\quad \sigma > 0$	—	$\exp(\mu + \sigma^2/2)$	$\exp(2\mu + \sigma^2)$ $\times[\exp(\sigma^2) - 1]$
χ^2	$f(z;n) = \dfrac{z^{n/2-1}e^{-z/2}}{2^{n/2}\Gamma(n/2)}\ ;\quad z \ge 0$	$(1-2iu)^{-n/2}$	n	$2n$

http://pdg.lbl.gov/2014/reviews/rpp2014-rev-probability.pdf

Joint probability distribution

The concept of probability density function can be generalized to several dimensions (joint probability distribution).

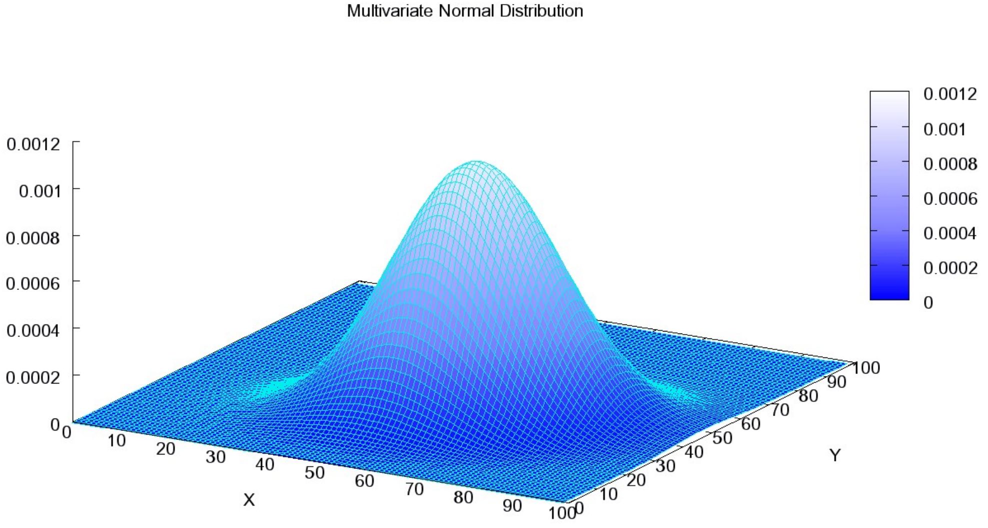
For instance in 2D, p(x,y) measures the probability density per unit area:

p(x,y)dxdy gives the probability that x is observed in [x, x + dx] and y in [y,y+dy]

Properties:

$$p(x,y) \geq 0$$

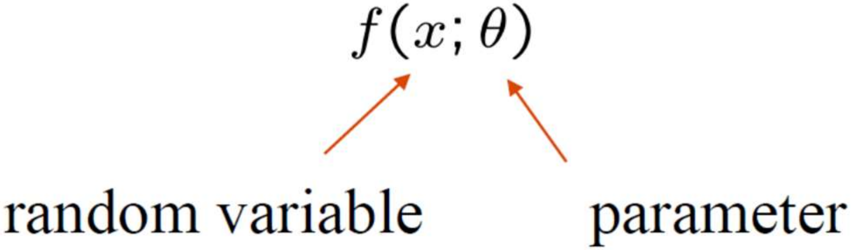
$$\int p(x,y)dxdy = 1$$



Maximum likelihood fits

Parameter estimation

The parameters of a pdf are constants that characterize its shape, e.g.



$$f(x;\theta)$$

random variable

parameter

Suppose we have a sample of observed values: $\vec{x} = (x_1, \ldots, x_n)$

We want to find some function of the data to estimate the parameter(s):

$$\boxed{\widehat{\theta}(\vec{x})} \quad \leftarrow \text{ estimator written with a hat}$$

Sometimes we say 'estimator' for the function of $x_1, ..., x_n$; 'estimate' for the value of the estimator with a particular data set.

The likelihood function

Suppose the entire result of an experiment (set of measurements) is a collection of numbers **x**$=\vec{x}=(x_1,\ldots x_n)$, and suppose the joint pdf for the data **x** is a function that depends on a set of parameters θ:

$$f(\vec{x};\vec{\theta})$$

Now evaluate this function with the data obtained and regard it as a function of the parameter(s). This is the likelihood function

$$L(\vec{\theta}) = f(\vec{x};\vec{\theta}) \qquad (\boldsymbol{x}\text{ constant})$$

The likelihood function gives for fixed data, the relative likelihood of various parameters.

The probability density function gives for fixed parameters, the probability density of various possible data.

Maximum likelihood estimators

If the hypothesized θ is close to the true value, then we expect a high probability to get data like that which we actually measured

So we define the maximum likelihood (ML) estimator(s) to be the parameter value(s) for which the likelihood is maximum

In practice, one prefer to minimize -ln L(θ) or -2ln L(θ)

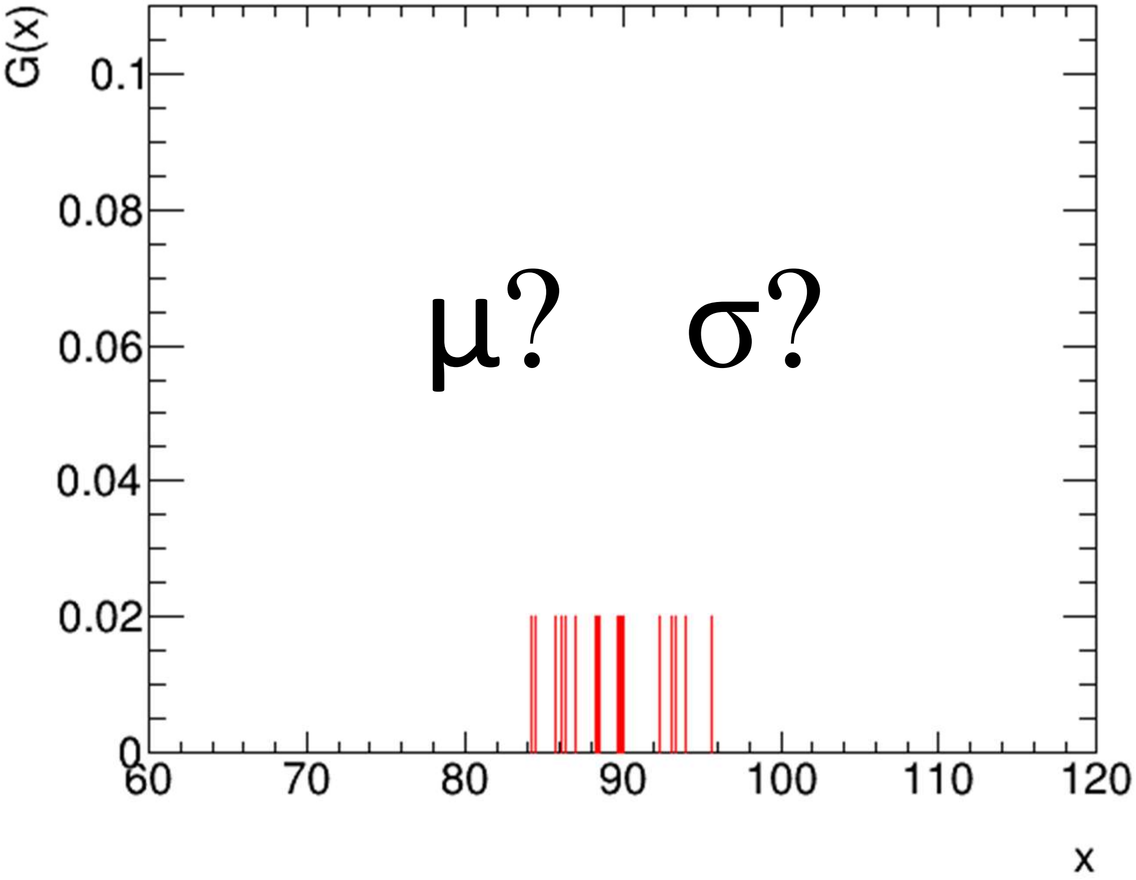
Maximum likelihood estimators (MLE) not guaranteed to have any ‘optimal’ properties (bias, variance) but in practice they’re very good.

MLE: the Gaussian example

Suppose we have a sample of N observed values $\{x_i\}$ and that the underlying distribution is a Gaussian

Measurements:

1. 94.0
2. 88.3
3. 93.1
4. 89.9
5. 93.3
6. 89.8
7. 86.4
8. 89.7
9. 90.0
10. 88.4
11. 95.6
12. 86.1
13. 89.8
14. 84.2
15. 85.8
16. 84.4
17. 93.1
18. 87.1
19. 92.3
20. 88.5



MLE: the Gaussian example

The likelihood to measure x_i for one measurement is :

$$\frac{e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

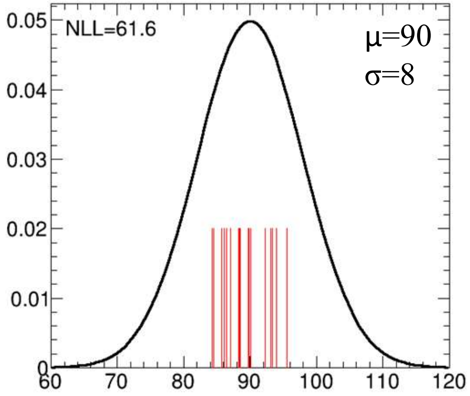
Likelihood for 1 measurement

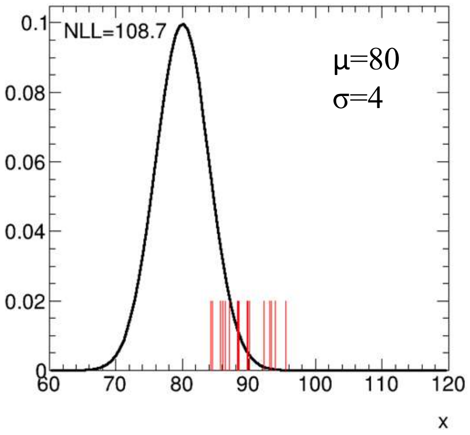
The likelihood to measure $(x_1,\ldots x_n)$ is the product of the individual likelihoods:

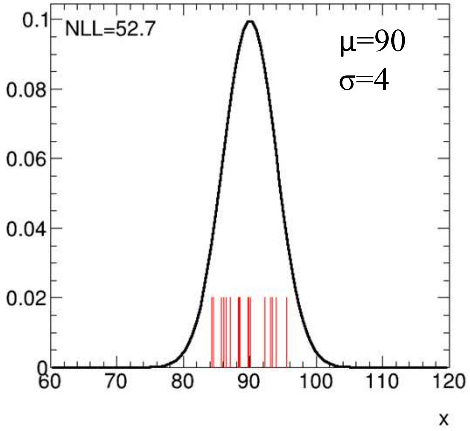
$$\prod_{i=1}^{N}\frac{e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

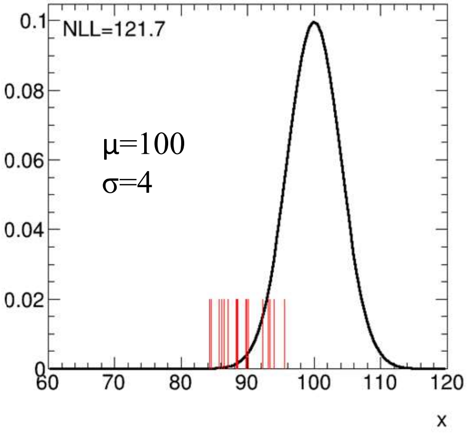
Likelihood for independent and identically distributed data

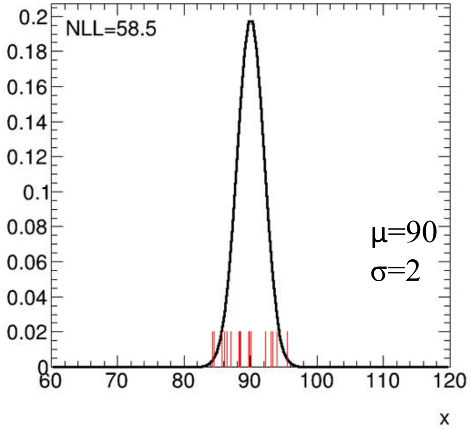
MLE: the Gaussian example











$$NLL = -\ln\Big(\prod_{i=1}^{N} \frac{e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\Big)$$

MLE: the Gaussian example

The likelihood:

$$L(\mu,\sigma)=\prod_{i=1}^{N}\frac{e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

Likelihood for independent
and identically distributed data

But it is more convenient to work with:

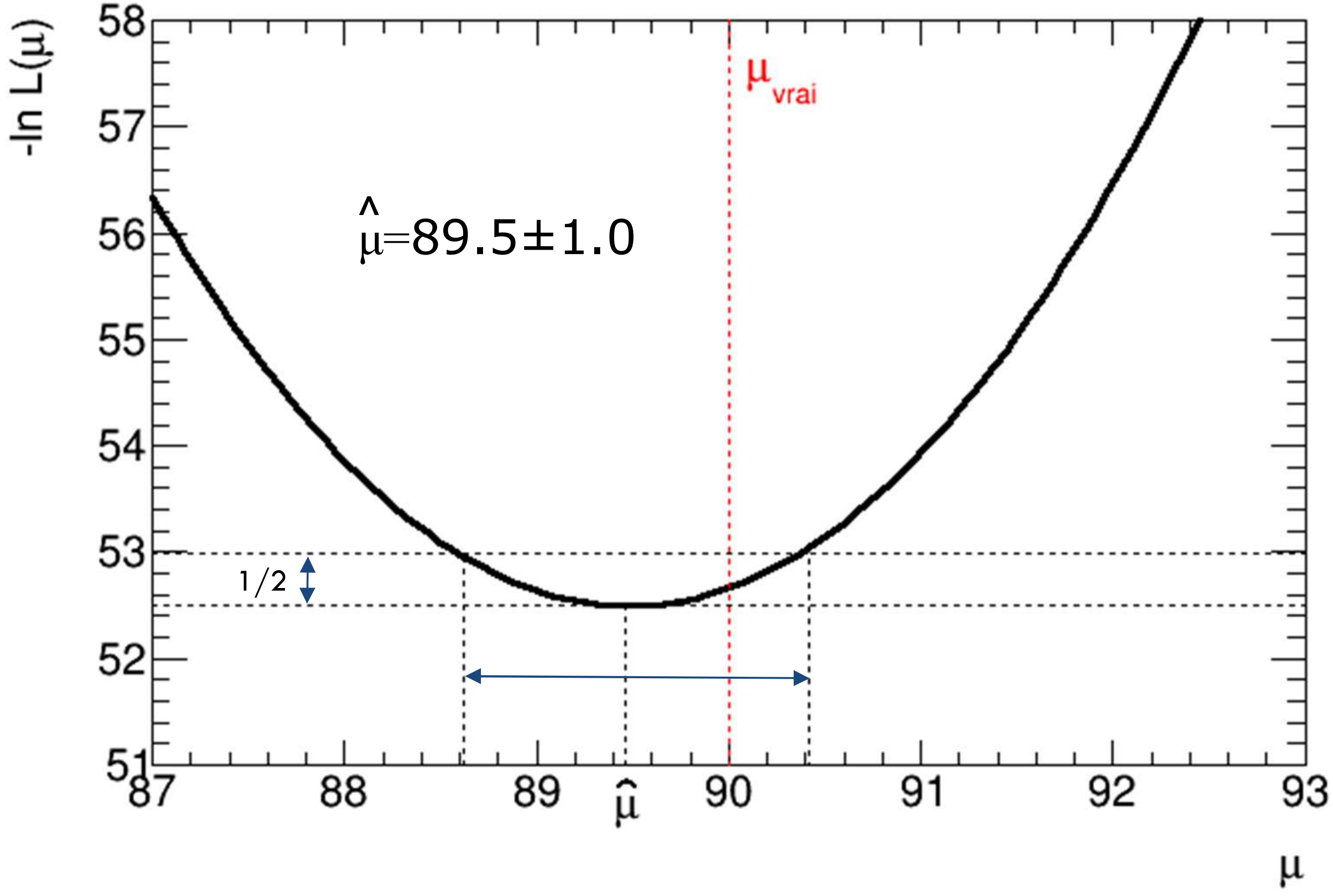
$$-\ln L(\mu,\sigma)=-N\ln\frac{1}{\sqrt{2\pi\sigma^2}}+\underbrace{\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\sigma^2}}_{2.\chi^2}$$

The minimization of –lnL(μ,σ) gives:

$$\left[\begin{array}{l}\dfrac{\partial-\ln L(\mu,\sigma)}{\partial\mu}=0\\[2em]\dfrac{\partial-\ln L(\mu,\sigma)}{\partial\sigma}=0\end{array}\right.\Longrightarrow\left[\begin{array}{l}\boxed{\hat{\mu}=\dfrac{1}{n}\sum_{i=0}^{n}x_i}\\[2em]\boxed{\hat{\sigma}=\sqrt{\dfrac{1}{n}\sum_{i=0}^{n}\left(x_i-\hat{\mu}\right)^2}}\end{array}\right.$$

MLE: the Gaussian example

Assuming σ known :

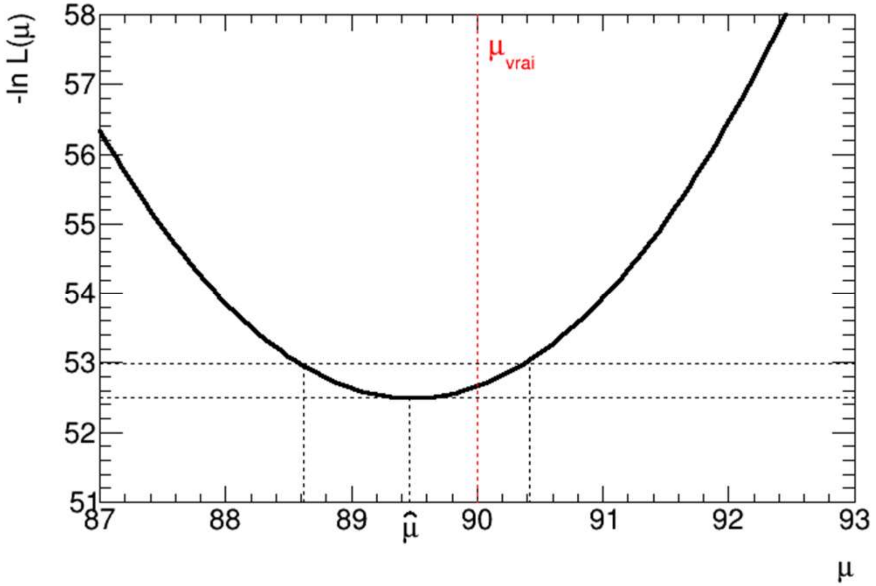


If the likelihood is Gaussian (true in the for large N), one can estimate the 1σ confidence interval for θ ("parameter uncertainty") by finding intersections $-\boldsymbol{\Delta}\mathbf{ln}\ \boldsymbol{L} = 1/2$ around minimum

If we repeat the experiment many times, [$\hat{\mu}$-σ,$\hat{\mu}$+σ] will contain the true value 68% of the time

