Basic concepts – part 2

SOS 2021 18-29 January, Online School

Julien Donini - UCA/LPC Clermont-Ferrand

Outline

Basics

- Sample measurements
- Error propagation
- Probabilities, Bayes Theorem
- Probability density function

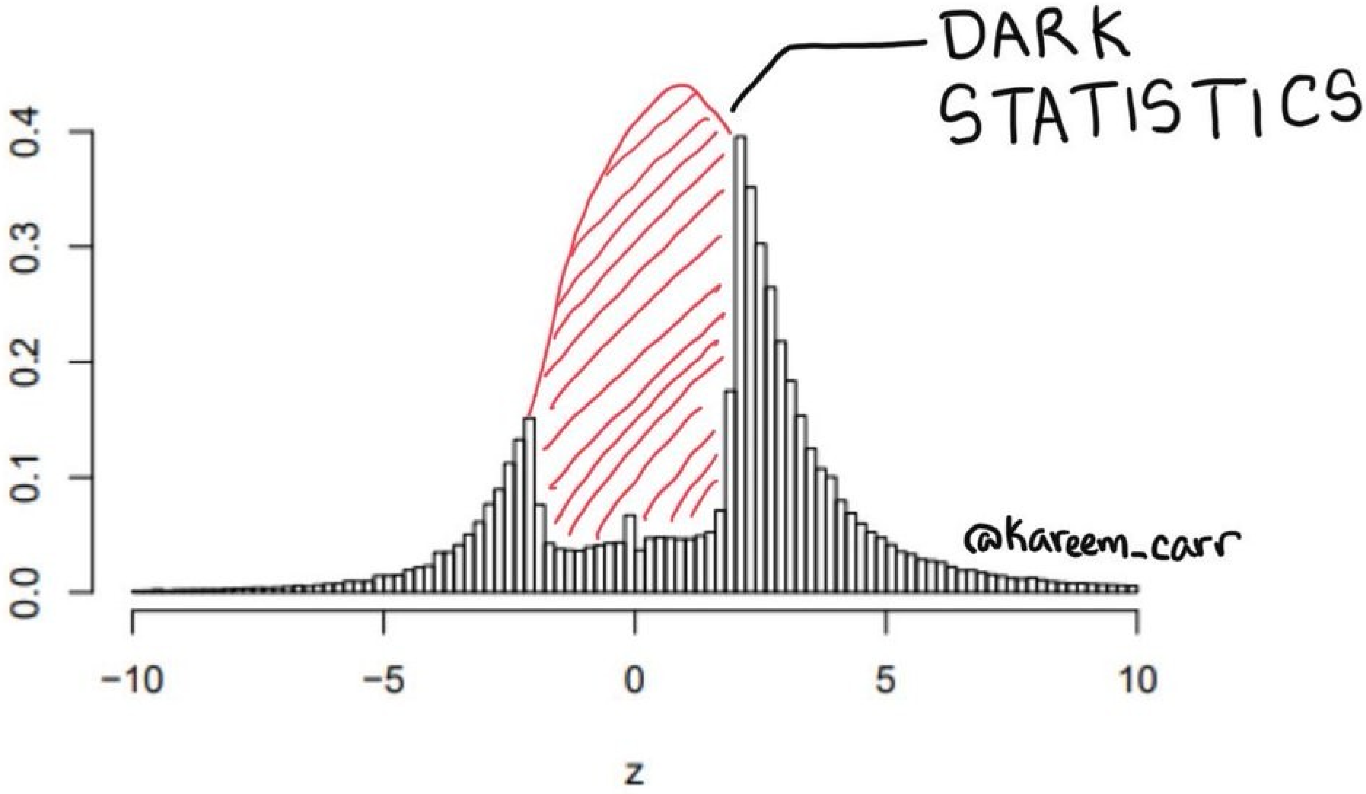
Parameter estimation

- Maximum likelihood method
- Linear regression
- Least square fit

Model testings

- p-value and test statistics
- Chi2 and KS tests
- Hypothesis testing

Samples and parameter estimation



Samples and parameter estimation

A random variable X can be described by its p.d.f *f(x)*

f depends of (generally unknown) **parameters** $\vec{\theta} = \{\theta_1, ..., \theta_p\} \rightarrow f(x, \vec{\theta})$

An **experiment** measuring X provides a **sample** of values $\vec{x} = \{x_1, \dots, x_N\}$

One can construct a function of $\vec{x}$ to **infer** the properties of the p.d.f

- This function is called an **estimator**
- The estimator for a parameter $\boldsymbol{\theta}$ is often written: $\widehat{\boldsymbol{\theta}}$
- **Parameter fitting:** estimate $\boldsymbol{\theta}$ using estimator $\widehat{\boldsymbol{\theta}}$ and data $\vec{x}$
- $\widehat{\boldsymbol{\theta}}(\vec{x})$ is itself a random variable following a p.d.f $g(\widehat{\boldsymbol{\theta}}; \boldsymbol{\theta})$

A **good estimator** should be

Consistent: $\widehat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}$ for infinite sample $(N \rightarrow +\infty)$

Unbiased: average of $\widehat{\boldsymbol{\theta}}$ for infinite number of measurements is $\boldsymbol{\theta}$

$\rightarrow$ that is: $E\left[\widehat{\boldsymbol{\theta}}(\vec{x})\right] - \boldsymbol{\theta} = \boldsymbol{b} = \mathbf{0}$

Basic estimators

Consider a **sample** of size N of a random variable X: $\vec{x} = \{x_1, \ldots, x_N\}$

X follows a p.d.f $f(x)$ of truth **mean μ and variance σ^2**

A simple estimator is the **arithmetic mean** of values x_i: $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

$$E[\bar{x}] = \frac{1}{N}\sum_{i=1}^{N} E[x_i] = \mu$$

→ Unbiased estimator of μ

$$V[\bar{x}] = \boldsymbol{E}\left[\bar{x}^2\right] - \boldsymbol{E}[\bar{x}]^2 = \frac{\boldsymbol{\sigma^2}}{\boldsymbol{N}}$$

This implies that the uncertainty on the sample mean $\bar{x}$ is: $\sigma/\sqrt{N}$

Estimator of the variance: $v = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2$

Expected value of the estimator: $E[v] = \sigma^2 - \frac{\sigma^2}{N} = \frac{N-1}{N}\sigma^2$

→ Biased estimator of σ^2!

Basic estimators

Consider a **sample** of size N of a random variable X: $\vec{x} = \{x_1, \ldots, x_N\}$

X follows a p.d.f $f(x)$ of truth **mean μ and variance σ^2**

A simple estimator is the **arithmetic mean** of values x_i: $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

$$E[\bar{x}] = \frac{1}{N}\sum_{i=1}^{N} E[x_i] = \mu$$

→ Unbiased estimator of μ

$$V[\bar{x}] = E\left[\bar{x}^2\right] - E[\bar{x}]^2 = \frac{\sigma^2}{N}$$

This implies that the uncertainty on the sample mean $\bar{x}$ is: $\sigma/\sqrt{N}$

Estimator of the variance: $v = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \frac{N}{N-1}\left(\overline{x^2} - \bar{x}^2\right)$

Expected value of the estimator: $E[v] = \sigma^2$

→ Unbiased estimator of σ^2 !

Maximum Likelihood estimator (ML)

Suppose a random variable **X** distributed according to a p.d.f $\boldsymbol{f(x;\vec{\theta})}$

- The form of f being know but not the parameters $\vec{\theta} = \{\theta_1, \ldots, \theta_P\}$
- Consider a **sample** of X of N values: $\vec{x} = \{x_1, \ldots, x_N\}$

The method of ML is a technique to estimate $\vec{\theta}$ given data $\vec{x}$

Joint **likelihood function**
(the x_i are fixed here)

$$\boldsymbol{L(\vec{\theta}) = \prod_{i=1}^{N} f(x_i;\vec{\theta})}$$

The **estimators $\widehat{\theta_i}$** are given by:

$$\boldsymbol{\frac{\partial L}{\partial \theta_i} = 0, i = 1 \ldots P}$$

Notes:

- maximizing the likelihood provides and estimate of parameters θ
- In practice the log of L (log likelihoood) is often used
- The likelihood is not a p.d.f !
- Bayesian do transform the likelihood in a p.d.f

Simple examples

Exponential distribution $f(x;\tau) = \frac{1}{\tau} e^{-\frac{x}{\tau}}$

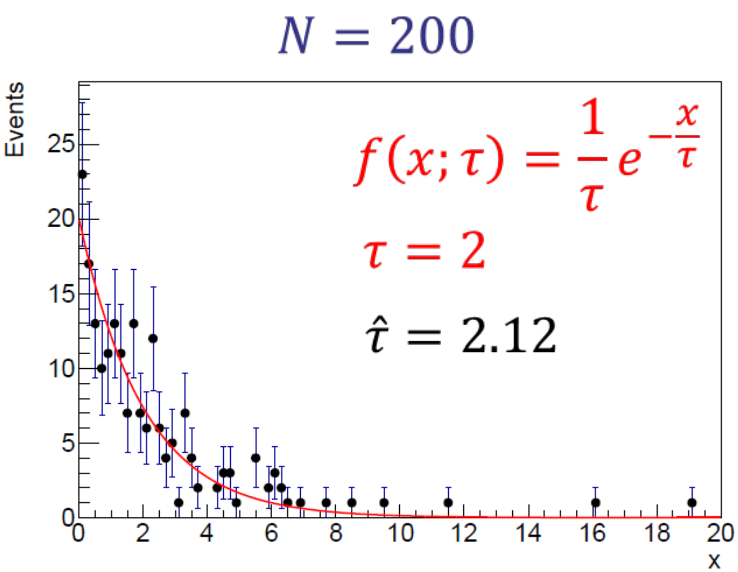
Likelihood: $L(\tau) = \prod_{i=1}^{N} \frac{1}{\tau} e^{-\frac{x_i}{\tau}}$

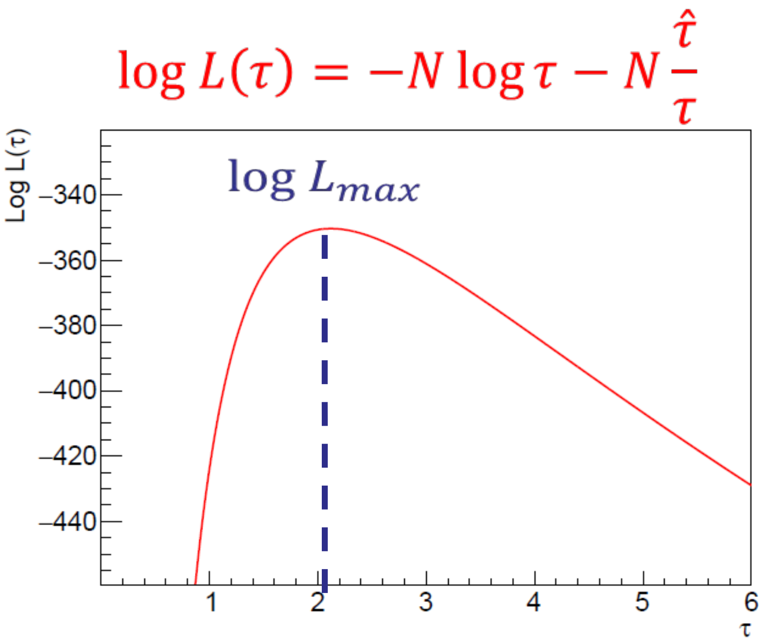
Log-likelihood:

$$\log L(\tau) = \sum_{i=1}^{N} \log f(x_i;\tau) = -N \log \tau - \sum_{i=1}^{N} \frac{x_i}{\tau}$$

Estimator: $\frac{d\log L}{d\tau} = 0 \Leftrightarrow \boldsymbol{\tau} = \boldsymbol{\hat{\tau}} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \boldsymbol{x_i}$

