Statistics for HEP, part I

Nicolas Berger (LAPP Annecy)

Statistics in Physics

“Statistics” might make you think of this:

Statistics in Physics

"Statistics" might make you think of this:

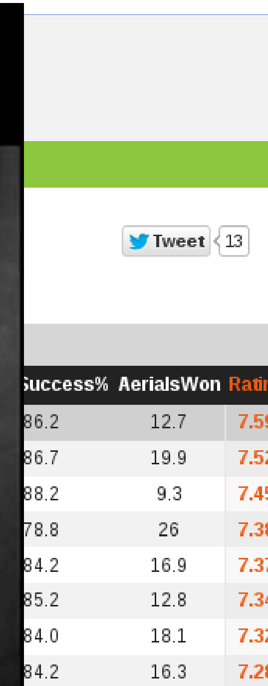
Or maybe this:

R	Team	Tournament	Shots pg	Discipline		Possession%	PassSuccess%	AerialsWon	Rating
1	Real Madrid	La Liga	18.7	40	3	58.0	86.2	12.7	7.59
2	Bayern Munich	Bundesliga	18.7	14	0	70.4	86.7	19.9	7.52
3	Barcelona	La Liga	16.6	35	1	70.2	88.2	9.3	7.45
4	Wolfsburg	Bundesliga	15.9	20	1	54.3	78.8	26	7.38
5	Chelsea	Premier League	15.7	40	2	55.8	84.2	16.9	7.37
6	Juventus	Serie A	17.1	35	4	60.2	85.2	12.8	7.34
7	Arsenal	Premier League	15.9	47	2	56.3	84.0	18.1	7.32
8	Marseille	Ligue 1	15.9	30	3	58.5	84.2	16.3	7.28

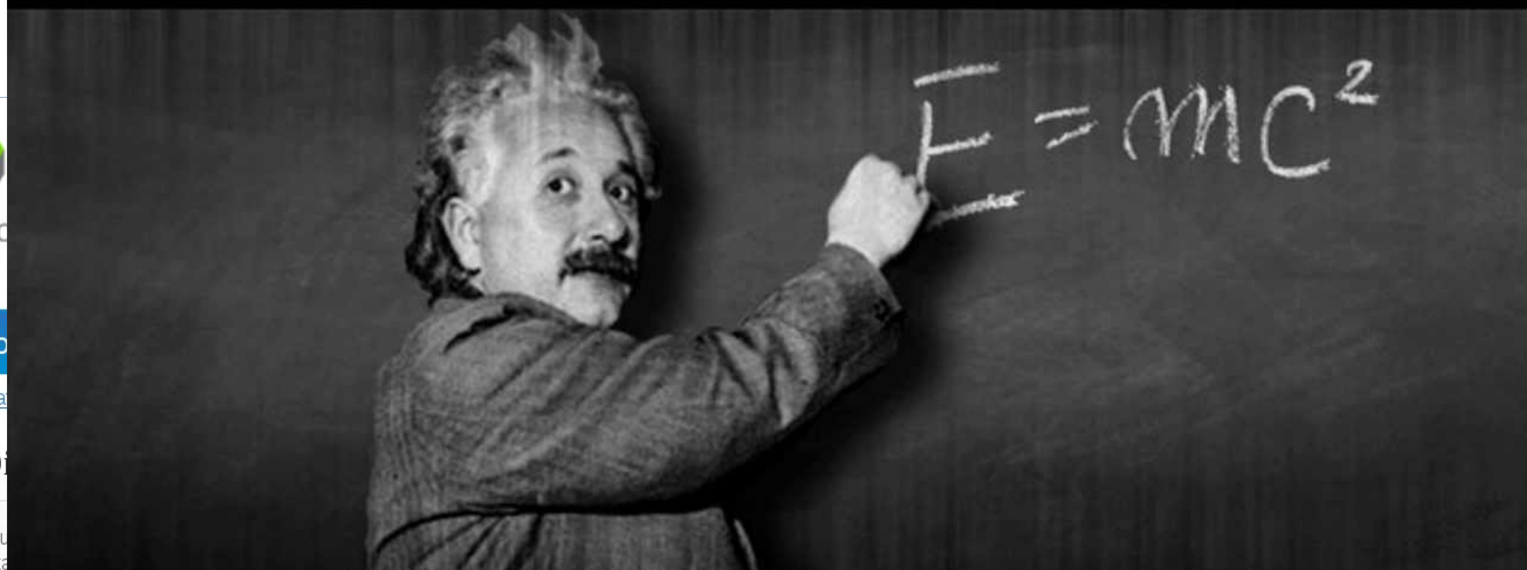
Statistics in Physics

"Statistics" might make you think of this:

Or maybe this:



But probably not this

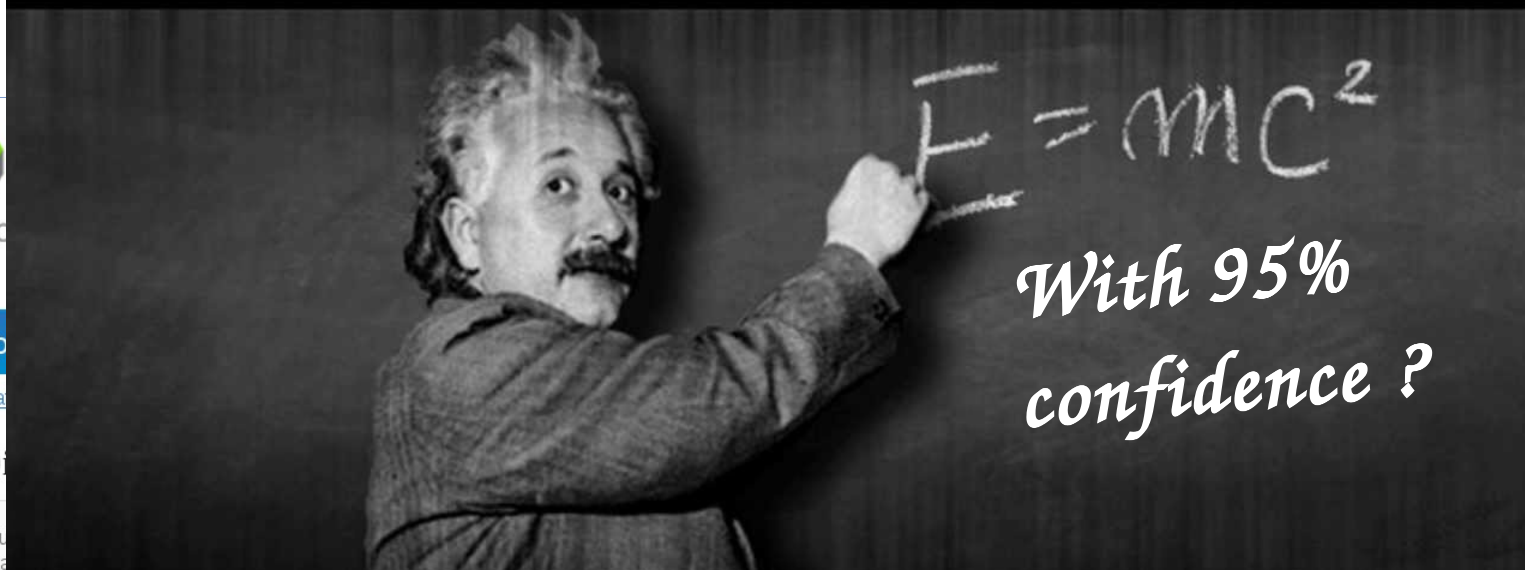


Statistics in Physics

"Statistics" might make you think of this:

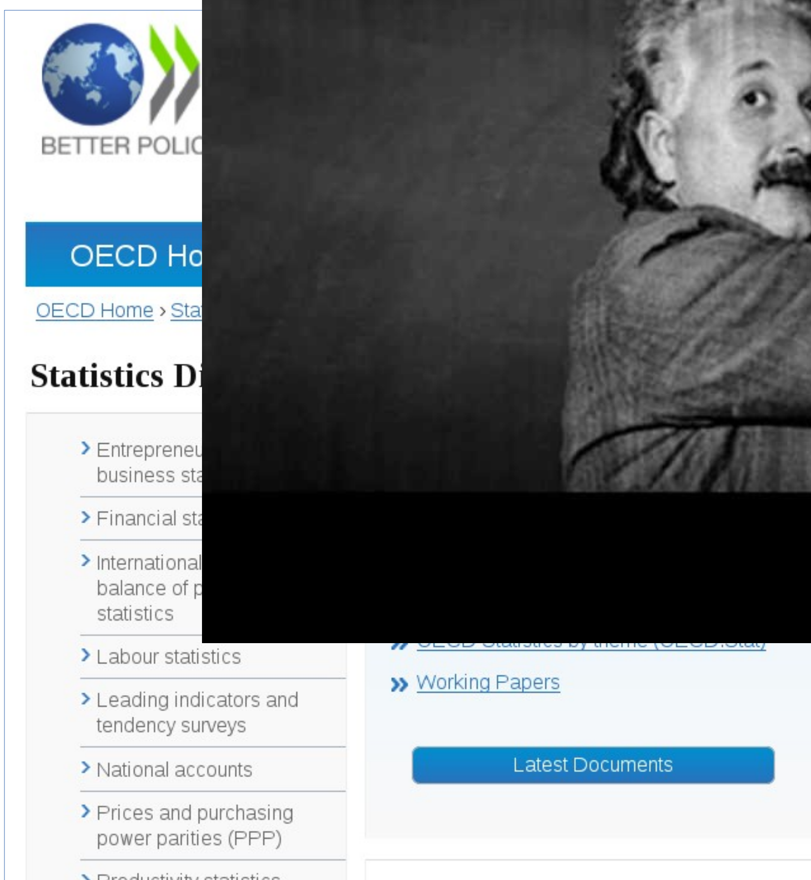
Or maybe this:

But probably not this

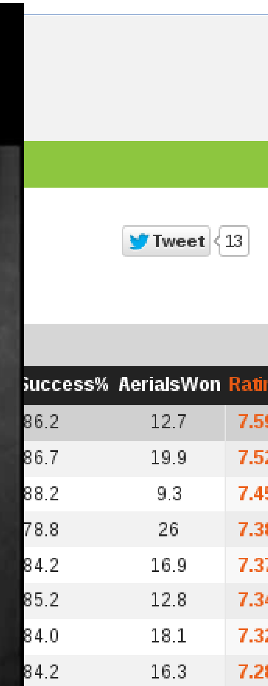


Statistics in Physics

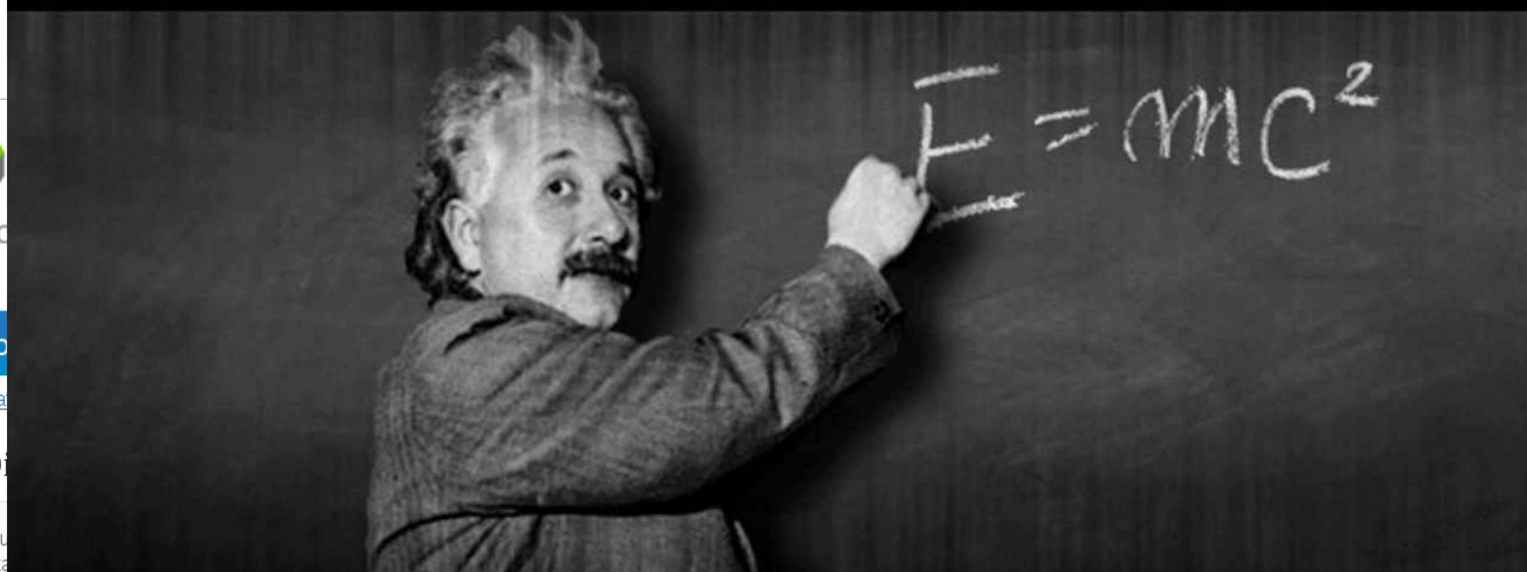
"Statistics" might make you think of this:



Or maybe this:



But probably not this



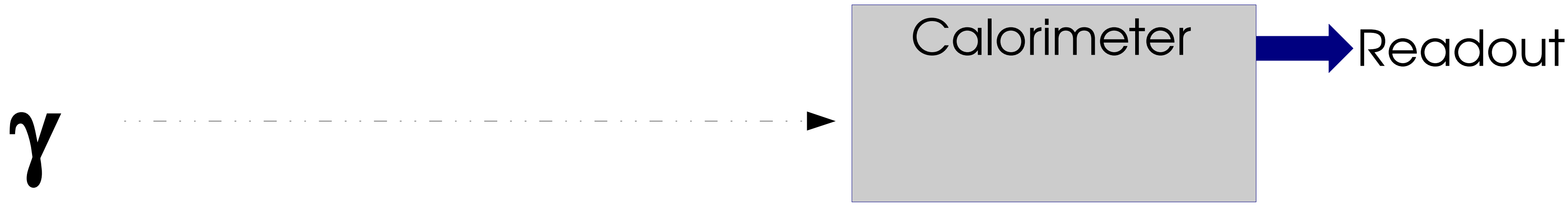
Random Processes

- Statistics is the description of **random** processes. Where does this come into HEP ?
  - **Measurement errors**