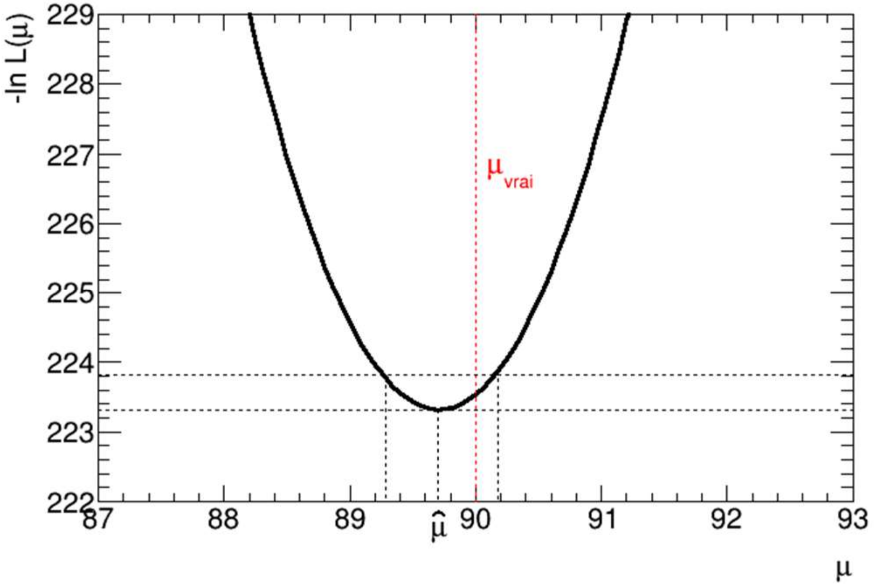
MLE: the Gaussian example

N=20



$\hat{\mu}$=89.5±1.0

N=80

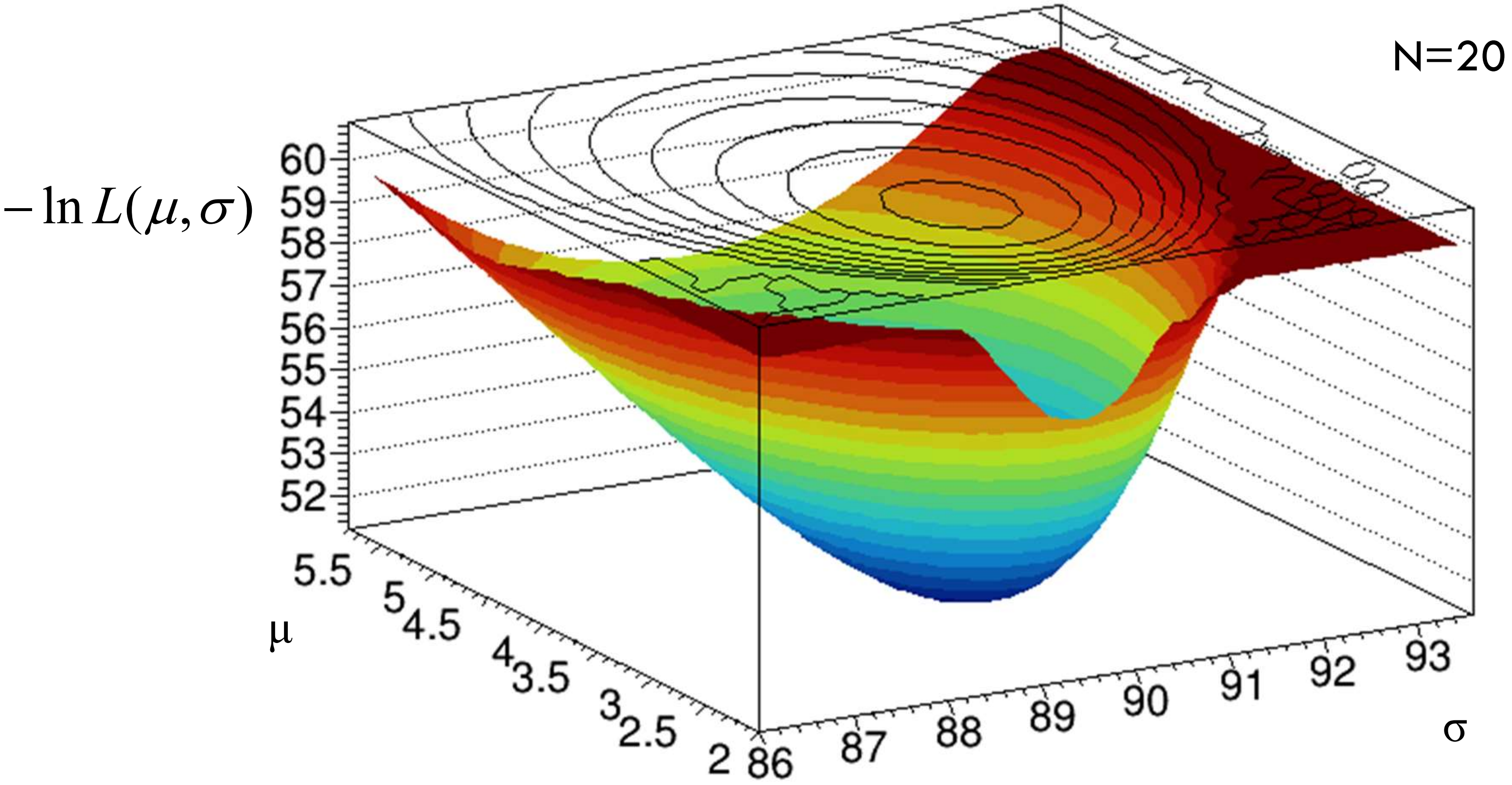


$\hat{\mu}$=89.7±0.5

Uncertainty decreases as $1/\sqrt{N}$

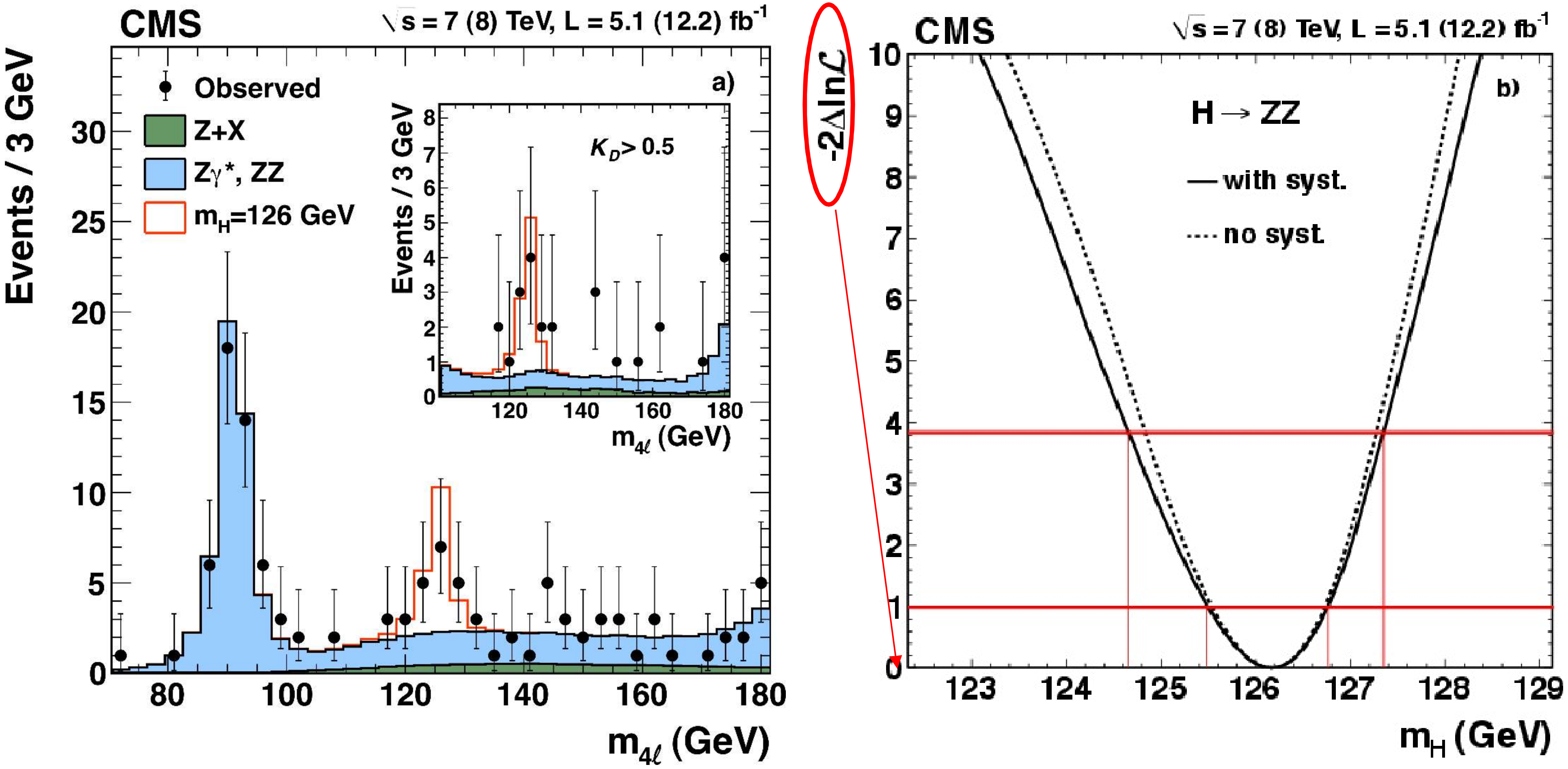
MLE: the Gaussian example

σ and μ unknown



In most of the realistic cases, the minimization is performed with numerical methods implemented as computer algorithms (ex: Minuit)

A real example



Modeling the data

which likelihood should I use?

Counting experiment

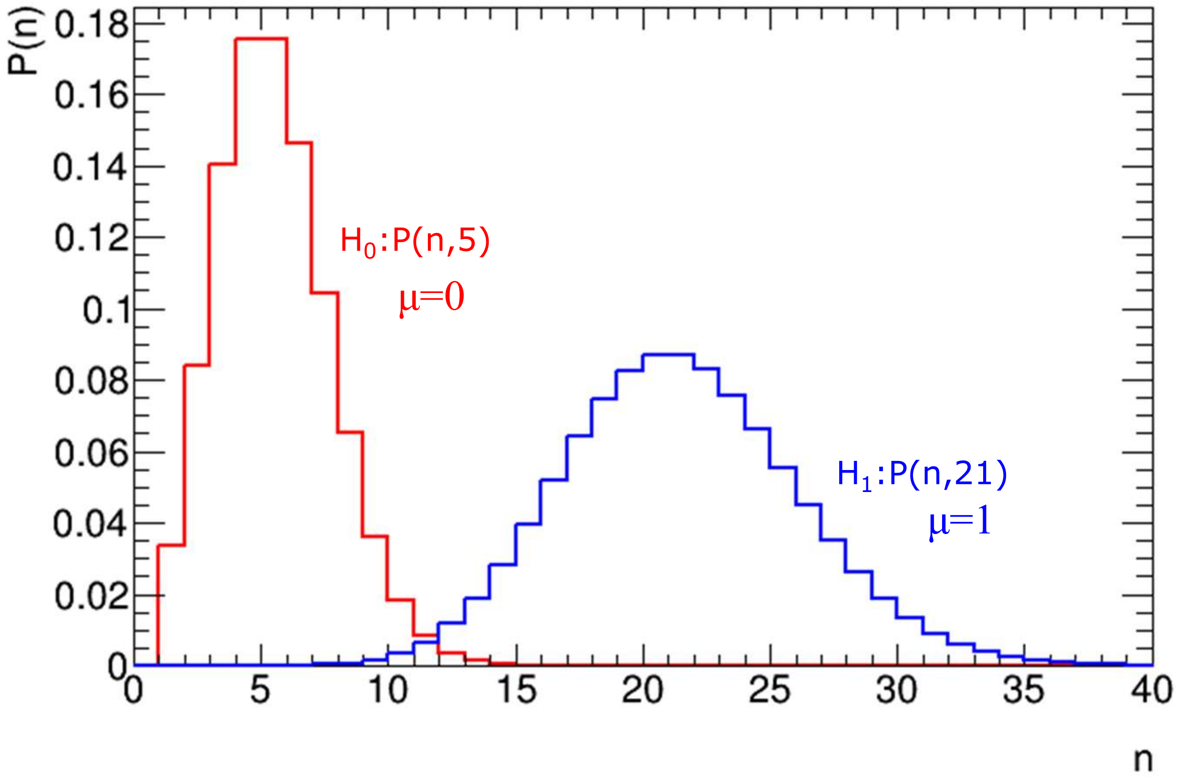
Observable: number of events (n)

PDF: Poisson distribution

$$P(n,\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$

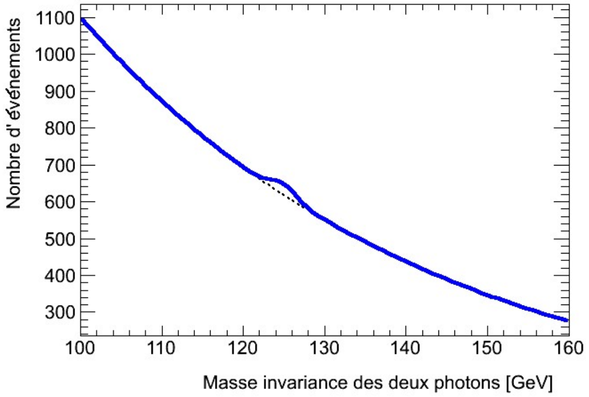
$\lambda = \mu s + b$

s is the expected nb of signal events
b is the expected nb of bkg events
μ is called signal strengh parameter



Shape analysis

Observable: set of values $m_1,\ldots,m_n$

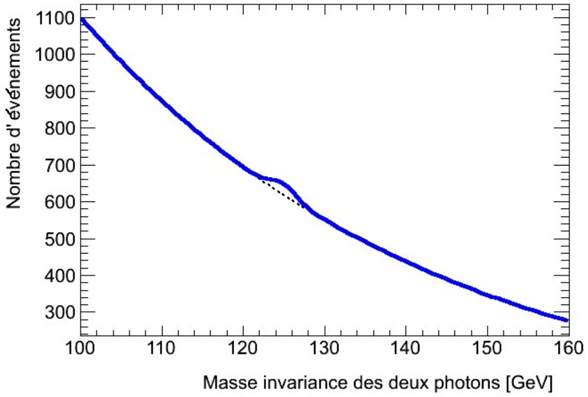


$$\underbrace{f(m_i \mid \mu)}_{\text{Probability to measure } m_i}$$

μ is called signal strengh parameter

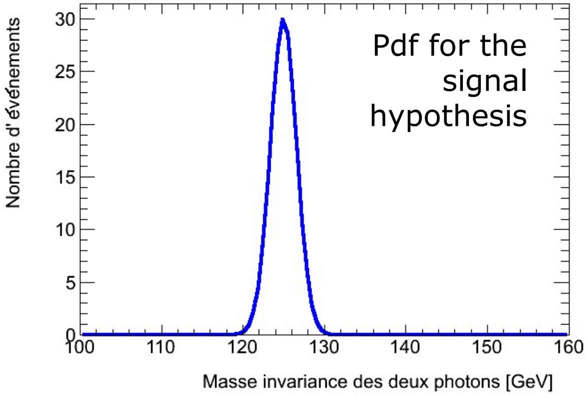
Shape analysis

Observable: set of values $m_1,\ldots,m_n$

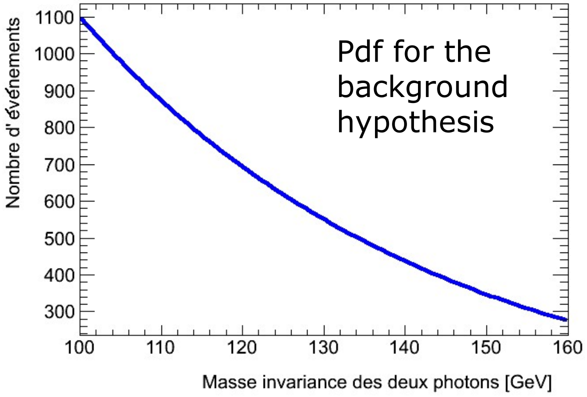


$$\underbrace{f(m_i\,|\,\mu)}_{\text{Probability to measure } m_i}$$

$$\frac{\mu s}{\mu s+b}f_s(m_i)+\frac{b}{\mu s+b}f_b(m_i)$$



Pdf for the signal hypothesis

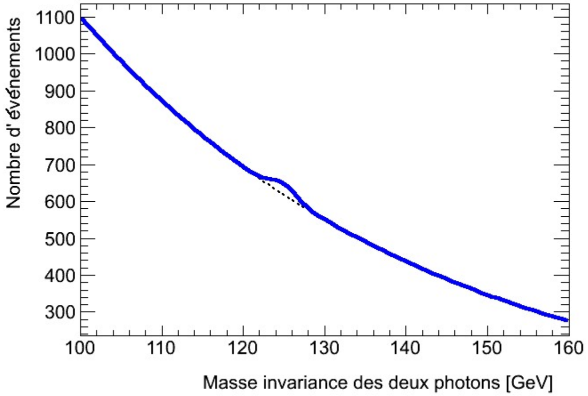


Pdf for the background hypothesis

μ is called signal strengh parameter

Shape analysis

Observable: set of values $m_1,\ldots,m_n$

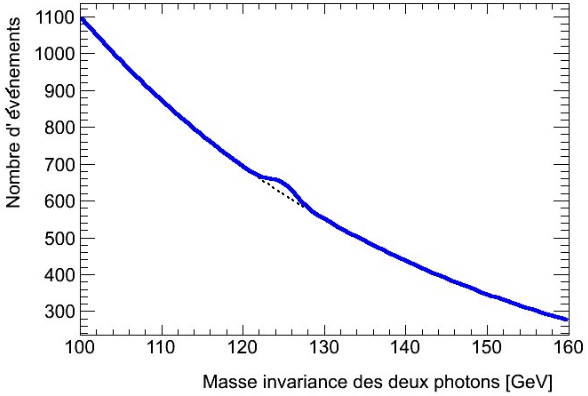


$$L(m_1,\ldots m_n \mid \mu) = \underbrace{\prod_{i=1}^{n} \underbrace{f(m_i \mid \mu)}_{\text{Probability to measure } m_i}}_{\text{Likelihood for independent and identically distributed data}}$$

μ is called signal strengh parameter

Shape analysis

Observable: set of values $m_1,\ldots,m_n$



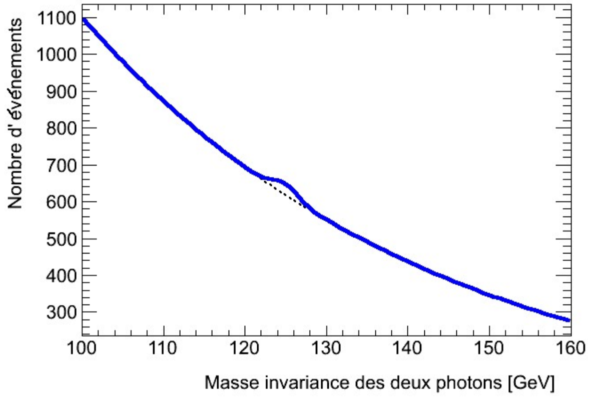
$$\underbrace{L(m_1,\ldots m_n \mid \mu) = \underbrace{P(n \mid \mu)}_{\text{Probability to observe n events}} . \prod_{i=1}^{n} \underbrace{f(m_i \mid \mu)}_{\text{Probability to measure } m_i}}_{\text{Likelihood for independent and identically distributed data}}$$

$$\frac{(b+\mu s)^n e^{-(b+\mu s)}}{n!}$$

μ is called signal strengh parameter

Shape analysis

Observable: set of values $m_1,\ldots,m_n$



$$L(m_1,\ldots m_n\,|\,\mu)=\underbrace{P(n\,|\,\mu)}_{\text{Probability to observe n events}}.\prod_{i=1}^{n}\underbrace{f(m_i\,|\,\mu)}_{\text{Probability to measure } m_i}$$

Using:

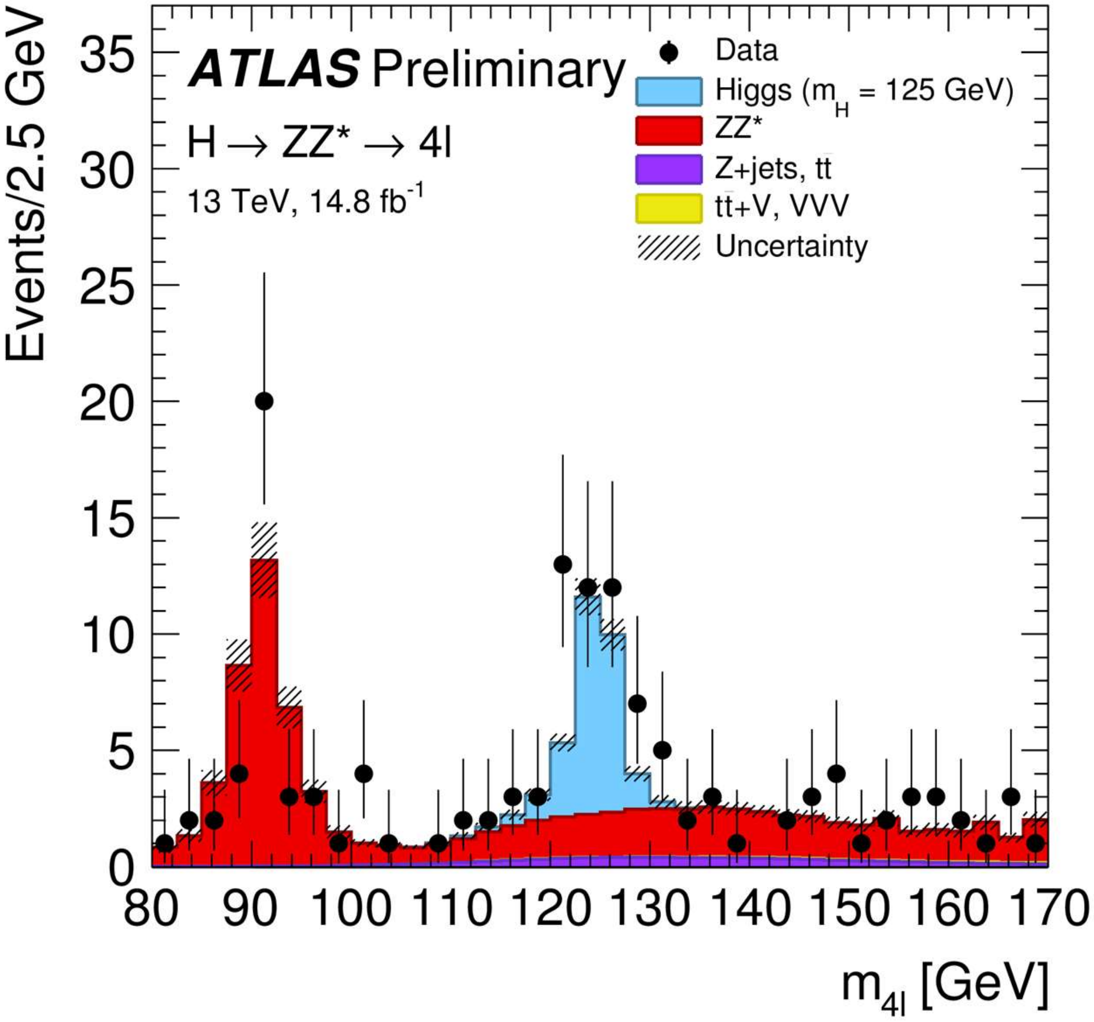
$$f(m_i|\mu)=\frac{\mu s}{\mu s+b}f_s(m_i)+\frac{b}{\mu s+b}f_b(m_i)$$

One get:

$$L(m_1,\ldots m_n\,|\,\mu)=\frac{e^{-(b+\mu s)}}{n!}\prod_{i=1}^{n}(\mu sf_s(m_i)+bf_b(m_i))$$

Binned shape analysis

Observable: number of events in the bins of an histogram



Note that in this case, there are several background components.

Binned shape analysis

Observable: number of events in the bins of an histogram

$$L(n_1,\dots,n_{Nbins}|\mu) = \prod_{i=1}^{Nbins} P(ni|\,\lambda_i\,) = \prod_{i=1}^{Nbins} P(n_i|\underset{\uparrow}{f_i^{\,b}}b + \underset{\nearrow}{f_i^{\,s}}\mu.s) = \prod_{i=1}^{Nbins} \frac{(f_i^{\,b}b + f_i^{\,s}\mu s)^n e^{-(f_i^{\,b}b+f_i^{\,s}\mu.s)}}{n_i!}$$

Per-bin fractions (shape) of signal and background

Nbins=1: counting analysis

Nbins=∞: unbinned shape analysis (the fraction becomes pdf values)

Faster to work with binned likelihood compared to unbinned likelihood

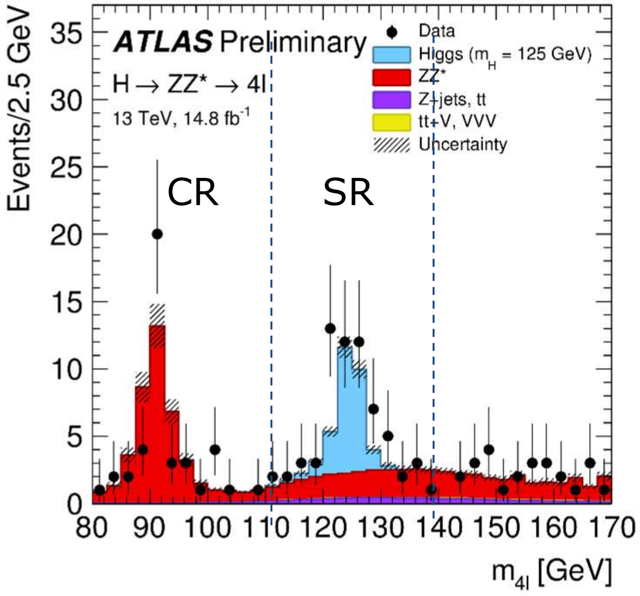
Introducing nuisance parameters

1) The background can be constrained by the data using a control region where the number of events is noted m

$$L(n,m|\mu,b) = \underbrace{P(n|b+\mu s)}_{\text{SR}}.\underbrace{P(m|b_{cr})}_{\text{CR}} = \frac{(b+\mu s)^n e^{-(b+\mu s)}}{n!}.\frac{(\tau b)^m e^{-\tau b}}{m!} \qquad \tau = b/b_{CR}$$

Here b is treated as a nuisance parameter. If $b_{CR}=\tau b \neq m_{meas}$, need to adjust b to maximize the likelihood.

In general, there should also be also an uncertainty on τ which is in general relatively smaller than the uncertainties on b and b_{cr}



s is the expected nb of signal events in SR

b is the expected nb of bkg events in SR

b_{CR} is the expected nb of bkg events in CR

μ is called signal strengh parameter

Introducing nuisance parameters

1) The background can be constrained by the data using a control region where the number of events is noted m

$$L(n,m|\mu,b) = \underbrace{P(n|b+\mu s)}_{\text{SR}}.\underbrace{P(m|b_{cr})}_{\text{CR}} = \frac{(b+\mu s)^n e^{-(b+\mu s)}}{n!}.\frac{(\tau b)^m e^{-\tau b}}{m!} \qquad \tau = b/b_{CR}$$

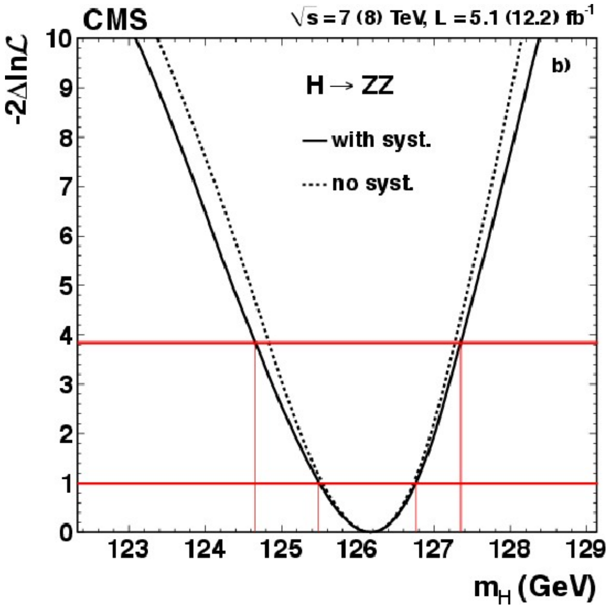
Here b is treated as a **nuisance parameter.** If $\tau b \neq m_{meas}$, need to adjust b to maximize the likelihood.