$E[\hat{\tau}] = \tau$ (unbiased estimator)





Simple examples

Gaussian distribution $f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, $\log L(\vec{\theta})=\sum_{i=1}^{N}\log f(x_i;\mu,\sigma)$

Estimators:

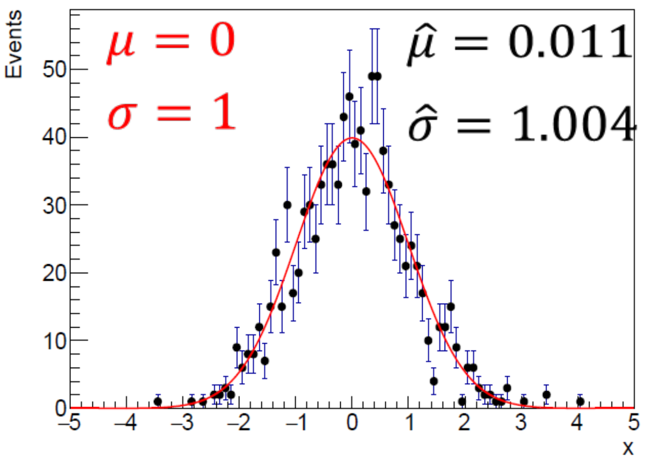
$$\begin{cases} \dfrac{\partial \log L}{\partial\mu}=0 \Leftrightarrow \boldsymbol{\hat{\mu}}=\dfrac{\mathbf{1}}{\boldsymbol{N}}\displaystyle\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}}\boldsymbol{x_i} \\ \dfrac{\partial \log L}{\partial\sigma^2}=0 \Leftrightarrow \boldsymbol{\widehat{\sigma^2}}=\dfrac{\mathbf{1}}{\boldsymbol{N}}\displaystyle\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}}(\boldsymbol{x_i}-\boldsymbol{\hat{\mu}})^{\mathbf{2}} \end{cases}$$

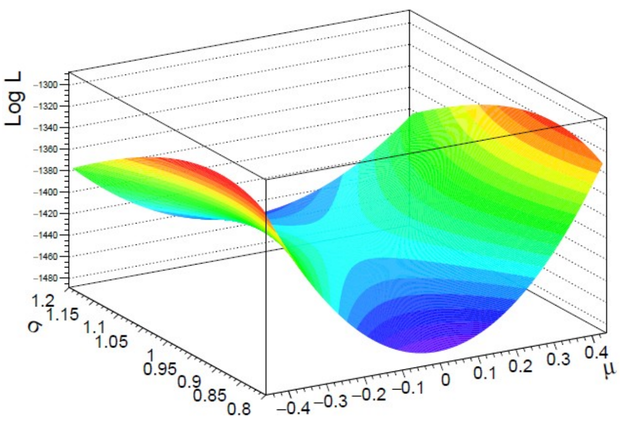
$E[\hat{\mu}]=\mu$ (unbiased)

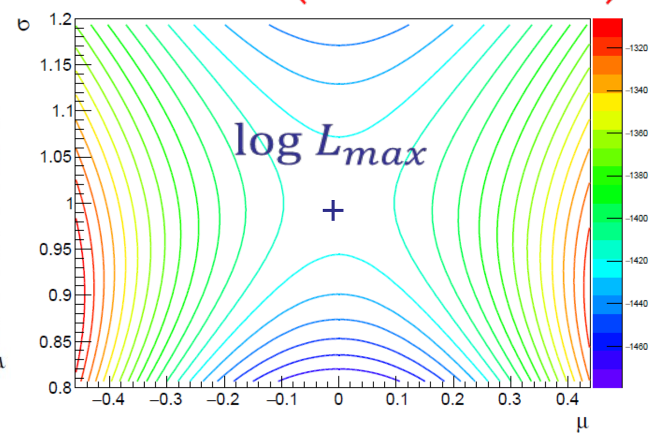
$E\left[\widehat{\sigma^2}\right]=\frac{N-1}{N}\sigma^2$ (biased)

$N=1000$

$$\log L(\mu,\sigma)=-N\log\left(\sqrt{2\pi}\sigma\right)-\frac{1}{2\sigma^2}\left(\sum x_i^2-N\mu^2\right)$$







Uncertainty of ML estimator

Variance of estimator, $V[\hat{\tau}]$ can be tricky to estimate. Several methods exist:

1) Analytical method

For example for the previous exponential distribution

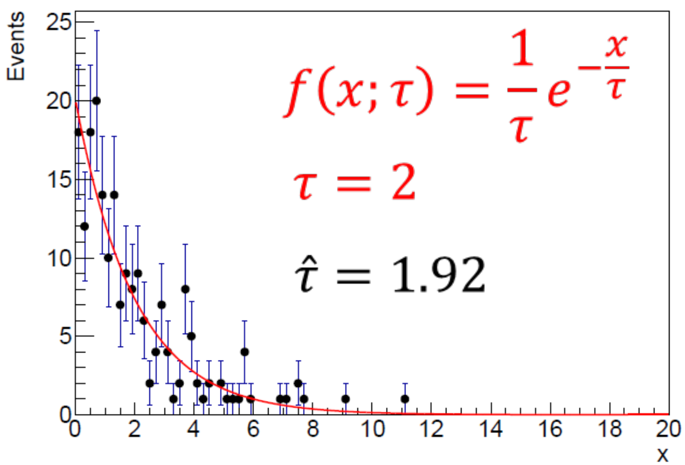
$$\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{and} \qquad V[\hat{\tau}] = (\dots) = \frac{\tau^2}{N}$$

2) Monte-Carlo method

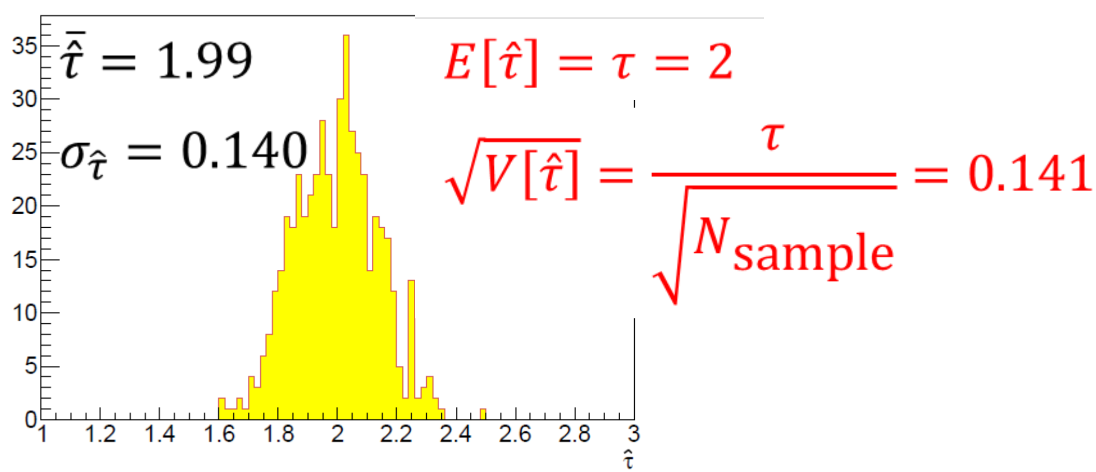
Very useful for complex cases (multiparameters, systematic uncertainties)

Ex: generate samples distributed exponentially

$N_{\text{sample}} = 200$



$N_{\text{experiments}} = 500$



Uncertainty of ML estimator

3) Cramér-Rao bound

Gives a lower bound on any estimator variance (not only ML)

$$\boxed{V[\theta] \geq \frac{\left(1 + \frac{\partial b}{\partial \theta}\right)^2}{E\left[-\frac{\partial^2 \log L}{\partial \theta^2}\right]}, (b\text{: bias})}$$

Equality: estimator is **efficient**

ML are asymptotically efficient

For multiple parameters $\vec{\theta} = \{\theta_1, \ldots, \theta_P\}$: $(V^{-1})_{ij} = E\left[-\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}\right]$

(and assuming efficiency and b=0)

For large samples: an estimate of the inverse covariant matrix V^{-1} is: $\left(\widehat{V^{-1}}\right)_{ij} = -\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}(\theta = \hat{\theta})$

1 parameter: $\boxed{\widehat{\sigma^2} = \frac{-1}{\frac{\partial^2 \log L}{\partial \theta^2}(\hat{\theta})}}$

4) Graphical method

Taylor expansion of log L on estimate :

$$\log L(\theta) = \log L(\hat{\theta}) + (\theta - \hat{\theta}) \frac{\partial \log L}{\partial \theta}(\hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^2 \frac{\partial^2 \log L}{\partial \theta^2}(\hat{\theta})$$

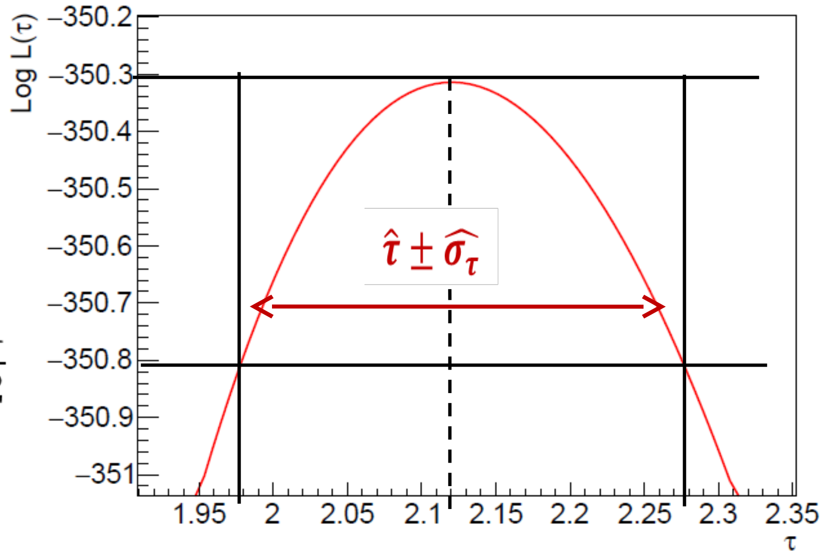
$$= \log L_{\max} - \frac{1}{2\widehat{\sigma^2}}(\theta - \hat{\theta})^2$$

$$\boxed{\Rightarrow \log L(\hat{\theta} \pm \hat{\sigma}) = \log L_{\max} - \frac{1}{2}}$$

$\hat{\tau} \pm \widehat{\sigma_\tau}$ **corresponds to a 68.3% confidence interval**