  - **Quantum Uncertainty**

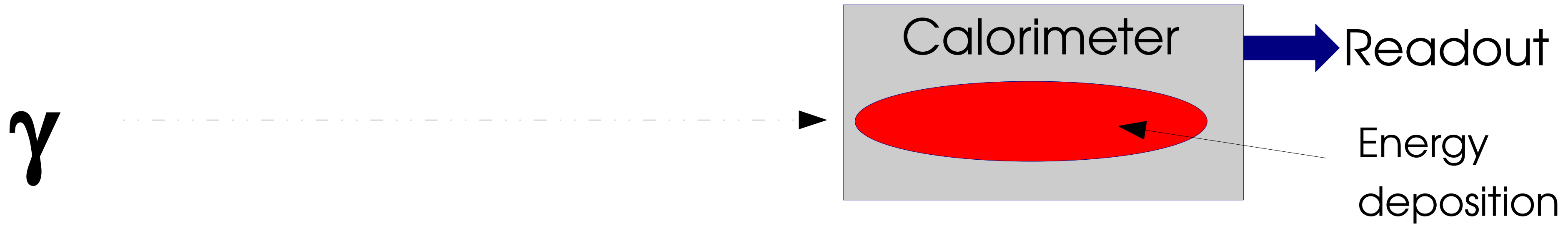
Measurement Errors : Example

Example: measuring the energy of a photon in a calorimeter

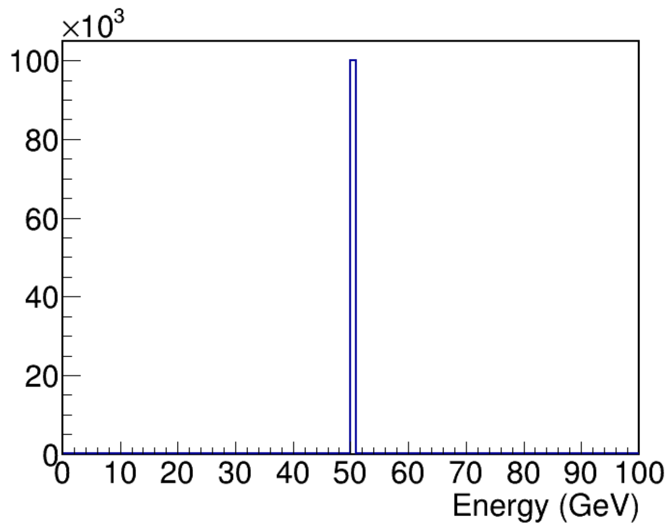


Measurement Errors : Example

Example: measuring the energy of a photon in a calorimeter

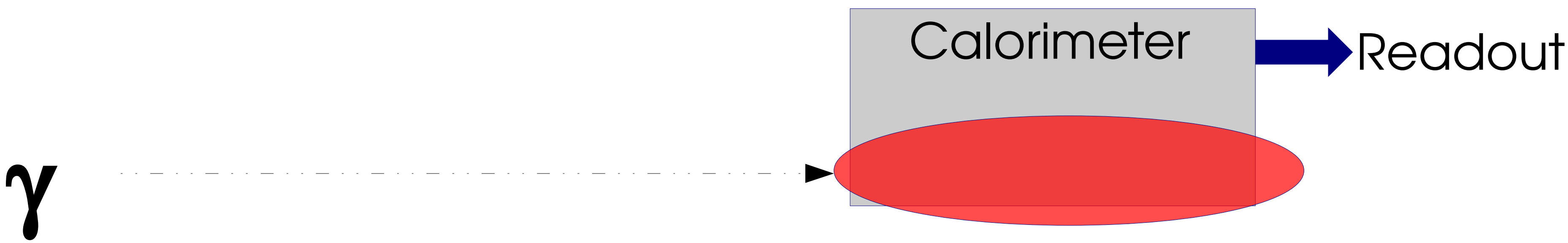


Perfect case

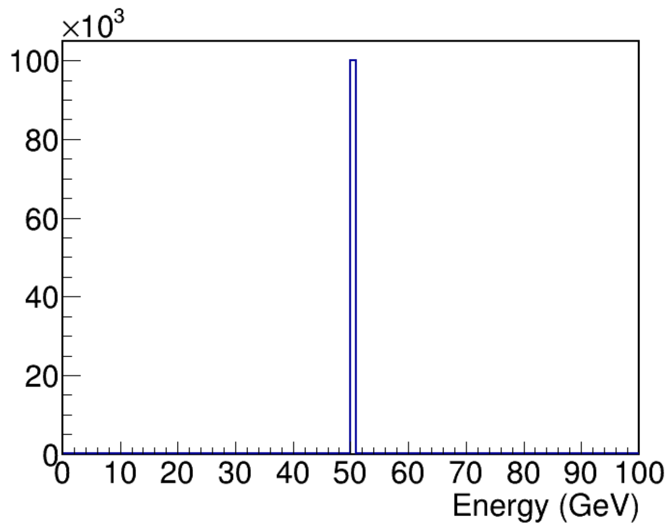


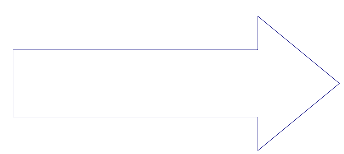
Measurement Errors : Example

Example: measuring the energy of a photon in a calorimeter

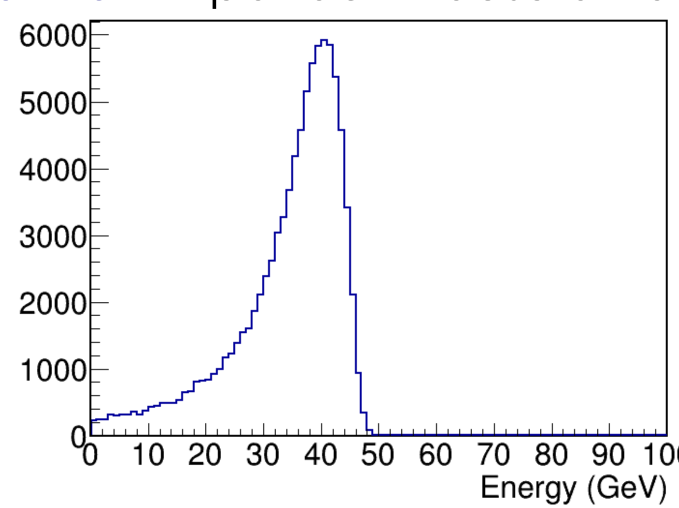


Perfect case



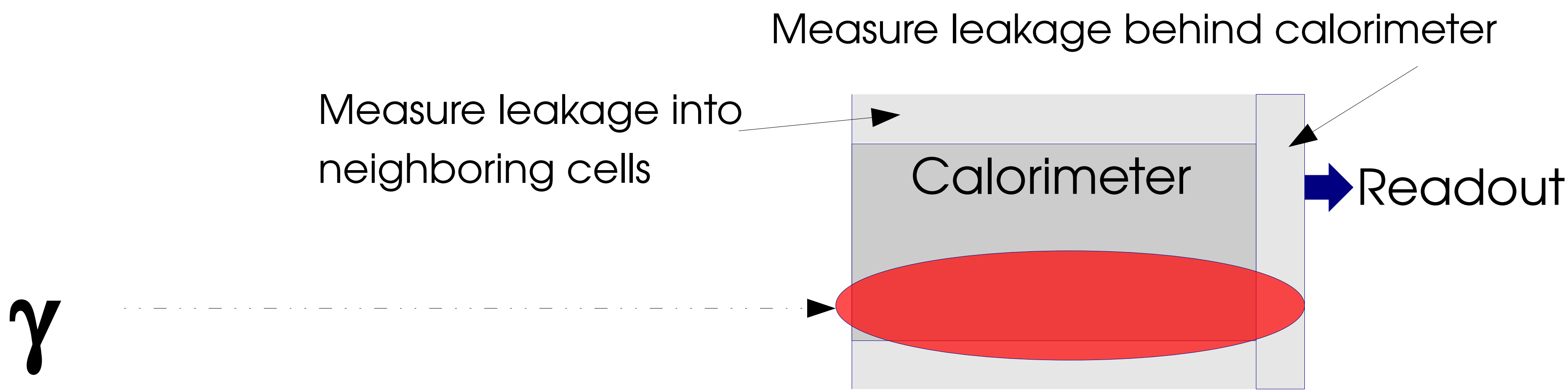


Real life : imperfect measurement

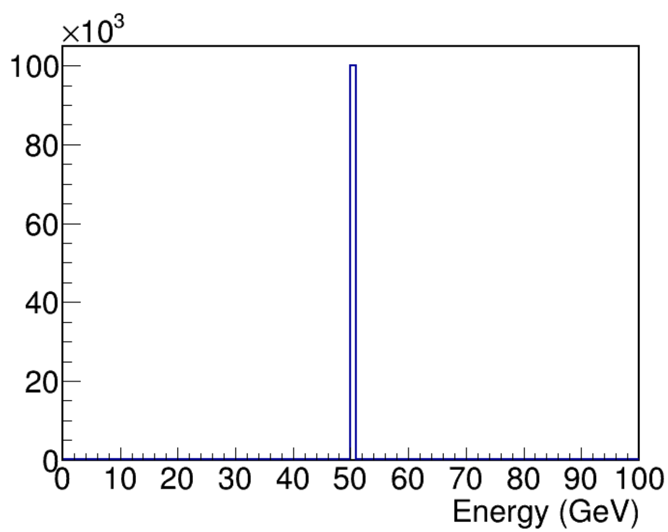


Measurement Errors : Example

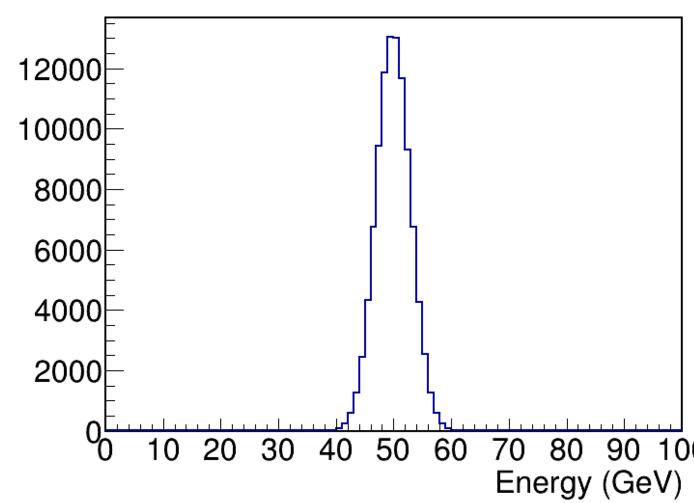
Example: measuring the energy of a photon in a calorimeter



Perfect case



Real life : imperfect measurement



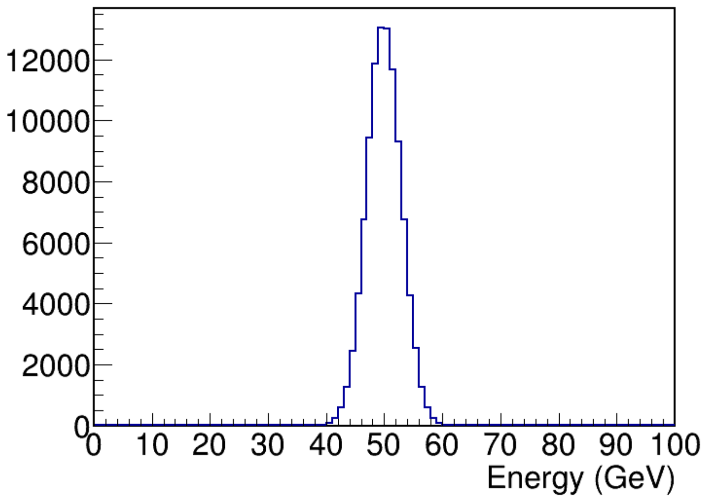
Can correct for main effects, but never perfectly

Measurement Errors

- **Best case**: measurements imperfections ("**bias**") can be determined
  - Apply correction "event by event", **remove effect**
- **Not always possible**
  - Too small to be measured reliably
  - Impossible to measure
- **Next-best solution**: describe **overall distribution** of imperfections
  - Typical size
  - Probability to reach a given value