In general, there should also be also an uncertainty on τ which is in general relatively smaller than the uncertainties on b and b_{cr}

2) Counting experiment with systematic uncertainty on b (ex: uncertainty on the bkg cross-section):

$$L(n\,|\,\mu,\theta) = \frac{(\theta b+\mu s)^n e^{-(\theta b+\mu s)}}{n!}.Gaus(\theta;1,\sigma_\theta)$$

where θ is a **nuisance parameter** constrained to $\theta=1$ within σ_θ by a Gaussian PDF (penalty for $\theta \neq 1$)



Nuisance parameters

More generally, we write the likelihood as

$$L(\mu,\boldsymbol{\theta}) = L_{meas}(\mu,\boldsymbol{\theta}).\underset{\substack{\uparrow \\ \text{NP constraint term} \\ \Rightarrow \text{ penalty for } \theta \neq \theta_{nominal}}}{C}(\boldsymbol{\theta})$$

μ is a **parameter of interest**. In some cases, there can be several (signal strengh parameter, mass,…)

θ represent the **nuisance parameters** needed to define the model (ex: syst. uncertainties)

NPs must be either

⟶ known a priori (possibly within systematics)

⟶ constrained by the data (e.g. in sidebands)

Combining analyses

The combined likelihood is obtained by multiplying the likelihood functions of individual channels in order to

$$L(\mu,\boldsymbol{\theta}) = \prod_{i=1}^{Nanalysis} L_i(\mu,\boldsymbol{\theta}_i)$$

The main challenge is to properly deal with the correlation of the nuisance parameters

- Ex: luminosity is fully correlated between analysis, theory uncertainties could be very tricky

Combination can be done within one experiment or between different experiments

Hypothesis testing: discovery case

Hypothesis testing

A key task in most of the experiments is to discriminate between two hypothesis on the basis of the observed experimental data ($\vec{x}$)

- H_0, null hypothesis that we want to disprove (eg, SM background only)
- H_1, alternative hypothesis (eg, SM background + new physics)

The goal of a hypothesis test is to determine whether the observed data sample better agrees with H_0 or rather with H_1

Test statistics: a scalar variable (called t($\vec{x}$)) computed from the data that discriminates between the two hypotheses H_0 and H_1. Usually a 'summary' of the information available in the sample

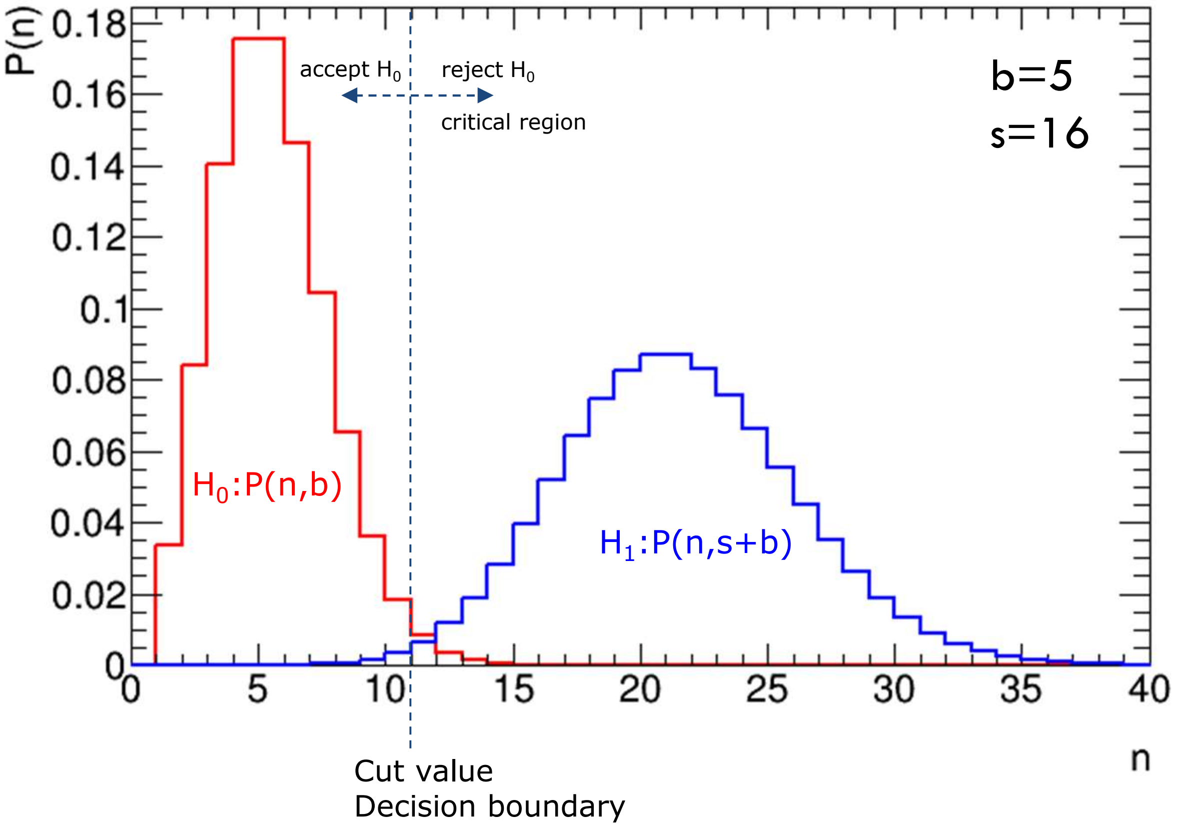
A simple example: counting experiment

Observable: number of events (n)

PDF: Poisson distribution

$$P(n,\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$

Test statistics: number of events (t(n)=n)



Hypothesis testing

Significance $\boldsymbol{\alpha}$: Type-1 error rate:

$$\alpha = \int_{y(x) > \text{cut}} P(x|H_0)dx$$

should be small

α is the probability to reject the null hypothesis when it is true

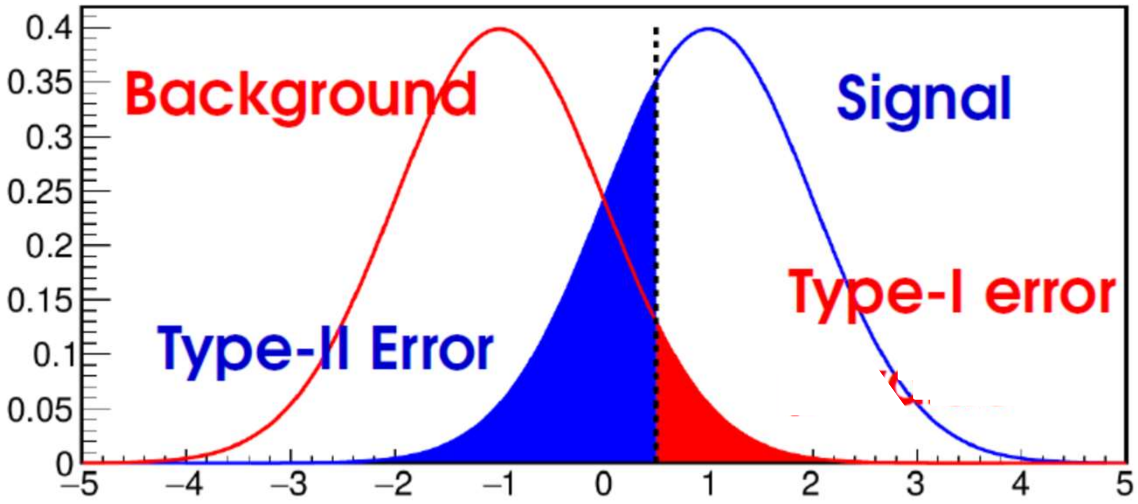
Size $\boldsymbol{\beta}$: Type-2 error rate:

$$\beta = \int_{y(x) < \text{cut}} P(x|H_1)dx$$

should be small

β is the probability to accept the null hypothesis when the alternative is true

e”



Hypothesis testing

Significance $\boldsymbol{\alpha}$: Type-1 error rate:

$$\alpha = \int_{y(x)>\text{cut}} P(x|H_0)dx$$

should be small

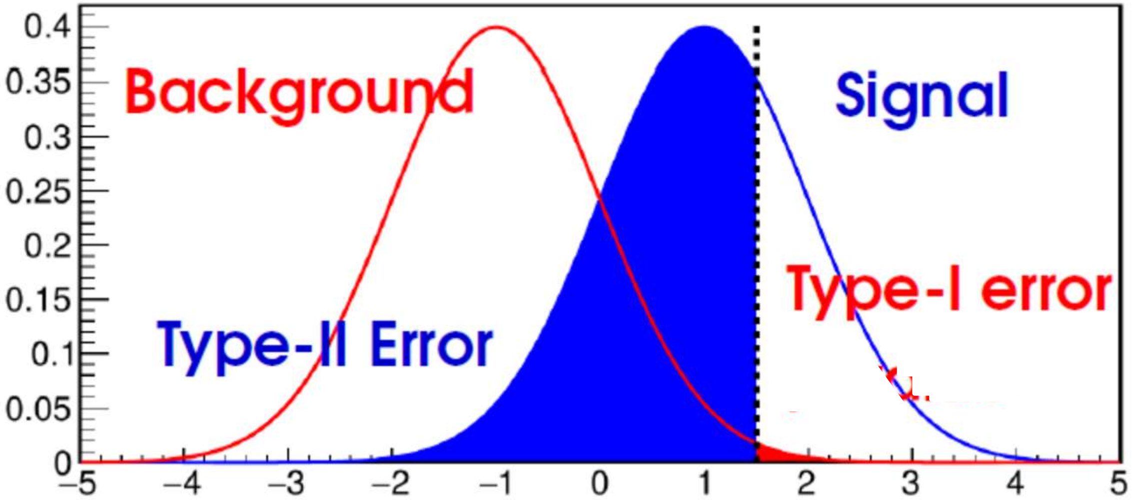
α is the probability to reject the null hypothesis when it is true

Size $\boldsymbol{\beta}$: Type-2 error rate:

$$\beta = \int_{y(x)<\text{cut}} P(x|H_1)dx$$

should be small

β is the probability to accept the null hypothesis when the alternative is true



Difficult to minimize the two at the same time!

Neyman-Pearson lemma

When comparing two simple hypotheses H0 and H1, the optimal discriminator is the Likelihood ratio (LR):

$$t(\vec{x}) = \frac{L(\vec{x} \mid H_1)}{L(\vec{x} \mid H_0)}$$

It minimizes Type-II uncertainties (β) for a given level of Type-I uncertainties (α)

Any monotonic function of the likelihood ratio is also optimal (ex: $q(\vec{x})=-2 \ln t(\vec{x})$)

Caveat: Neyman-Pearson Lemma holds strictly only for simple hypotheses without free parameters (ex: Higgs boson search, the mass is a free parameter)

However: the likelihood ratio is a very convenient test statistic (*probably* close to optimal) and therefore commonly used in experimental particle physics

Different versions of the likelihood ratio are used in statistical tests

Procedure

Specify the null hypothesis that you want to disprove and the alternate hypothesis

- Ex for discovery: H_0=SM background only, H_1=BSM

Build you test statistic: t(x) using for instance the Neyman-Person lemma

- Ex: counting experiment → number of events (demonstration later)

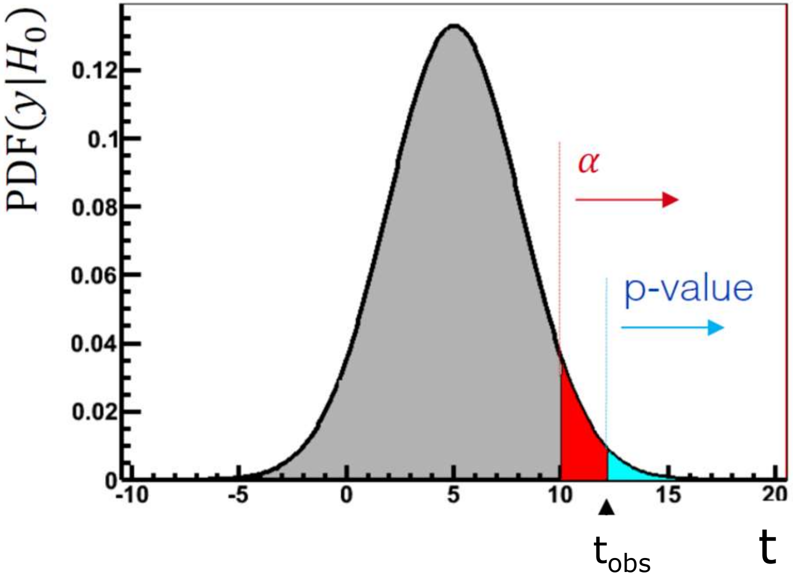
Specify the significance α of the test (how likely you are willing to claim a false discovery)

- Set to $2.9.10^{-7}$ (5σ) for the discovery or 0.05 for exclusion

Take the measurement: t_{obs}

Check whether t_{obs} lies inside or outside of critical region → decide on H_0

Compute p-value of H_0 to see how deep it lies in the critical region



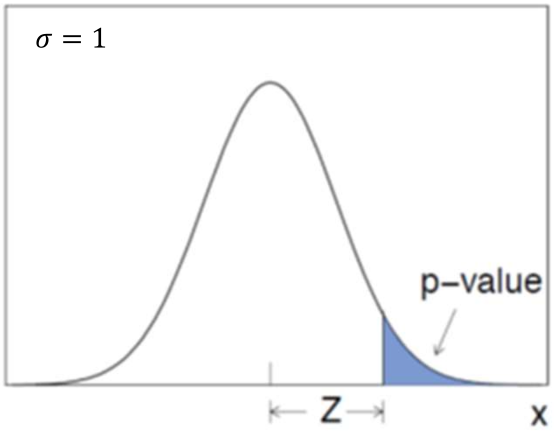
$$p-value=\int_{t_{obs}}^{\infty}pdf(t\,|\,H_0)$$

p-value : fraction of outcomes that are at least as signal-like (H1-like) as data, when H_0 is assumed to be true (no signal present).

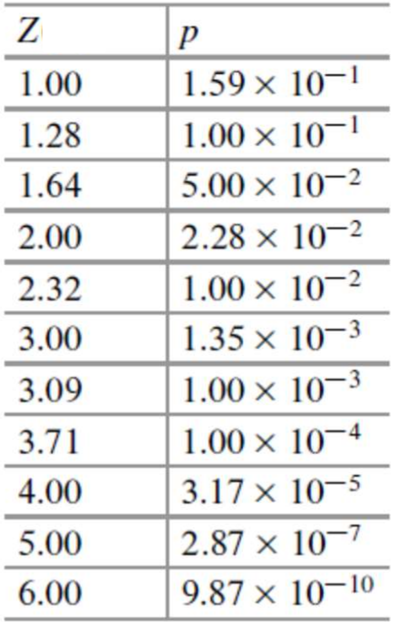
Significance and p-values

It is convenient to express the observed p-values in terms of a Gaussian σ

arXiv:1007.1727



$$p = \int_Z^{\infty} G(x; 0,1)dx = 1 - \Phi(Z) = \Phi(-Z)$$



Z	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}

with $$\Phi(x) = \int_{-\infty}^{x} G(x'; 0,1)d\,x'$$

Gaussian cumulative distribution function

$$Z = \Phi^{-1}(1-p)$$

Application to counting experiments

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

In this case, the likelihood ratio is (using the Neyman-Person lemma):

$$t(n) = \frac{L(n|H_1)}{L(n|H_0)} = \frac{L(n|\mu=1)}{L(n|\mu=0)} = \frac{\frac{(b+s)^n e^{-(b+s)}}{n!}}{\frac{b^n e^{-b}}{n!}} = (1+\frac{s}{b})^n e^{-s}$$

where μ is the signal strength parameter (proportional to the cross section for the signal process whose existence is not yet established)

And the negative log likelihood (NLL) ratio is

$$q(n) = -\ln t(n) = s - n\ln(1+\frac{s}{b})$$

Since t(n), q(n) and n are monotonic, they conveys the same level of information → can use n as a test statistics

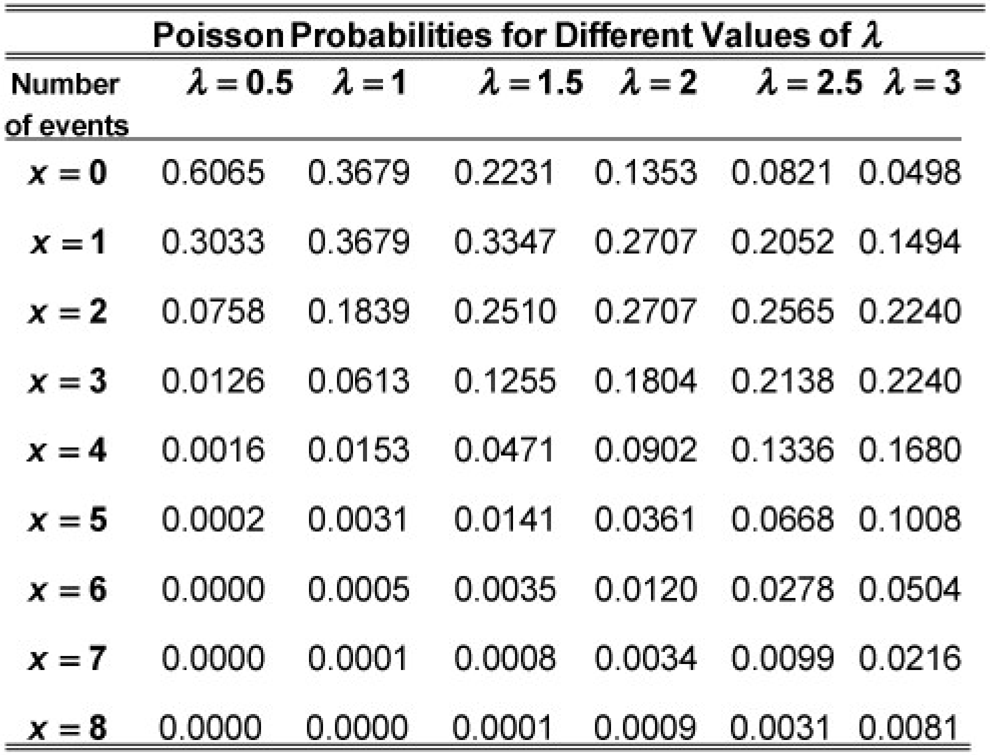
Exercise 1

Counting experiment with 1.5 expected background events

7 events are observed in the data

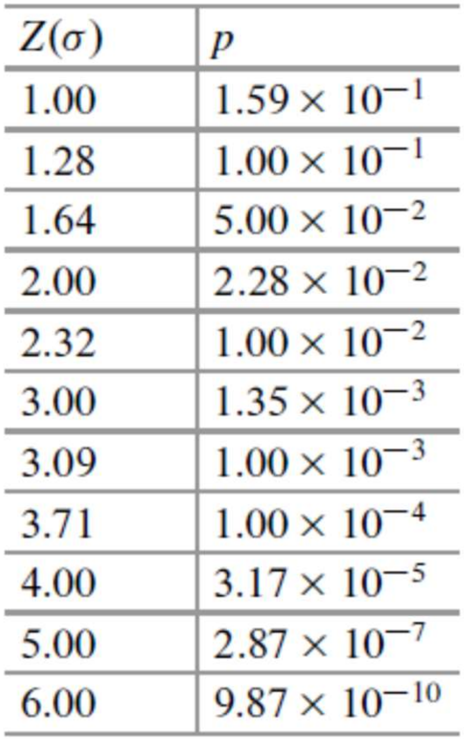
What is the corresponding p-value?

Is it a discovery, an evidence or nothing?



Poisson Probabilities for Different Values of λ

Number of events	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$
$x = 0$	0.6065	0.3679	0.2231	0.1353	0.0821	0.0498
$x = 1$	0.3033	0.3679	0.3347	0.2707	0.2052	0.1494
$x = 2$	0.0758	0.1839	0.2510	0.2707	0.2565	0.2240
$x = 3$	0.0126	0.0613	0.1255	0.1804	0.2138	0.2240
$x = 4$	0.0016	0.0153	0.0471	0.0902	0.1336	0.1680
$x = 5$	0.0002	0.0031	0.0141	0.0361	0.0668	0.1008
$x = 6$	0.0000	0.0005	0.0035	0.0120	0.0278	0.0504
$x = 7$	0.0000	0.0001	0.0008	0.0034	0.0099	0.0216
$x = 8$	0.0000	0.0000	0.0001	0.0009	0.0031	0.0081



$Z(\sigma)$	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}

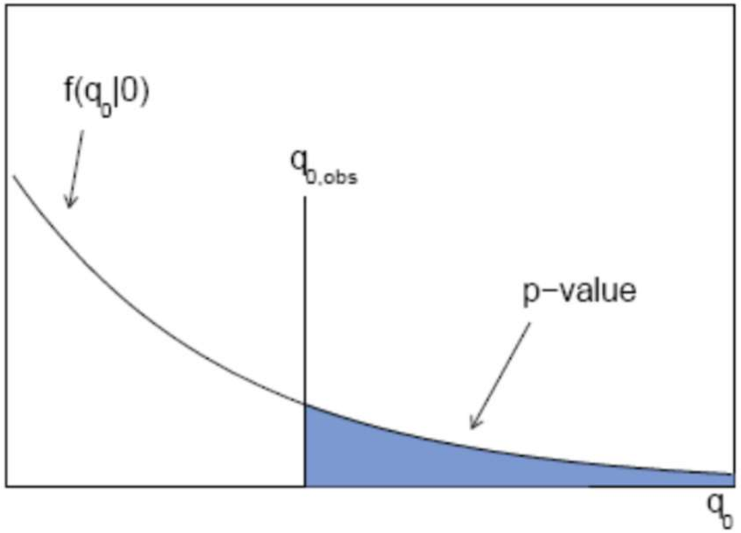
Application to shape analyses

The likelihood ratio is given by:

$$t(m_1,\ldots m_n) = \frac{L(m_1,\ldots m_n|\mu = 1)}{L(m_1,\ldots m_n|\mu = 0)} = e^{-s}\prod_{i=1}^{n}(\frac{sf_s(m_i)}{bf_b(m_i)} + 1)$$

and the negative log likelihood (NLL) ratio is

$$q(m_1,...m_n) = -\ln(t(m_1,...m_n) = s - \sum_{i=1}^{n}(\frac{sf_s(m_i)}{bf_b(m_i)} + 1)$$



q

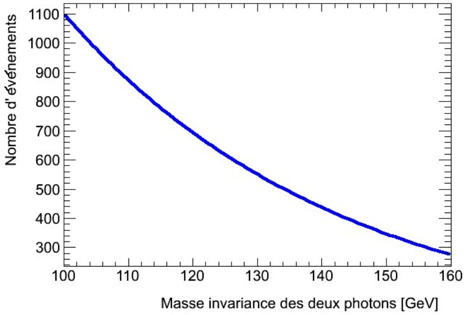
The critical region lies on the higher side of the q distribution

The p-value (blue area) can be computed as follow:

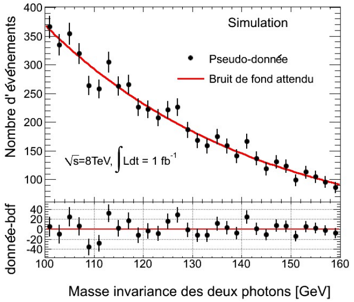
$$p - value = \int_{q_{obs}}^{+\infty} pdf(q\,|\,H_0)$$

Toy MC simulation to compute p-value

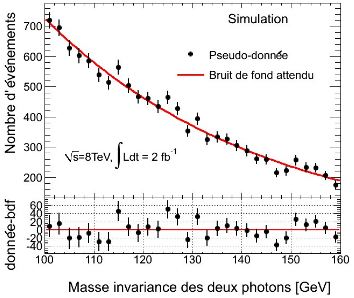
Generate *pseudo data* (toys) using the PDF under the tested hypothesis



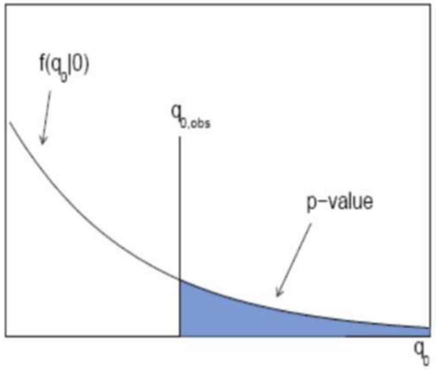
Toy
simulation



……



Compute q for
each toy
simulation



Compute the p-value as the fraction of toy events giving a value larger than q_{obs}

Precision limited by the number of toys events

- Small p-values (5σ : $p \sim 10^{-7}$) → Need a very large number of toys

Analytical computation is preferred when available and fortunately there is a solution…

Profile likelihood ratio

In the presence of nuisance parameters, one used the **profile likelihood ratio** (PLR) as the test statistics instead of the likelihood ratio (LR):

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}$$

Conditional maximum given the signal strength parameter value μ

maximize *L*

By definition, $\lambda(\mu)$ lies between 0 and 1

Higher values indicating greater compatibility between the data and the hypothesized value of μ