$\log L_{\max}$

$\log L_{\max} - \frac{1}{2}$

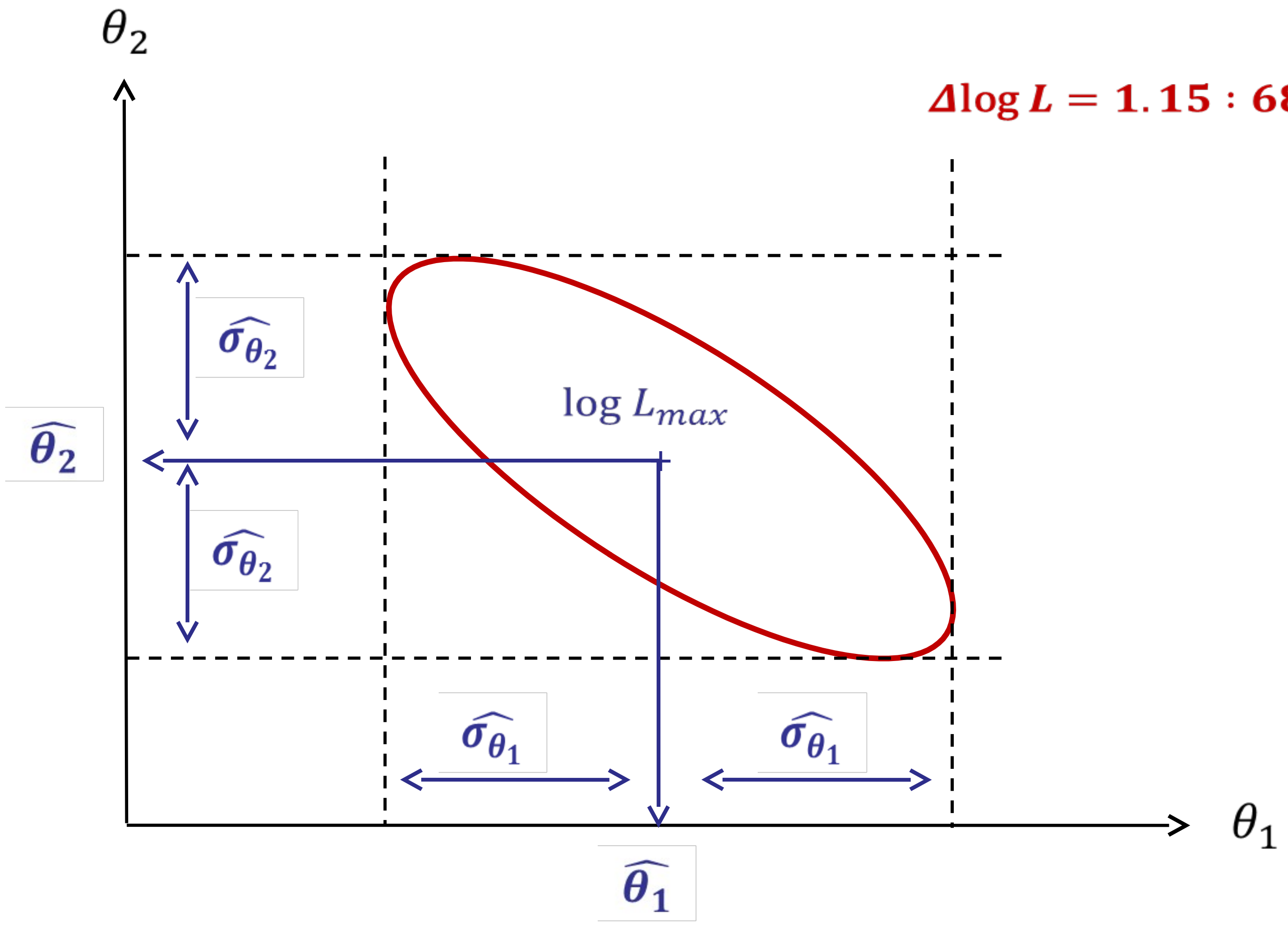


$\Delta \log L = 0.5 : 68.3\%$ **CI**

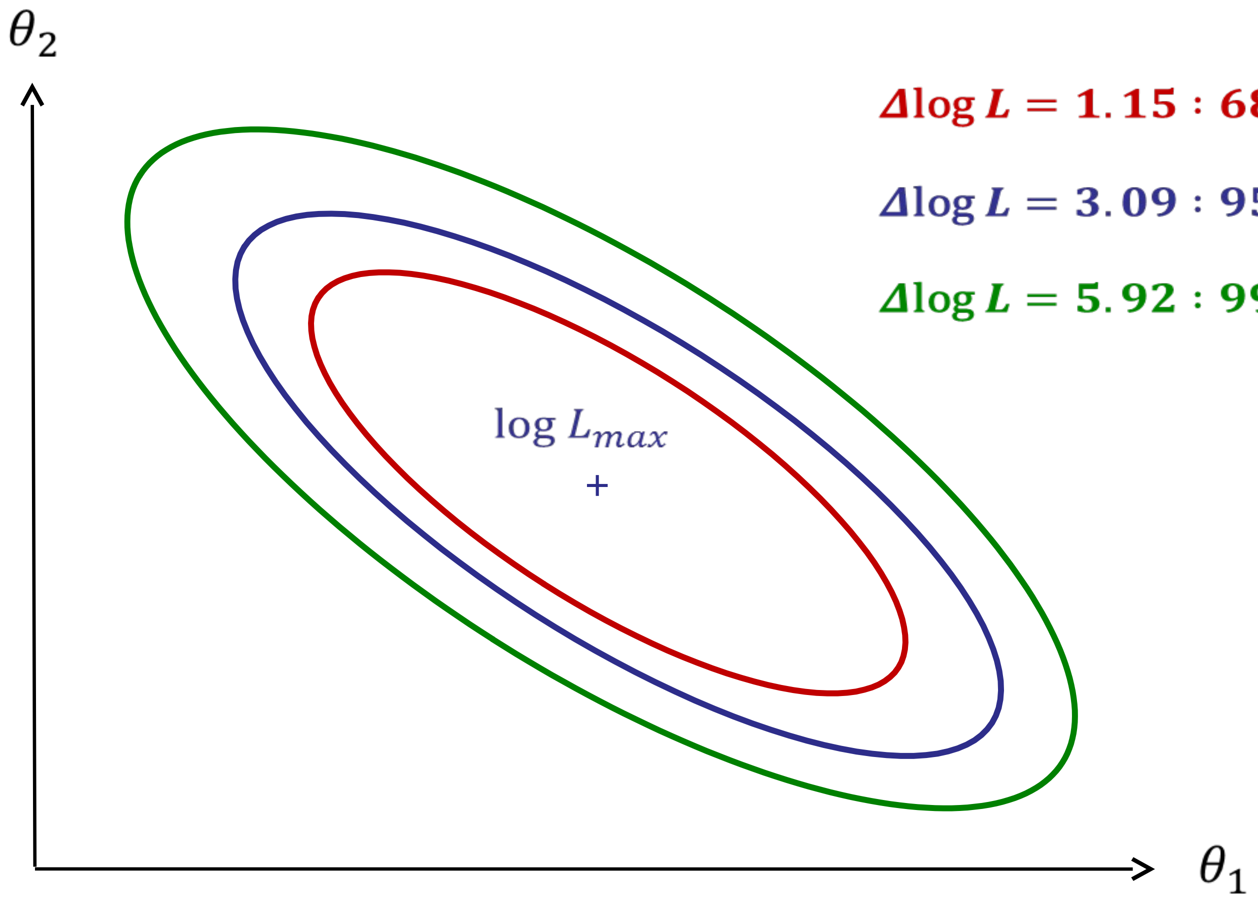
$\Delta \log L = 2 : 95.4\%$ **CI**

$\Delta \log L = 4.5 : 99.7\%$ **CI**

Case for 2 parameters θ_1 and θ_2:

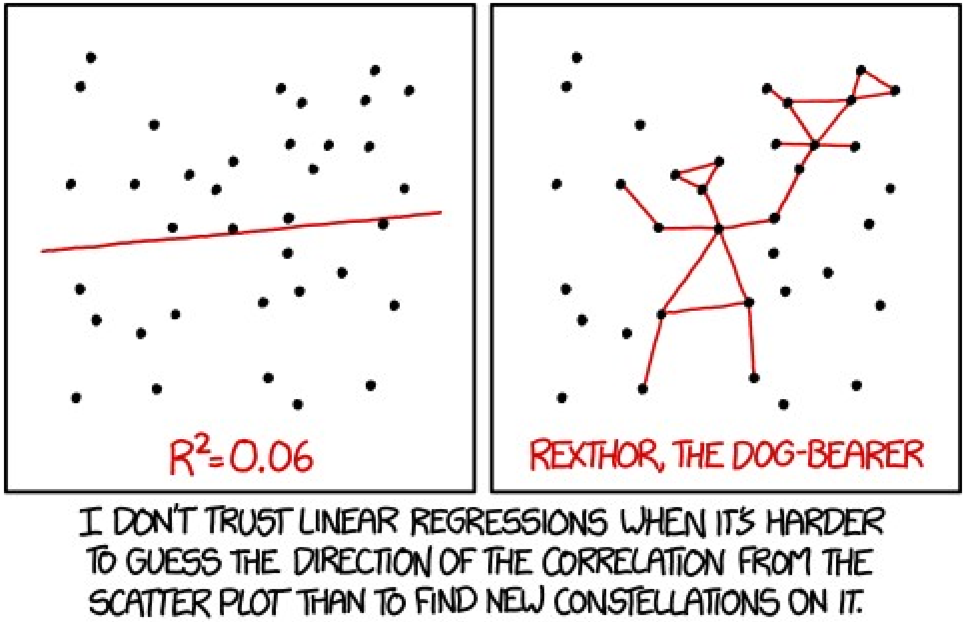


Case for 2 parameters θ_1 and θ_2:



Likelihood, regression, chi2-method

[xkcd.com]



N observations of **variable** $\mathbf{x} \in \mathbb{R}^p$ and **target** $\mathbf{y} \in \mathbb{R}^q$

Linear model: $\mathbf{y} = \mathbf{f}(\mathbf{W}, \mathbf{x}) = \mathbf{W}^T \mathbf{x} + \mathbf{w_0}$ $\mathbf{W}$: parameters

$$\text{for q=1:} \;\; f(\mathbf{x}) = \sum_{i=1}^{p} w_i x_i + w_0$$

Linear basis function models: apply **M** non-linear functions ϕ to features **x**

$$\mathbf{x} \longrightarrow \begin{pmatrix} \phi_1(\mathbf{x}) \\ \vdots \\ \phi_M(\mathbf{x}) \end{pmatrix}$$

$\phi_j(\mathbf{x})$: **basis function**

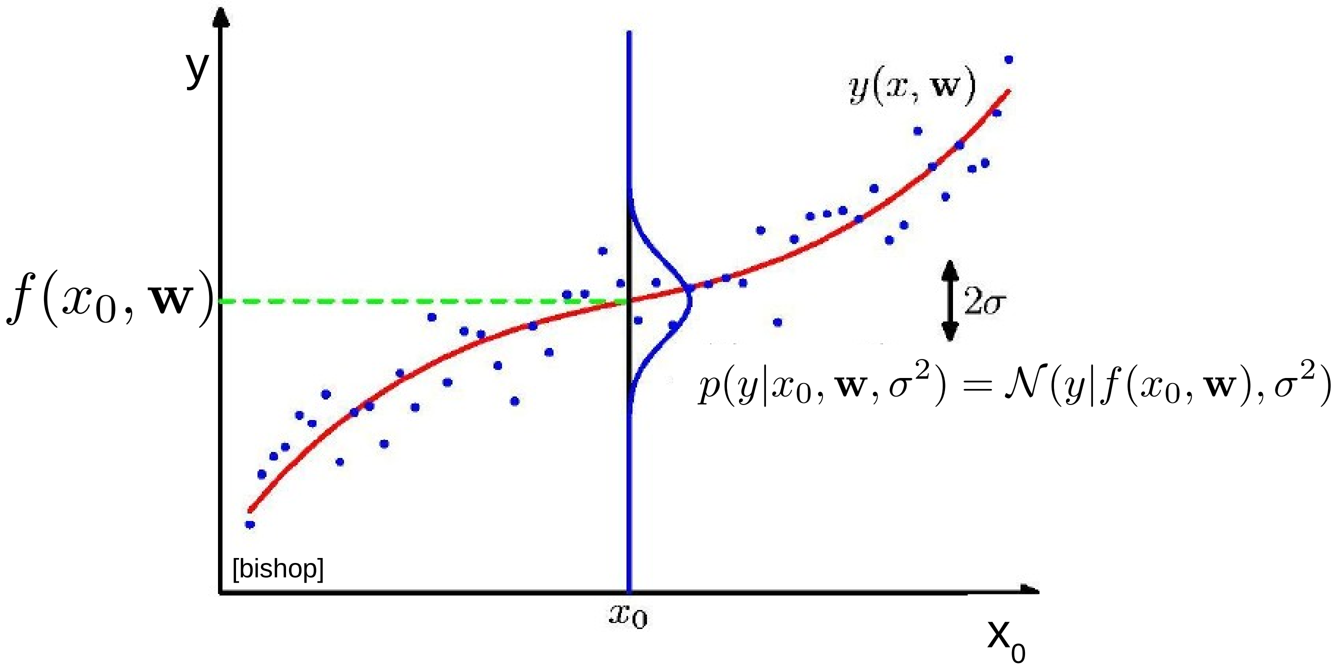
$\phi_j : \mathbb{R}^p \to \mathbb{R}$

$$\mathbf{f}(\mathbf{W}, \mathbf{x}) = \mathbf{W}^T \mathbf{\Phi}(\mathbf{x}) + \mathbf{w_0}$$

$$\text{for q=1:} \;\; f(\mathbf{x}, \mathbf{w}) = \sum_{j=1}^{M} w_j \phi_j(\mathbf{x}) + w_0$$