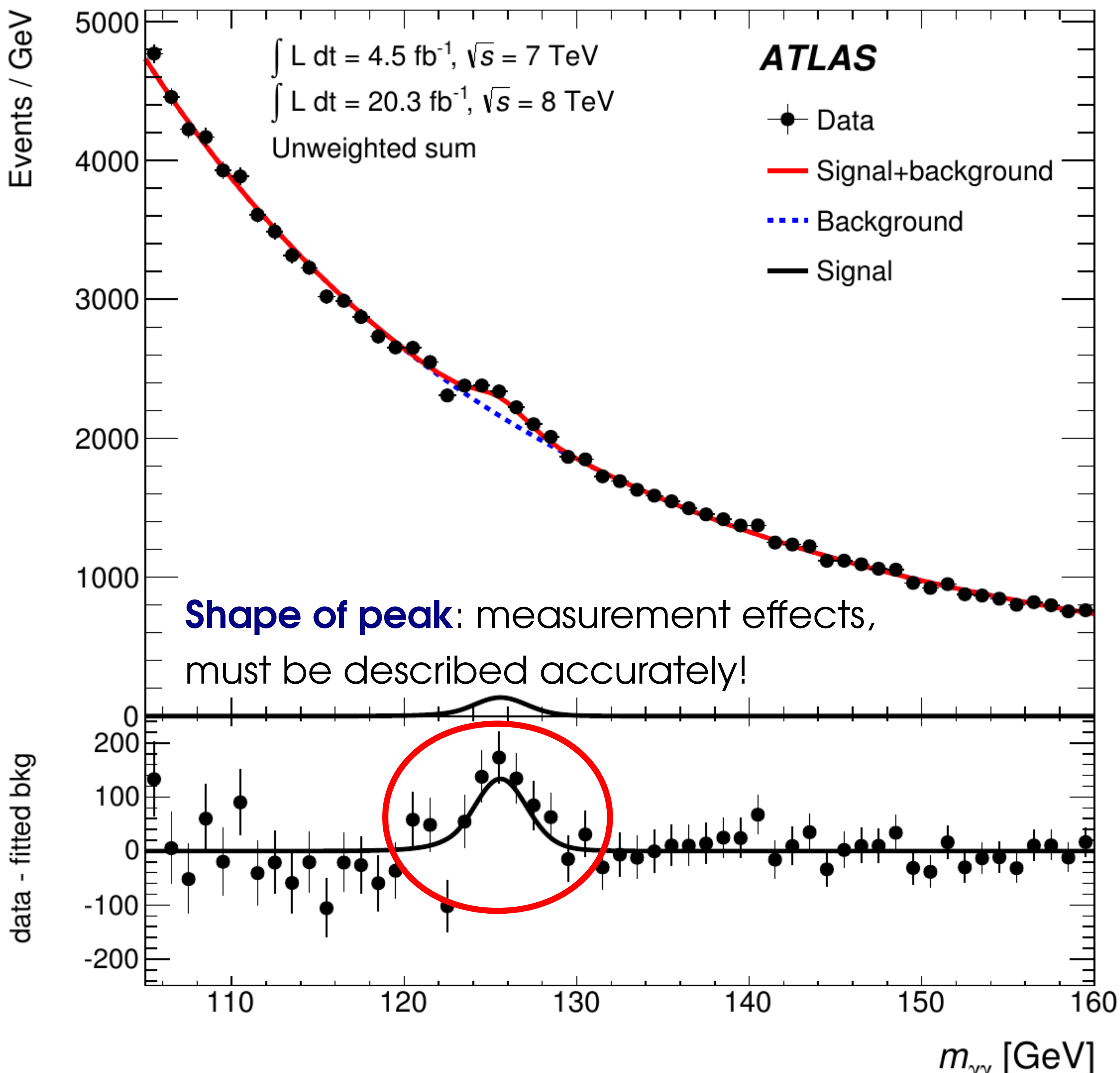
    ⇒ not m_H = 125 GeV but

    m_H = 125.36 +/- 0.40 GeV
- **Need to precisely quantify our uncertainty**



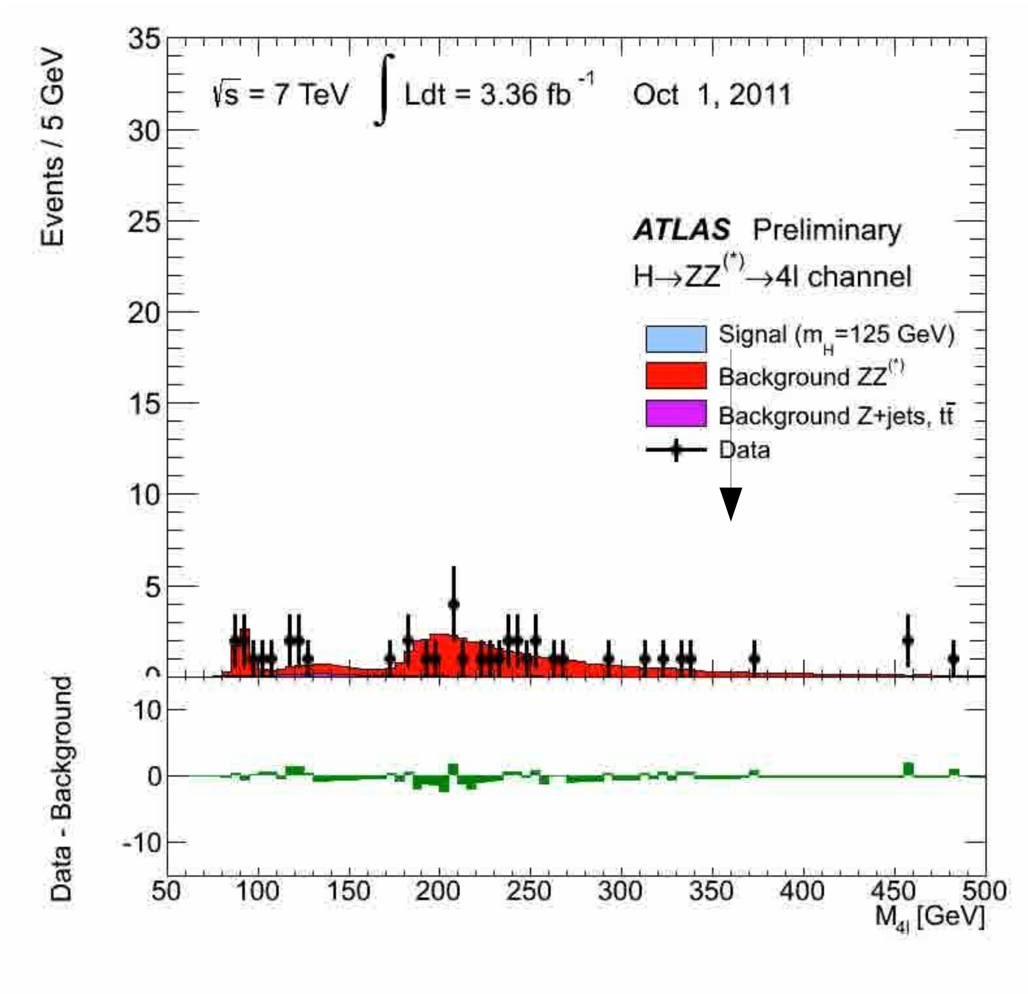
H→γγ Example

Phys. Rev. D **90**, 112015

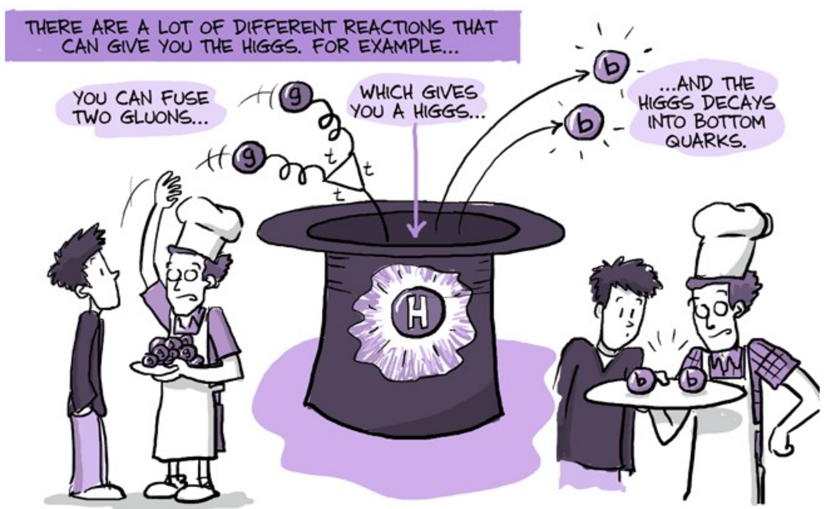


Another Example: H→ZZ*→4l

Phys. Rev. D **91**, 012006



Rare process: Expect 1 signal event every **~6 days**



http://www.phdcomics.com/comics/archive.php?comicid=1489

Quantum randomness: "Will I get an event today ?" - only probabilistic answer

Event counts must be described in a probabilistic way

Contents of the Lectures

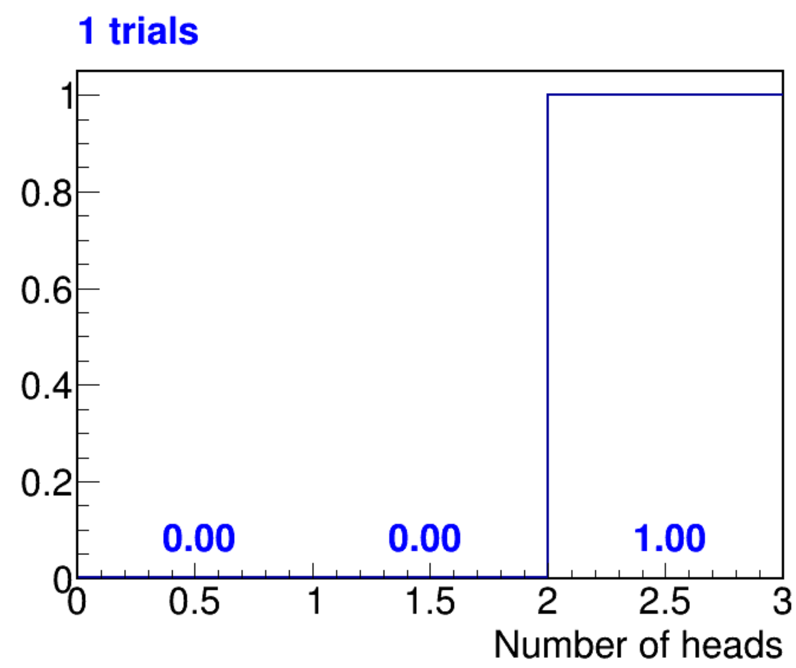
- **Probability Distributions (short reminder)**
- **How to build a statistical model**
- **How to Estimate a parameter value**
- **How to compute Confidence Intervals (uncertainties on parameters**
- **Tomorrow**:
  - Computing a discovery significance
  - Setting limits

Probability Distributions

Probability Distribution

Probabilistic treatment of possible outcomes ⇒ **Probability Distribution**

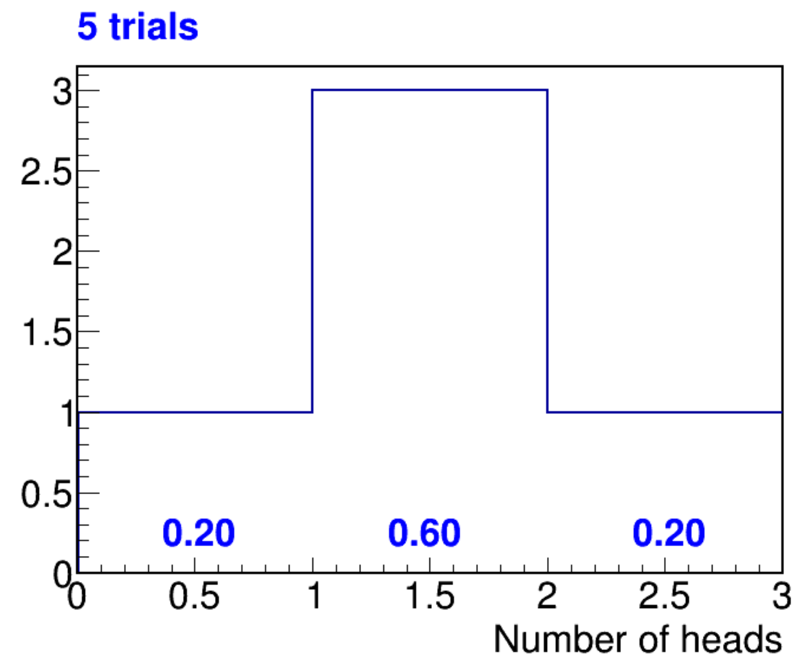
- **Example**: two-coin toss
  - Fractions of events in each bin converge to a limit



- **Probability distribution** : p_i, i=0,1,2
- **Properties**
  - $p_i > 0$
  - $\sum p_i = 1$

Probability Distribution

Probabilistic treatment of possible outcomes ⇒ **Probability Distribution**



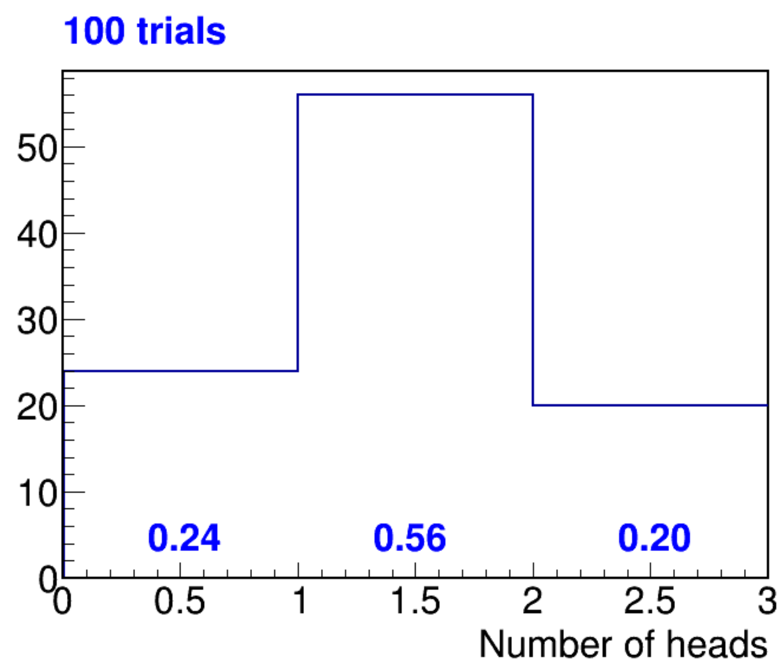
- **Example**: two-coin toss
  - Fractions of events in each bin converge to a limit
- **Probability distribution** : p_i, i=0,1,2
- **Properties**
  - $p_i > 0$
  - $\sum p_i = 1$



Probability Distribution

Probabilistic treatment of possible outcomes ⇒ **Probability Distribution**

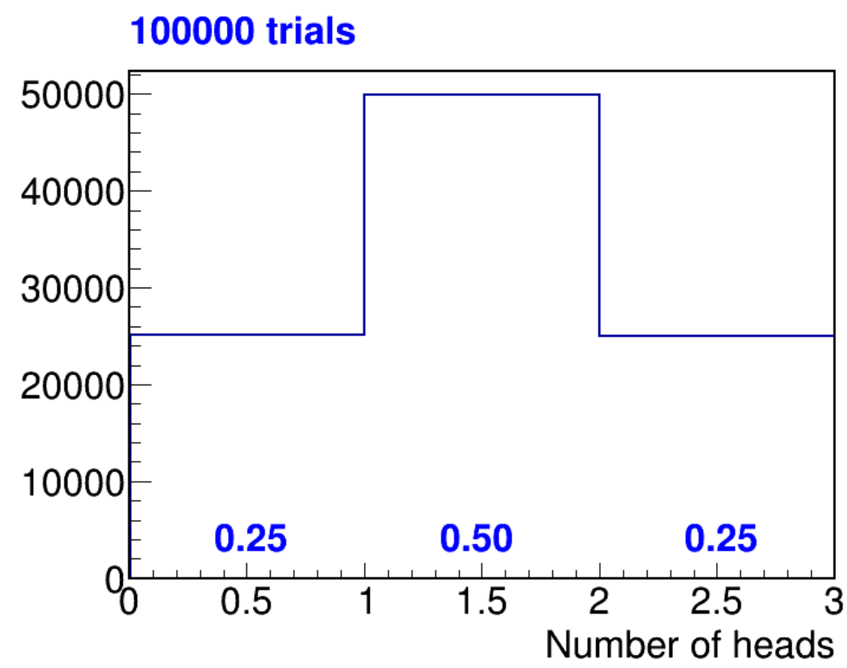
- **Example**: two-coin toss
  - Fractions of events in each bin converge to a limit
- **Probability distribution** : p_i, i=0,1,2
- **Properties**
  - $p_i > 0$
  - $\sum p_i = 1$



Probability Distribution

Probabilistic treatment of possible outcomes ⇒ **Probability Distribution**

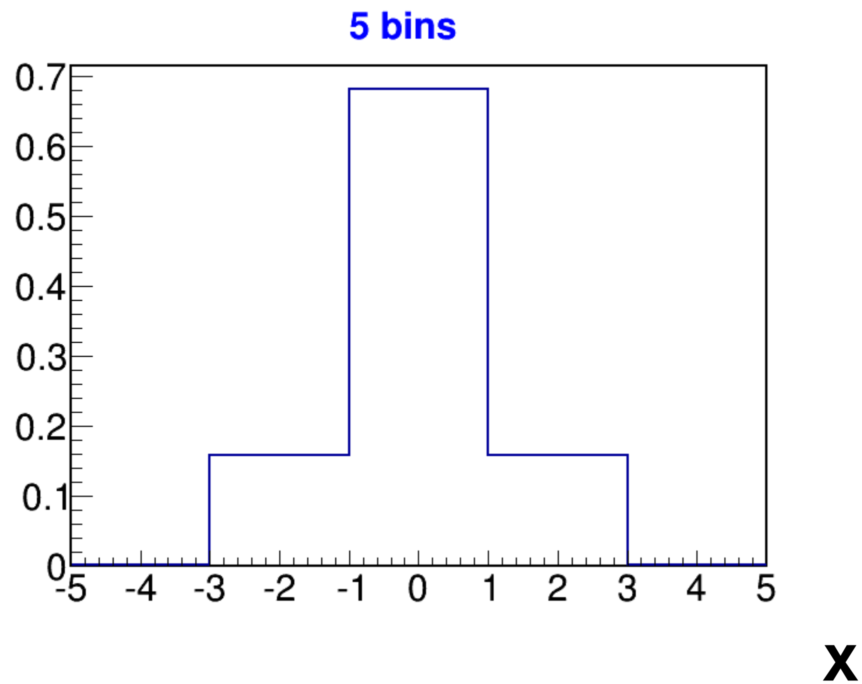
- **Example**: two-coin toss
  - Fractions of events in each bin converge to a limit



- **Probability distribution** : p_i, i=0,1,2
- **Properties**
  - $p_i > 0$
  - $\sum p_i = 1$

Continuous Variables: PDFs

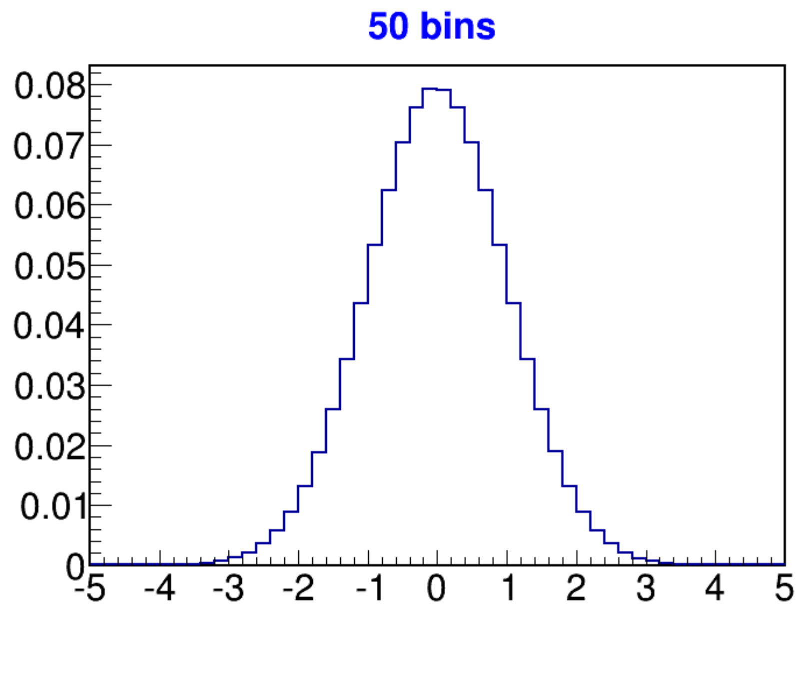
- **Continuous variable**, can consider **binned** probability distribution
  p_i, $i=1.. n_{bins}$



- Bin size → 0 :
  Probability distribution **function** **p(x)**
  - High values ⇔ high chance to get a measurement here
  - $p(x) > 0$
  - $\int p(x)\, dx = 1$
- Generalizes to **multiple variables** : $\int p(x,y)\, dx\, dy = 1$

Continuous Variables: PDFs

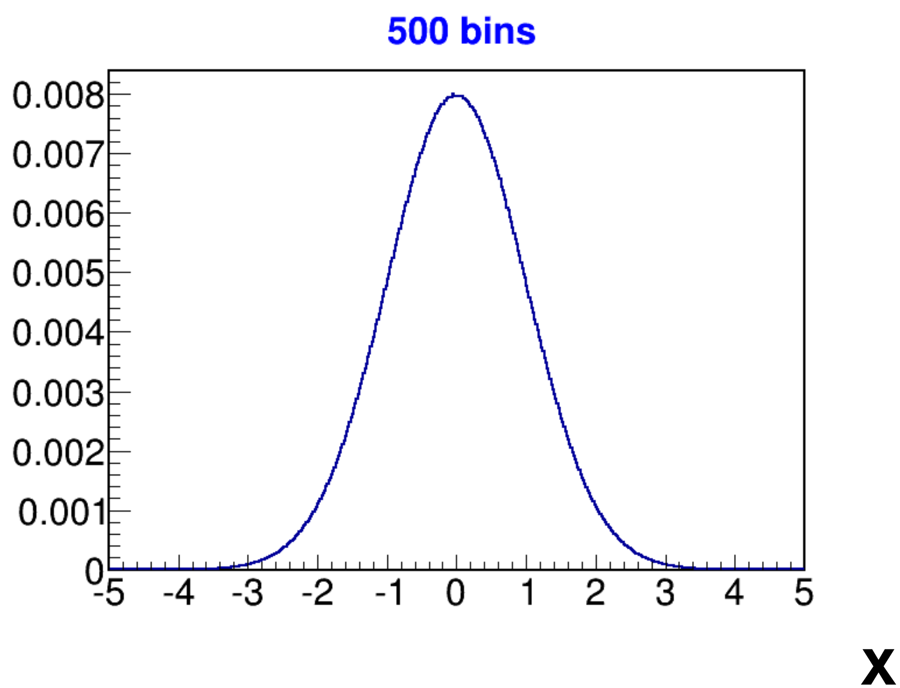
- **Continuous variable**, can consider **binned** probability distribution
  p_i, $i=1..\ n_{bins}$



- Bin size $\rightarrow$ 0 :
  Probability distribution function p(x)
  - High values $\Leftrightarrow$ high chance to get a measurement here
  - $p(x) > 0$
  - $\int p(x)\, dx = 1$
- Generalizes to **multiple variables** : $\int p(x,y)\, dx\, dy = 1$

Continuous Variables: PDFs

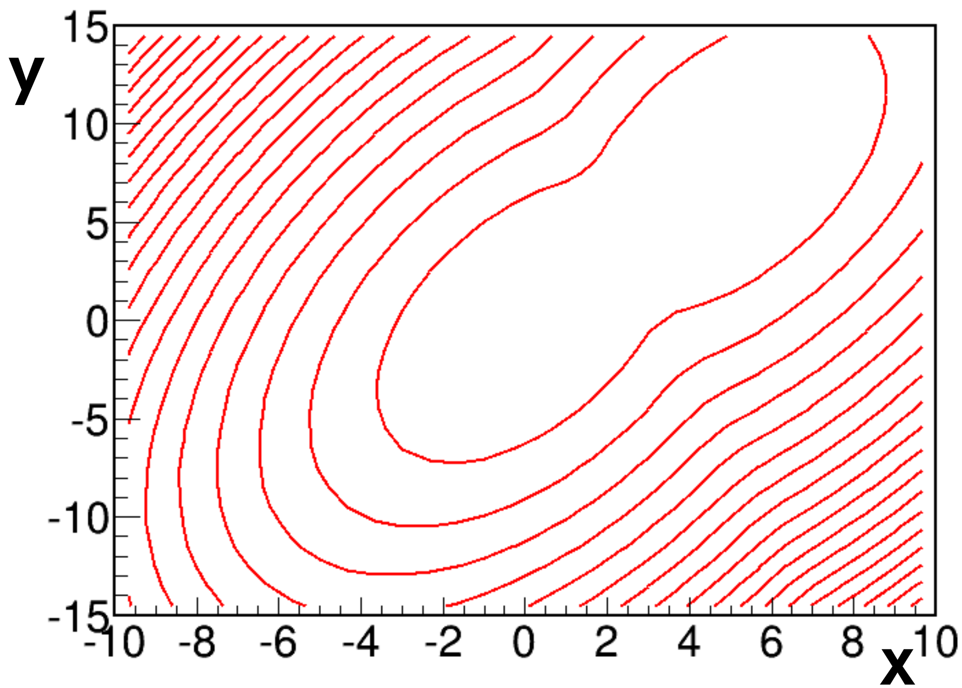
- **Continuous variable**, can consider **binned** probability distribution
  p_i, $i=1..\ n_{bins}$