Good properties in the large sample limit allowing **analytical computation**

Test statistics for discovery

Try to reject background-only (μ = 0) hypothesis using

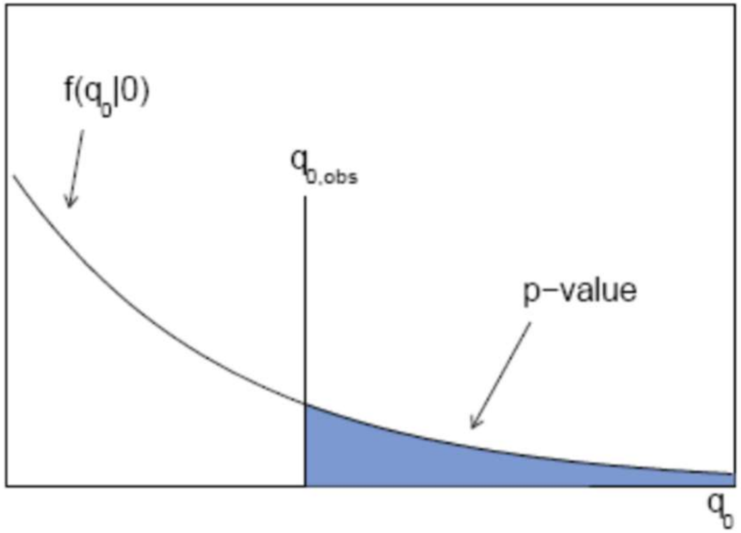
$$q_0 = \begin{cases} -2\ln\lambda(0) \\ \end{cases}$$

$$\lambda(0) = \frac{L(0,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}$$

Physically μ ≥ 0 (*) but $\hat{\mu}$ could be negative due to downward background fluctuation

q_0 increases when $\hat{\mu}$ deviates from 0

But we don’t want to reject the background only hypothesis if $\hat{\mu}<0$



More background-like ←

(*) unless negative interference

Test statistics for discovery

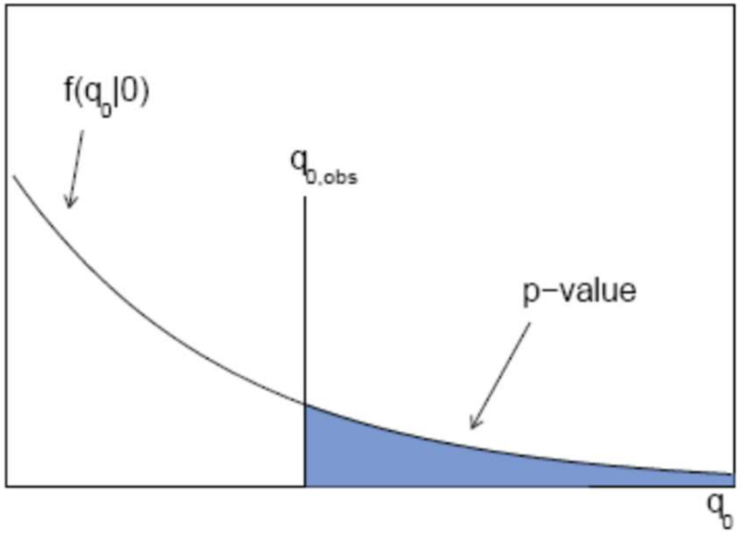
Try to reject background-only ($\mu = 0$) hypothesis using

$$\lambda(\mu) = \frac{L(0,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}$$

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

Less events than predicted bkg

i.e. here only regard upward fluctuation of data as evidence against the background-only hypothesis. This is a "one-sided" definition (only claim signal for $\hat{\mu}$>0)



More background-like ←

Test statistics for discovery

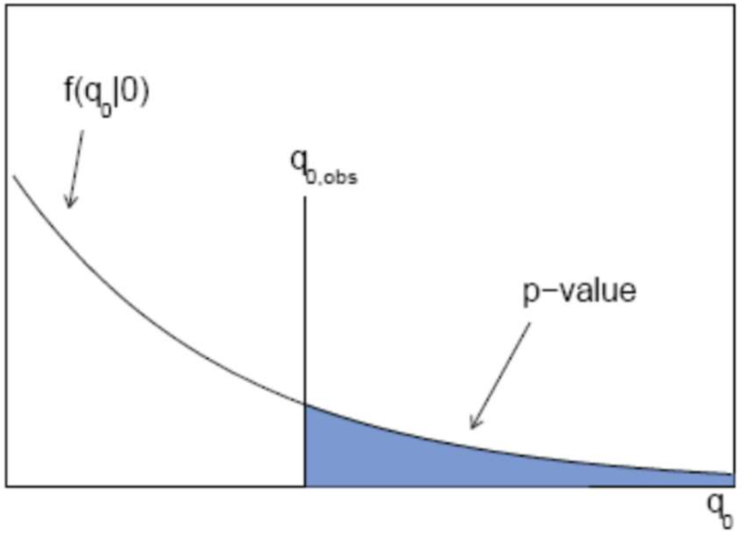
Try to reject background-only ($\mu = 0$) hypothesis using

$$\lambda(\mu) = \frac{L(0, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}$$

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

Less events than predicted bkg

i.e. here only regard upward fluctuation of data as evidence against the background-only hypothesis. This is a "one-sided" definition (only claim signal for $\hat{\mu}>0$)



More background-like ←

$$p_0 = \int_{q_{0,\mathrm{obs}}}^{\infty} f(q_0|0)\, dq_0$$

In the large sample (asymptotic) limit, one has this simple relation:

$$Z = \Phi^{-1}(1 - p_0) = \sqrt{q_0}$$

↑

Inverse gaussian cumulative distribution function

Test statistics for discovery

In the large sample (asymptotic) limit, one has this simple relation when $\hat{\mu} > 0$:

$$Z = \sqrt{-q_0} = \sqrt{-2\ln\lambda(0)} = \sqrt{-2\ln\frac{L(0,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}}$$

Example for counting experiment

Poisson likelihood with p.o.i μ (n events observed):

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

No nuisance parameter

$$Z = \sqrt{-2\ln\frac{L(0)}{L(\hat{\mu})}} = \sqrt{-2\ln\frac{(b)^n e^{-(b)}}{(\hat{\mu}s + b)^n e^{-(\hat{\mu}s + b)}}} = \sqrt{2(n.\ln(\hat{\mu}s / b + 1) - \hat{\mu}s)}$$

using $n = \hat{\mu}s + b$

$\hat{\mu}$ maximized the likelihood

$$Z = \sqrt{2\left(n.\ln\left(\frac{n}{b}\right) + b - n\right)}$$

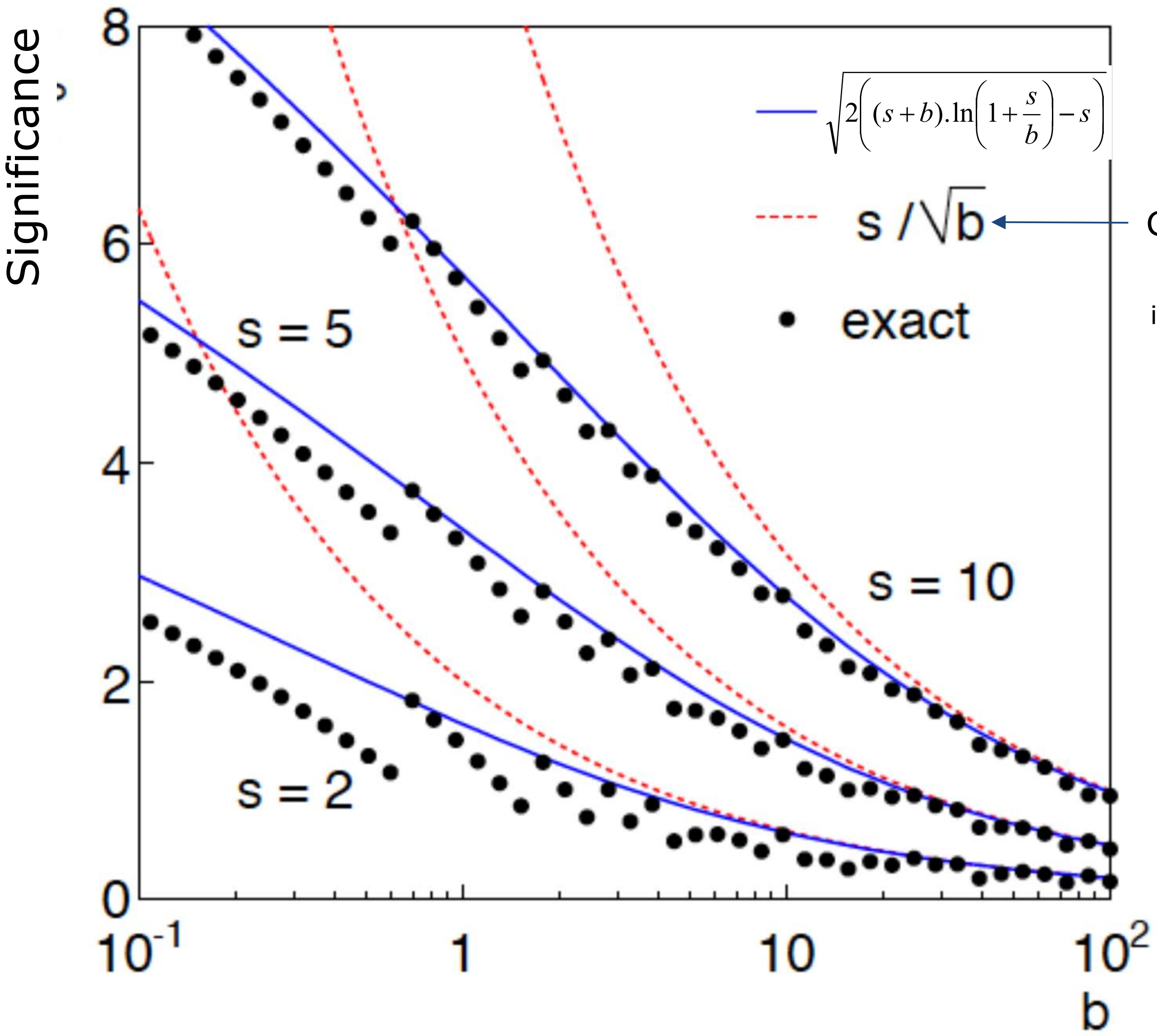
Only correct for n>b
Z=0 otherwise

Or alternatively

assuming $\hat{\mu}$ =1

$$Z = \sqrt{2\left((s + b).\ln\left(1 + \frac{s}{b}\right) - s\right)}$$

Example for counting experiment

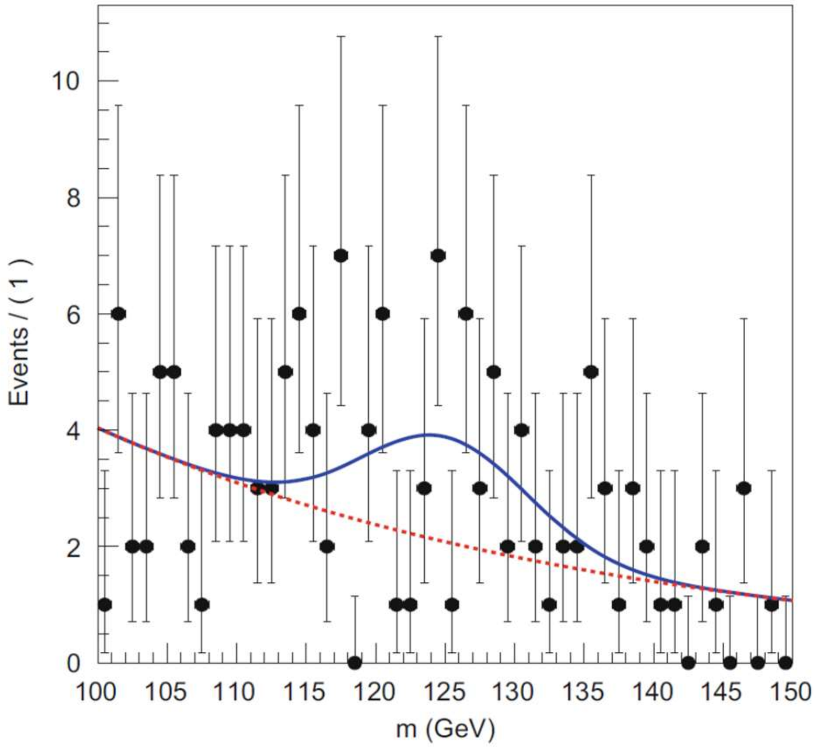


Gaussian approx.

The significance increase as the square root of the luminosity

Example for shape analysis

True parameters: b=100 and s=40



Assume background and signal shape perfectly known

Uncertainty on background normalisation modelled by θ

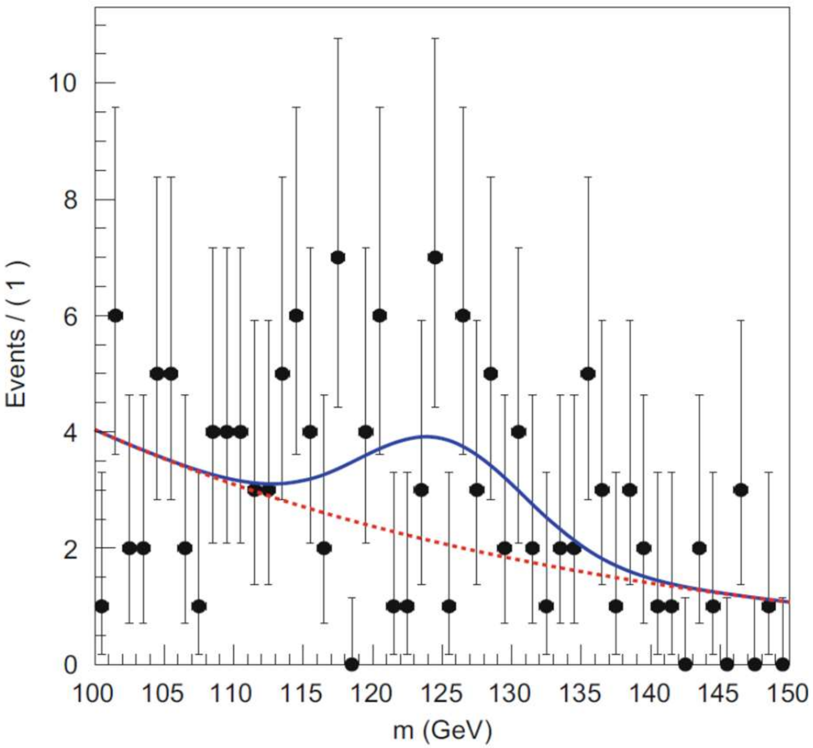
$$L(m_1,\ldots m_n \mid \mu,\theta) = \frac{e^{-(\theta b+\mu s)}}{n!}\prod_{i=1}^{n}(\mu.s.f_s(m_i)+\theta.b.f_b(m_i))G(\theta;1,\sigma_b)$$

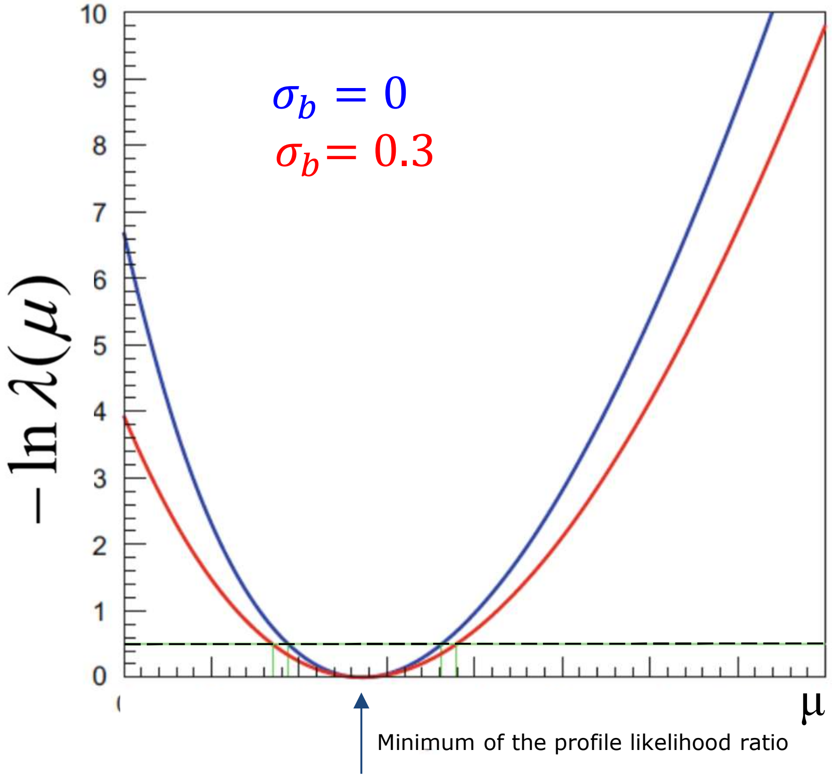
$$\lambda(0) = \frac{L(0,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}$$

The likelihood is minimized two times

Example for shape analysis

True parameters: b=100 and s=40





Minimum of the profile likelihood ratio

$\hat{\mu} = 0.67 \pm 0.2$

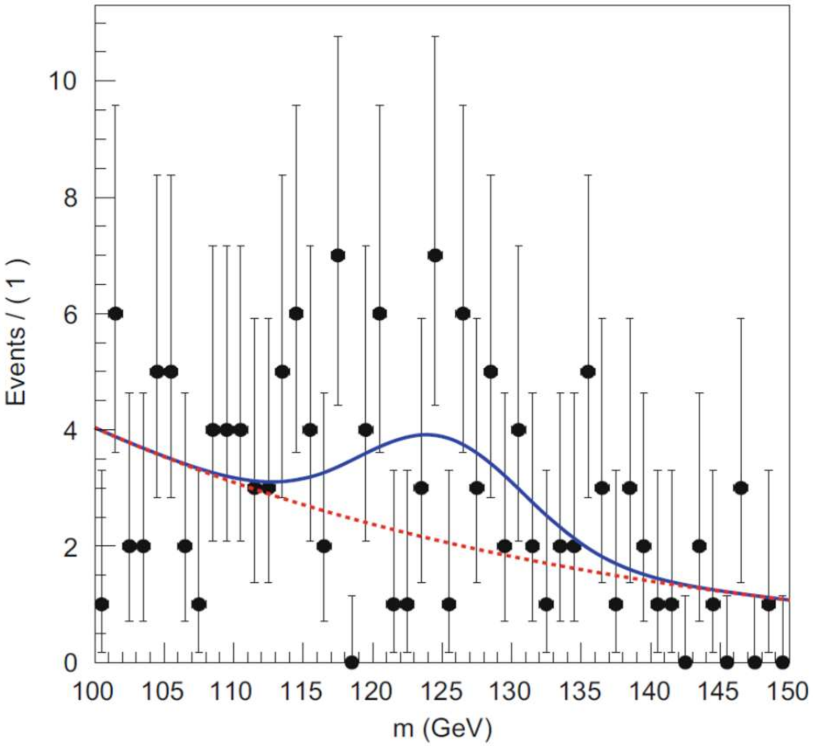
Excess observed!

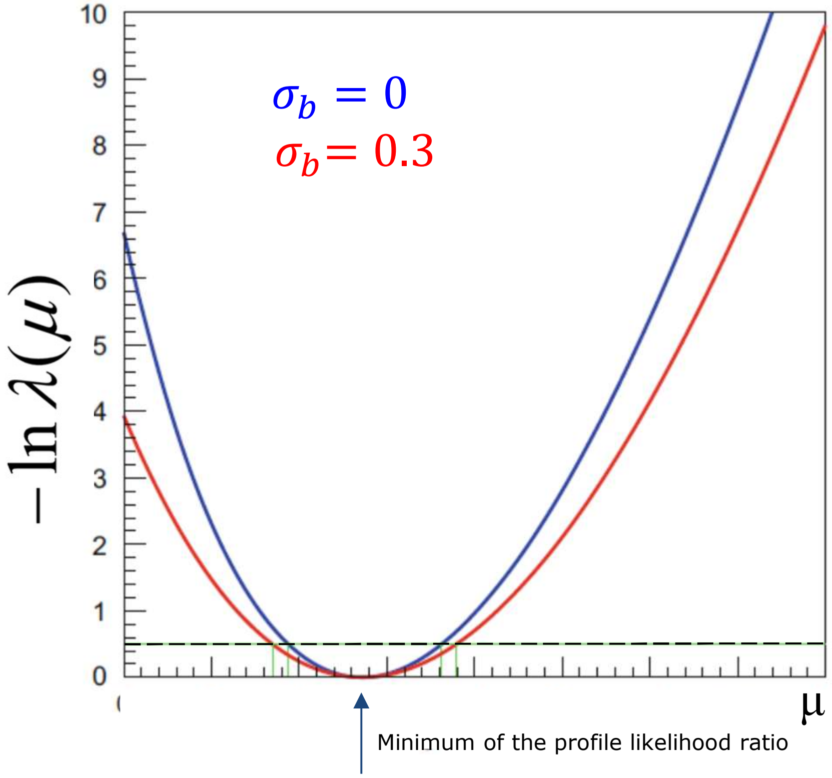
What is the significance for σ_b=0.3?

$$Z = \sqrt{-q_0} = \sqrt{-2\ln\lambda(0)} = \sqrt{-2\ln\frac{L(0,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}}$$

Example for shape analysis

True parameters: b=100 and s=40





Minimum of the profile likelihood ratio

$\hat{\mu} = 0.67 \pm 0.2$ Excess observed!