Likelihood and regression

Say we want to fit the **data** shown below using a **linear model** $f(x, \mathbf{w})$



Let's assume target value, **y**, is subject to **Gaussian noise** σ

We can construct a **predictive probabilistic model** as:

$$p(y|x, \mathbf{w}, \sigma^2) = \mathcal{N}(y|f(x, \mathbf{w}), \sigma^2)$$

Likelihood and regression

The **model parameters** (**w**, σ) are determined by maximizing the **likelihood**

$$\log(\mathbf{w}, \sigma^2) = \sum_{i=1}^{N} \log \mathcal{N}(y_i | f(x_i, \mathbf{w}), \sigma^2)$$

This is equivalent as minimizing the **sum of square error** E(**w**) to determine $\mathbf{w}_{ML}$

$$E(\mathbf{w}) = \frac{1}{N} \sum_{i=1}^{N} (f(x_i, \mathbf{w}) - y_i)^2$$

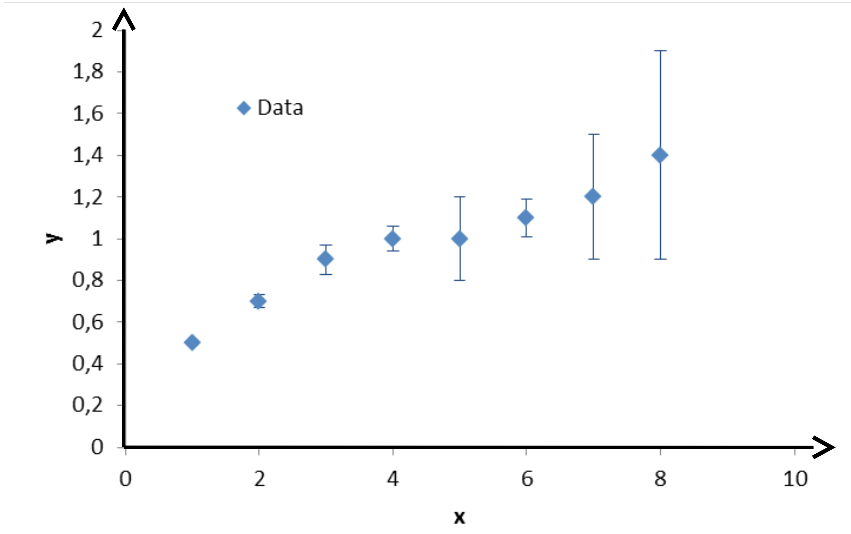
And the **noise** parameter is given by:

$$\sigma_{ML}^2 = \frac{1}{N} \sum_{i=1}^{N} (f(x_i, \mathbf{w_{ML}}) - y_i)^2$$

Chi-square method

Consider N independent variables $\mathbf{y_i}$ function of a another variable $\mathbf{x_i}$

- The y_i are **Gaussian** distributed of mean $\boldsymbol{\mu_i}$ and (known) std $\boldsymbol{\sigma_i}$
- Suppose that $\boldsymbol{\mu} = \boldsymbol{f}(\boldsymbol{x}; \vec{\boldsymbol{\theta}})$ with unknow parameters $\vec{\boldsymbol{\theta}}$



Likelihood: $L(\vec{\theta}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\left(\frac{y_i - f(x_i;\vec{\theta})}{\sigma_i}\right)^2}$

Maximizing $\log L(\vec{\theta})$ to estimate parameters $\vec{\theta}$ is equivalent to **minimize**:

$$\boxed{\chi^2(\vec{\theta}) = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i;\vec{\theta})}{\sigma_i}\right)^2}$$

Simple example

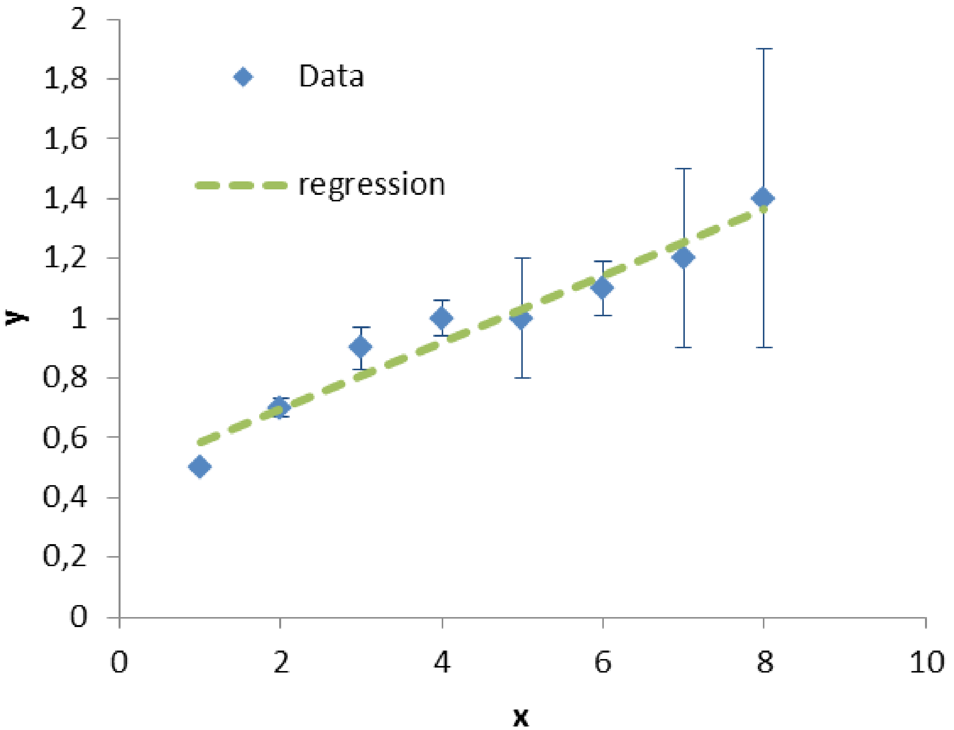
Fit data with a line $f(x;a,b) = ax + b$

Simple **linear regression**: minimize the variance of $y_i - f(x_i;a,b)$

$$w(a,b) = \sqrt{\frac{1}{n}\sum_i (y_i - (ax_i + b))^2}$$

$$\begin{cases} \dfrac{\partial w(a,b)}{\partial a} = 0 \\ \dfrac{\partial w(a,b)}{\partial b} = 0 \end{cases}$$

$$\begin{cases} a = \dfrac{\mathrm{cov}(x,y)}{\mathrm{var}(x)} = r\dfrac{\sigma(y)}{\sigma(x)} \\ b = \bar{y} - r\dfrac{\sigma(y)}{\sigma(x)}\bar{x} \end{cases}$$



(r: correlation factor between x and y)

Simple example

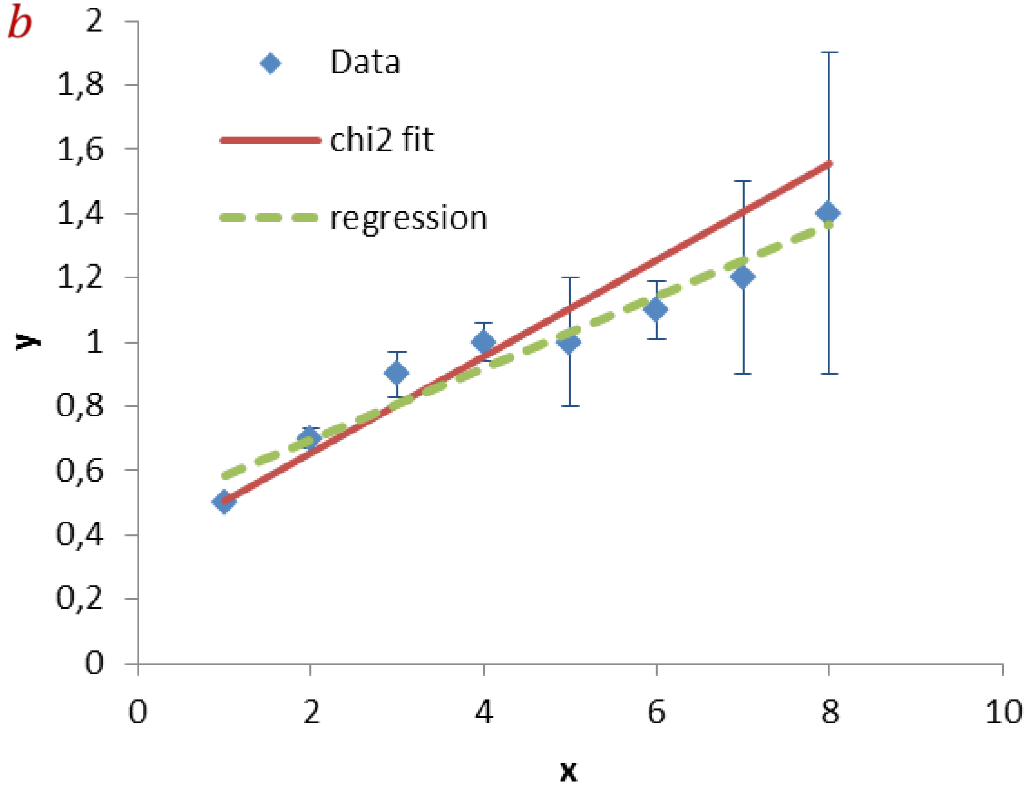
Fit data with a line $f(x; a, b) = ax + b$

Chi-square fit: minimize $\chi^2(a, b)$

$$\chi^2(a,b) = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i; a, b)}{\sigma_i} \right)^2$$

$$\frac{\partial \chi^2}{\partial a} = 0 \qquad \frac{\partial \chi^2}{\partial b} = 0$$

$$a = \frac{AE - DC}{BE - C^2} \quad b = \frac{DB - AC}{BE - C^2}$$



$$A = \sum_i \frac{x_i y_i}{(\Delta y_i)^2}, \quad B = \sum_i \frac{x_i^2}{(\Delta y_i)^2}, \quad C = \sum_i \frac{x_i}{(\Delta y_i)^2}, \quad D = \sum_i \frac{y_i}{(\Delta y_i)^2}, \quad E = \sum_i \frac{1}{(\Delta y_i)^2}$$

Chi-square: generalization

If $\mathbf{y}_i$ measurements are not independent but related by their cov. matrix V_{ij}

$$\log L(\vec{\theta}) = -\frac{1}{2}\sum_{i,j=1}^{N}(y_i - f(x_i;\vec{\theta}))(V^{-1})_{ij}(y_j - f(x_j;\vec{\theta})) + \text{additive terms}$$

$\log L(\vec{\theta})$ is maximized by minimizing:

$$\chi^2(\vec{\theta}) = \sum_{i,j=1}^{N}(y_i - f(x_i;\vec{\theta}))(V^{-1})_{ij}(y_j - f(x_j;\vec{\theta}))$$