- Bin size → 0 :
  Probability distribution function p(x)
  - High values ⇔ high chance to get a measurement here
  - $p(x) > 0$
  - $\int p(x)\, dx = 1$
- Generalizes to **multiple variables** : $\int p(x,y)\, dx\, dy = 1$

Continuous Variables: PDFs

- **Continuous variable**, can consider **binned** probability distribution
  p_i, $i=1..\ n_{bins}$



Contours:

p(x,y)

- Bin size → 0 :
  Probability distribution function p(x)
  - High values ⇔ high chance to get a measurement here
  - $p(x) > 0$
  - $\int p(x)\, dx = 1$
- Generalizes to **multiple variables** : $\int p(x,y)\, dx\, dy = 1$

PDF Properties: Mean

- Expectation values = expected outcome **on average**
- E(X) = **Mean** of X

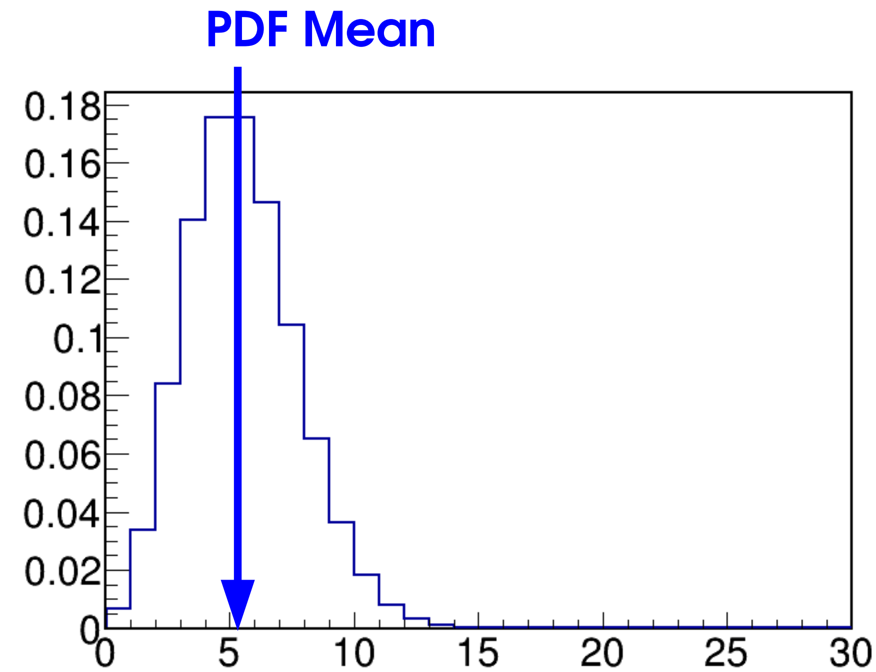
$$E(X)=\sum_i X_i p_i \quad \text{or}$$

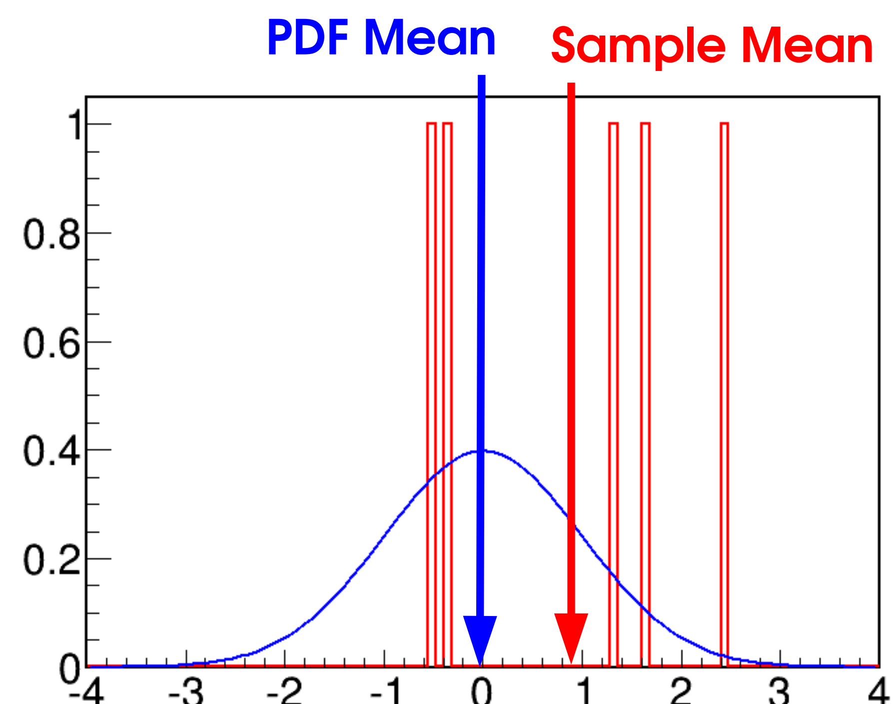
$$E(X)=\int X\, p(X)\, dX$$

  - Property of the PDF
- If one has a sample $x_1 \ldots x_n$, then can compute **Sample Mean**:

$$\bar{x}=\frac{1}{n}\sum_i x_i$$

  - Property of the sample
  - Should approximate PDF mean.

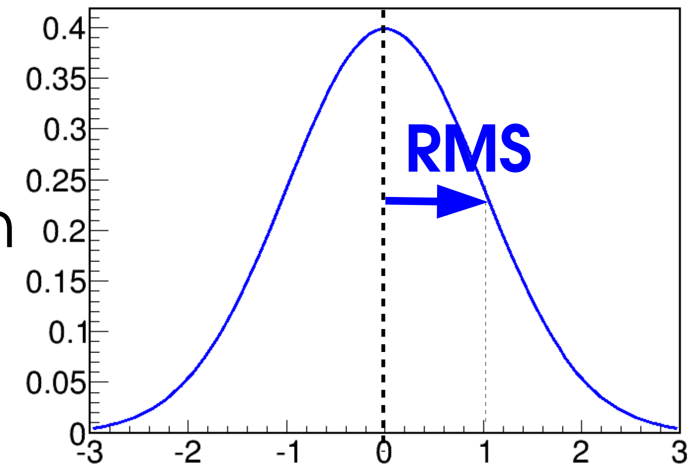


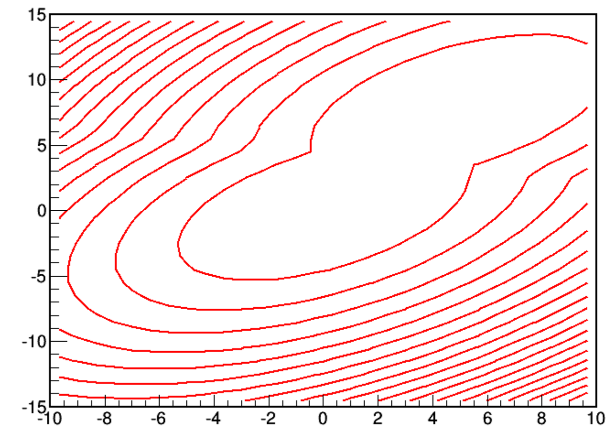


PDF Properties: Variance

- **Var(X) = E([X – E(X)]²)** = **Variance** of X
  - Average square of deviation from mean
  - **RMS(X) = √Var(X)** "**root mean square**"
  - Can be approximated by **sample variance**:

$$\sigma^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$$



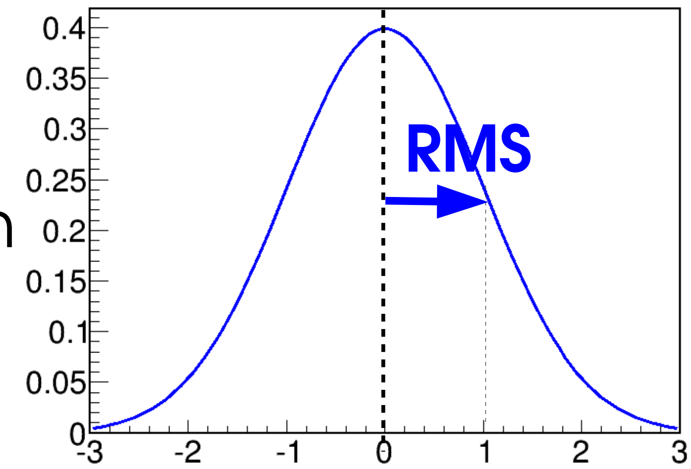


- **Covariance of X and Y:**
  Cov(X,Y) = E([X - E(X)] [Y - E(Y)])
  - Large if variations of X, Y are "synchronized"
  - **Cov(X,Y) > 0** if X and Y vary in the **same** direction
  - **Cov(X,Y) < 0** if X and Y vary in **opposite** direction
  - **Cov(X,Y) = 0** if X and Y vary **independently**

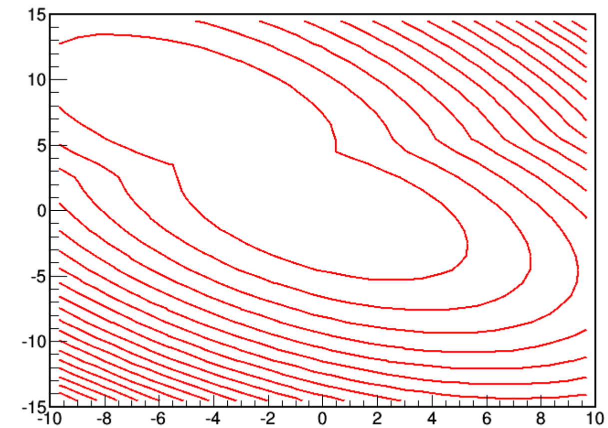
PDF Properties: Variance

- **Var(X) = E([X – E(X)]²)** = **Variance** of X
  - Average square of deviation from mean
  - **RMS(X) = √Var(X)** "**root mean square**"
  - Can be approximated by **sample variance**:

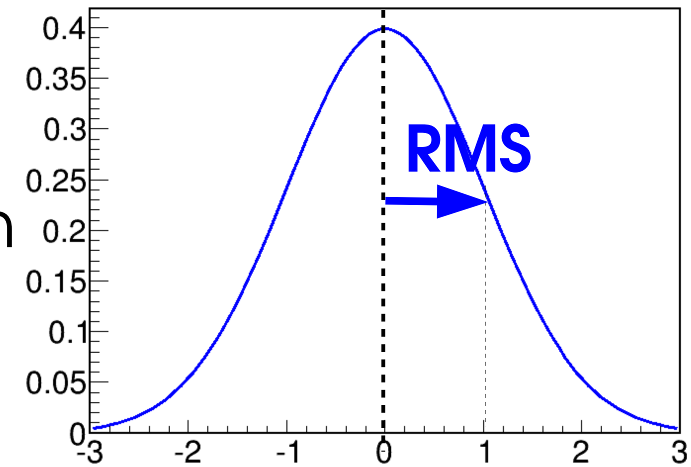
$$\sigma^2 = \frac{1}{n-1} \sum_i \left(x_i - \bar{x} \right)^2$$



- **Covariance of X and Y:**
  Cov(X,Y) = E([X - E(X)] [Y - E(Y)])
  - Large if variations of X, Y are "synchronized"
  - **Cov(X,Y) > 0** if X and Y vary in the **same** direction
  - **Cov(X,Y) < 0** if X and Y vary in **opposite** direction
  - **Cov(X,Y) = 0** if X and Y vary **independently**

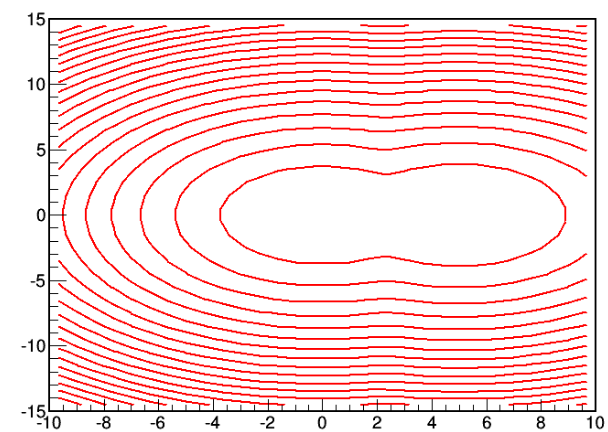


PDF Properties: Variance



- **Var(X) = E([X – E(X)]^2) =** **Variance** of X
  - Average square of deviation from mean
  - **RMS(X) = √Var(X)** "**root mean square**"
  - Can be approximated by **sample variance**:

$$\sigma^2 = \frac{1}{n-1} \sum_i \left(x_i - \bar{x}\right)^2$$



- **Covariance of X and Y:**
  Cov(X,Y) = E([X - E(X)] [Y - E(Y)])
  - Large if variations of X, Y are "synchronized"
  - **Cov(X,Y) > 0** if X and Y vary in the **same** direction
  - **Cov(X,Y) < 0** if X and Y vary in **opposite** direction
  - **Cov(X,Y) = 0** if X and Y vary **independently**

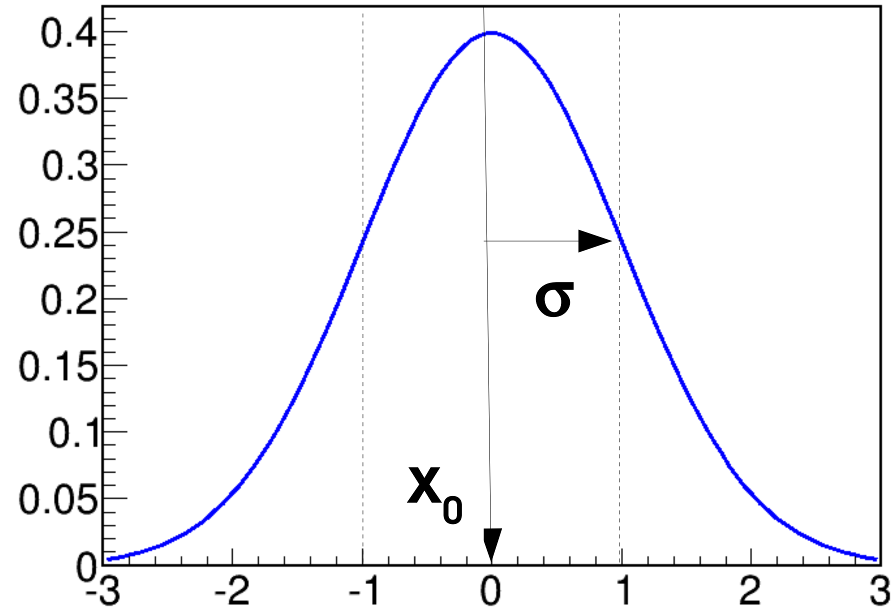
Example 1 : Gaussian

Gaussian distribution:

$$G(x; x_0, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-x_0)^2}{2\sigma^2}}$$

Mean : x_0

Variance : σ^2 ($\Rightarrow$ RMS=σ)



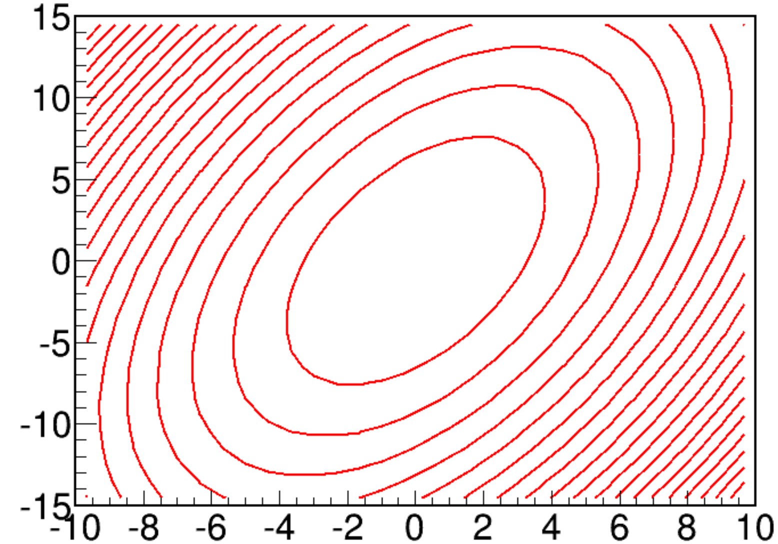
- Generalize to **N dimensions**:

$$G(X; X_0, C) = \frac{1}{(2\pi|C|)^{n/2}} e^{-\frac{1}{2}(X-X_0)^T C^{-1}(X-X_0)}$$

Mean (vector) = X_0

Covariance matrix

$$\begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix} = C$$



Central Limit Theorem

- For a random variable X with **any distribution**, one has

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \overset{n\to\infty}{\sim} G\left(\bar{x};\ E(X),\ \frac{RMS(X)}{\sqrt{n}}\right)$$