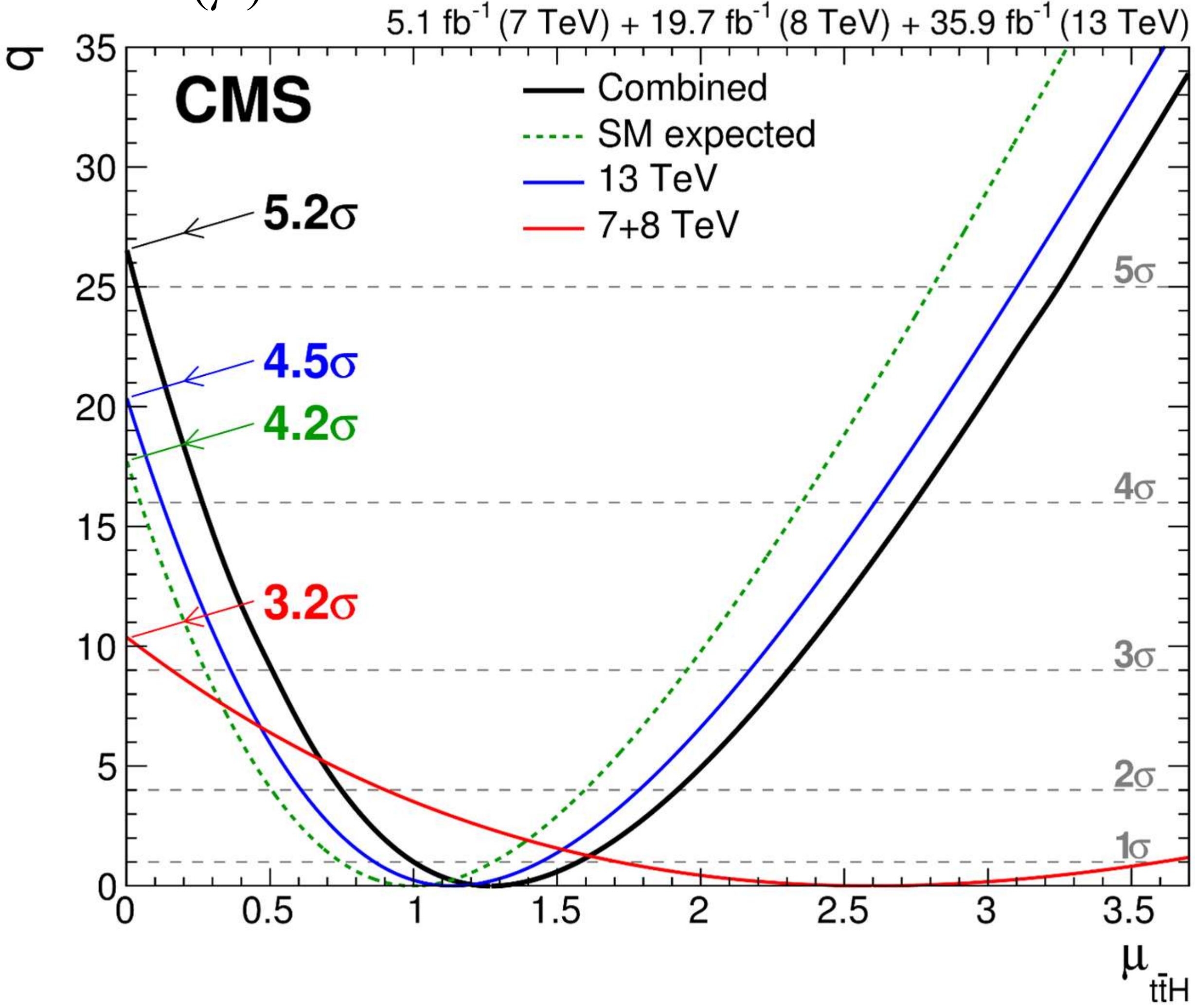
Profile likelihood broadened by nuisance parameters θ (loss of information)

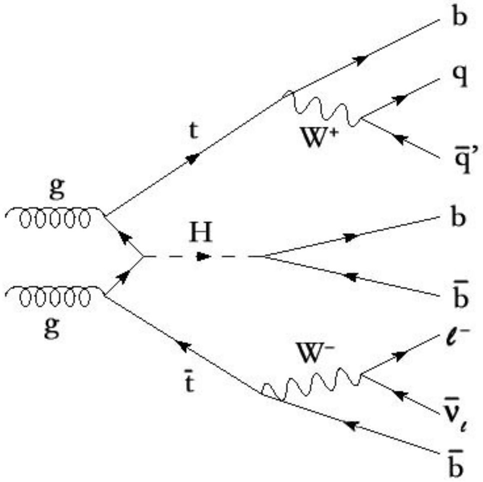
Since $Z = \sqrt{-2\ln(\lambda(0))}$, one could compute the significance directly from the right plot using the intercept of the curves Z=$\sqrt{(2*4)}$=2.8 and Z= $\sqrt{(2*6.6)}$=3.6 with and without uncertainties respectively

A realistic example

Observation of ttH production

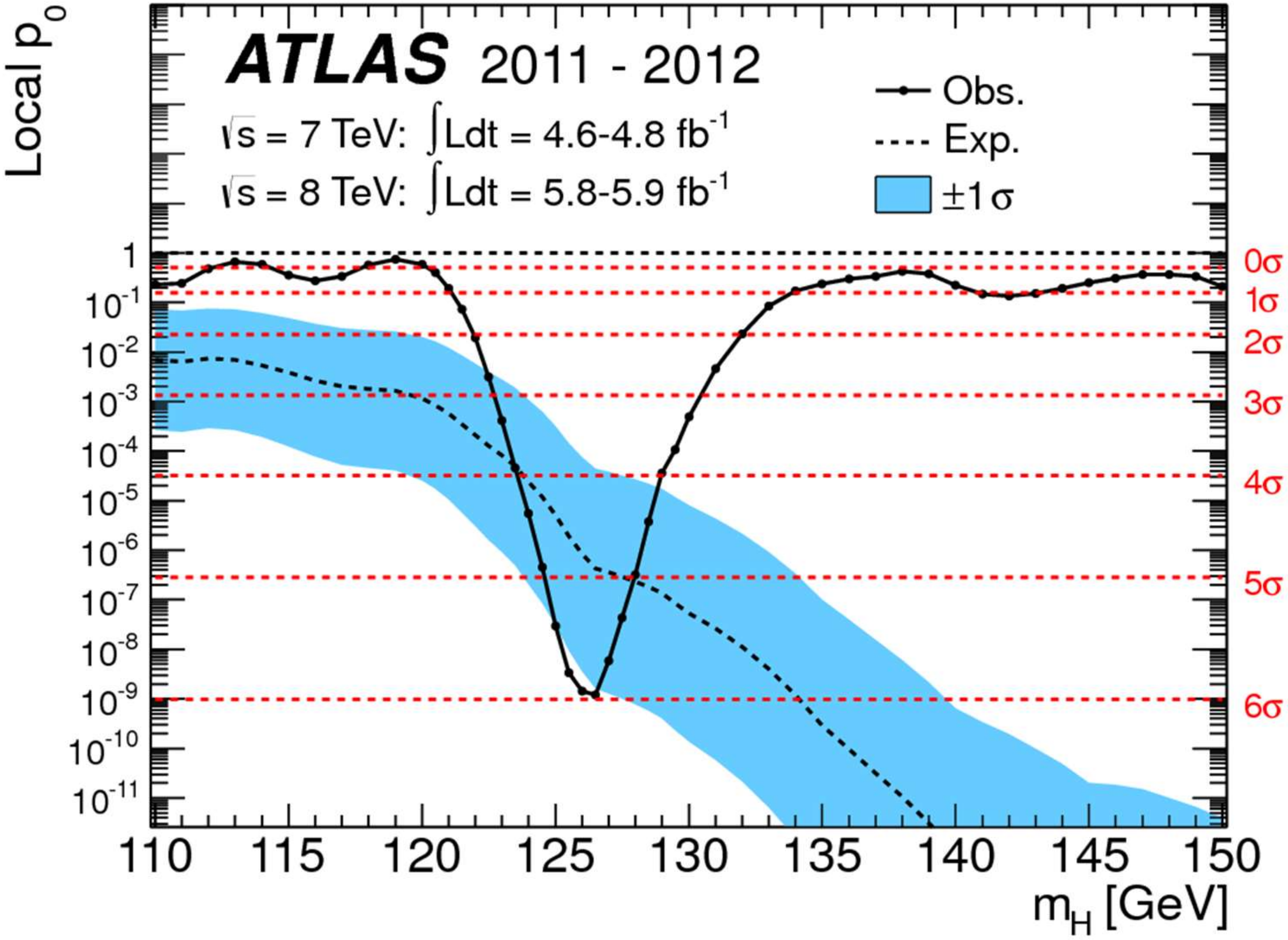
$q = -2 \ln \lambda(\mu)$





arXiv:1804.02610

The Higgs p_0 plot



The "local" p_0 means the p-value of the background-only hypothesis obtained from the test of $\mu = 0$ at each individual m_H.

Hypothesis testing: exclusion

Exclusion

Procedure similar to the discovery case except that the hypothesis are now inverted

- H_0= signal + background hypothesis
- H_1= background only hypothesis

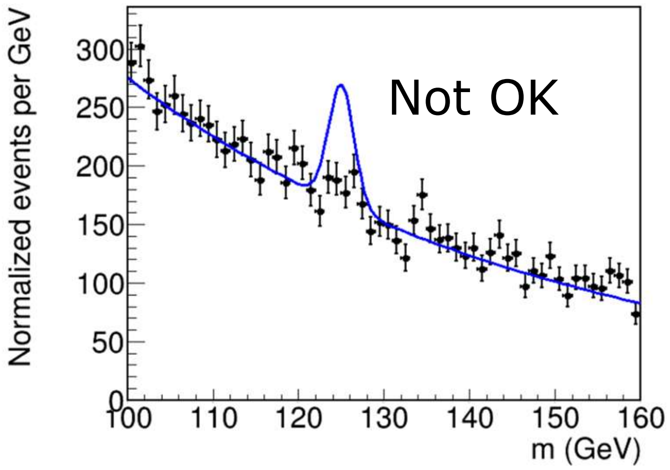
Goal: disprove $\boldsymbol{H_0}$ by estimating the probability of downward fluctuation of signal + background

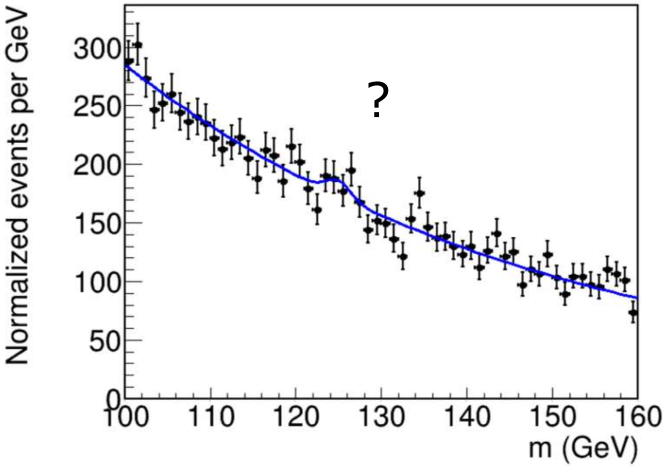
Size of the test less stringent than for the discovery case: α =5%

Confidence level of the test is 1-α =95% confidence level

Upper limit: find minimal signal, for which $\boldsymbol{H_0}$ can be excluded at specified confidence Level

- Smaller signal level satisfying p-value> α





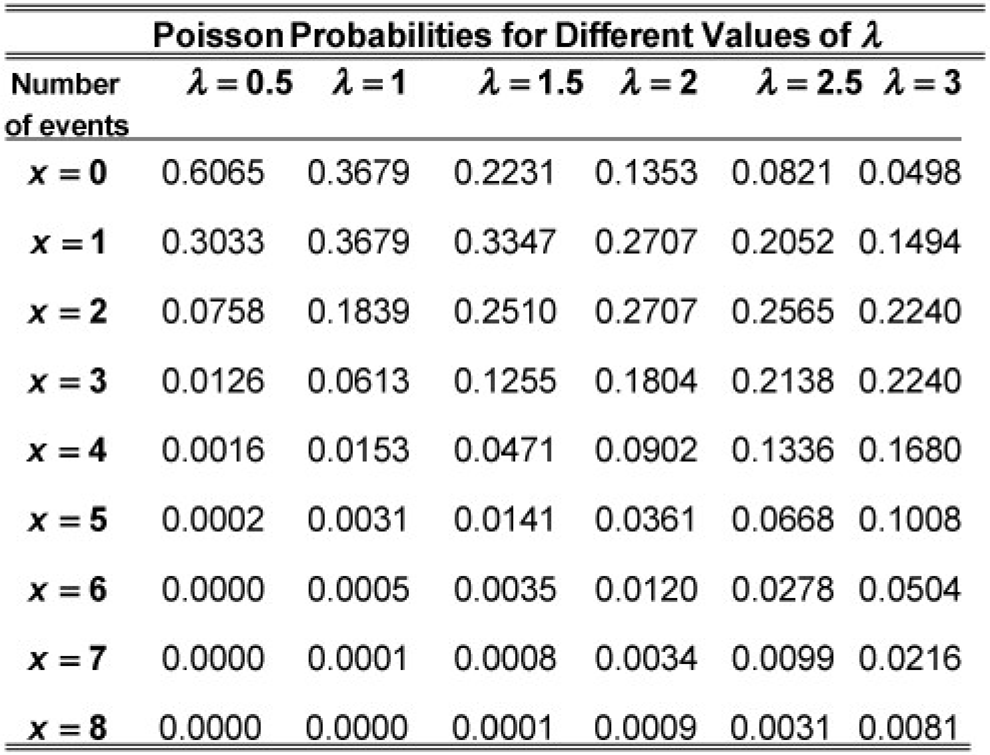
Exercise 2

Counting experiment with 0 expected background events and 2.5 expected signal events

0 events are observed in the data

Is the signal hytpothesis excluded?

What is the upper limit on the signal?



Poisson Probabilities for Different Values of λ

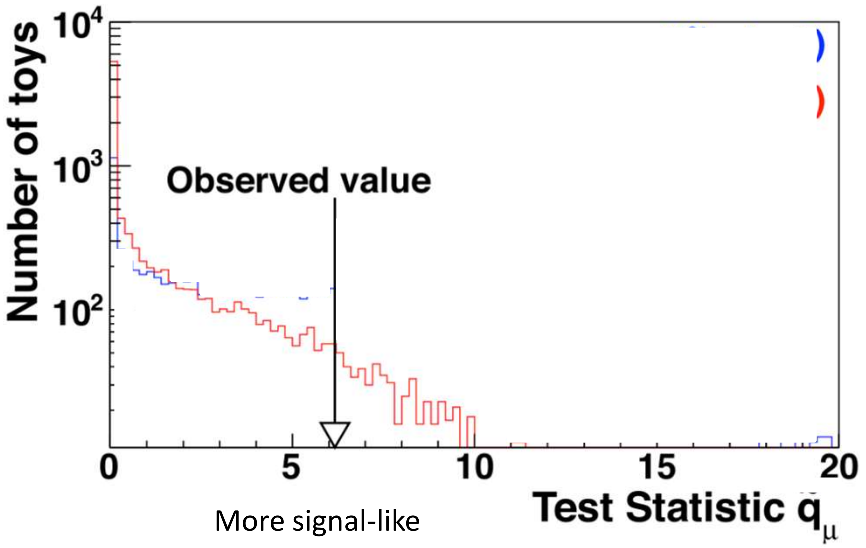
Number of events	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$
$x = 0$	0.6065	0.3679	0.2231	0.1353	0.0821	0.0498
$x = 1$	0.3033	0.3679	0.3347	0.2707	0.2052	0.1494
$x = 2$	0.0758	0.1839	0.2510	0.2707	0.2565	0.2240
$x = 3$	0.0126	0.0613	0.1255	0.1804	0.2138	0.2240
$x = 4$	0.0016	0.0153	0.0471	0.0902	0.1336	0.1680
$x = 5$	0.0002	0.0031	0.0141	0.0361	0.0668	0.1008
$x = 6$	0.0000	0.0005	0.0035	0.0120	0.0278	0.0504
$x = 7$	0.0000	0.0001	0.0008	0.0034	0.0099	0.0216
$x = 8$	0.0000	0.0000	0.0001	0.0009	0.0031	0.0081

Test statistics for exclusion

For purposes of setting an upper limit on μ one may use

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu \end{cases} \qquad \text{where} \qquad \lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}$$

Note for purposes of setting an upper limit, one does not regard an upwards fluctuation of the data as representing incompatibility with the hypothesized μ.



More signal-like

a.k.a CL_{s+b}

$$p_\mu = \int_{q_{\mu,\mathrm{obs}}}^{\infty} f(q_\mu|\mu)\, dq_\mu$$

In the large sample (asymptotic) limit, one has this simple relation:

$$p_\mu = 1 - \Phi\left(\sqrt{q_\mu}\right)$$

Gaussian cumulative distribution function

CL_s

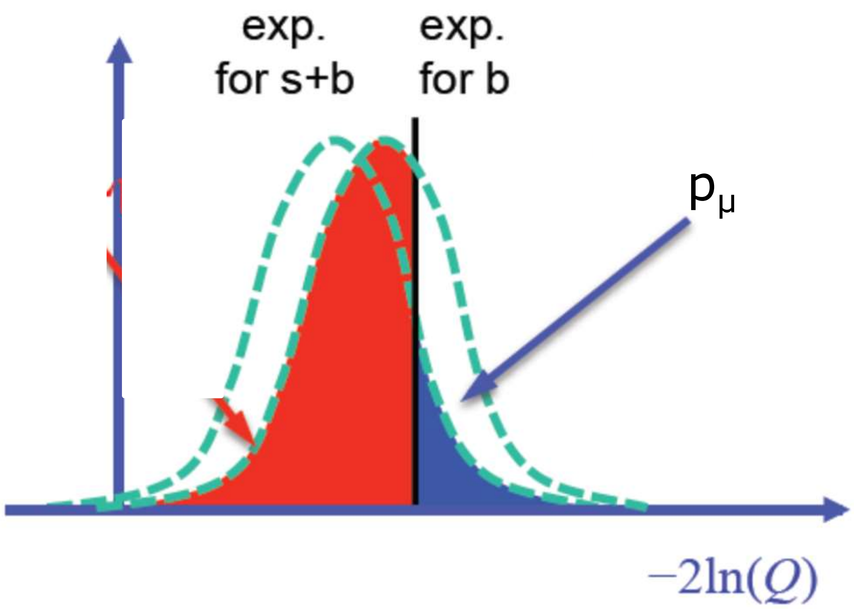
The problem:

Consider the case of low sensitivity: $f(q_\mu|\mu)$ and $f(q_\mu|0)$ very similar

(example: B=10 and S=0.001, S is true but you measure 6)

By construction the probability to reject μ if μ is true is α (e.g., 5%)

→ spurious exclusion in 5% of the case



CL_s

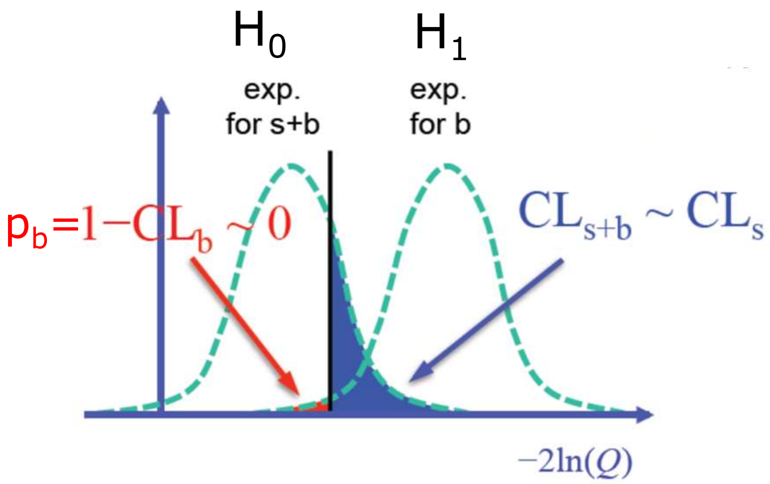
$$CL_s = \frac{p_\mu}{1-p_b} = \frac{CL_{s+b}}{CL_b}$$

"approximation to the confidence in the signal hypothesis, one might have obtained if the experiment had been performed in the complete absence of background"

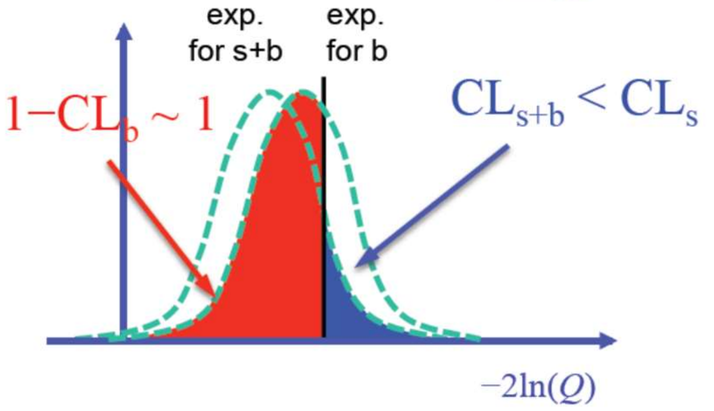
$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} \,|\, \text{signal+background}) = \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu^{\text{obs}})\, d\tilde{q}_\mu$$

$$1-p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} \,|\, \text{background-only}) = \int_{q^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|0,\hat{\theta}_0^{\text{obs}})\, d\tilde{q}_\mu$$

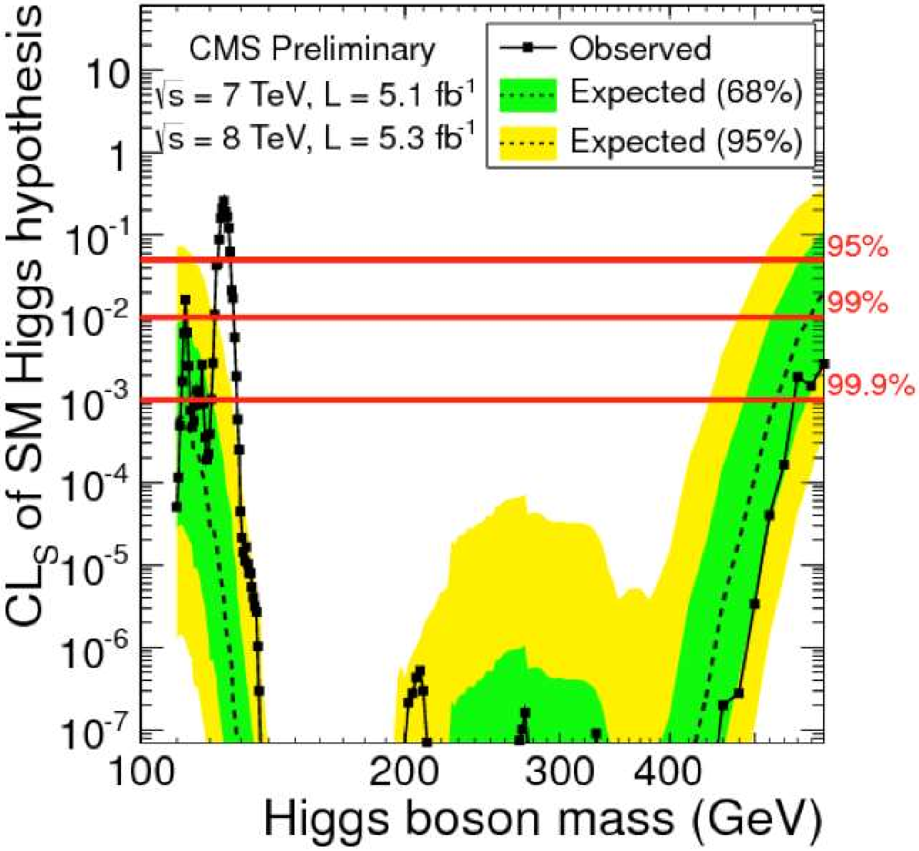
If the two distributions are very well separated then 1−CL_b will be very small ⇒ CL_b ~1 and CL_s ~ CL_{s+b} , i.e: the ordinary *p*-value of the s+b hypothesis



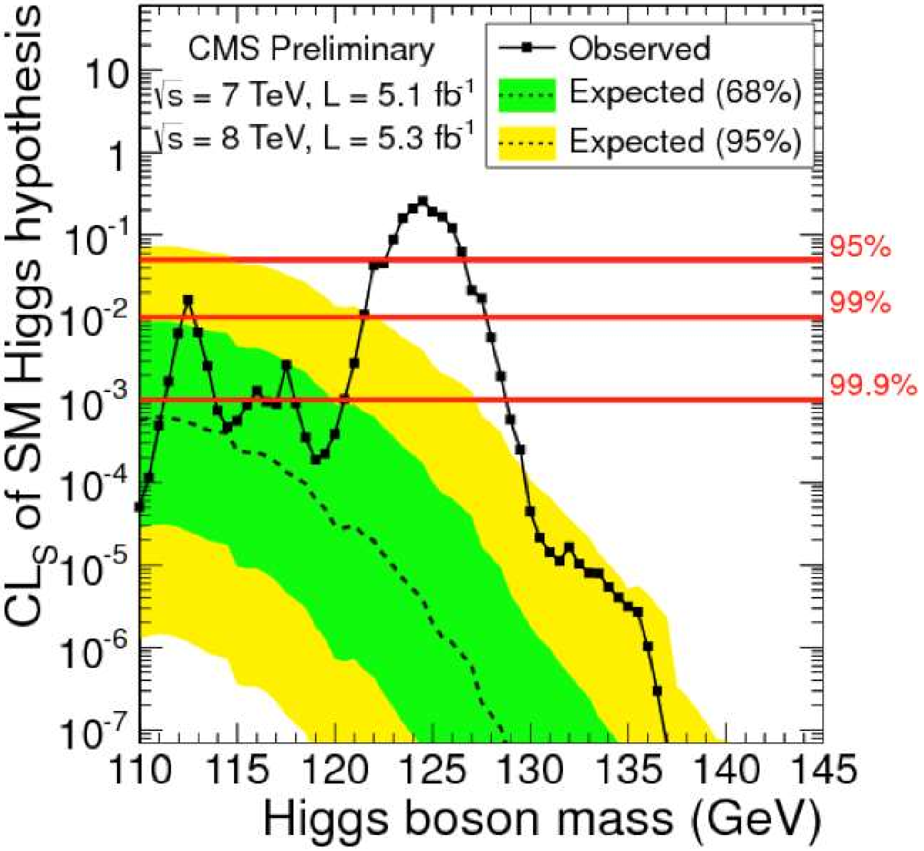
If the two distributions are very close than 1−CL_b will be large ⇒ CL_b small, preventing CLs to become very small



CL_s



CMS Preliminary
√s = 7 TeV, L = 5.1 fb⁻¹
√s = 8 TeV, L = 5.3 fb⁻¹
Observed
Expected (68%)
Expected (95%)
95%
99%
99.9%
CL_S of SM Higgs hypothesis
10
1
10⁻¹
10⁻²
10⁻³
10⁻⁴
10⁻⁵
10⁻⁶
10⁻⁷
100
200
300
400
Higgs boson mass (GeV)