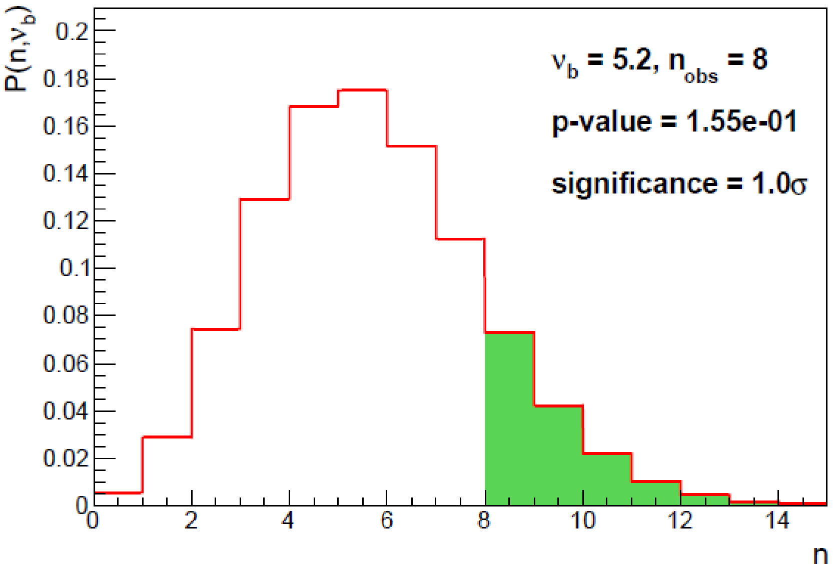
Written in matrix notation: $\boxed{\chi^2(\vec{\theta}) = \left(\vec{y} - \vec{f}\right)^T V^{-1}(\vec{y} - \vec{f})}$

If $f(x_i;\vec{\theta})$ is linear in the parameters $\vec{\theta}$: 1-σ uncertainty contour given by:

$$\boxed{\chi^2(\vec{\theta}) = \chi^2\left(\hat{\vec{\theta}}\right) + 1 = \chi^2_{min} + q}$$

N param.	1	2	3
q	1.00	2.30	3.53

Test hypothesis



Testing compatibility of observed data against a model

- **model** = background predictions (for simplicity)
  → n_b events: follows **Poisson** distribution of mean ν_b
- **data: n_{obs} observed** events

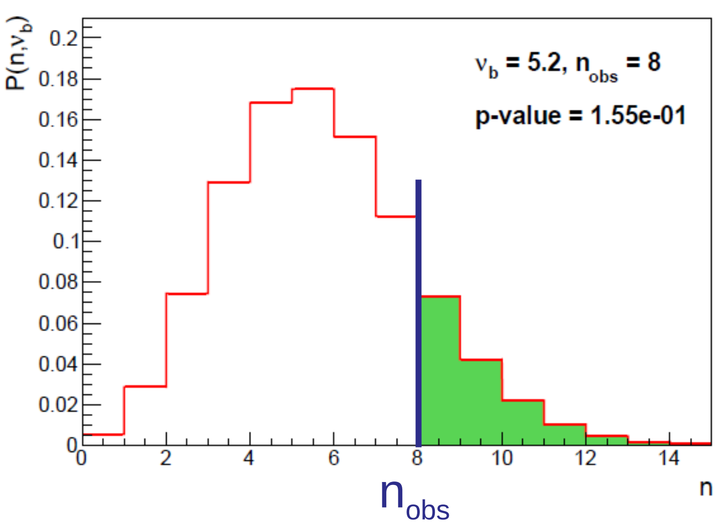
To quantify **degree of compatibility** of n_{obs} with the background-only hypothesis we calculate how likely it is to find n_{obs} or more events of background

p-value: probability that the expected number of event (background) is at least as high as the number of observed data

$$\boxed{\text{p-value} = P(n \geq n_{obs}) = 1 - P(n < n_{obs})}$$

$$= \sum_{n=n_{obs}}^{+\infty} \frac{e^{-\nu_b}\nu_b^n}{n!} = 1 - \sum_{n=0}^{n_{obs}-1} \frac{e^{-\nu_b}\nu_b^n}{n!}$$

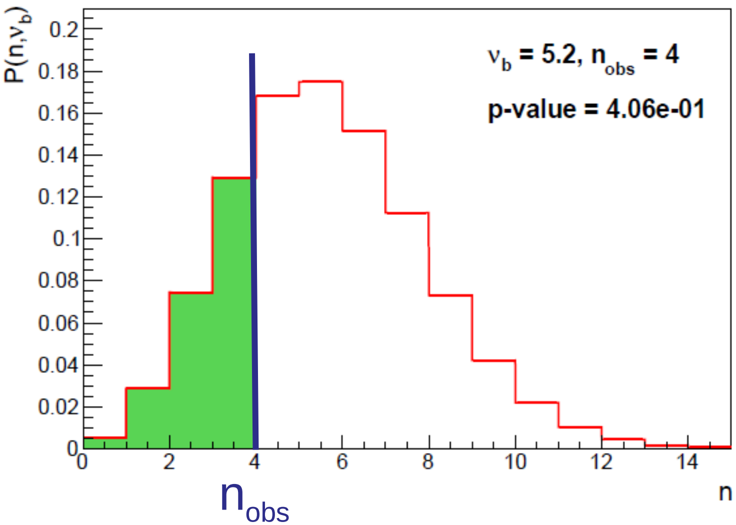
$$[\text{for } \nu_b < n_{obs}]$$



Test hypothesis

For the case where ν_b > n_{obs} one can define:

$$\text{p-value} = \sum_{n=0}^{n_{obs}} \frac{e^{-\nu_b}\nu_b^n}{n!}$$



The previous sums can be **simplified** using incomplete **Gamma** functions:

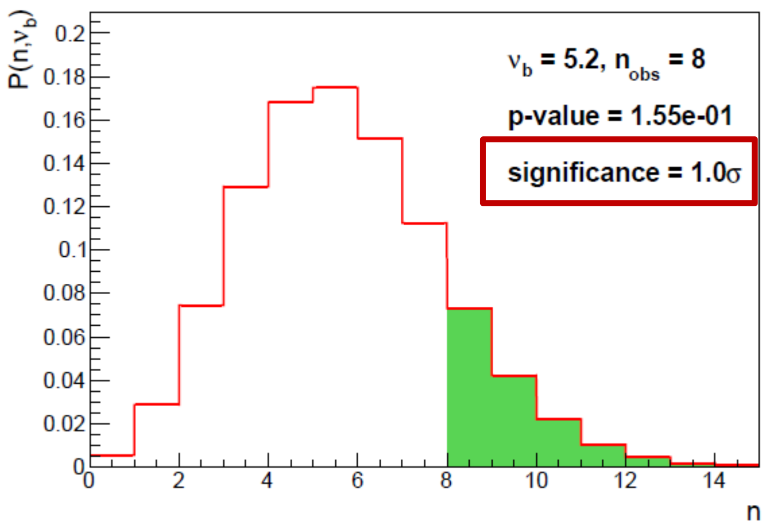
$$\boxed{\sum_{n=n_{obs}}^{+\infty} \frac{e^{-\nu_b}\nu_b^n}{n!} = \frac{1}{\Gamma(n_{obs})}\int_0^{\nu_b} t^{n_{obs}-1}e^{-t}dt = \Gamma(\nu_b, n_{obs})}$$

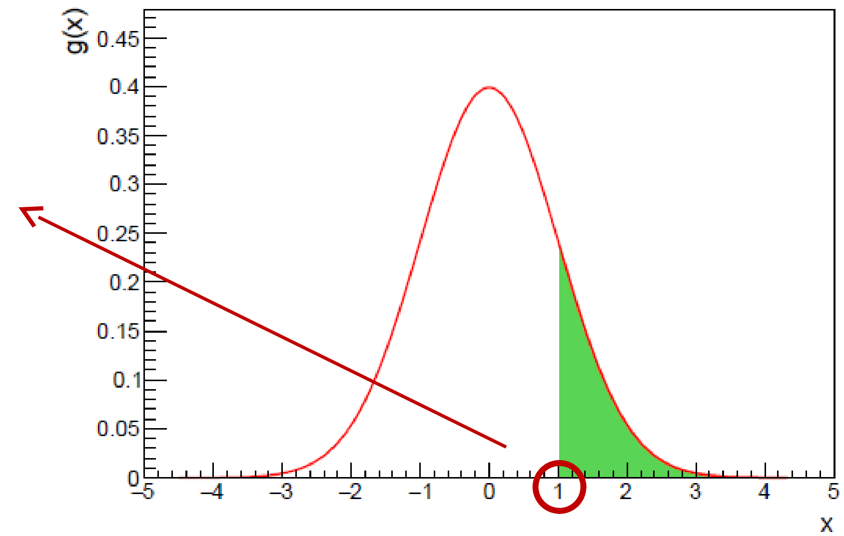
$$\text{with } \Gamma(n_{obs}) = \int_0^{\infty} t^{n_{obs}-1}e^{-t}dt = (n_{obs}-1)! \quad (\text{if } n_{obs} \text{ integer})$$

It is customary to transform the p-value into a **Z-value** using the integral of the Gaussian distribution:

$$\int_{-\infty}^{Z} \mathrm{Gaus}(x, \mu = 0, \sigma = 1)dx = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \ = 1 - \mathrm{pvalue}$$

Z-value = number of standard deviation, used as a measure of the **significance** of an excess (or a deficit) w.r.t the (background) hypothesis.

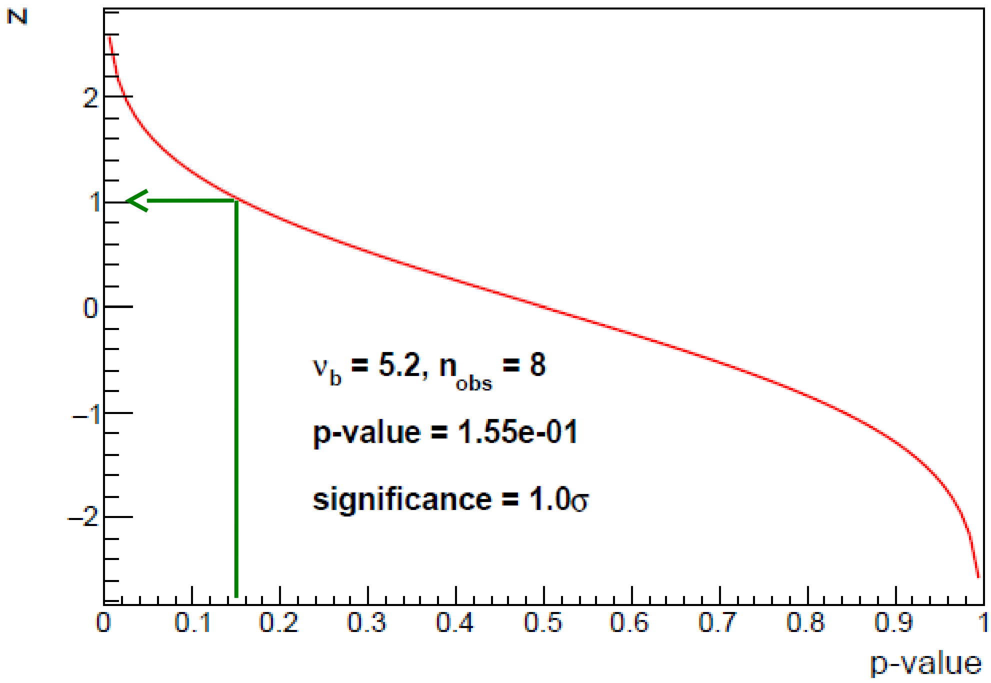




Significance

In practice one uses the **inverse cumulative distribution function** of the Gaussian distribution to compute the significance:

$$Z = \sqrt{2}\mathrm{Erf}^{-1}(1 - 2 \times \text{p-value})$$



p-value	Z	
0.159	1σ	
2.28×10^{-2}	2σ	
1.35×10^{-3}	**3σ**	(evidence)
3.15×10^{-5}	4σ	
2.85×10^{-7}	**5σ**	(discovery)