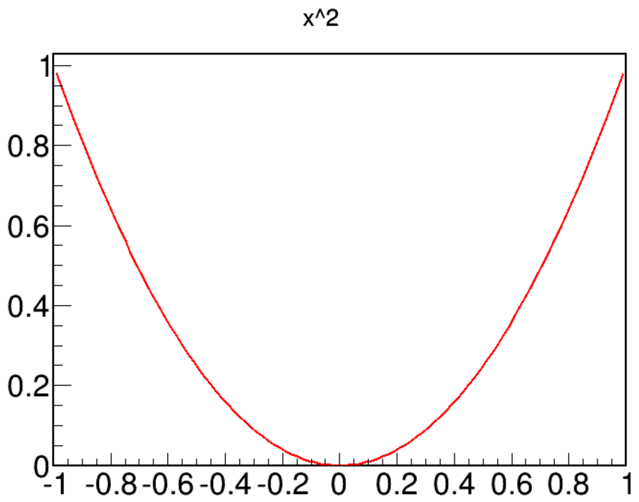
- What this means:
  - **The average of many measurements is always Gaussian,** whatever the distribution for a single measurement
  - **The mean of the Gaussian is the mean of the single measurements**
  - **the RMS of the Gaussian decreases as √n** : less fluctuations when averaging over many measurements
- Another version, for the sum:

$$\sum_{i=1}^{n} x_i \overset{n\to\infty}{\sim} G\left(\bar{x};\ n\,E(X),\ \sqrt{n}\,\sigma(X)\right)$$

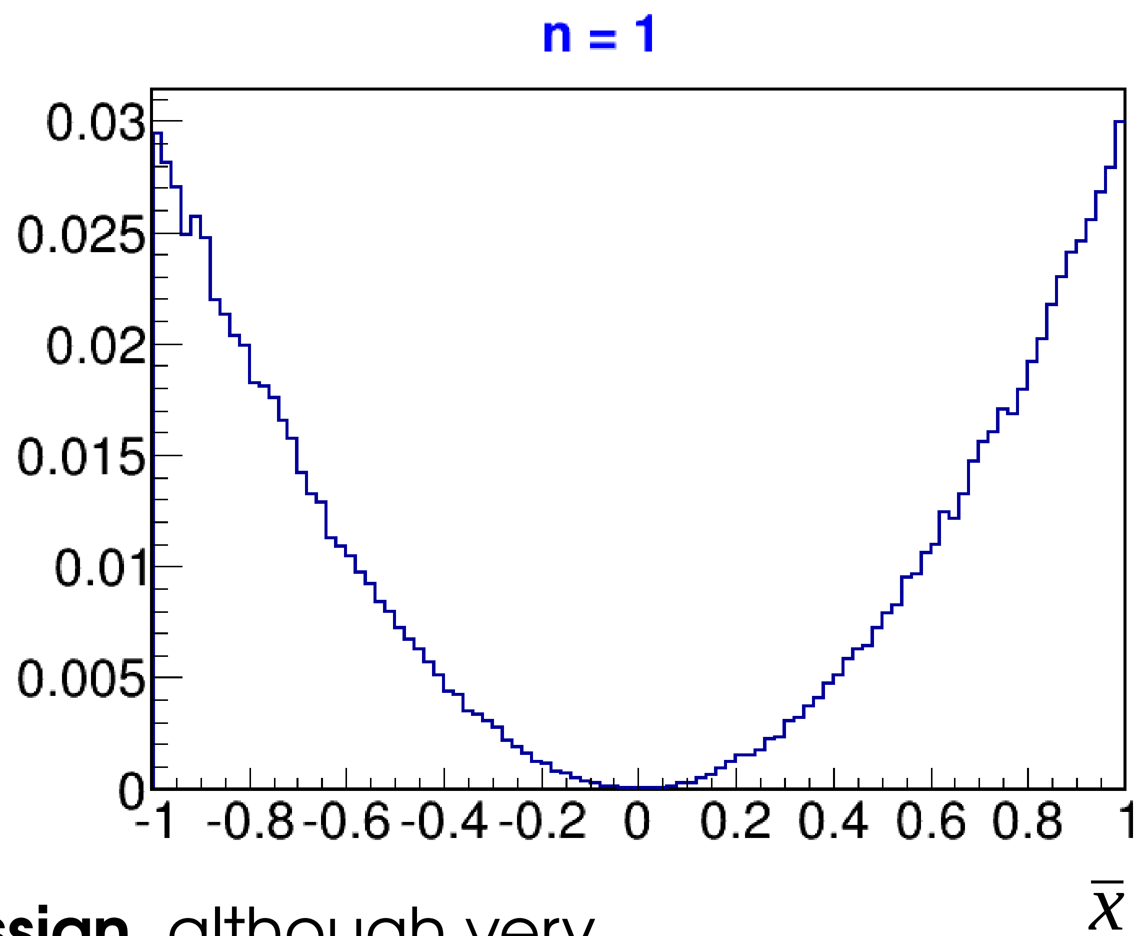
- **Mean scales like n, but RMS only like √n**

Central Limit Theorem Example

Draw events from a x^2 distribution (for illustration only)



$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

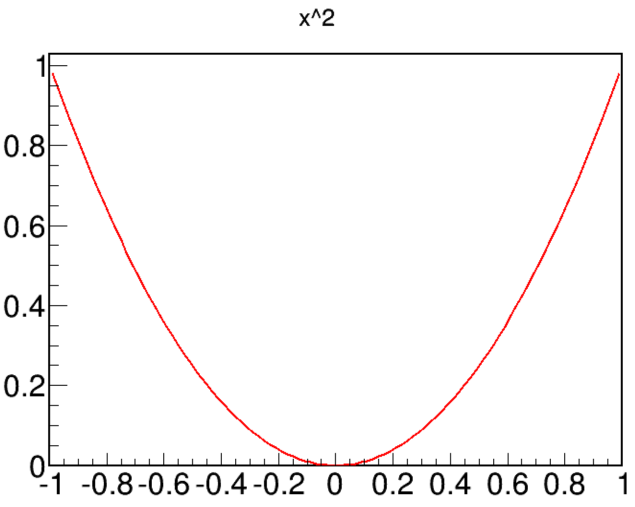


Distribution becomes Gaussian, although very non-Gaussian originally

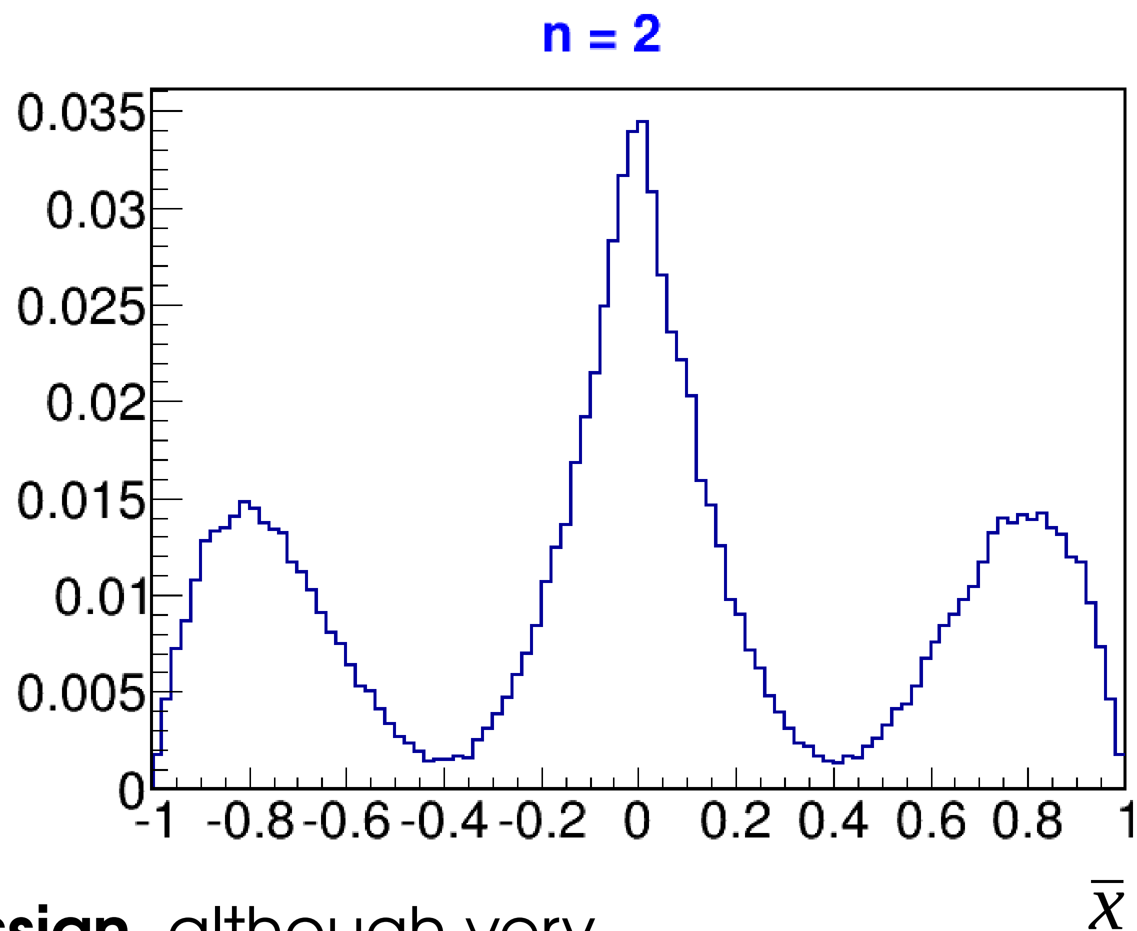
Distribution becomes narrower as expected (as 1/√n)

Central Limit Theorem Example

Draw events from a x^2 distribution (for illustration only)



$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

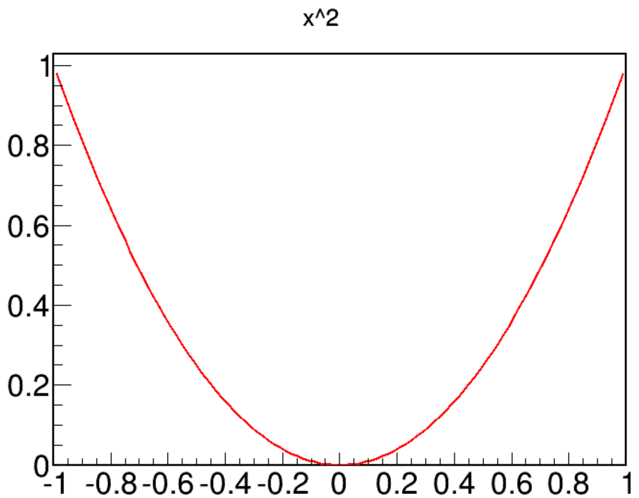


Distribution becomes Gaussian, although very non-Gaussian originally

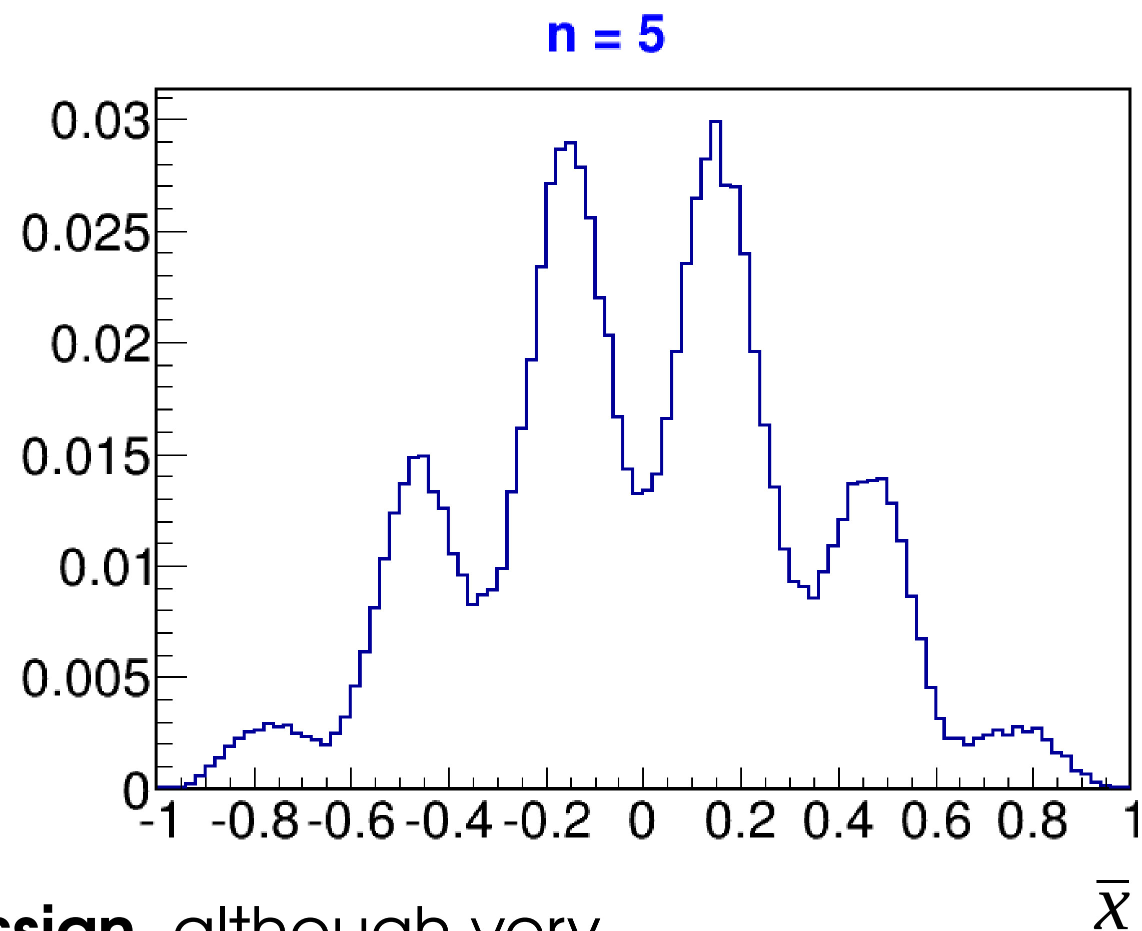
Distribution becomes narrower as expected (as 1/√n)

Central Limit Theorem Example

Draw events from a x^2 distribution (for illustration only)



$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

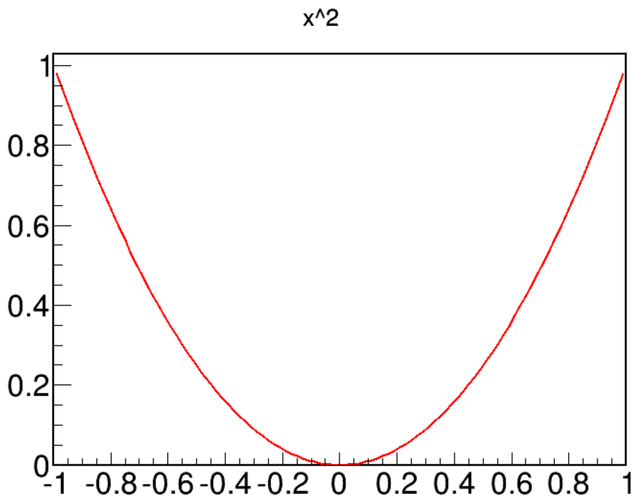


Distribution becomes Gaussian, although very non-Gaussian originally

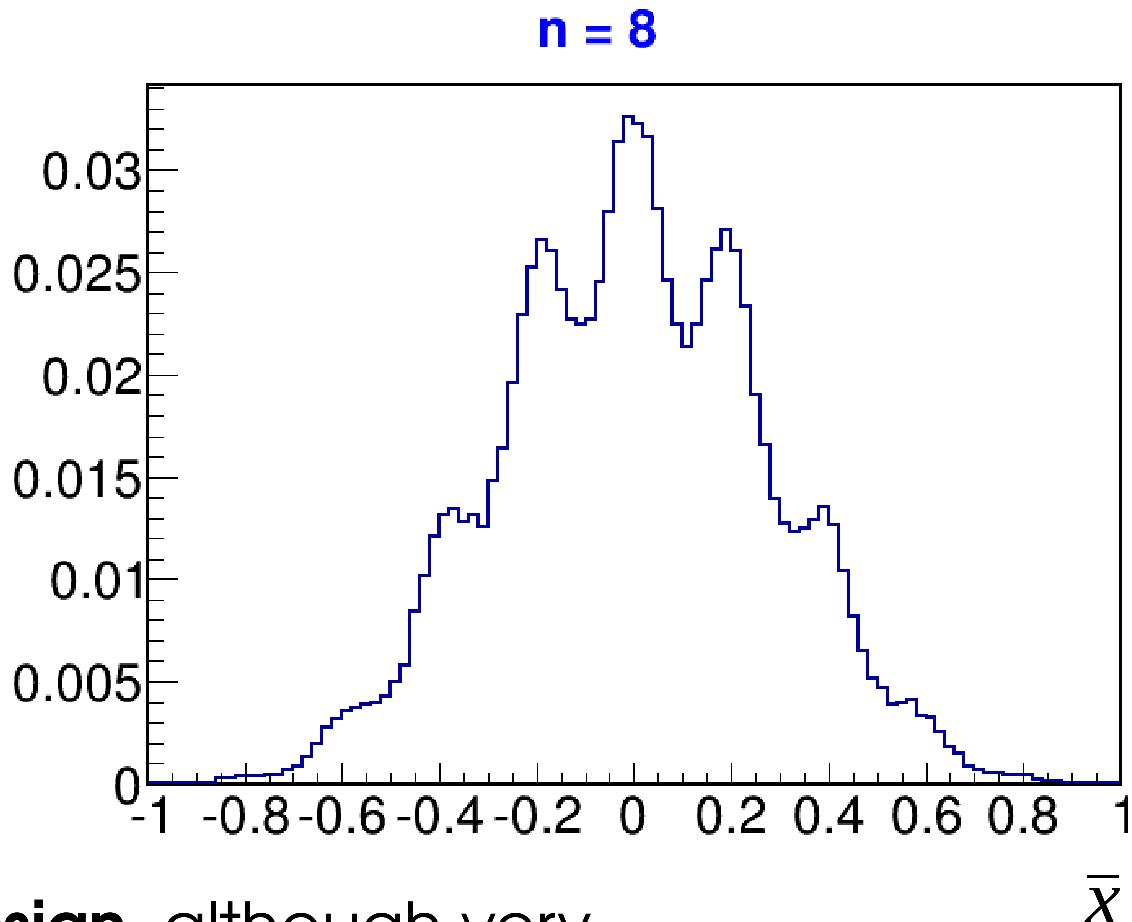
Distribution becomes narrower as expected (as $1/\sqrt{n}$)