CMS Preliminary
√s = 7 TeV, L = 5.1 fb⁻¹
√s = 8 TeV, L = 5.3 fb⁻¹
Observed
Expected (68%)
Expected (95%)
95%
99%
99.9%
CL_S of SM Higgs hypothesis
10
1
10⁻¹
10⁻²
10⁻³
10⁻⁴
10⁻⁵
10⁻⁶
10⁻⁷
110
115
120
125
130
135
140
145
Higgs boson mass (GeV)

Conclusion: statistics for BSM searches

- Build the likelihood that represents the measurements
  - Observables: counting experiment, unbinned shape analysis or binned analysis
  - Main parameters that we want to measure: parameter of interests
    - ex: signal strengh parameter (μ) or mass
  - The other parameters are called nuisance parameters (θ)
    - ex: syst. uncertainties or auxiliary measurements

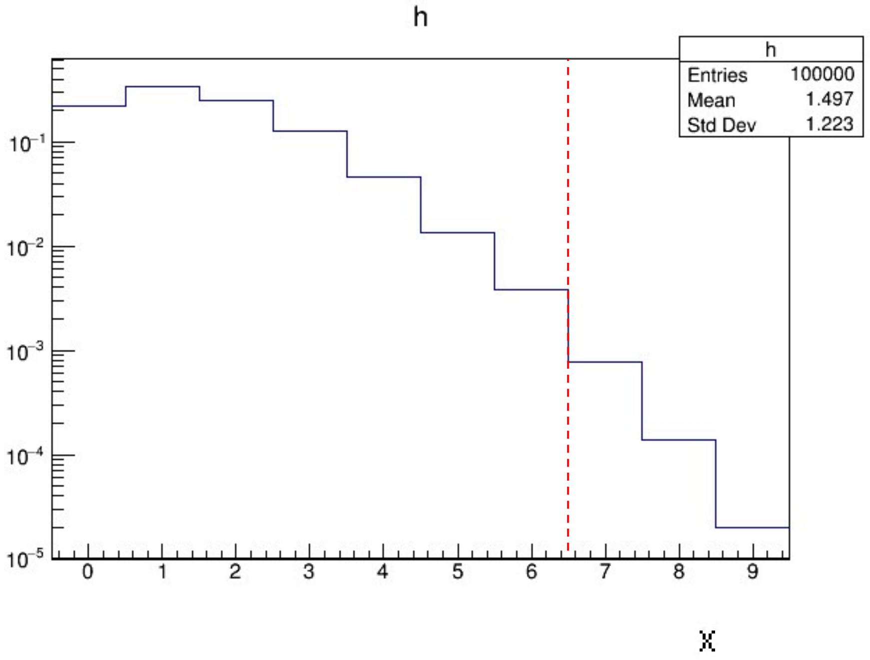
- Parameter estimation via likelihood maximisation

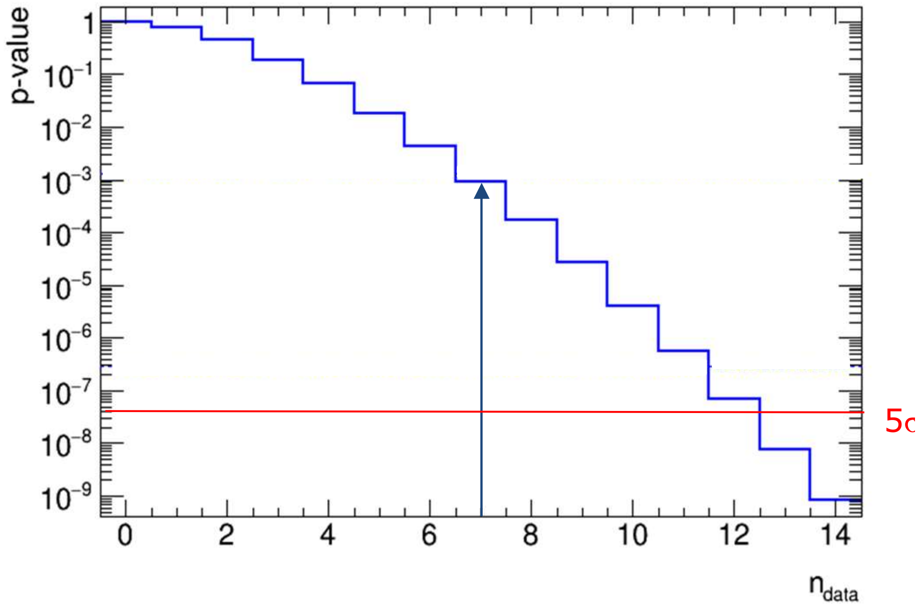
- Hypothesis testing:
  - Specify the null hypothesis that you want to disprove and the alternate hypothesis
    - Ex for discovery: H_0=SM background and H_1=BSM
  - Build you test statistic: t(x)
    - Often based on likelihood ratio
    - Counting experiment: number of events
  - Specify the significance α of the test (how likely you are willing to claim a false discovery)
    - Set to $2.9.10^{-7}$ (5σ) for the discovery or 0.05 for exclusion
  - Compute the p-value: probability of obtaining test results at least as extreme as the results actually observed, under the assumption that the null hypothesis is correct
  - If the p-value is smaller than α then the hypothesis H_0 is rejected


That's all Folks!

Solution 1

p_0=1 -(0.2231+0.3347+0.2510+0.1255+0.0471+0.0141+0.0035)~0.001 → Z=3.09





Z=s/√b=(7-1.5)/√1.5=4.49 → p=3.5×10^{-6} Gaussian approximation not applicable in this case

$$Z = \sqrt{2\left[(s+b)\log\left(1+\frac{s}{b}\right)-s\right]} \;=3.25 \rightarrow \text{p}=0.0006$$

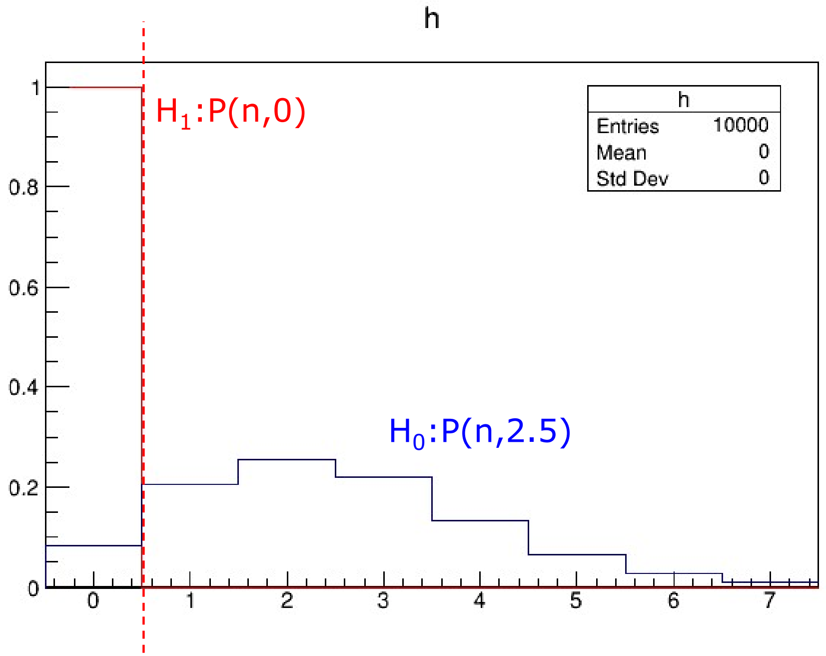
Solution 2

H_0 with b=0: $$P(n,s) = \frac{s^n e^{-s}}{n!}$$

For n=0, we have:

$$p = P(0\,;s) = e^{-s}$$

exp(-2.5)~0.08 => not excluded



The upper limit is defined by $p<\alpha=0.05$ gives

$$e^{-s}<0.05$$

$$s > -\ln 0.05 \approx 3$$ is excluded

Covariance and correlation

Expectation Variance

Recall, for 1D PDF $\boldsymbol{p_x(x)}$ we had: $E[x] = \mu_x; \;\; V[x] = \sigma_x^2$

For a 2D PDF $\boldsymbol{p_{xy}(x,y)}$, one correspondingly has: $\mu_x, \mu_y, \sigma_x, \sigma_y$

How do $\boldsymbol{x}$ and $\boldsymbol{y}$ co-vary ? → $\boxed{\mathrm{C}_{xy} = \text{covariance}_{xy} = E\big[(x-\mu_x)(y-\mu_y)\big] = E[xy] - \mu_x\mu_y}$

From this define the scale / dimension invariant *correlation coefficient*:

$$\boxed{\rho_{xy} = \frac{\mathrm{C}_{xy}}{\sigma_x\sigma_y}}, \quad \text{where } \rho_{xy} \subset [-1,+1]$$

- If x, y are independent: $\rho_{xy} = 0$, ie, they are *uncorrelated* (or they *factorise*)

  Proof: $E[xy] = \iint xy \cdot p_{xy}\,(x,y)dxdy = \iint xy \cdot p_x(x)p_y(y)dxdy = \int x \cdot p_x(x)dx \cdot \int y \cdot p_y(y)dy = \mu_x\mu_y$

- Note that the contrary is not always true: non-linear correlations can lead to $\rho_{xy} = 0$,

2D Gaussian (μ_x=μ_y=0)

general case recovers with x→x-μ_x and y → y-μ_y

$$g(x,\, y) = \frac{1}{2\pi\, \sigma_x \sigma_y} \exp\left[-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right]$$

2D Gaussian (μ_x=μ_y=0)

general case recovers with x→x-μ_x and y → y-μ_y

$$g(x,\,y) = \frac{1}{2\pi\,\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\exp\left[-\frac{1}{2(1-\rho_{xy}^2)}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}-\frac{2x\,y\,\rho_{xy}}{\sigma_x\sigma_y}\right)\right]$$

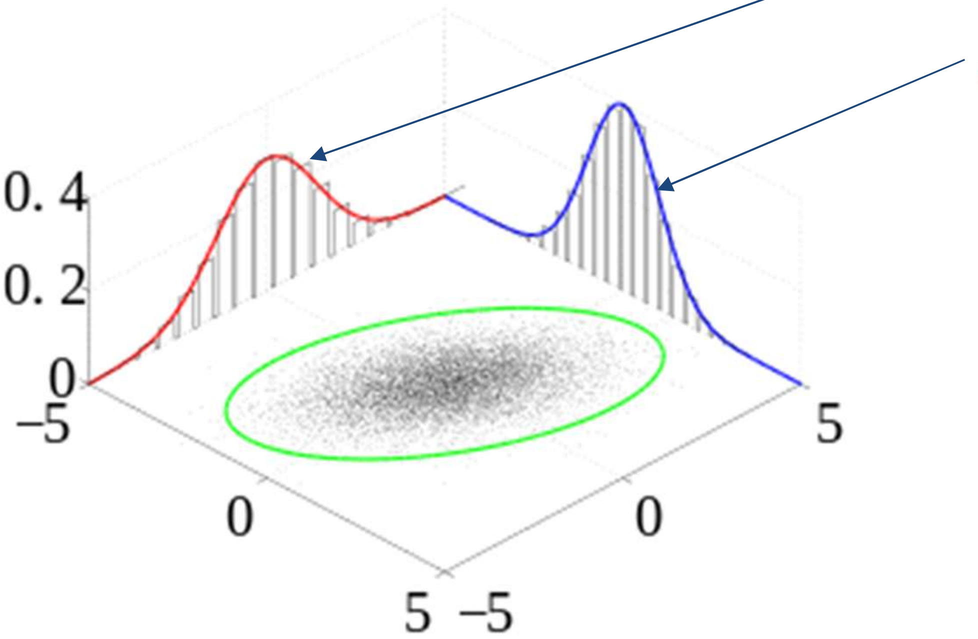
Covariance matrix:

$$C = \begin{pmatrix} \sigma_x^2 & \rho_{xy}\,\sigma_x\sigma_y \\ \rho_{xy}\,\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

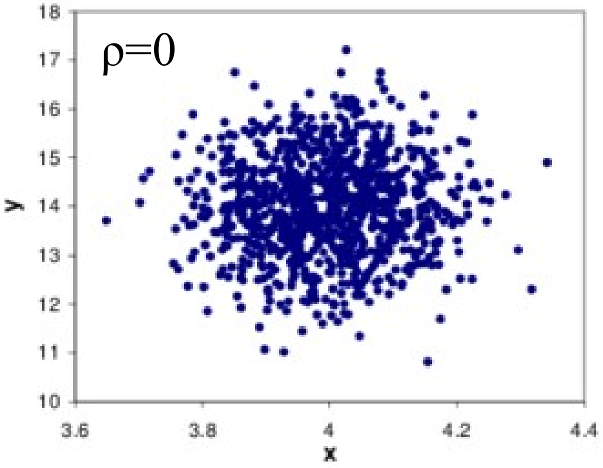
Marginal pdf:

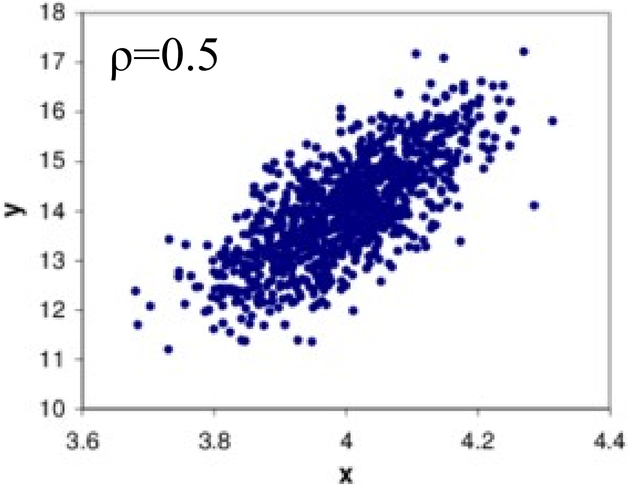
$$g_x(x) = \int_{-\infty}^{+\infty} g(x,\,y)\,\mathrm{d}y = \frac{1}{\sqrt{2\pi\,\sigma_x^2}}\,e^{-x^2/2\sigma_x^2}$$

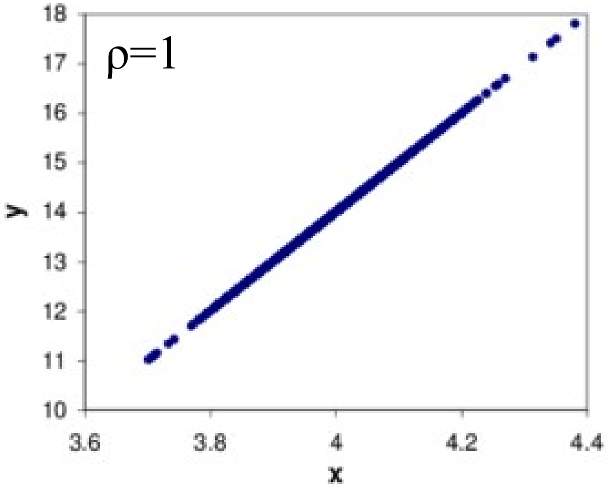
$$g_y(y) = \int_{-\infty}^{+\infty} g(x,\,y)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi\,\sigma_y^2}}\,e^{-y^2/2\sigma_y^2}$$

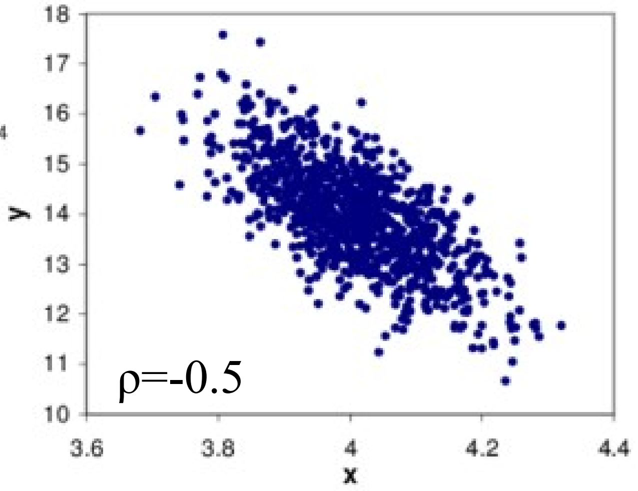


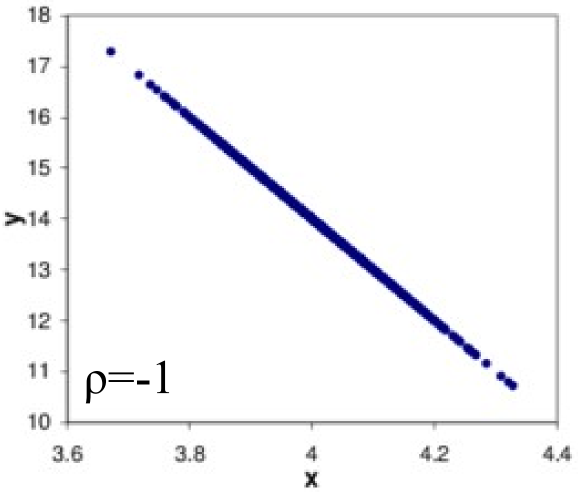
2D Gaussian


ρ=0
18
17
16
15
14
13
12
11
10
y
3.6
3.8
4
4.2
4.4
x


ρ=0.5
18
17
16
15
14
13
12
11
10
y
3.6
3.8
4
4.2
4.4
x

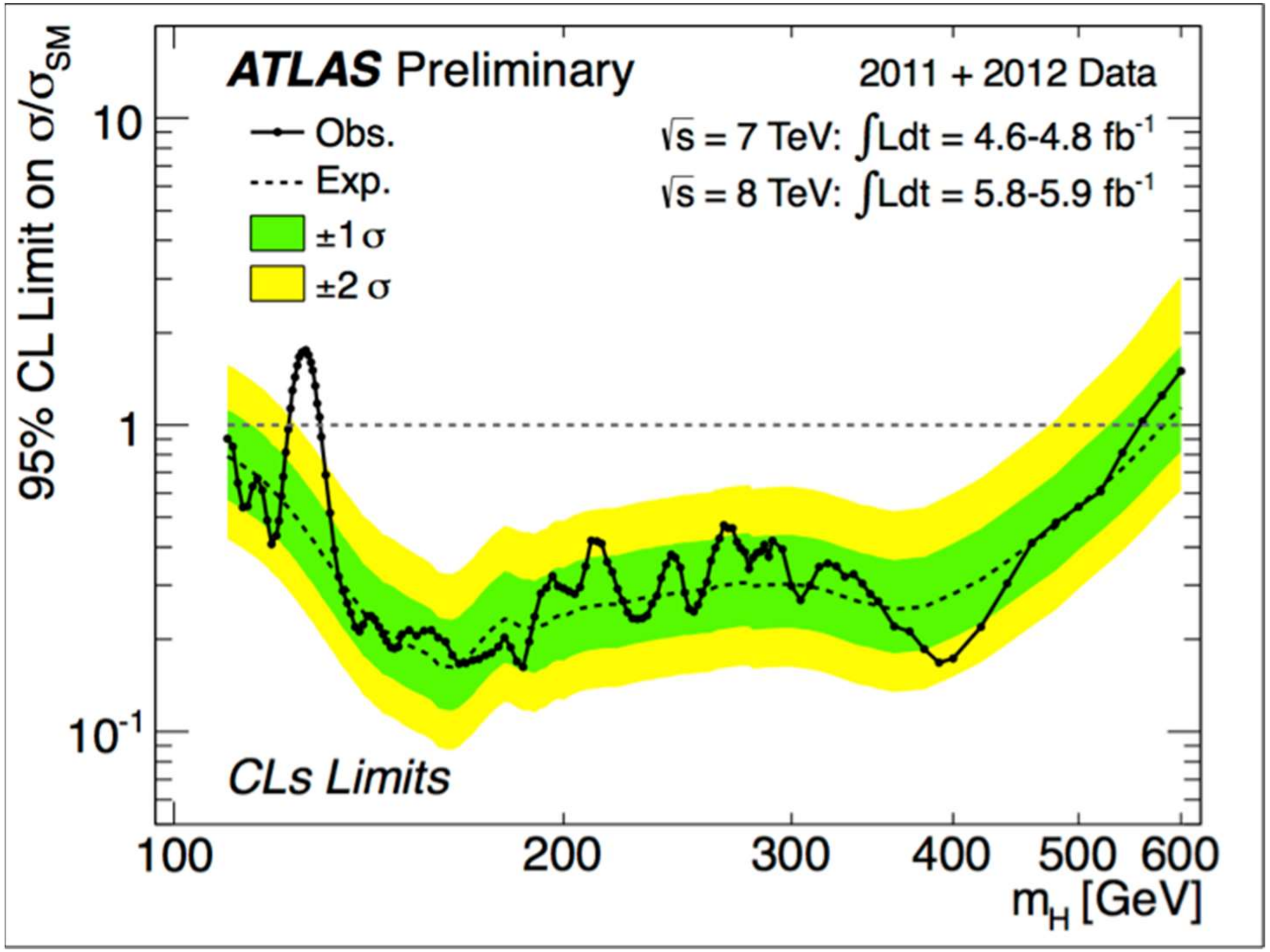

ρ=1
18
17
16
15
14
13
12
11
10
y
3.6
3.8
4
4.2
4.4
x


ρ=-0.5
18
17
16
15
14
13
12
11
10
y
3.6
3.8
4
4.2
4.4
x


ρ=-1
18
17
16
15
14
13
12
11
10
y
3.6
3.8
4
4.2
4.4
x

The Higgs brazil plot

For every value of m_H, find the CLs upper limit on μ (CLs(μ_{up})=0.05)



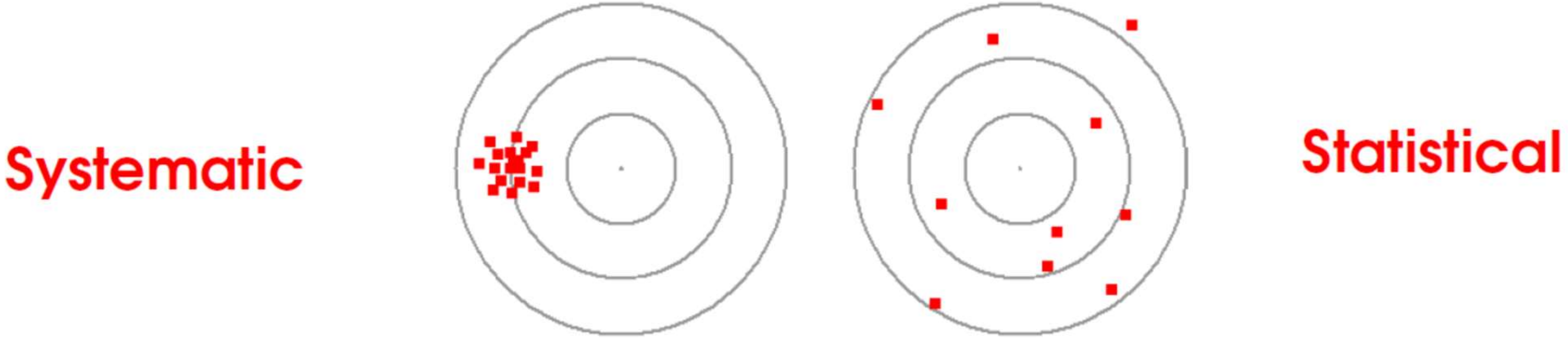
Systematic uncertainties

The uncertainties we have dealt with so far are statistical uncertainties

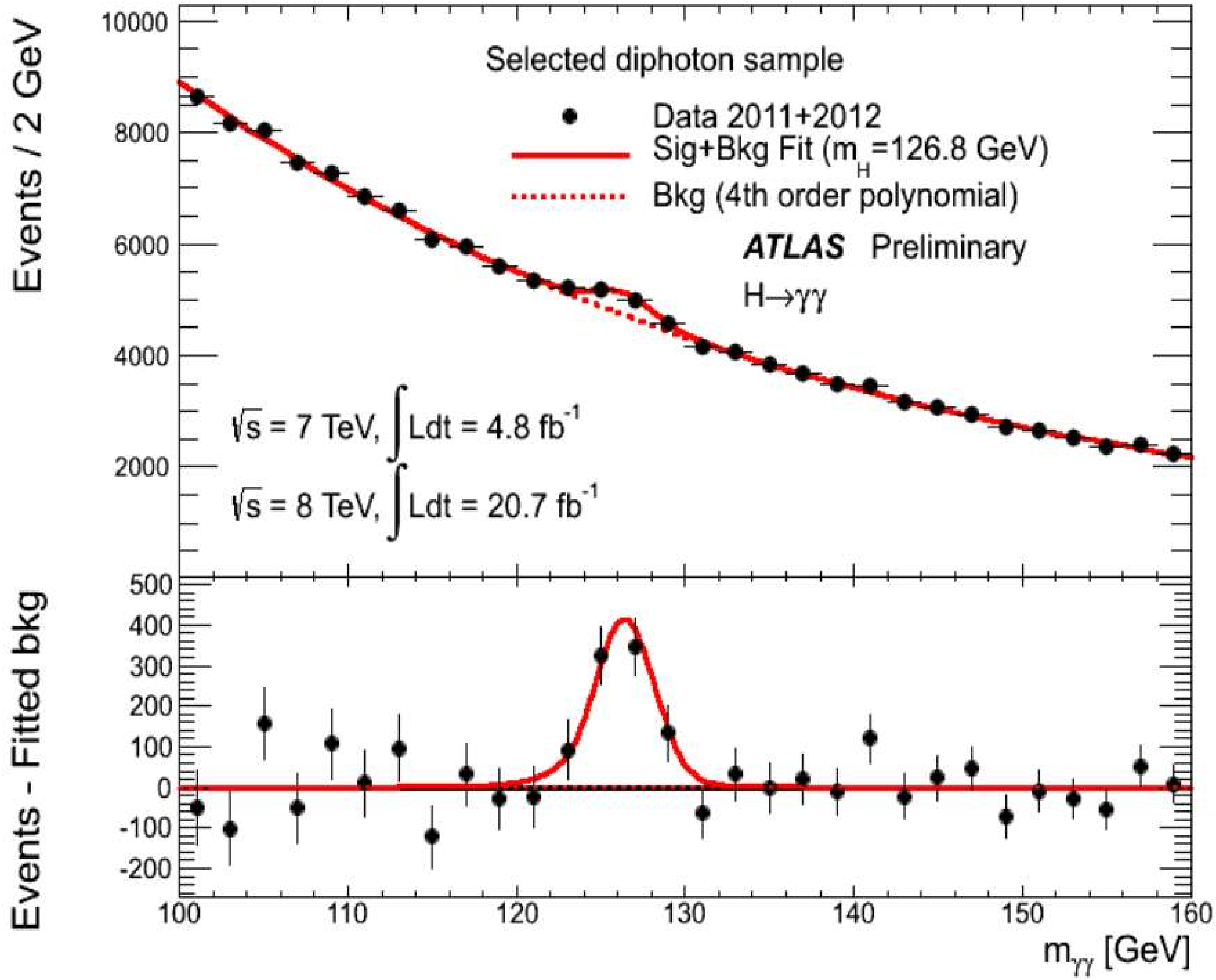
- “Random noise”, not correlated between events
- Decrease usually as 1/√n.