BumpHunter algorithm

Search for excess or deficit in a spectrum

G. Choudalakis
1101.0390

- **No assumptions** are made on the signal shape or yield
- Just test **data** against **background-only hypothesis**

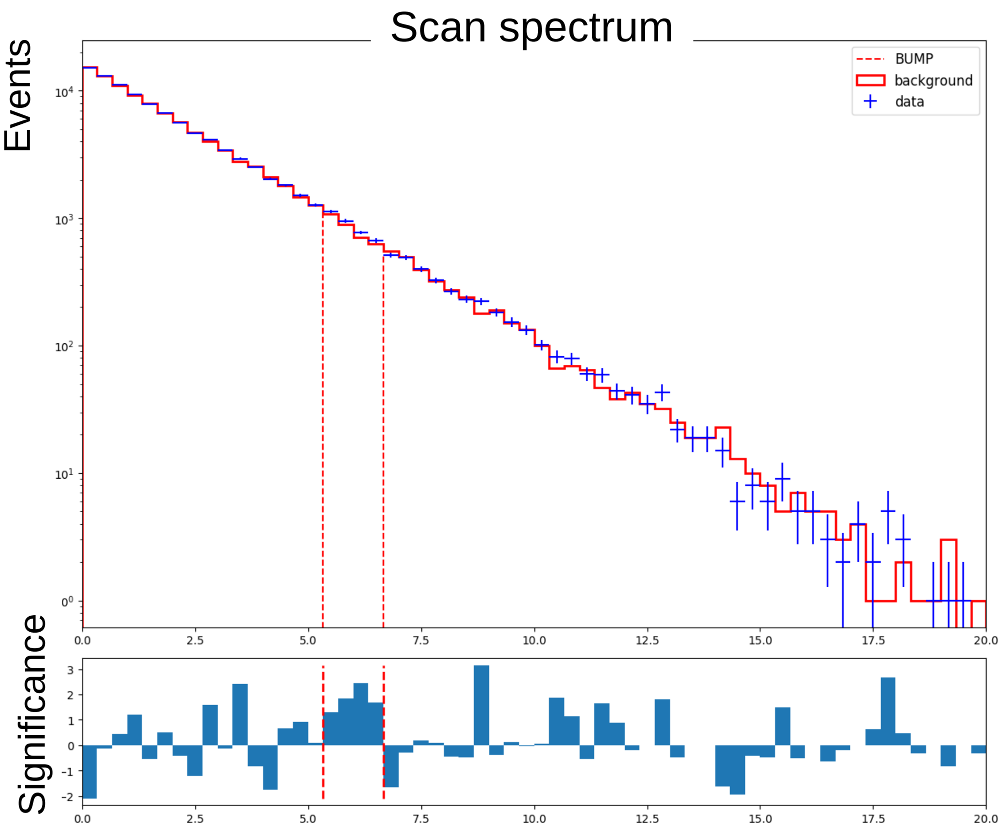
NEW

Python implementation (L. Vaslin): https://pypi.org/project/pyBumpHunter/

➔ Compute the p-value for **all** possible intervals.

➔ Select the **interval** with **smallest** p-value.

This gives the local p-value: $p_{\mathrm{min}}^{\mathrm{local}}$



BumpHunter algorithm

Search for excess or deficit in a spectrum

G. Choudalakis
1101.0390

- **No assumptions** are made on the signal shape or yield
- Just test **data** against **background-only hypothesis**

NEW

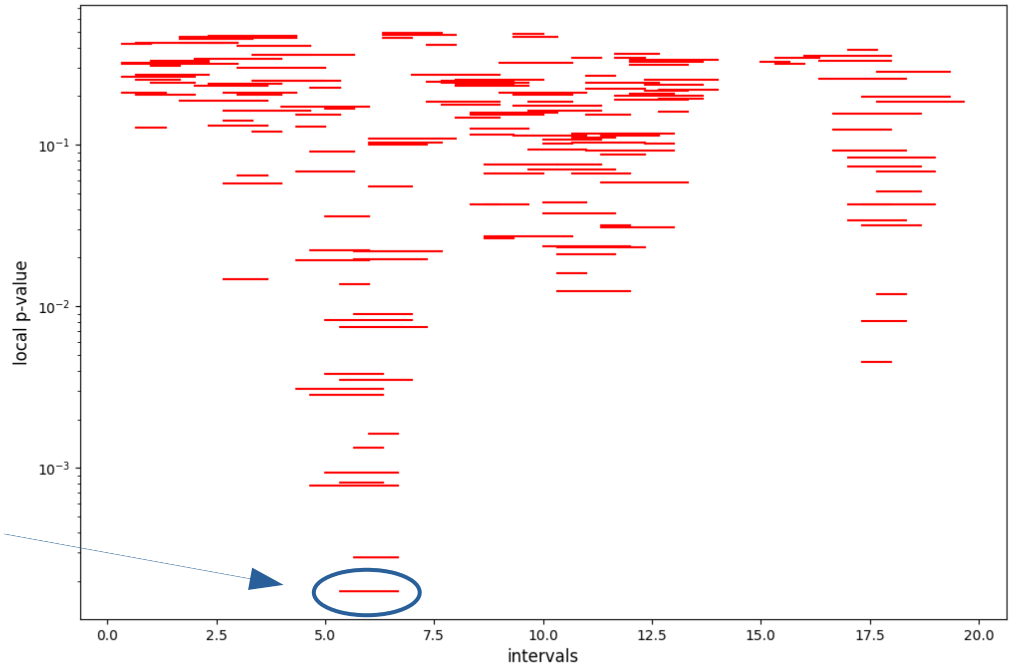
Python implementation (L. Vaslin): https://pypi.org/project/pyBumpHunter/

➔ Compute the p-value for **all** possible intervals.

➔ Select the **interval** with **smallest** p-value.

This gives the local p-value: $p_{\mathrm{min}}^{\mathrm{local}}$

Tomography scan

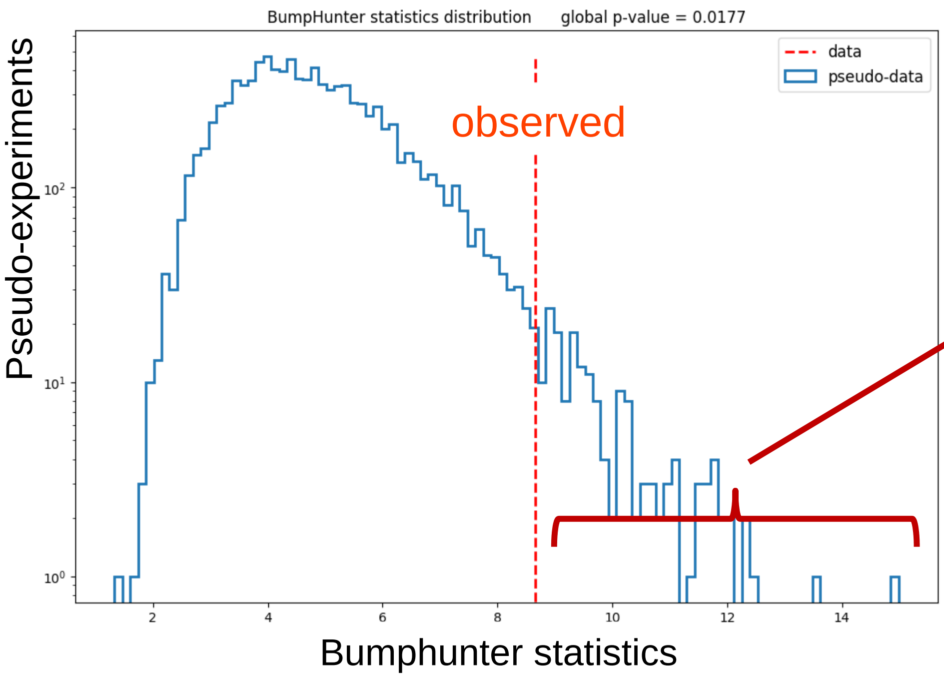


BumpHunter algorithm

Since **many intervals** are considered there is a increasing probability that an excess is **found** due to statistical **fluctuations**

- This is the (in)famous (and misnamed) **Look Elsewhere Effect: LEE**
- To cope for this effect a **global p-value** is calculated

➔ The global p-value is extracted by comparing $-\log(\mathrm{p}_{\mathrm{min}}^{\mathrm{local}})$ to a set of $-\log(\mathrm{p}_{\mathrm{min}}^{\mathrm{local}})$ generated using background-only **pseudo-experiments**



$\mathrm{p}^{\mathrm{global}}$: **fraction of PE** that gives a result higher than the one observed (p-value of p-value !)

$$\mathrm{p}^{\mathrm{global}} = \text{fraction of } (\mathrm{P}_{\mathrm{min}}^{\mathrm{PE}} > \mathrm{P}_{\mathrm{min}}^{\mathrm{obs}})$$

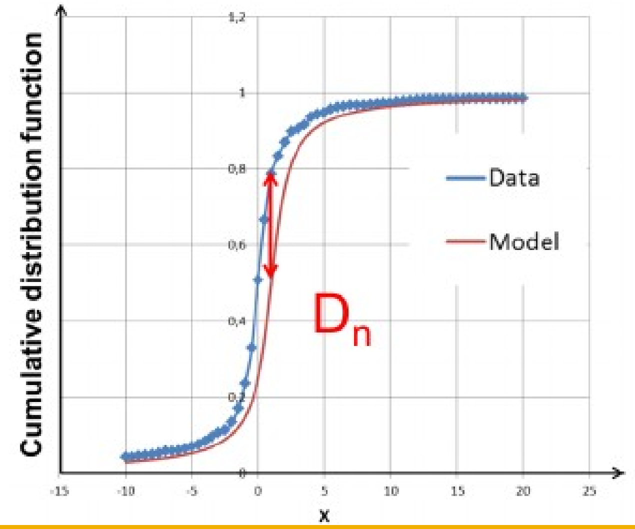
The **KS** test is an **unbinned** method that uses **all the measured values** of variable x to test the compatibility of the data to a model.