Central Limit Theorem Example

Draw events from a x^2 distribution (for illustration only)



$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

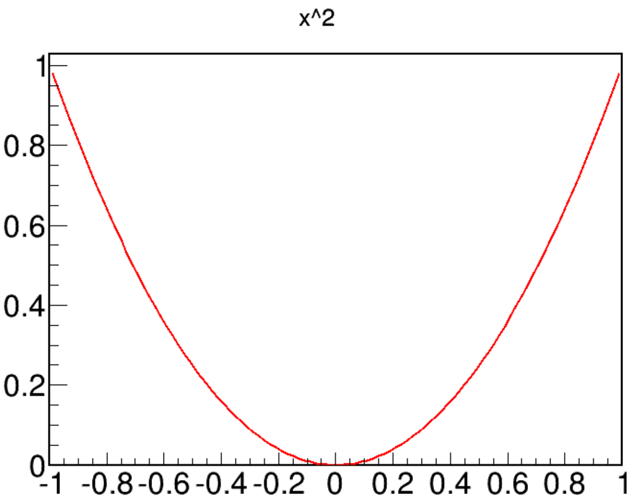


Distribution becomes Gaussian, although very non-Gaussian originally

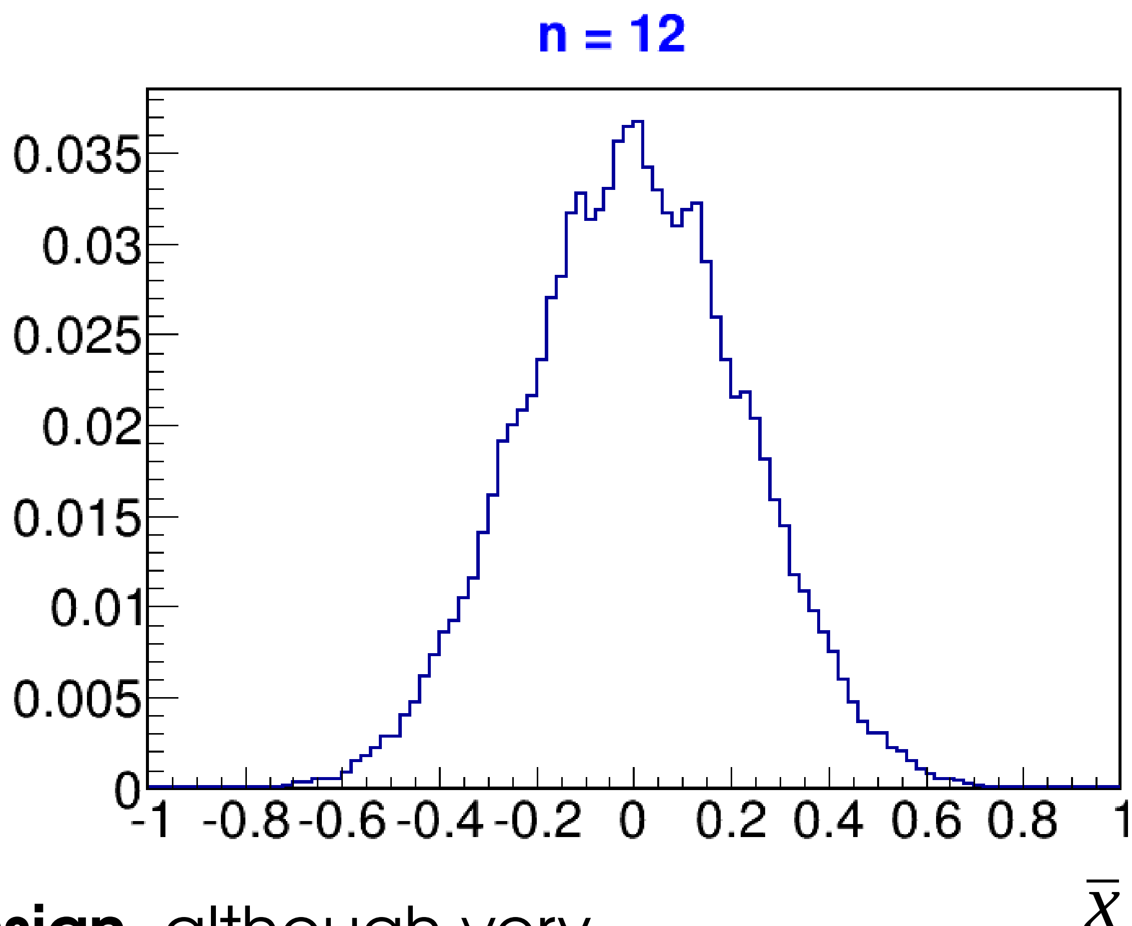
Distribution becomes narrower as expected (as $1/\sqrt{n}$)

Central Limit Theorem Example

Draw events from a x^2 distribution (for illustration only)



$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

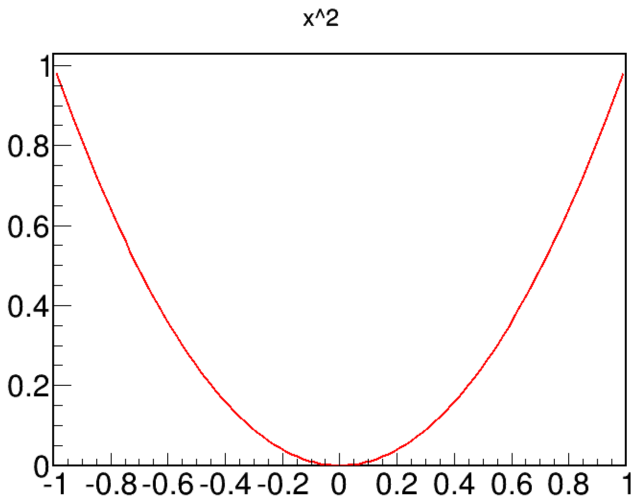


Distribution becomes Gaussian, although very non-Gaussian originally

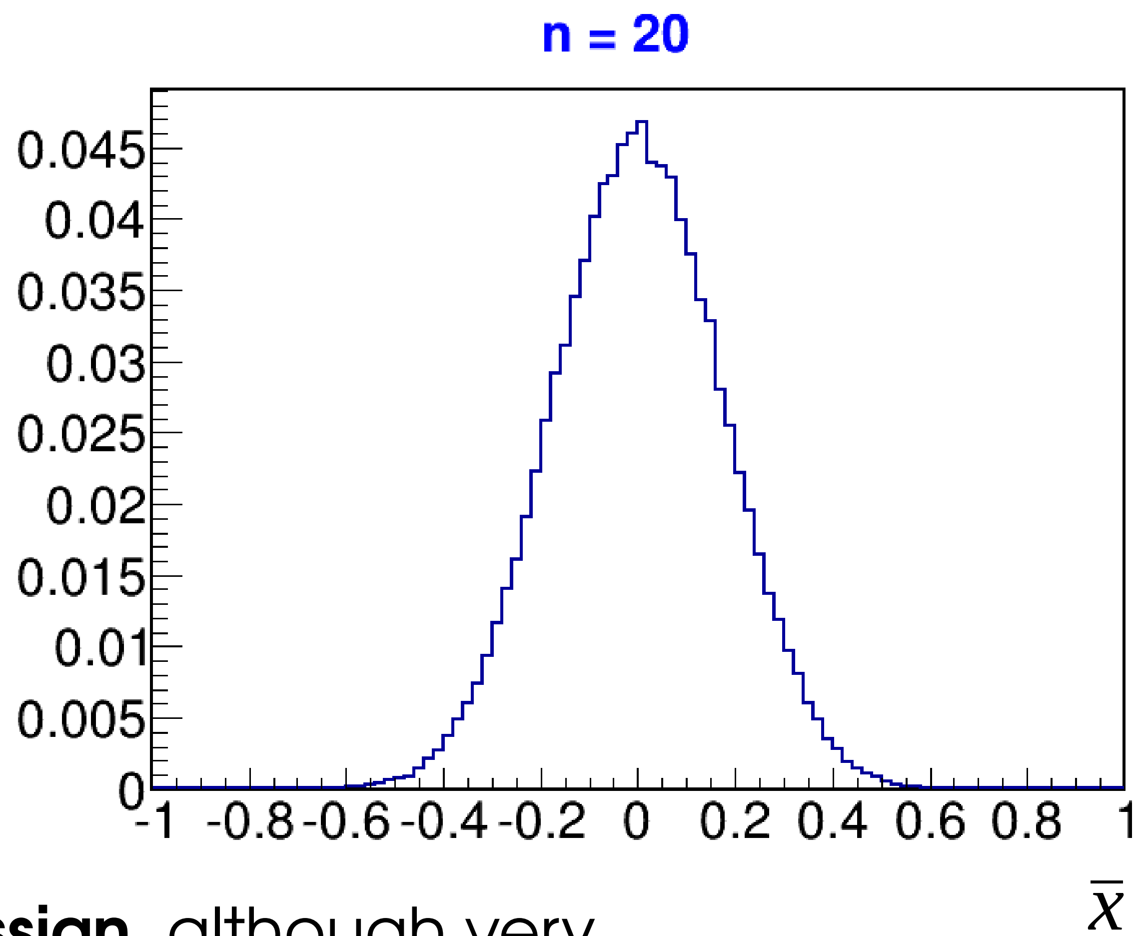
Distribution becomes narrower as expected (as 1/√n)

Central Limit Theorem Example

Draw events from a x^2 distribution (for illustration only)



$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

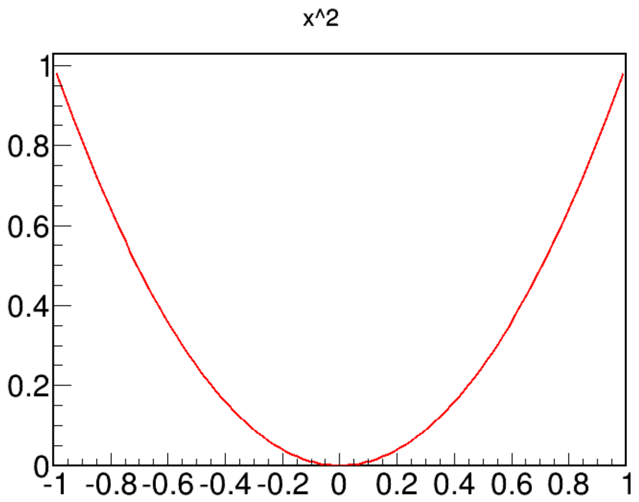


Distribution becomes Gaussian, although very non-Gaussian originally

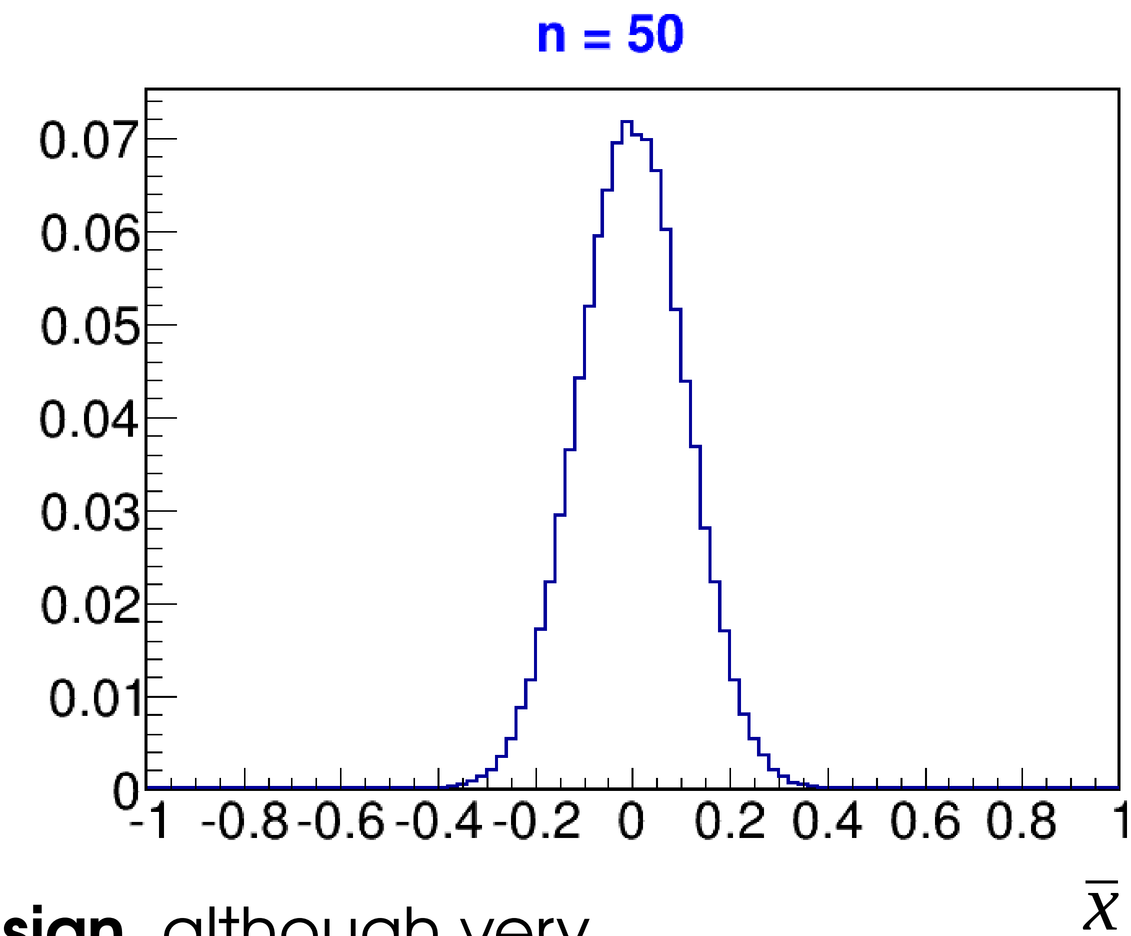
Distribution becomes narrower as expected (as $1/\sqrt{n}$)

Central Limit Theorem Example

Draw events from a x^2 distribution (for illustration only)



$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

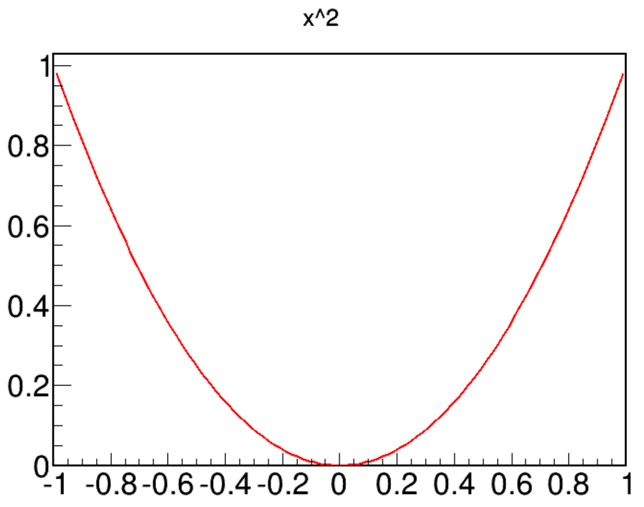


Distribution becomes Gaussian, although very non-Gaussian originally

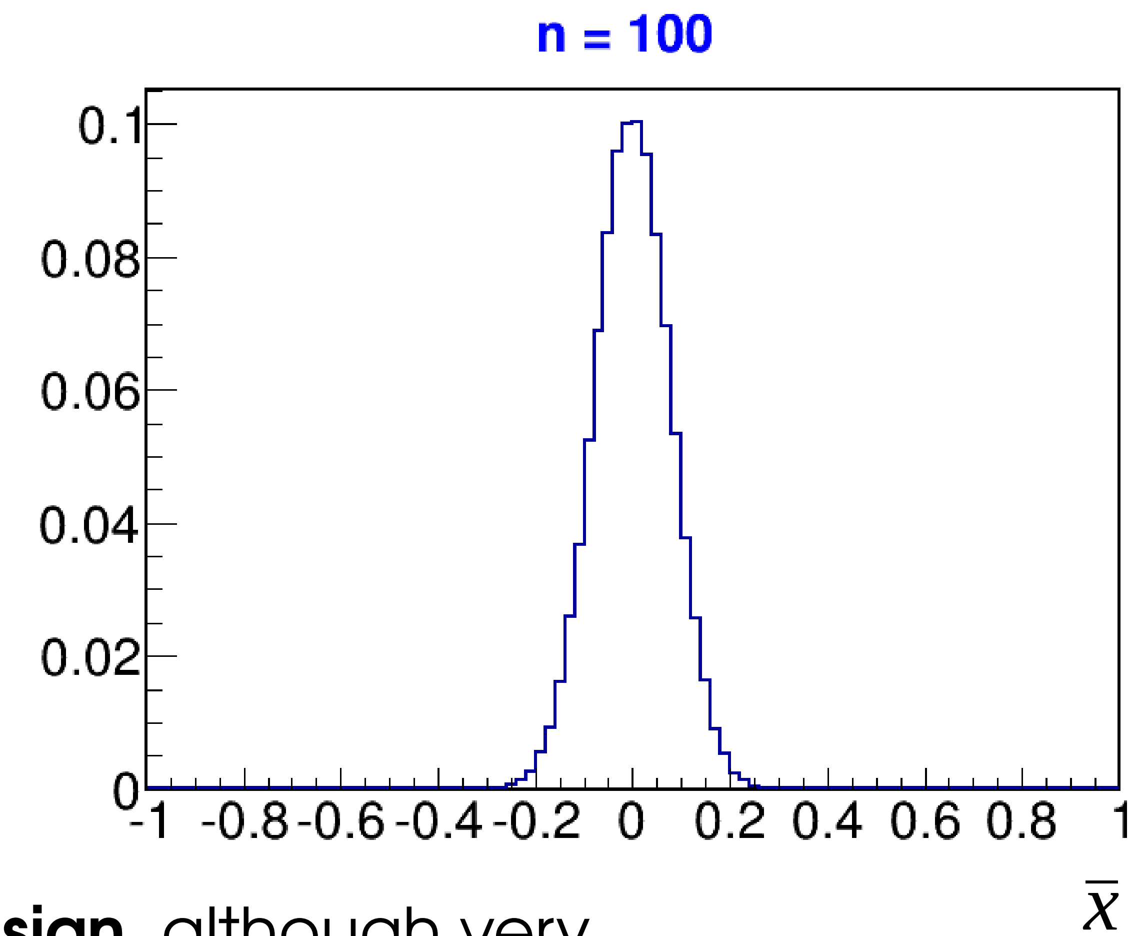
Distribution becomes narrower as expected (as $1/\sqrt{n}$)