Systematic uncertainties:

- Can have underlying bias in the measurement (ex: luminosity, energy calibration)
- Same for all the events : does not improve with more data
- Can be constrained from data or with an auxiliary measurement (ex: luminosity)



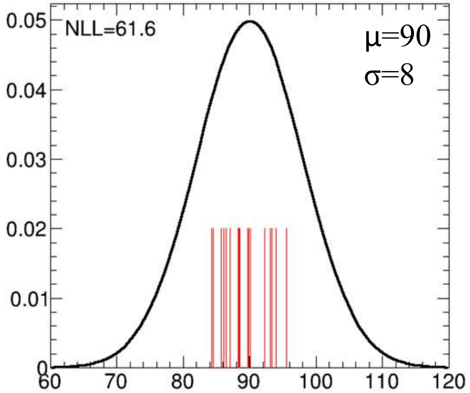
Shape analysis

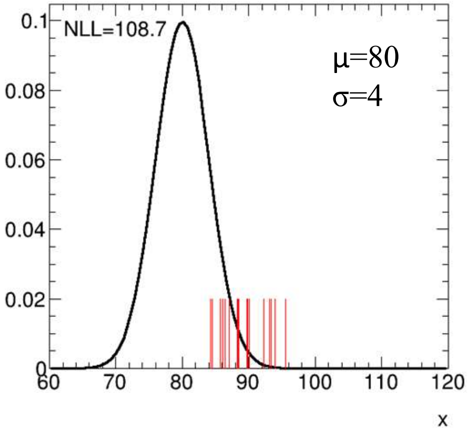


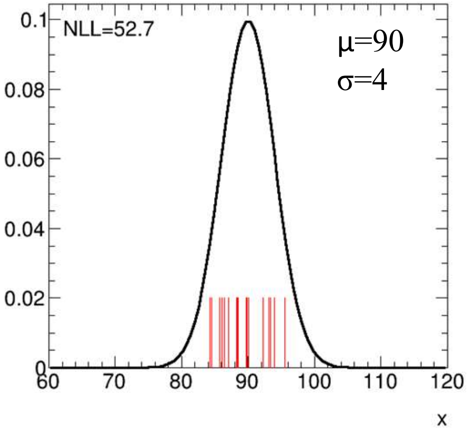
MLE: the Gaussian example

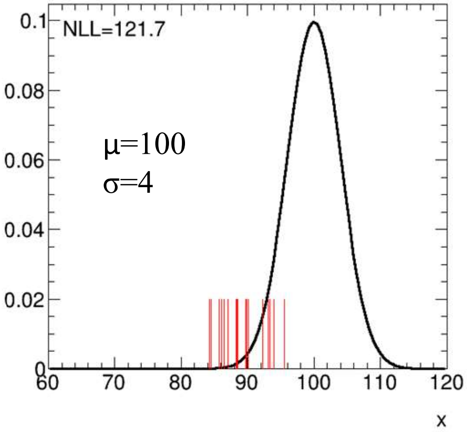
$$NLL = -\ln L(\mu,\sigma)$$

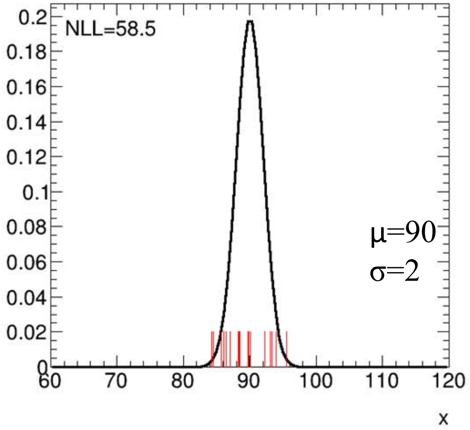
$$= -\ln\left(\prod_{i=1}^{N} \frac{e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\right)$$



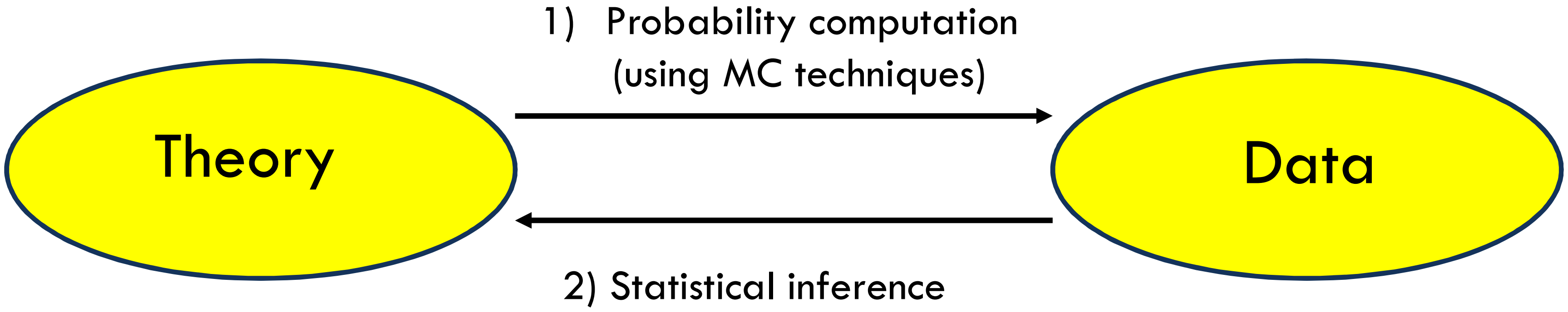








Statistics in particle physics



1) Probability computation (using MC techniques)

Theory

Data

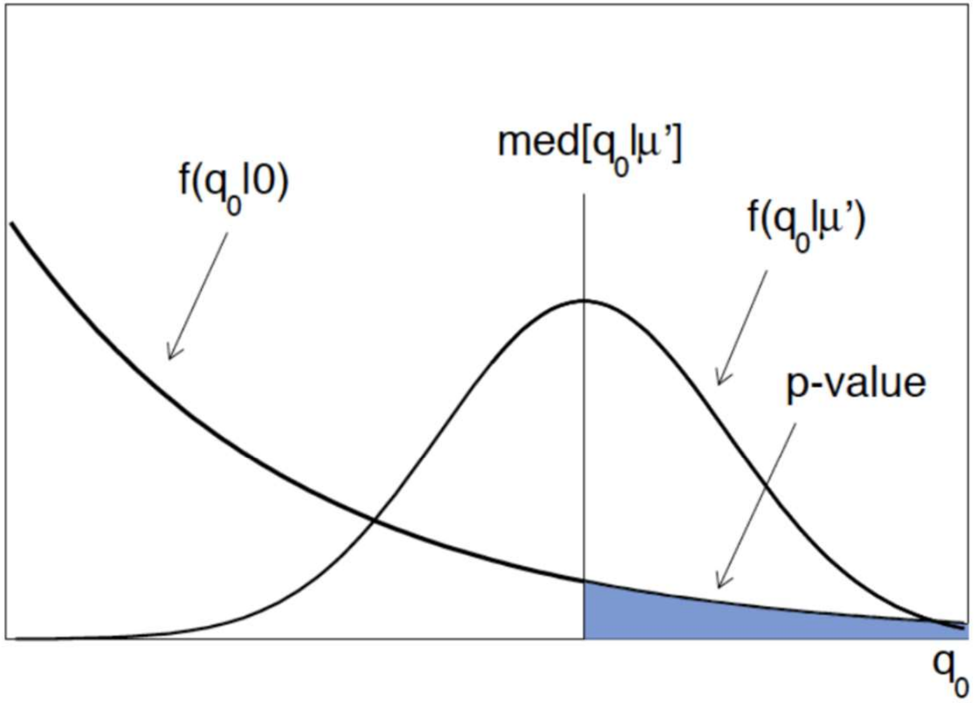
2) Statistical inference

In particle physics, we want to:

- Measure a quantity (ex: Higgs mass): parameter estimation
  - The best estimate of the true parameter with lowest uncertainty as possible based on the data
- Test a theory (ex: SUSY): hypothesis testing
  - Which model best describes the data: “a relative probability”

Expected significance

When planning the experiment, we want to quantify how sensitive we are to a potential discovery, e.g., by given median significance assuming some nonzero strength parameter μ'.



So for p-value, need $f(q_0|0)$, for sensitivity, will need $f(q_0|\mu')$,

The Central Limit Theorem

CLT: the sum of $\boldsymbol{n}$ independent samples $\boldsymbol{x_i}$ $(i = 1, \ldots, n)$ drawn from any PDF $\boldsymbol{D(x_i)}$ with well defined expectation value and variance is Gaussian distributed in the limit $n \to \infty$

$$\boldsymbol{D}:\ E_D[x_i] = \mu;\ V_D[x_i] = \sigma_D^2\text{, and: } y = \frac{1}{n}\sum_{i=1}^{n} x_i \ \Longrightarrow\ E_{\text{Gauss}}[y] = \mu;\ V_{\text{Gauss}}[y] = \underset{\uparrow}{\frac{\sigma_D^2}{n}}$$

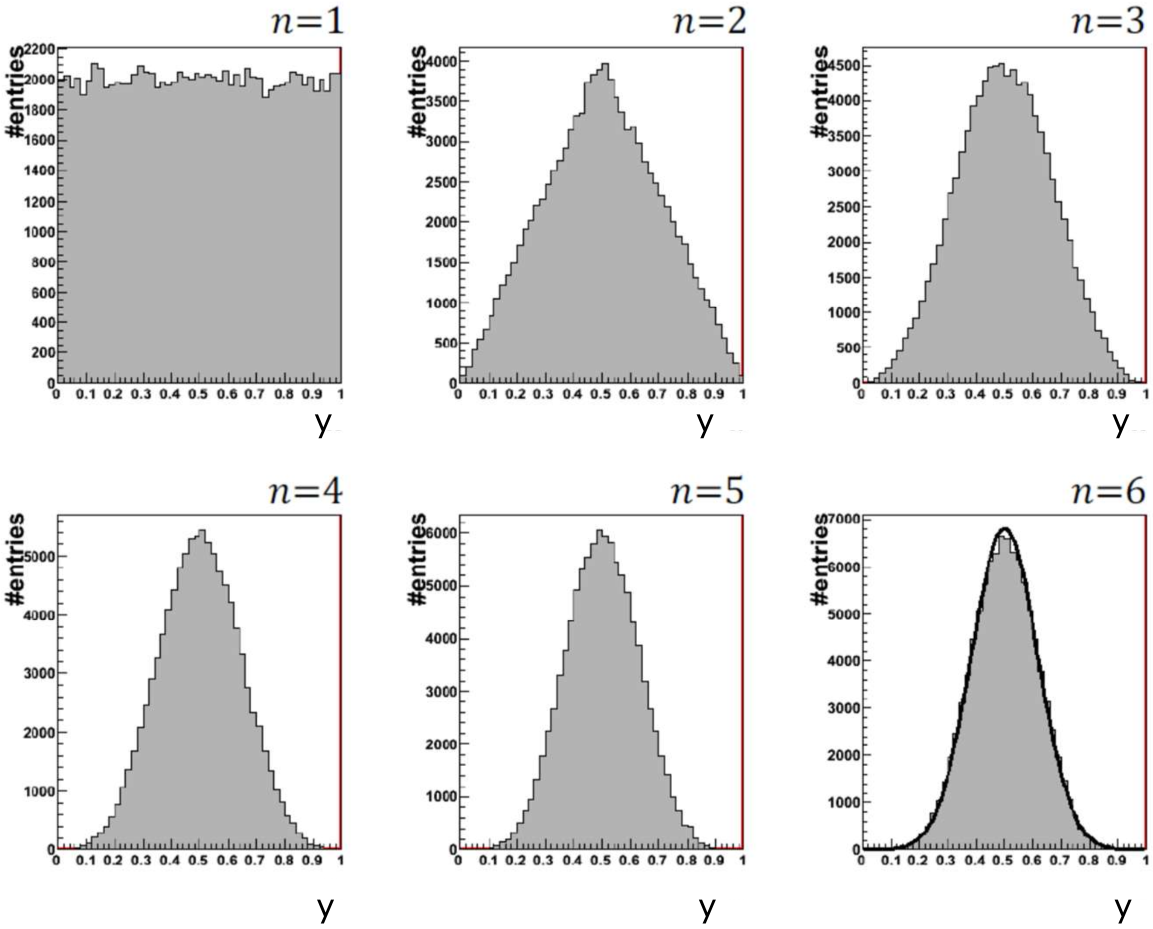
Averaging reduces the variance

The Central Limit Theorem

CLT: the sum of $\boldsymbol{n}$ independent samples $\boldsymbol{x_i}$ $(i = 1, \ldots, n)$ drawn from any PDF $\boldsymbol{D(x_i)}$ with well defined expectation value and variance is Gaussian distributed in the limit $n \to \infty$

$$\boldsymbol{D}\text{: } E_D[x_i] = \mu;\ V_D[x_i] = \sigma_D^2\text{, and: } y = \frac{1}{n}\sum_{i=1}^{n} x_i \ \Longrightarrow\ E_{\text{Gauss}}\,[y] = \mu;\ V_{\text{Gauss}}[y] = \frac{\sigma_D^2}{n}$$

Averaging reduces the variance

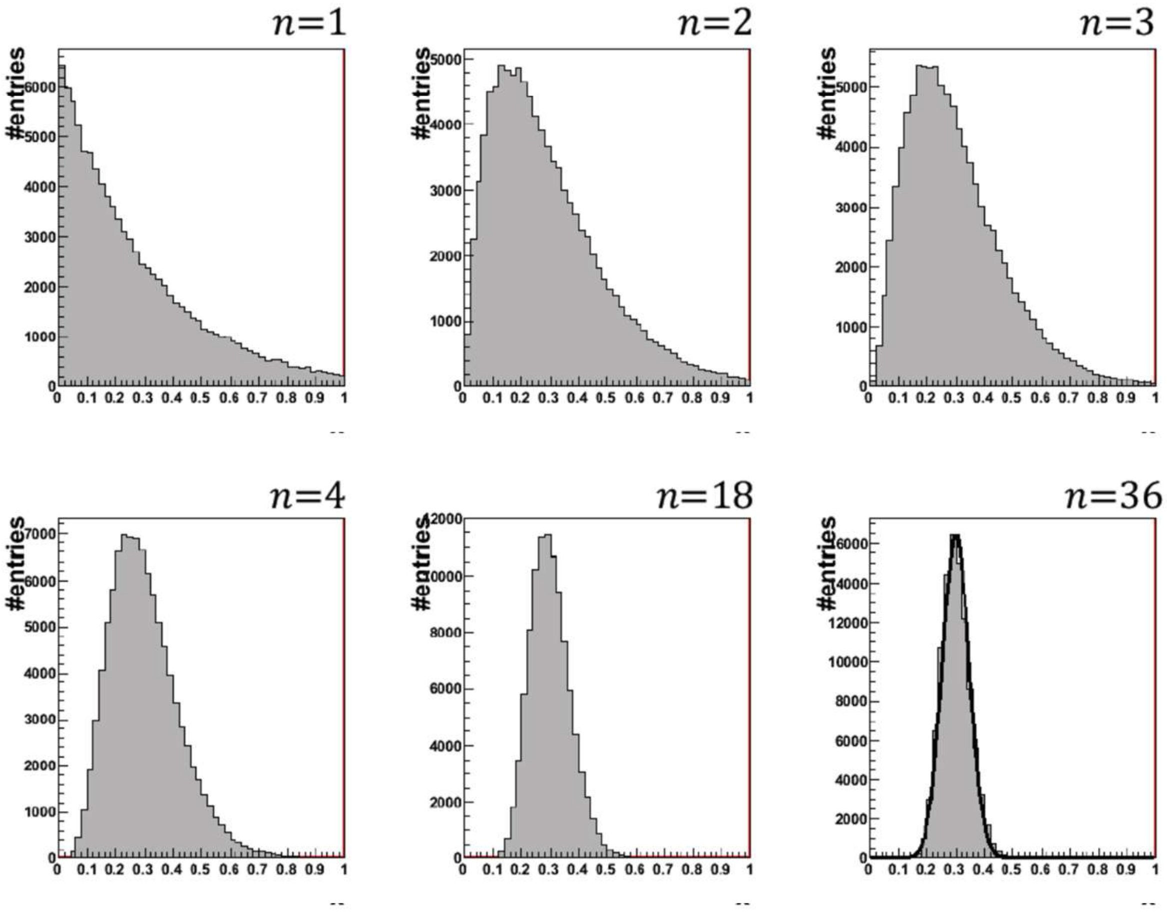


Example: summing up uniformly distributed ensembles within [0,1]

The Central Limit Theorem

CLT: the sum of $\boldsymbol{n}$ independent samples $\boldsymbol{x_i}$ $(i = 1, \ldots, n)$ drawn from any PDF $\boldsymbol{D}(\boldsymbol{x_i})$ with well defined expectation value and variance is Gaussian distributed in the limit $\boldsymbol{n} \to \infty$

$$\boldsymbol{D}\text{: } E_D[x_i] = \mu;\ V_D[x_i] = \sigma_D^2\text{, and: } y = \frac{1}{n}\sum_{i=1}^{n} x_i \ \Rightarrow\ E_{\text{Gauss}}[y] = \mu;\ V_{\text{Gauss}}[y] = \frac{\sigma_D^2}{n}$$



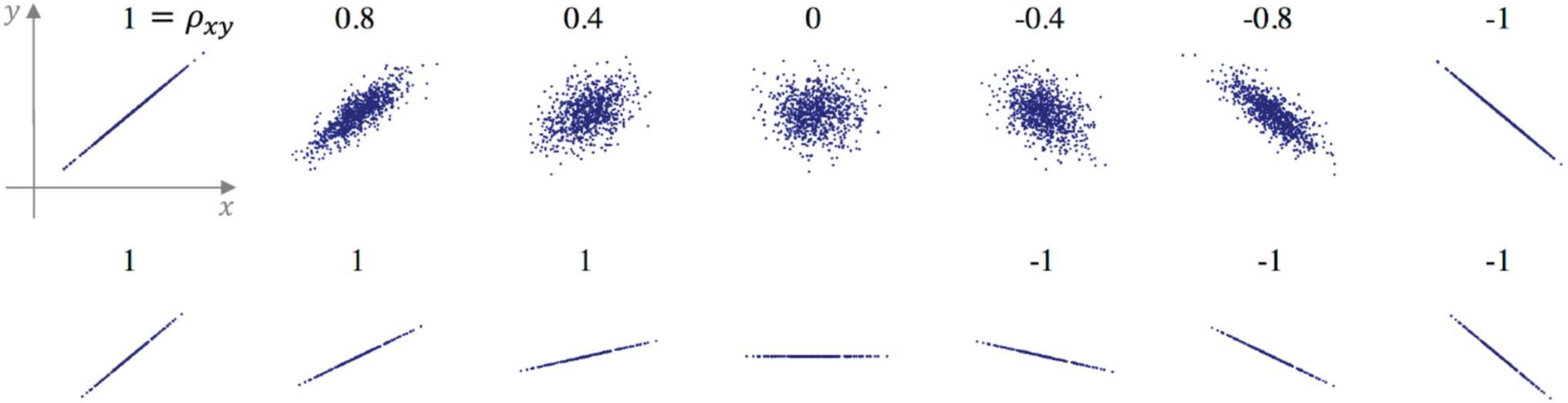
Example: summing up exponential distributions

Central Gaussian limit works even if ***D*** *doesn't look Gaussian at all*

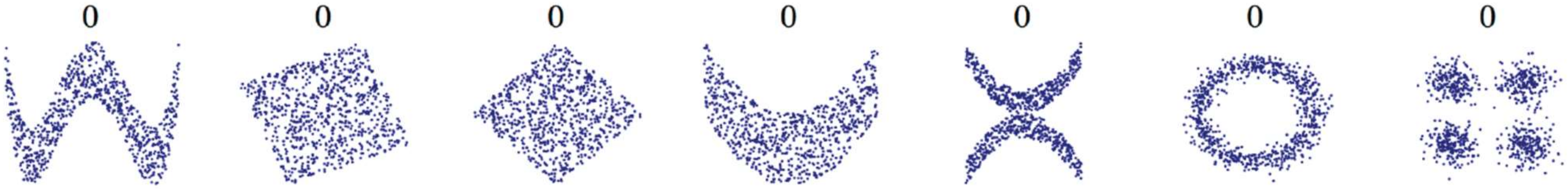
A. Hoecker

Correlation

The correlation coefficient measures the noisiness and direction of a linear relationship:



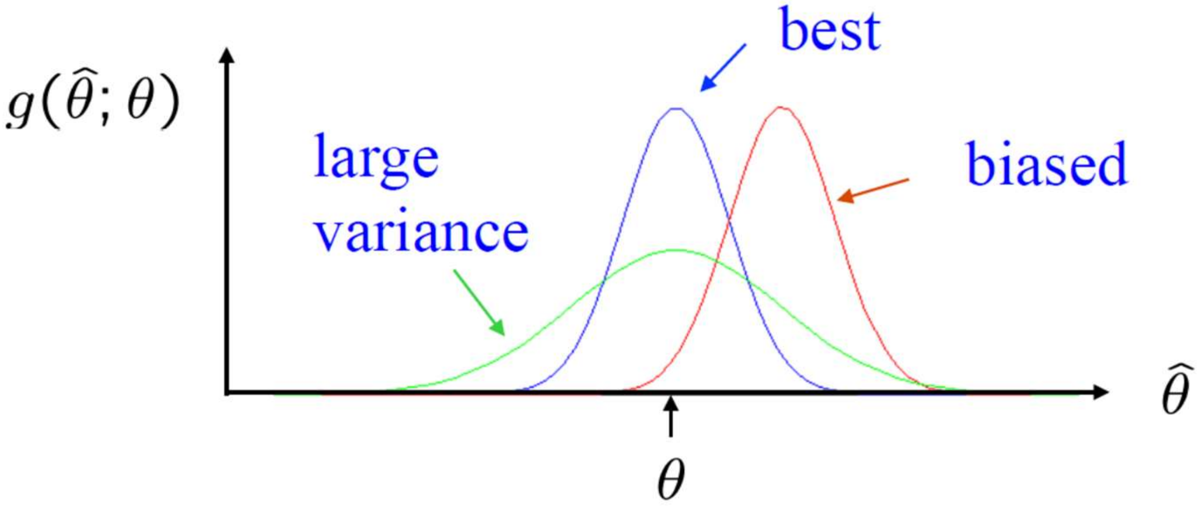
…it does not measure the slope ρ_{xy} (see above figures)



…and non-linear correlation patterns are not or only approximately captured by ρ_{xy} (see above figures)

Properties of estimators

If we were to repeat the entire measurement, the estimates $\hat{\theta}_i(\vec{x})$ from each would follow a pdf:



We want small (or zero) bias (systematic error): $b = E[\hat{\theta}] - \theta$

→ average of repeated measurements should tend to true value.