- The **M** measured values $\mathbf{x_i}$ are first sorted in ascending order: $x_1<x_2<\ldots<x_M$
- The sample **cumulative distribution** is calculated as:

$$F_{\text{data}}(x) = \begin{cases} 0 \text{ if } x \le x_1 \\ \text{i/M if } x_i \le x < x_{i+1} \\ 1 \; if \; x \ge x_M \end{cases}$$

The test compares **cumulative distribution** of the sample to that of the model. The **maximum distance $\mathbf{D_n}$** between the two is the test statistics:

$$D_n = \sup_x |F_{\text{model}}(x) - F_{\text{data}}(x)|$$



The **p-value** of the KS test is given (for large M) by:

$$\text{p-value} = 2\sum_{r=1}^{+\infty}(-1)^{r-1}e^{-2Mr^2D_n^2}$$

Kolmogorov-Smirnov test - Example

Exponential p.d.f

$$f(x;\lambda) = \lambda e^{-\lambda x}, x > 0$$

- Data: λ=0.4 (500 events)
- Model: λ=0.35

$$F_{\text{data}}(x) = \begin{cases} 0 & x \leq x_1 \\ i/n & x_i \leq x < x_{i+1} \\ 1 & x \geq x_M \end{cases}$$

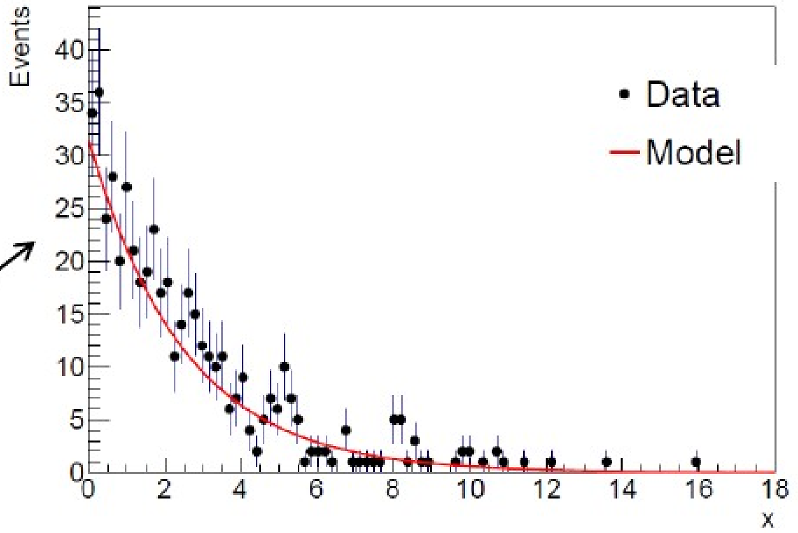
$$F_{\text{model}}(x) = 1 - e^{-\lambda x}$$

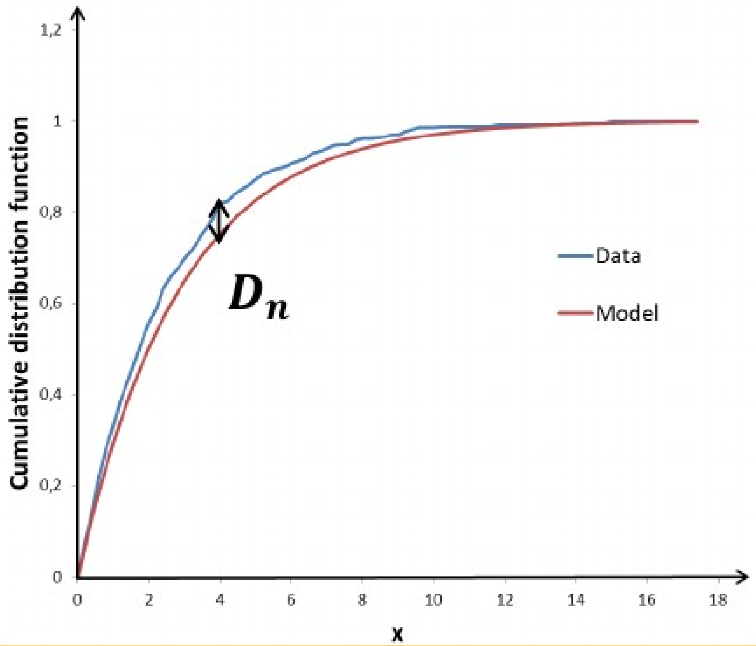
Max distance between cumulative distributions:

$$D_n = 0{,}0646$$

➔ p-value = 0,03

x_i
0,011401647
0,017623018
0,018095279
0,020447056
0,02616019
0,026849926
0,029898988
0,044689801
0,045548065
0,048410584
0,058308293
0,062655827
0,065376242
…
9,312545995
9,335461119
9,378006281
9,40176752
9,450497283
10,04570365
11,78017539
13,57118477
15,80234274





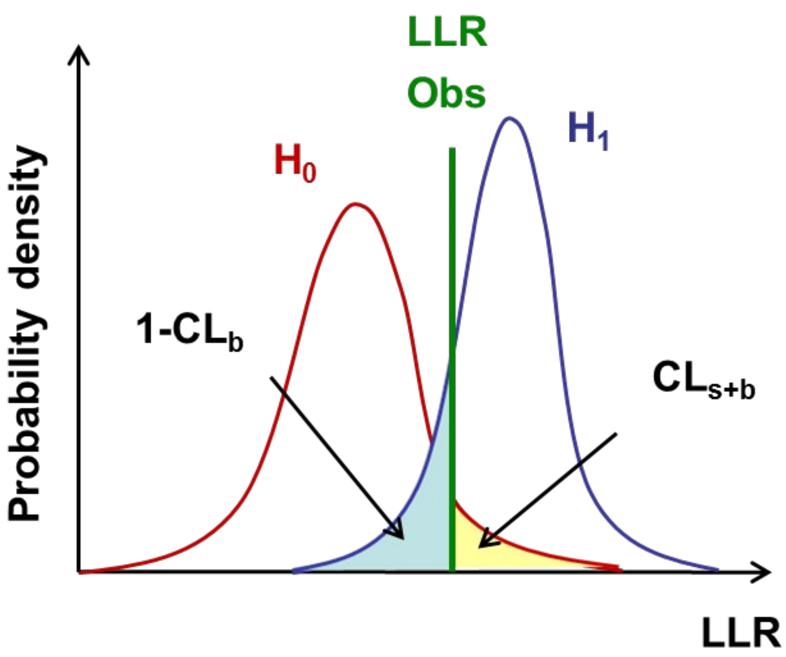
Hypothesis test: CLs method

Test of two hypothesis **H_0** and **H_1** using data

- **Likelihood** of data given an hypothesis: L(data|H_0) or L(data|H_1)

Neyman-Pearson lemma: optimal **test statistics** for hypothesis testing is given by (log) **likelihood ratio**

$$\mathrm{LLR} = -2\log\frac{L(\mathrm{data}|H_0)}{L(\mathrm{data}|H_1)}$$



$$\int_{LLR_{obs}}^{\infty} f(t|H_0)dt = \mathrm{CL}_{s+b}$$

$$\int_{-\infty}^{LLR_{obs}} f(t|H_1)dt = 1 - \mathrm{CL}_b$$

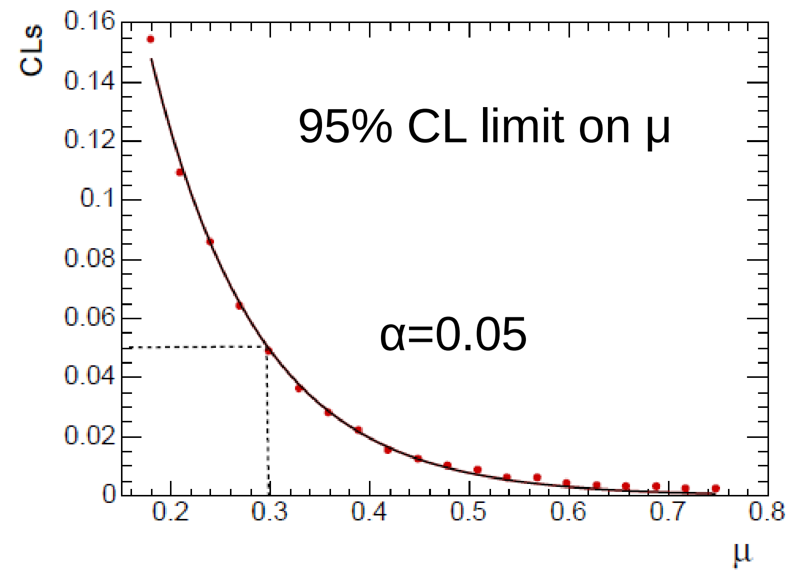
H_0 rejected at (1-α) **confidence level** if $\mathrm{CL}_{s+b} < \alpha$

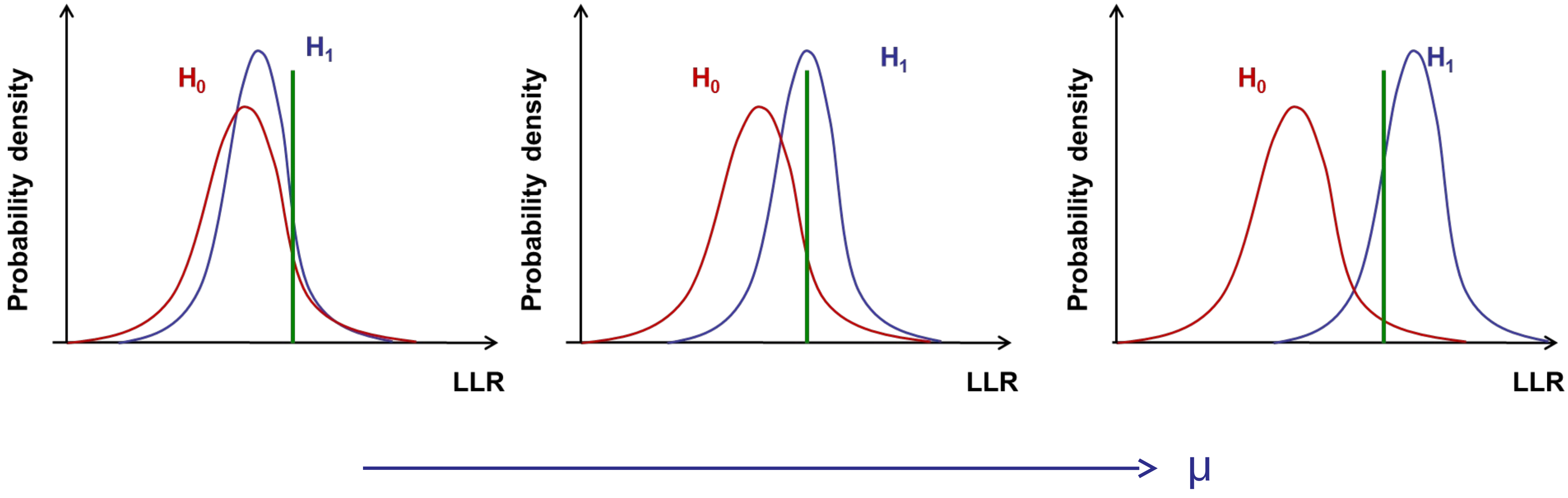
More robust test $\mathrm{CL}_s = \dfrac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} < \alpha$

Hypothesis test: CLs method

Testing **signal strenght** (μ):