Central Limit Theorem Example

Draw events from a x^2 distribution (for illustration only)



$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$



Distribution becomes Gaussian, although very non-Gaussian originally

Distribution becomes narrower as expected (as $1/\sqrt{n}$)

Gaussian Integrals

- Probability to be "less than nσ" away from the mean:

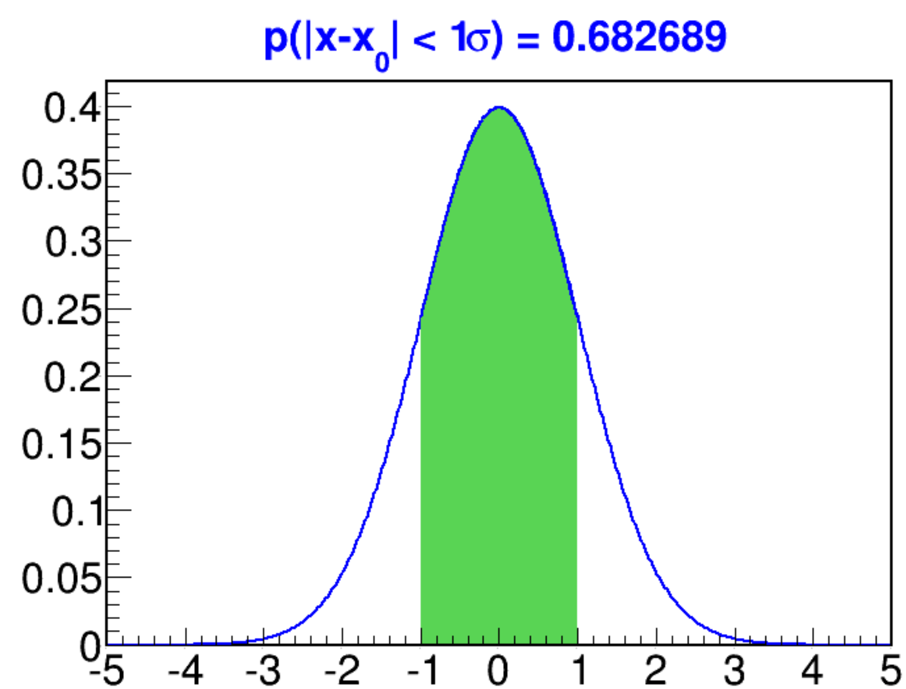
$$P(|x-x_0|<n\sigma)=\int_{x_0-n\sigma}^{x_0+n\sigma} G(x;x_{0,}\sigma)\,dx=\int_{-n}^{+n} N(x)\,dx$$

Standard Normal Distribution

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

- Used also for other distributions: "**1σ error**" for p=68%, etc.

Number of sigmas	Fraction inside	Fraction outside
1	0.68	0.32
2	0.955	0.045
3	0.997	0.003
5	0.999999	6×10^{-7}



Gaussian Integrals

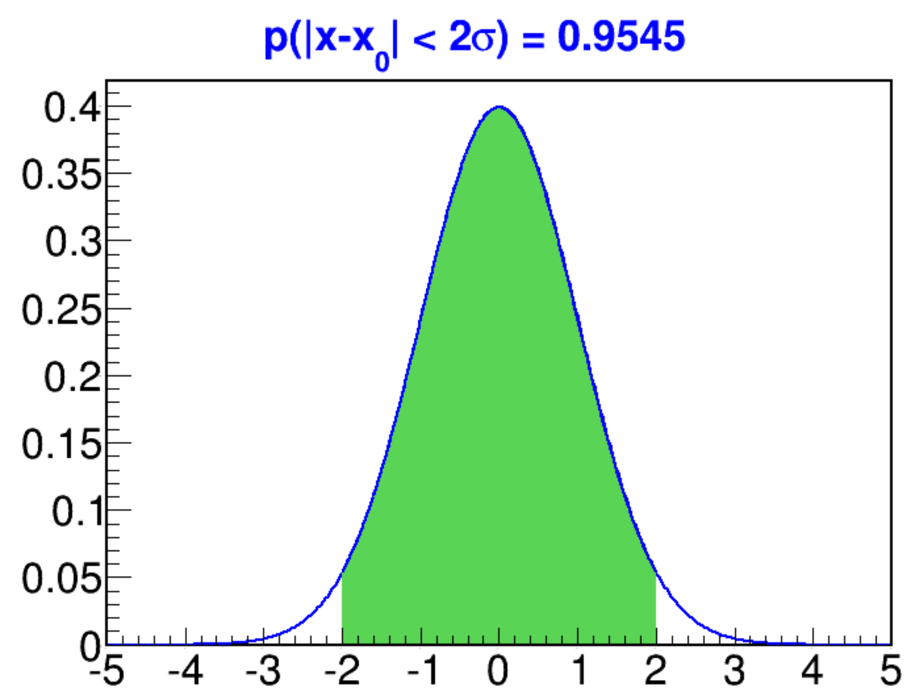
- Probability to be "less than nσ" away from the mean:

$$P(|x-x_0|<n\sigma)=\int_{x_0-n\sigma}^{x_0+n\sigma} G(x;x_0,\sigma)\,dx=\int_{-n}^{+n} N(x)\,dx$$

Standard Normal Distribution $N(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

- Used also for other distributions: "**1σ error**" for p=68%, etc.

Number of sigmas	Fraction inside	Fraction outside
1	**0.68**	**0.32**
2	**0.955**	**0.045**
3	**0.997**	**0.003**
5	**0.999999**	**6×10^{-7}**



Gaussian Integrals

- Probability to be "less than nσ" away from the mean:

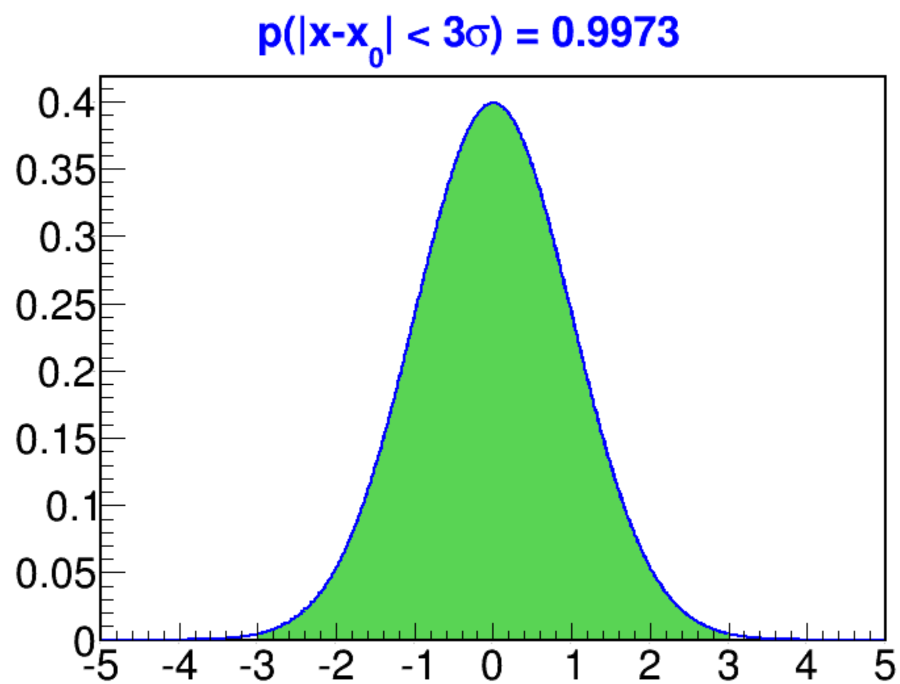
$$P(|x-x_0|<n\sigma)=\int_{x_0-n\sigma}^{x_0+n\sigma} G(x;x_{0,}\sigma)\,dx=\int_{-n}^{+n} N(x)\,dx$$

Standard Normal Distribution

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

- Used also for other distributions: "**1σ error**" for p=68%, etc.

Number of sigmas	Fraction inside	Fraction outside
1	**0.68**	**0.32**
2	**0.955**	**0.045**
3	**0.997**	**0.003**
5	**0.999999**	**6 x 10^{-7}**



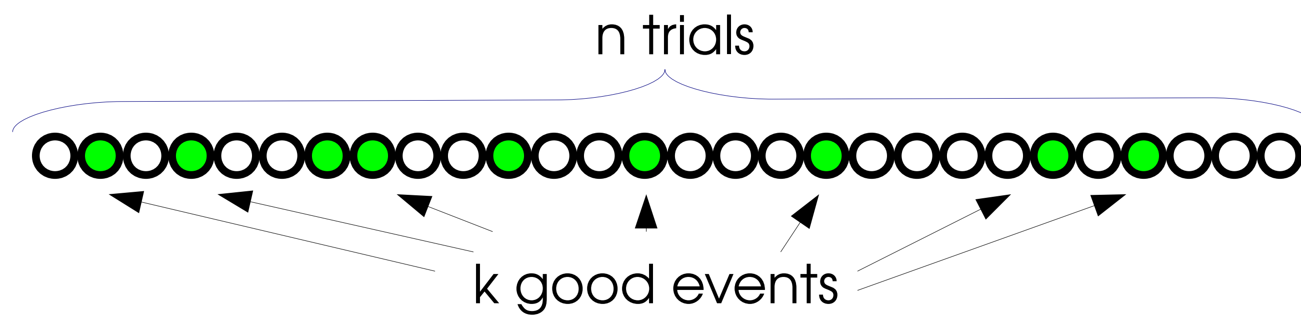
Example 2 : Counting events

- Consider n trials with probability p. Prob. to get k good events ?

Binomial distribution : $$P(k;n,p)=C_k^n p^k (1-p)^{n-k}$$

Mean = np

Variance = np(1-p)



- Not widely used because :

- Suppose **p≪1, n≫1**, let **λ = np**
  - i.e. **very rare** process, but **many trials** so still expect events

⇒ **Poisson distribution**

Mean = λ

Variance = λ

$$p(k;\lambda)=e^{-\lambda}\frac{\lambda^k}{k!}$$

$$(1-p)^{n-k} \sim \left(1-\frac{\lambda}{n}\right)^n \sim e^{-\lambda}$$

Rare Processes ?

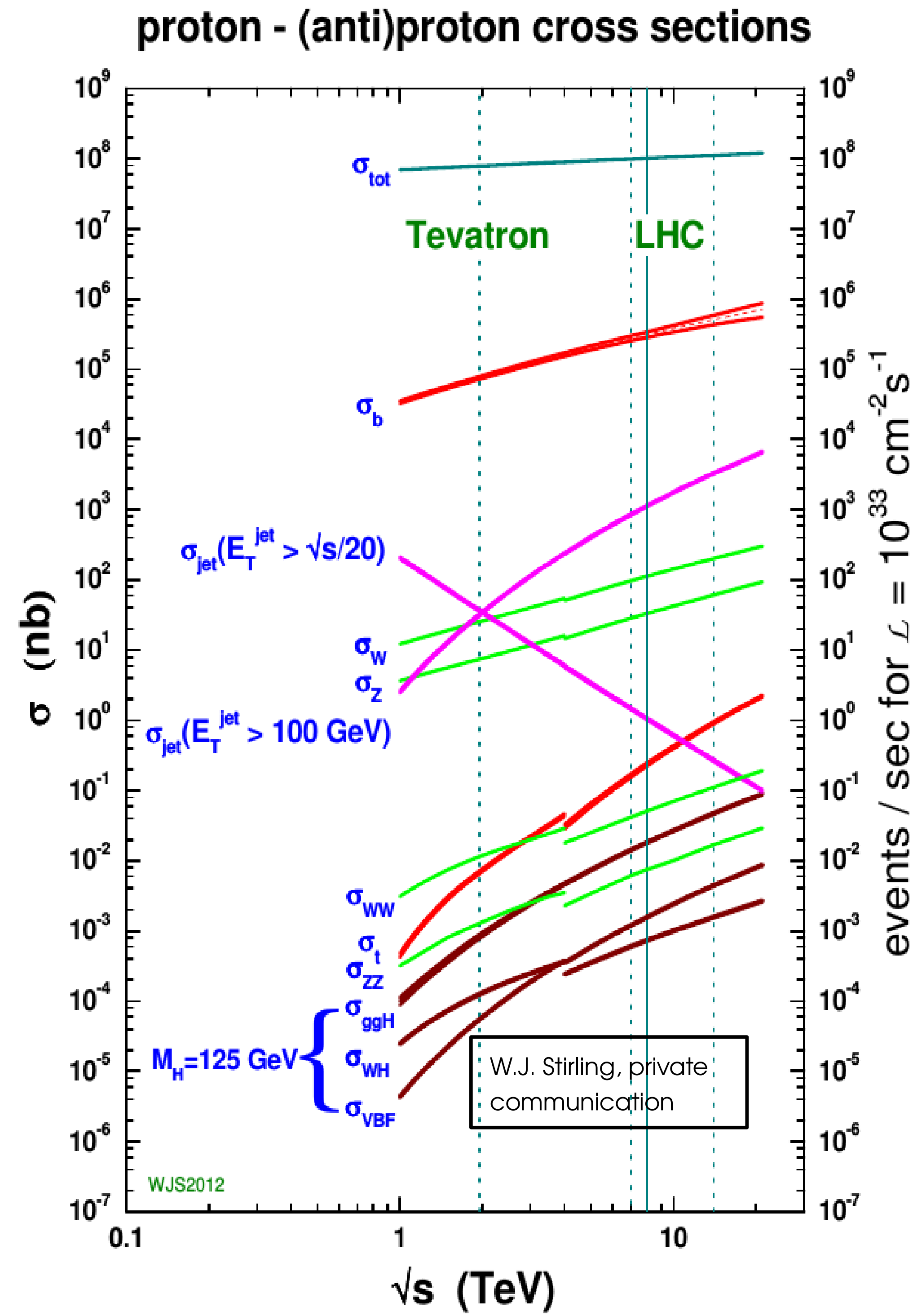
- **HEP** : almost always use Poisson distributions.
- **ATLAS** :
  - **Collision event rate ~ 1 GHz**

    ($L \sim 10^{34}\ cm^{-2}s^{-1} \sim 10\ nb^{-1}/s$, $\sigma_{tot} \sim 10^{8}$ nb,)
  - **Trigger rate ~ 1 kHz**

    (Higgs rate **~ 0.1 Hz**)
- **$p \sim 10^{-6}$** ($p_{Higgs} \sim$ **10^{-10}**)
- A year of data: **$n \sim 10^{16}$**

  ⇒ **Poisson regime!**

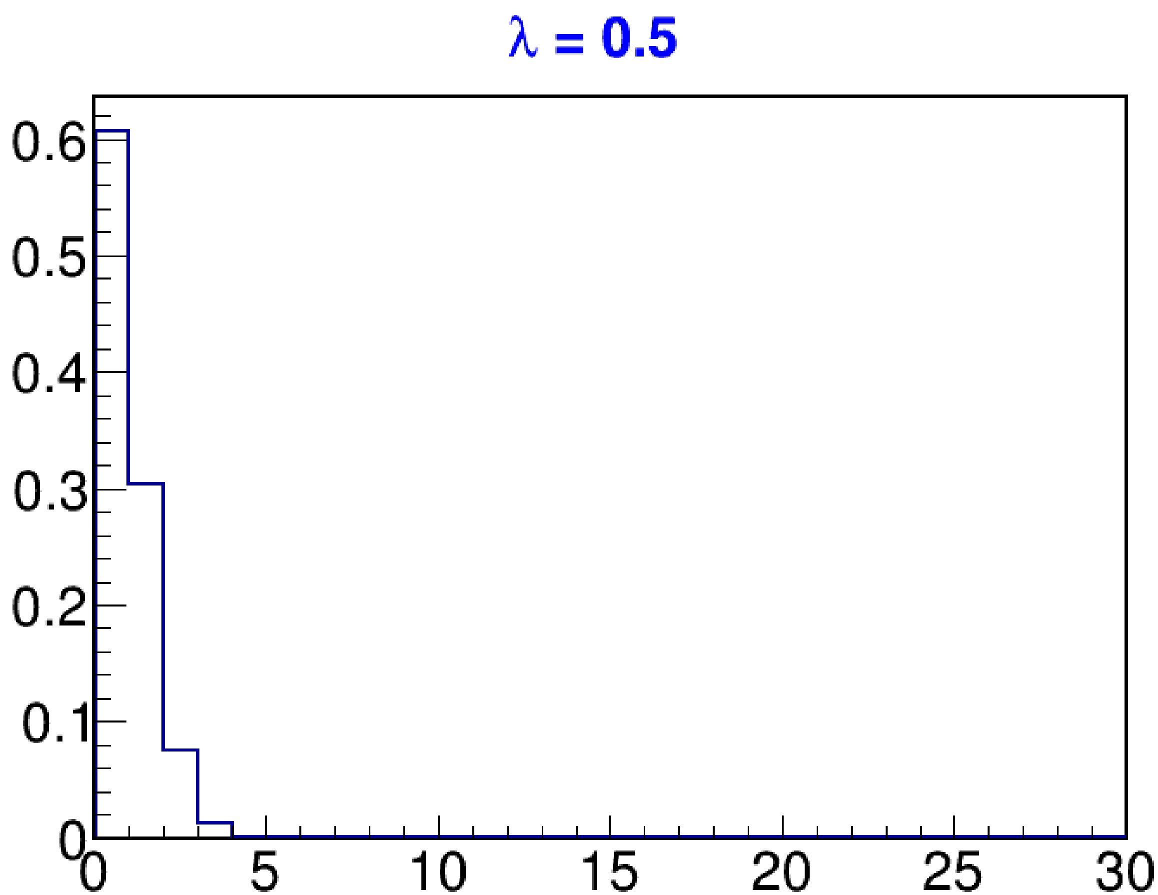
  (Large n = design requirement, to get not-too-small λ=np...)