And we want a small variance (statistical error): $V[\hat{\theta}]$

→ small bias & variance are in general conflicting criteria

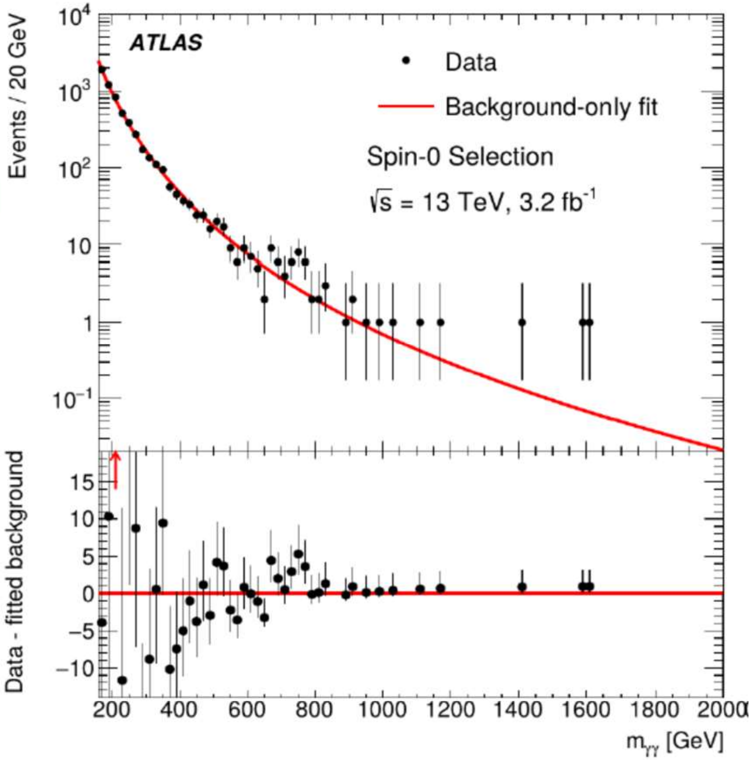
Look-Elsewhere effect

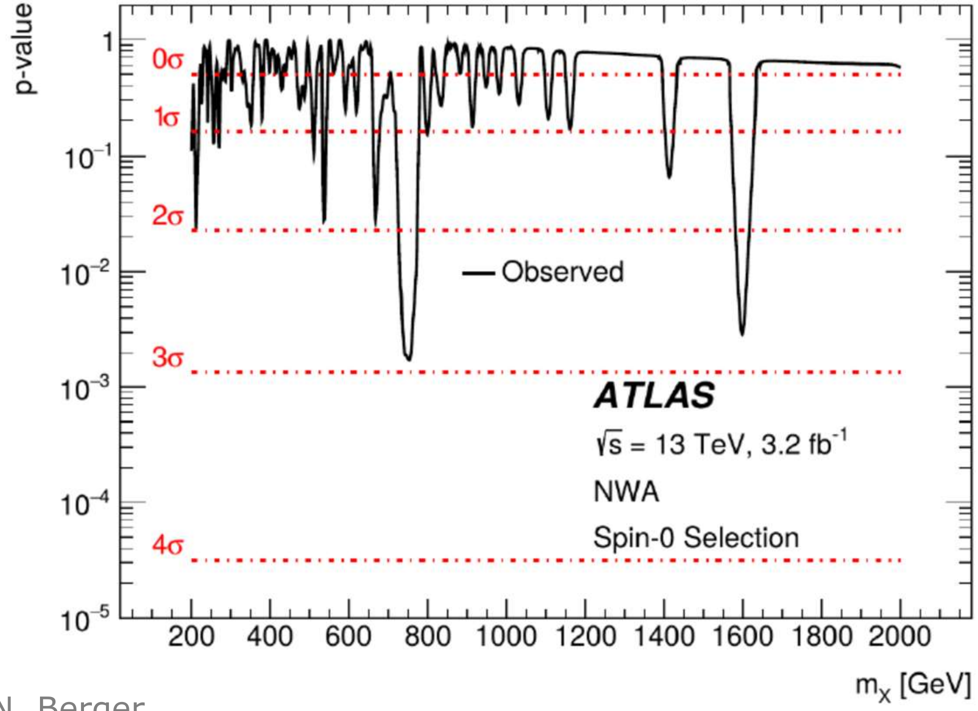
Sometimes, unknown parameters in signal model

e.g. p-values as a function of m_X

⇒ Effectively performing **multiple, simultaneous searches**

→ If e.g. small resolution and large scan range, **many independent experiments**





→ More likely to find an excess **anywhere in the range**, rather than in a **predefined** location

⇒ ***Look-elsewhere effect*** (LEE)

Testing the same H_0, but against different alternatives

⇒ different p-values

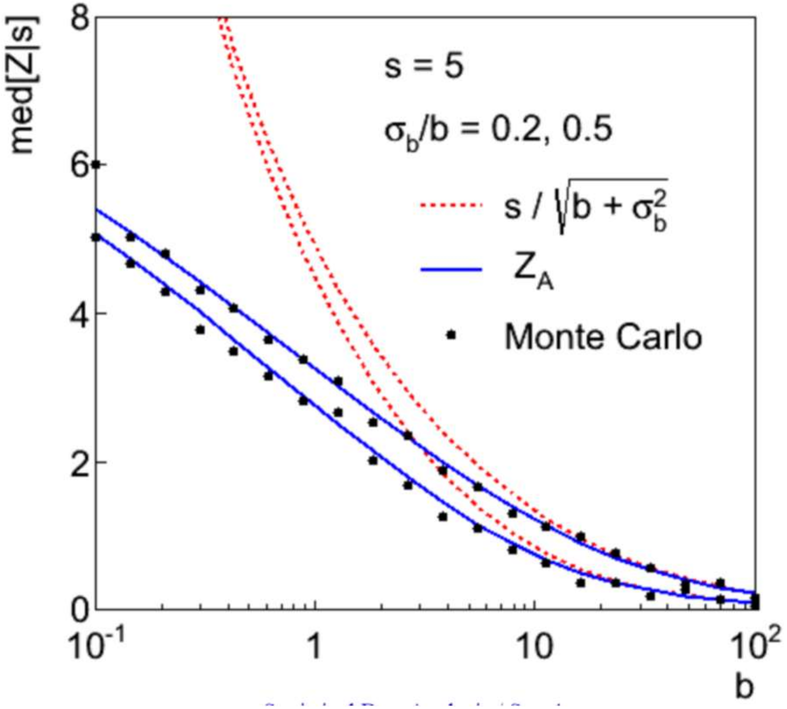
Now with uncertainty on b

Since standard deviations add in quadrature, one has:

$$Z = \frac{s}{\sqrt{b}} \qquad \text{becomes} \qquad Z = \frac{s}{\sqrt{b + \sigma_b^2}}$$

A better approximation is given by:

$$Z_{\mathrm{A}} = \left[2 \left((s+b) \ln \left[\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{1/2}$$

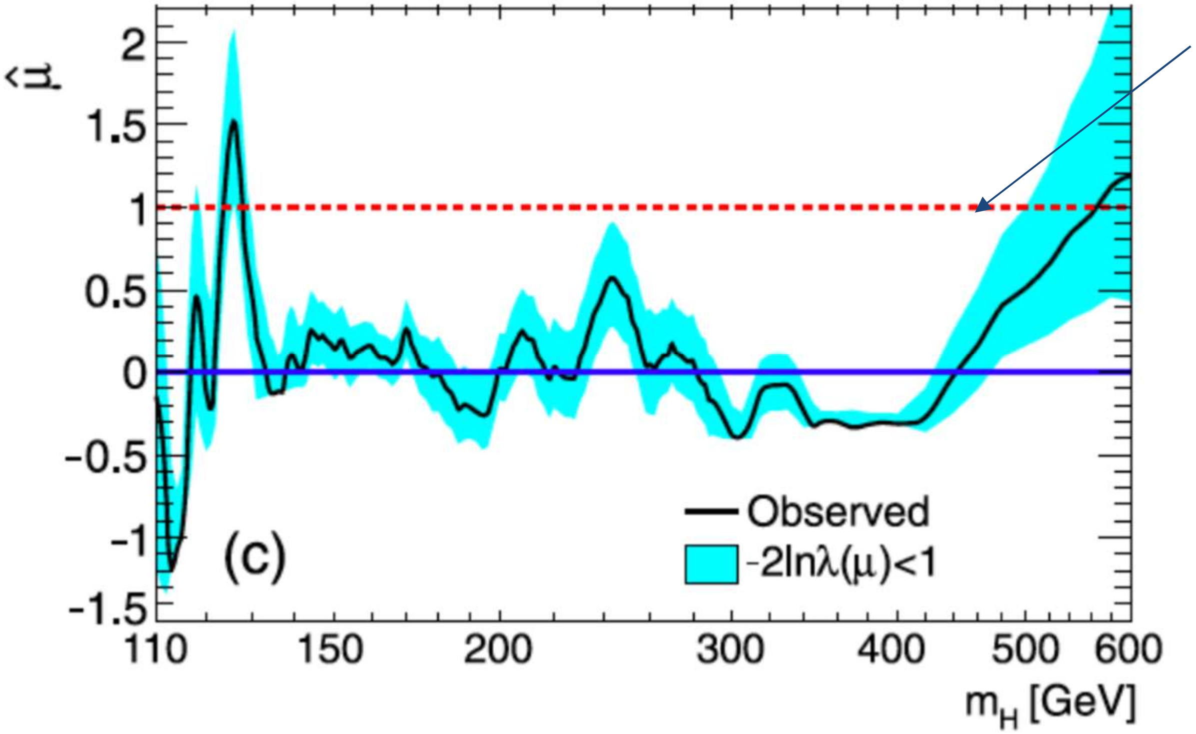


The Higgs $\hat{\mu}$ plot

On the plot of $\hat{\mu}$ versus m_{H}, the blue band is defined by

$$-2\ln\lambda(\mu) = -2\ln(L(\mu)/L(\hat{\mu})) < 1 \text{ i.e., } \ln L(\mu) > \ln L(\hat{\mu}) - \frac{1}{2}$$

i.e., it approximates the 1-sigma error band (68.3% CL conf. int.)



Standard Model

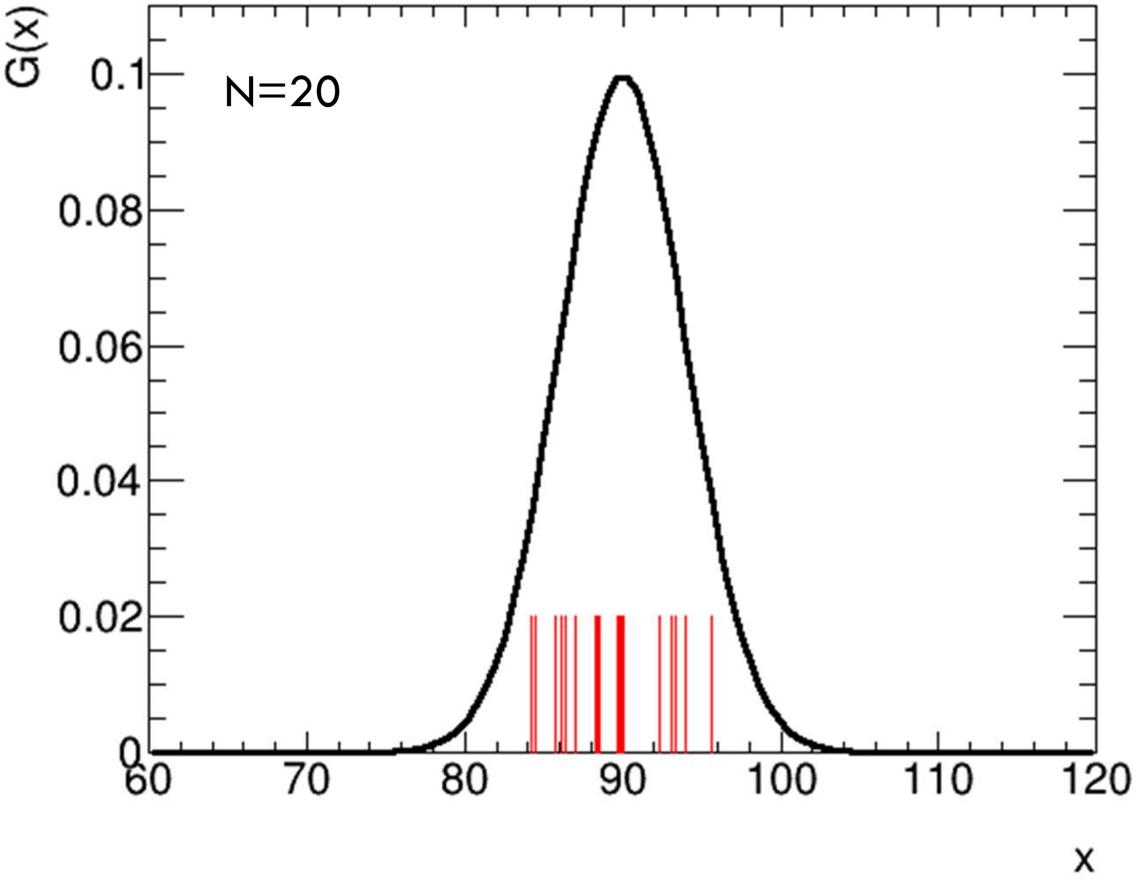
ATLAS, Phys. Lett. B 710 (2012) 49-66

MLE: the Gaussian example

Suppose we have a sample of N observed values $\{x_i\}$ and that the underlying distribution is a Gaussian

Measurements:

1. 94.0
2. 88.3
3. 93.1
4. 89.9
5. 93.3
6. 89.8
7. 86.4
8. 89.7
9. 90.0
10. 88.4
11. 95.6
12. 86.1
13. 89.8
14. 84.2
15. 85.8
16. 84.4
17. 93.1
18. 87.1
19. 92.3
20. 88.5



MLE: the Gaussian example

$$-2\log L(\vec{\theta}) = -2\log L_{\max} + Z^2$$

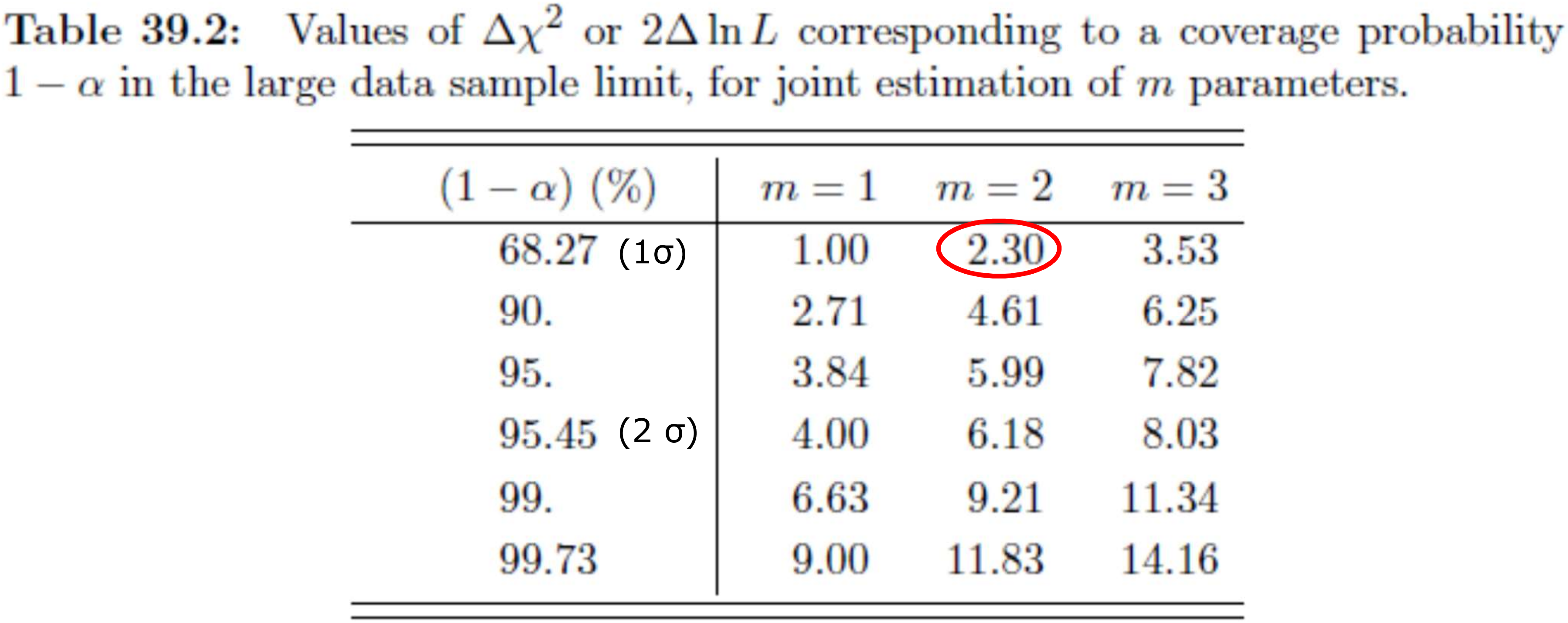
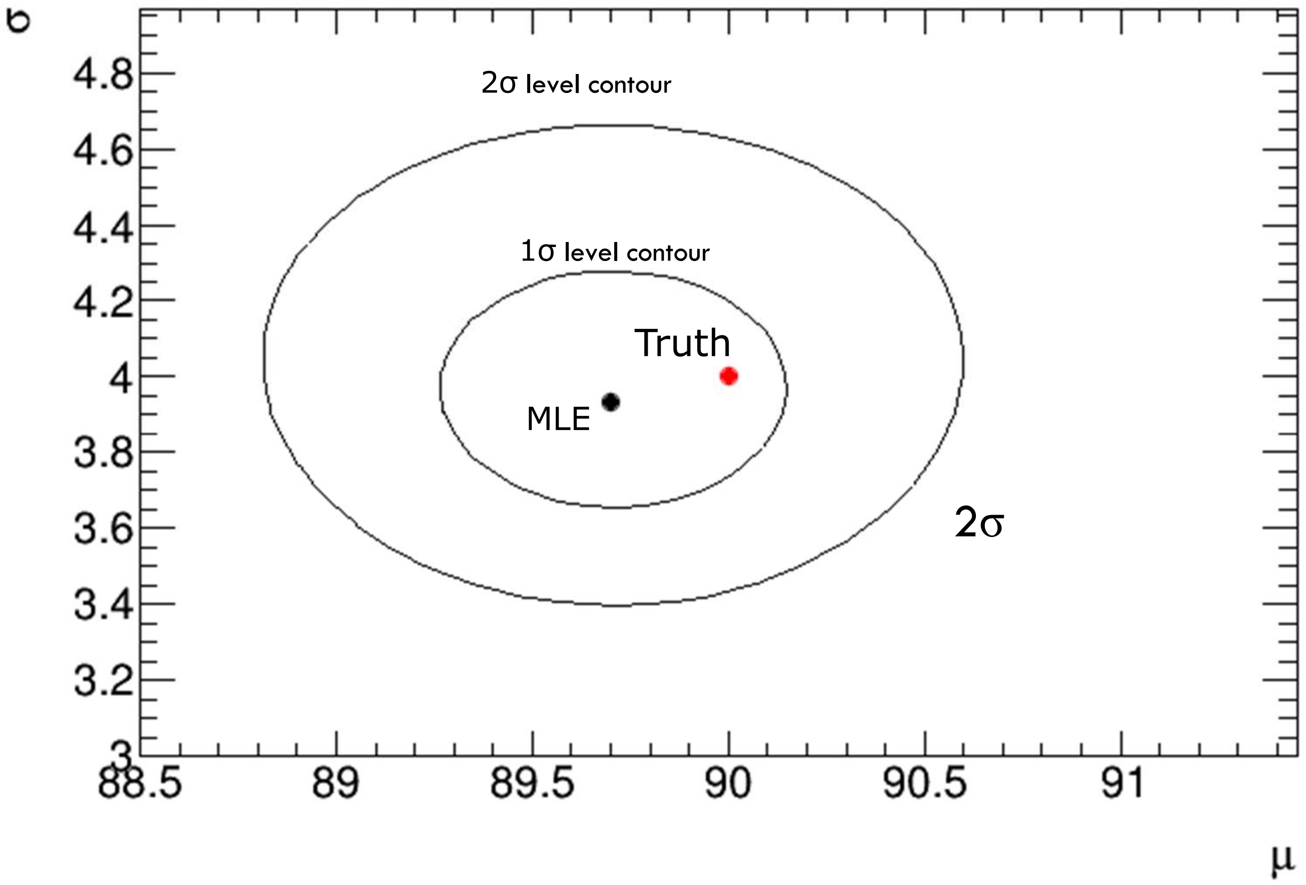


Table 39.2: Values of $\Delta\chi^2$ or $2\Delta\ln L$ corresponding to a coverage probability $1-\alpha$ in the large data sample limit, for joint estimation of m parameters.

$(1-\alpha)$ (%)	$m=1$	$m=2$	$m=3$
68.27 (1σ)	1.00	2.30	3.53
90.	2.71	4.61	6.25
95.	3.84	5.99	7.82
95.45 (2 σ)	4.00	6.18	8.03
99.	6.63	9.21	11.34
99.73	9.00	11.83	14.16

MLE: the Gaussian example



Procedure

Specify the null hypothesis that you want to disprove

- Ex: H_0=SM background only for discovery

Build yo

Specif

Take th

Check

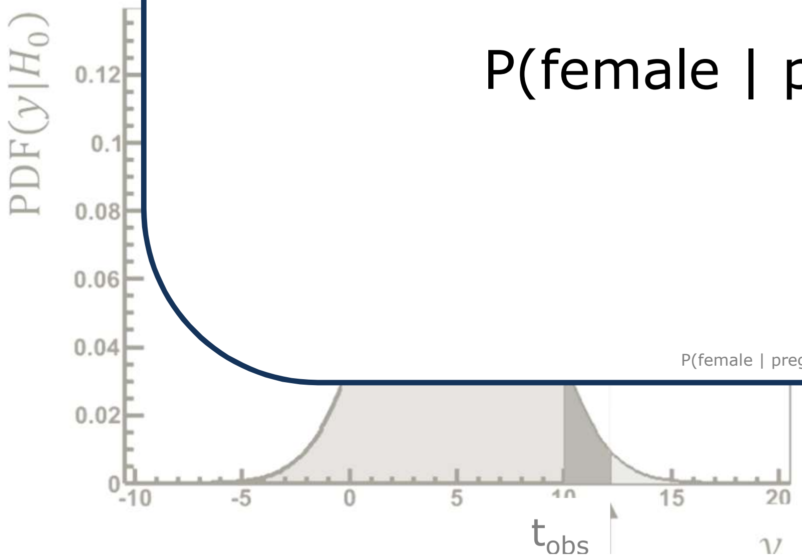
Compu

P(data|theory) != P(theory|data)

Example:

Theory = (male or female)
Data = (pregnant | not pregnant)
P(pregnant | female) ~ 3%
BUT
P(female | pregnant) ≠ 3%

P(female | pregnant) = P(pregnant | female)*P(female)/P(pregnant)



east as H1-like (signal-like) as data, when H0 is true (no signal present).