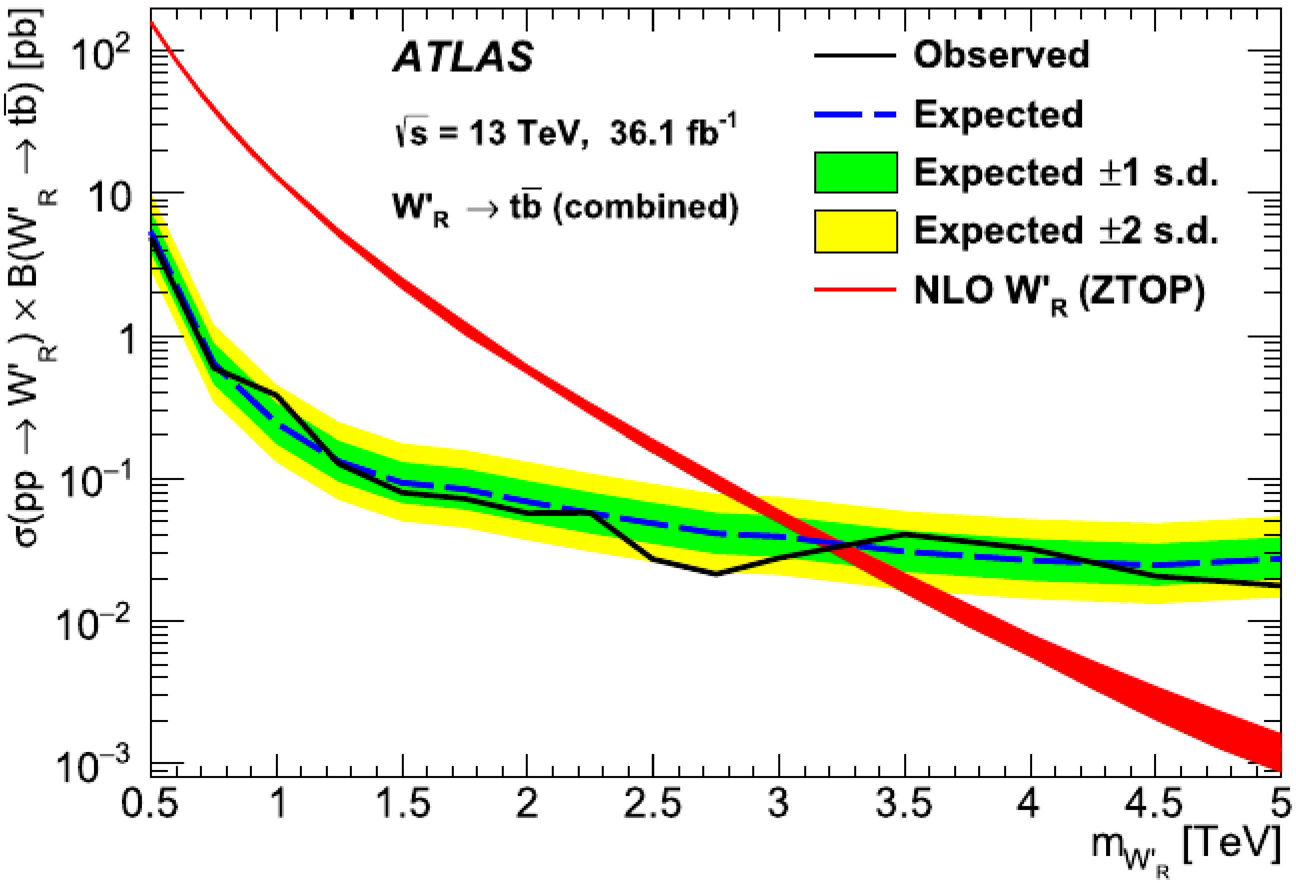
- Express number of event of **signal** as $s = \mu \times s_{nominal}$
- CLs test can be performed for increasing values of μ
- Exclusion limit on μ when CLs<α





Hypothesis test: typical HEP result

Here **95% C.L exclusion limit** on cross-section is calculated for each **signal mass** hypothesis from 0.5 to 5 TeV, for both **observed** data and **expected** background



Conclusion

In this lecture we saw **basic notions** of probability and statistics.

First step towards data **analysis** and statistical **learning**.

Simple notions in general but **easy to forget** !

Easy to **misunderstand** or **mishandle** as well ...

And some concepts are **more complex** than it seem.

Practice these notions by making your **own** calculations and coding !

Bonus track

Combining measurements

Best Linear Unbiased Estimator: L.Lyons et al. NIM A270 (1988) 110

- Find linear (unbiased) combination of results: $x = \Sigma\, w_i x_i$
  with weights w_i that give minimum possible variance σ_x^2
- Account properly of correlations between measurements
- For Gaussian errors: method equivalent to χ^2 minimization

- Two measurements: $x_1 \pm \sigma_1$, $x_2 \pm \sigma_2$ with correlation ρ
- The weights that minimize the χ^2:

Cov. matrix

$$\chi^2 = \begin{pmatrix} x_1 - x & x_2 - x \end{pmatrix} \boxed{\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}}^{-1} \begin{pmatrix} x_1 - x \\ x_2 - x \end{pmatrix}$$

are:

$$w_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \qquad w_2 = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \qquad (w_1+w_2=1)$$

BLUE method

Best Linear Unbiased Estimator: L.Lyons et al. NIM A270 (1988) 110

- Find linear (unbiased) combination of results: $x = \Sigma\, w_i x_i$
  with weights w_i that give minimum possible variance σ_x^2
- Account properly of correlations between measurements
- For Gaussian errors: method equivalent to χ^2 minimization

- Two measurements: $x_1 \pm \sigma_1$, $x_2 \pm \sigma_2$ with correlation ρ
- The combined result is: $x = w_1 x_1 + w_1 x_2$
- And the uncertainty on the combined measurement is:

$$\sigma_x = \sqrt{\frac{\sigma_1^2 \sigma_2^2 (1-\rho^2)}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}}$$

Iterative method

- Biases could appear when uncertainties depend on central value of each measurement (L. Lyons et al., Phys. Rev. D41 (1990) 982985)
- Reduced if covariance matrix determined as if the central value is the one obtained from combination
  - Rescale uncertainties to combined value

    ex: for measurement 1, and category i: $\sigma_{i,1}^{rescaled} = \sigma_{i,1} \, . \, x_1/x_{blue}$
  - Iterate until central value converges to stable value

Single-top t-channel 8 TeV results

ATLAS [ATLAS-CONF-2012-132, 5.8 fb^{-1}]:

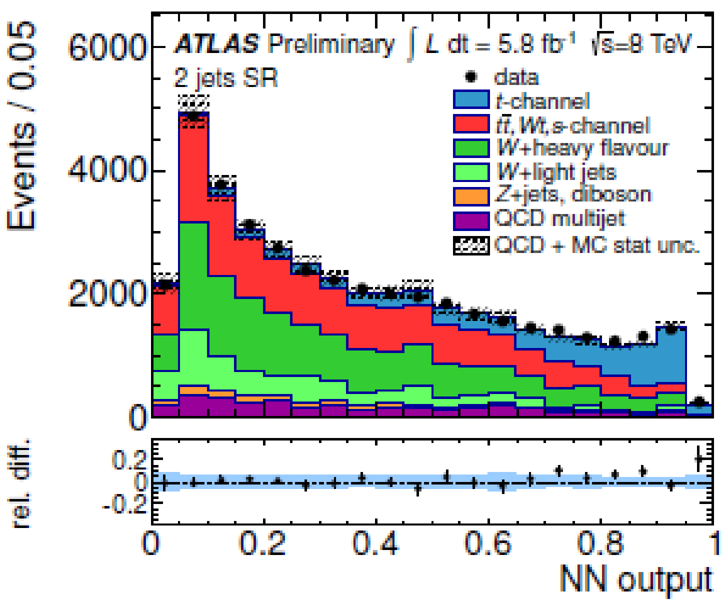
σ_t(t-ch.) = 95 ± 2 (stat.) ± 18 (syst.) pb = 95 ± 18 pb

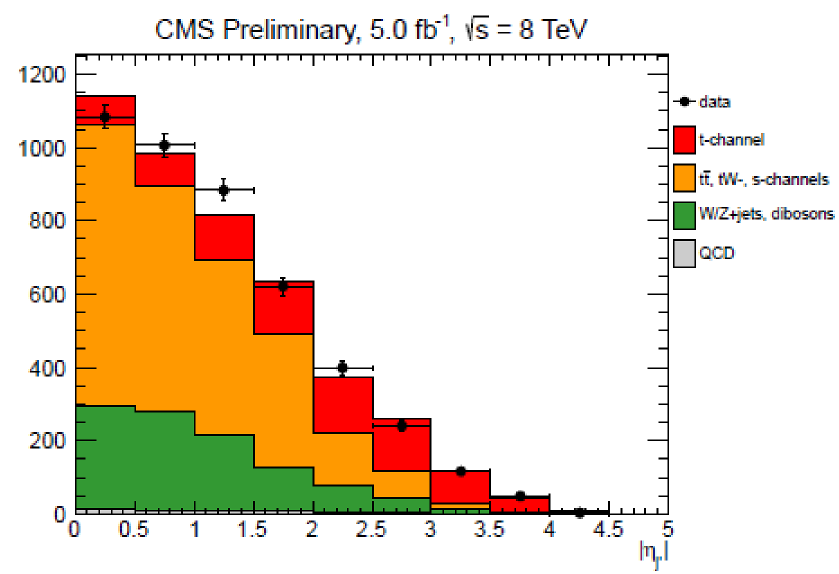
- Multivariate analysis with limited assumptions on simulations
- Fit of **NN distribution** in the data in **e/μ+2/3 jet events, with 1-btag**

CMS [CMS PAS TOP-12-011, 5.0 fb^{-1}]:

σ_t(t-ch.) = 80.1 ± 5.7(stat.) ± 11.0(syst.) ± 4.0(lumi.) pb = 80.1 ± 12.8 pb

- Cut-based analysis, data-driven background estimates (shapes, rates)
- Fit **|η| distribution of forward jet** in **μ+2 jet events, with 1-btag**





Uncertainties categories and correlations

6 categories of uncertainties. Correlation factor between ATLAS/CMS estimated for each.

Category	ATLAS		CMS		ρ
Statistics	Stat. data	2.4%	Stat. data	7.1%	0
	Stat. sim.	2.9%	Stat. sim.	2.2%	0
Total		**3.8%**		**7.5%**	**0**
Luminosity	Calibration	3.0%	Calibration	4.1%	1
	Long-term stability	2.0%	Long-term stability	1.6%	0
Total		**3.6%**		**4.4%**	**0.78**
Simulation and modelling	ISR/FSR	9.1%	Q^2 scale	3.1%	1
	PDF	2.8%	PDF	4.6%	1
	t-ch. generator	7.1%	t-ch. generator	5.5%	1
	$t\bar{t}$ generator	3.3%			0
	Parton shower/had.	0.8%			0
Total		**12.3%**		**7.8%**	**0.83**
Jets	JES	7.7%	JES	6.8%	0
	Jet res. & reco.	3.0%	Jet res.	0.7%	0
Total		**8.3%**		**6.8%**	**0**
Backgrounds	Norm. to theory	1.6%	Norm. to theory	2.1%	1
	Multijet (data-driven)	3.1%	Multijet (data-driven)	0.9%	0
			W+jets, $t\bar{t}$ (data-driven)	4.5%	0
Total		**3.5%**		**5.0%**	**0.19**
Detector modelling	b-tagging	8.5%	b-tagging	4.6%	0.5
	E_T^{miss}	2.3%	Unclustered E_T^{miss}	1.0%	0
	Jet Vertex fraction	1.6%			0
			pile up	0.5%	0
	lepton eff.	4.1%			0
			μ trigger + reco.	5.1%	0
	lepton res.	2.2%			0
	lepton scale	2.1%			0
Total		**10.3%**		**6.9%**	**0.27**
Total uncert.		**19.2%**		**16.0%**	**0.38**

Combined t-channel single-top cross section

Sum covariance matrices in each category to obtain total covariance matrix.

$$\mathbf{C} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$\Big\downarrow \Sigma$$

$$\mathbf{C} = \begin{pmatrix} 269 & 84 \\ 84 & 182 \end{pmatrix} \mathrm{pb}^2$$

Source	Uncertainty (pb)
Statistics	4.1
Luminosity	3.4
Simulation and modelling	7.7
Jets	4.5
Backgrounds	3.2
Detector modelling	5.5
Total systematics (excl. lumi)	11.0
Total systematics (incl. lumi)	11.5
Total uncertainty	12.2

Breakdown of uncertainties

$$\sigma_i^2 = w_1^2\sigma_{i,1}^2 + 2w_1w_2\rho_i\sigma_{i,1}\sigma_{i,2} + w_2^2\sigma_{i,2}^2$$

$$\boxed{\sigma_{\text{t-ch.}} = 85.3 \pm 4.1\,(\text{stat.}) \pm 11.0\,(\text{syst.}) \pm 3.4\,(\text{lumi.})\,\text{pb} = 85.3 \pm 12.2\,\text{pb}}$$

With $w_{ATLAS} = 0.35$ and $w_{CMS} = 0.65$, $\chi^2 = 0.79/1$

Overall correlation of measurements is $\rho_{tot} = 0.38$.

Summary plot