Poisson Distributions

$$P(k;\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}$$

λ : expected number of events



Mean = λ

Variance = λ

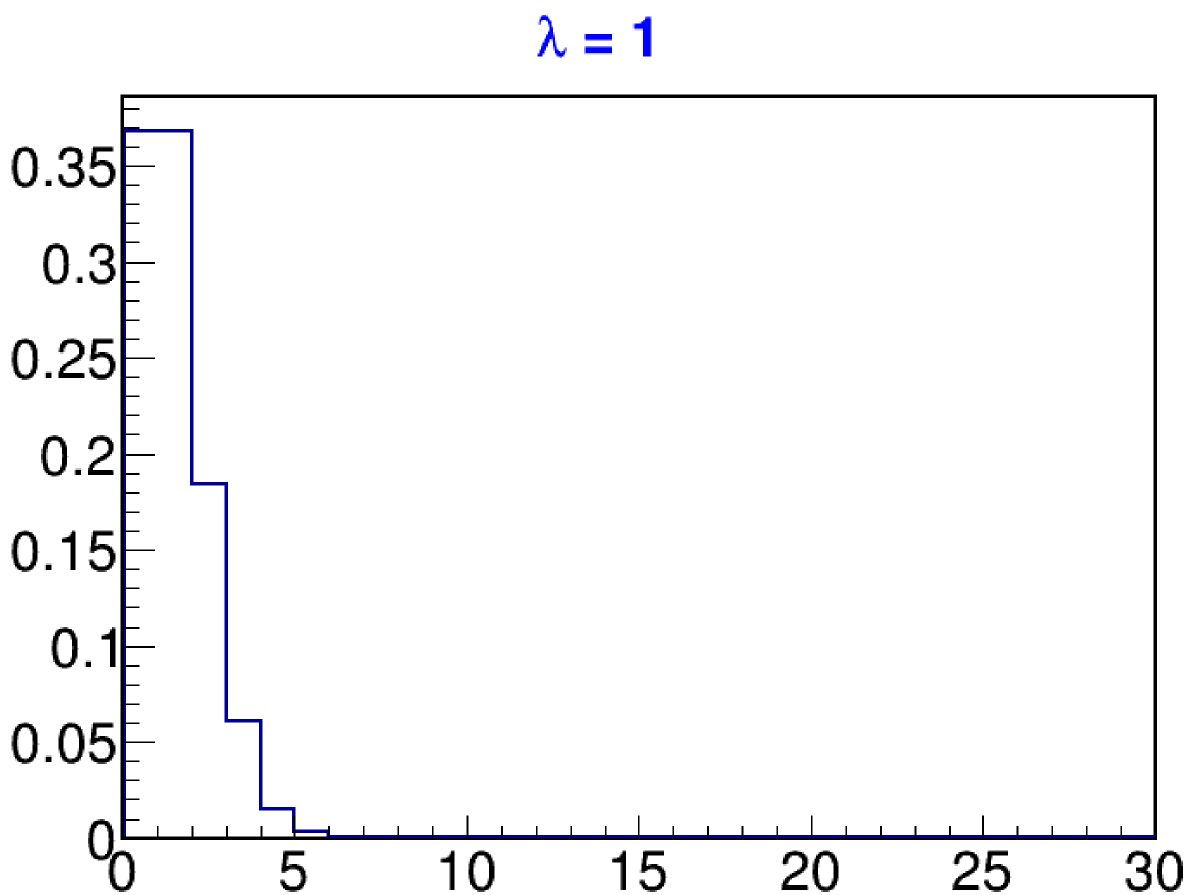
RMS = √λ

- **Discrete distribution** (integers only), **asymmetric for small λ**
- Central limit theorem : becomes **Gaussian for large λ**
- Typical uncertainty (RMS) on N events is **√N**, for large N

Poisson Distributions

$$P(k;\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}$$

λ : expected number of events



Mean = λ

Variance = λ

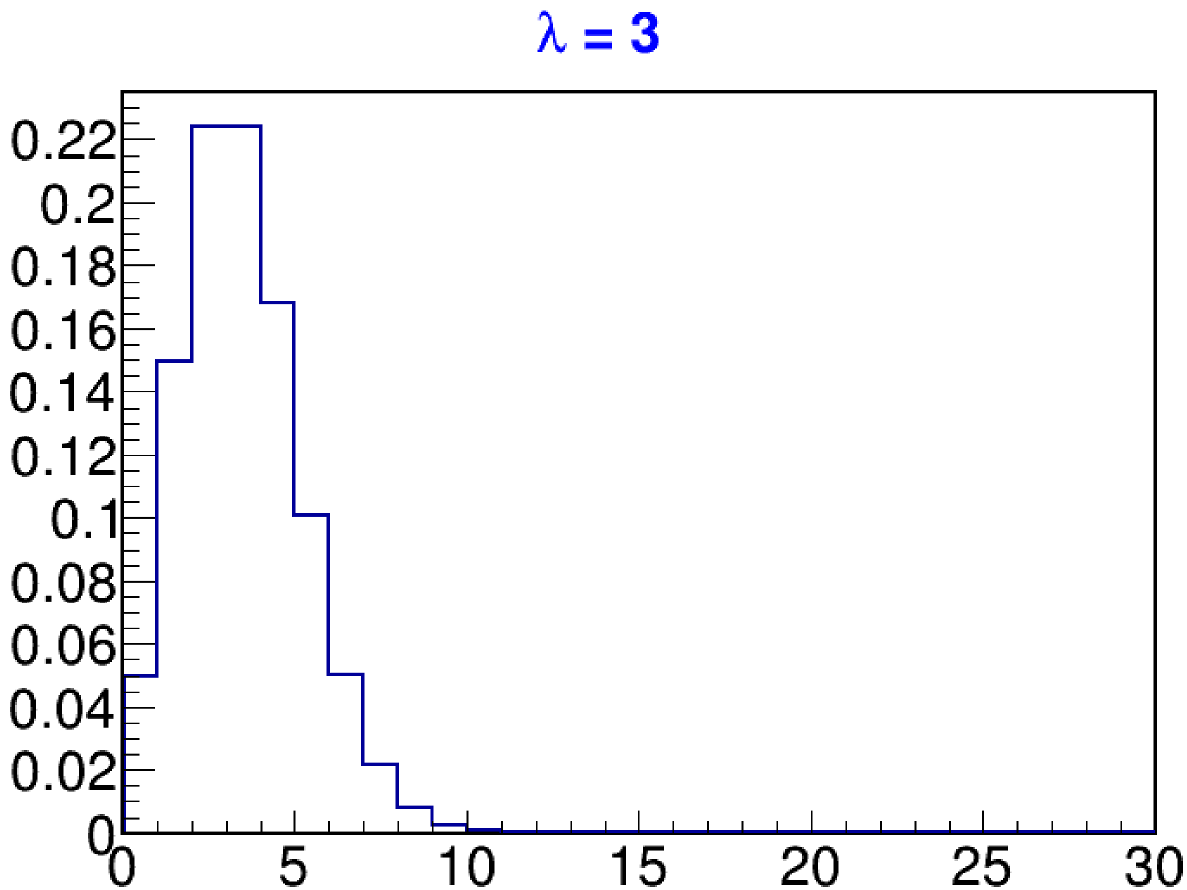
RMS = √λ

- **Discrete distribution** (integers only), **asymmetric for small λ**
- Central limit theorem : becomes **Gaussian for large λ**
- Typical uncertainty (RMS) on N events is **√N**, for large N

Poisson Distributions

$$P(k;\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}$$

λ : expected number of events



Mean = λ

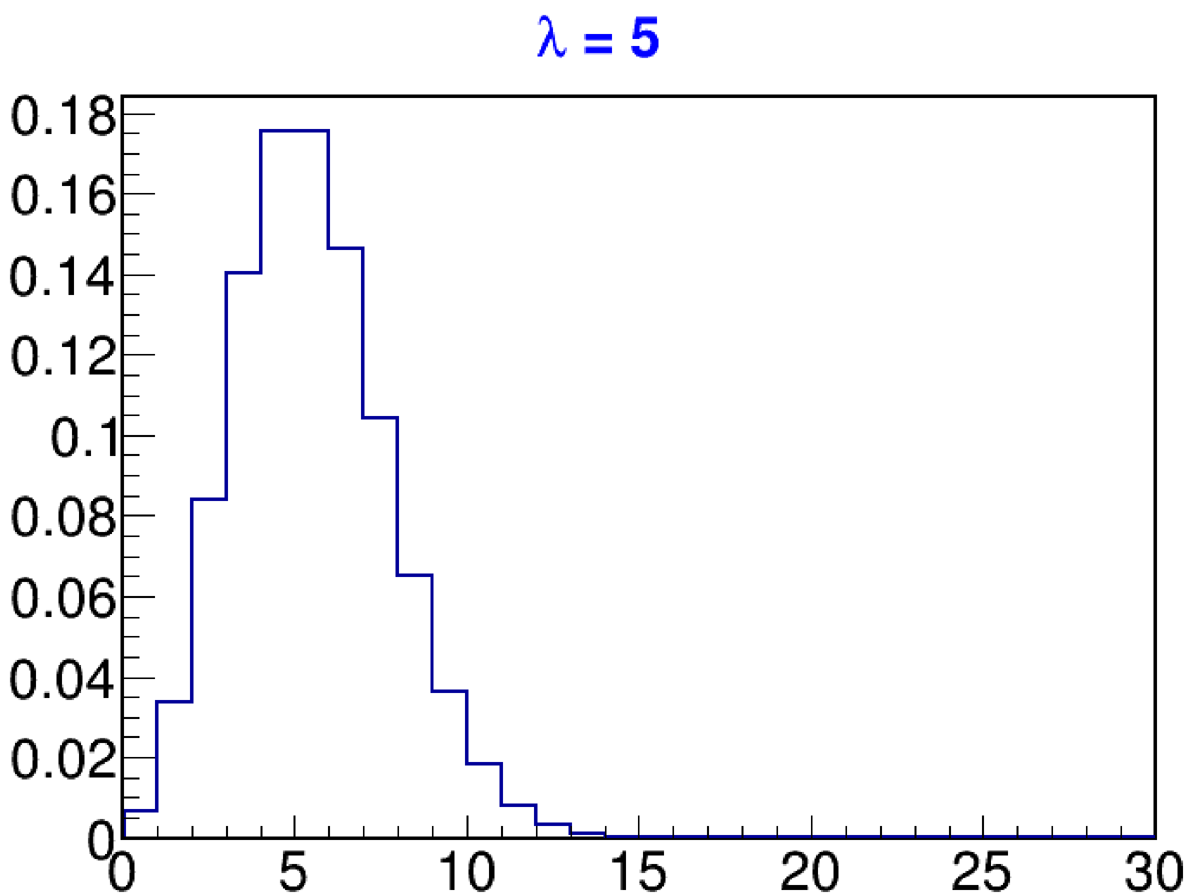
Variance = λ

RMS = √λ

- **Discrete distribution** (integers only), **asymmetric for small λ**
- Central limit theorem : becomes **Gaussian for large λ**
- Typical uncertainty (RMS) on N events is **√N**, for large N

Poisson Distributions

$$P(k;\lambda) = e^{-\lambda} \frac{\lambda^k}{k!}$$



λ : expected number of events

Mean = λ

Variance = λ

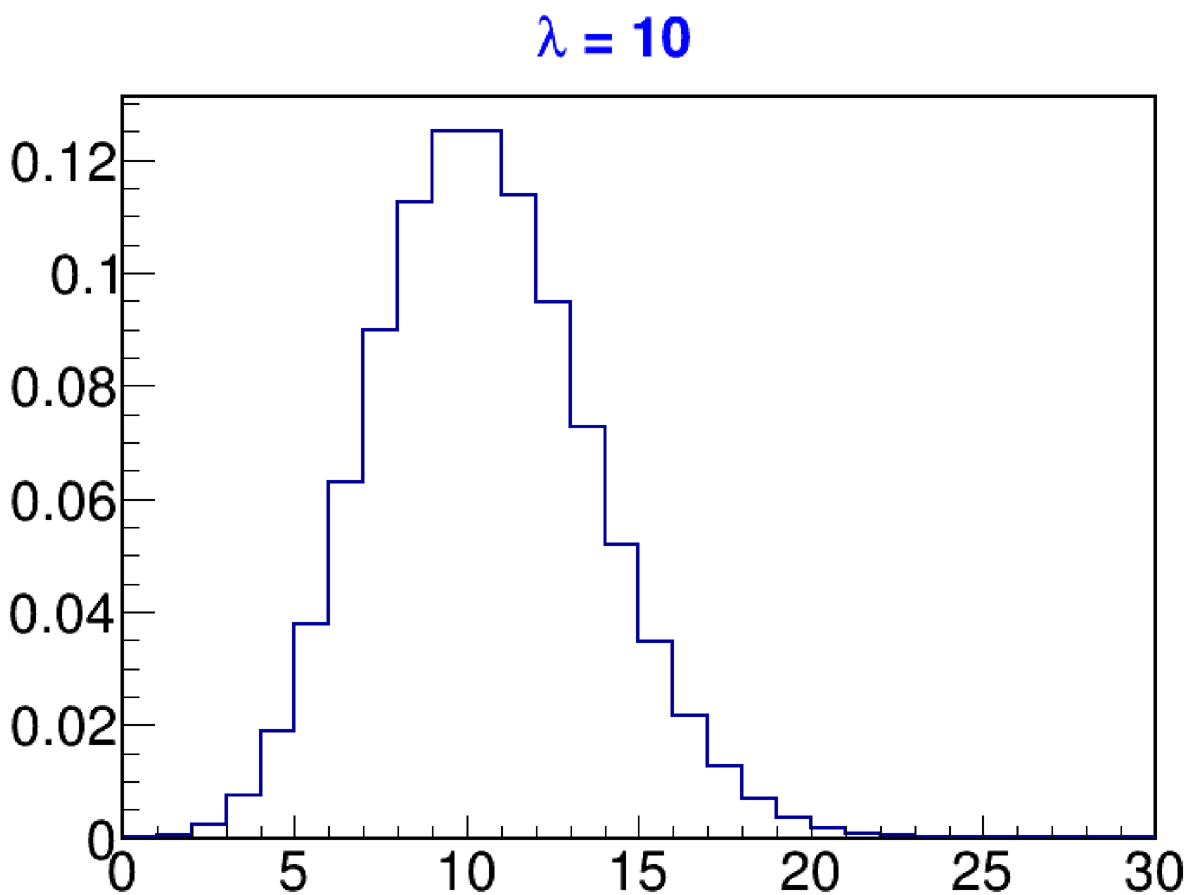
RMS = √λ

- **Discrete distribution** (integers only), **asymmetric for small λ**
- Central limit theorem : becomes **Gaussian for large λ**
- Typical uncertainty (RMS) on N events is **√N**, for large N

Poisson Distributions

$$P(k;\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}$$

λ : expected number of events



Mean = λ

Variance = λ

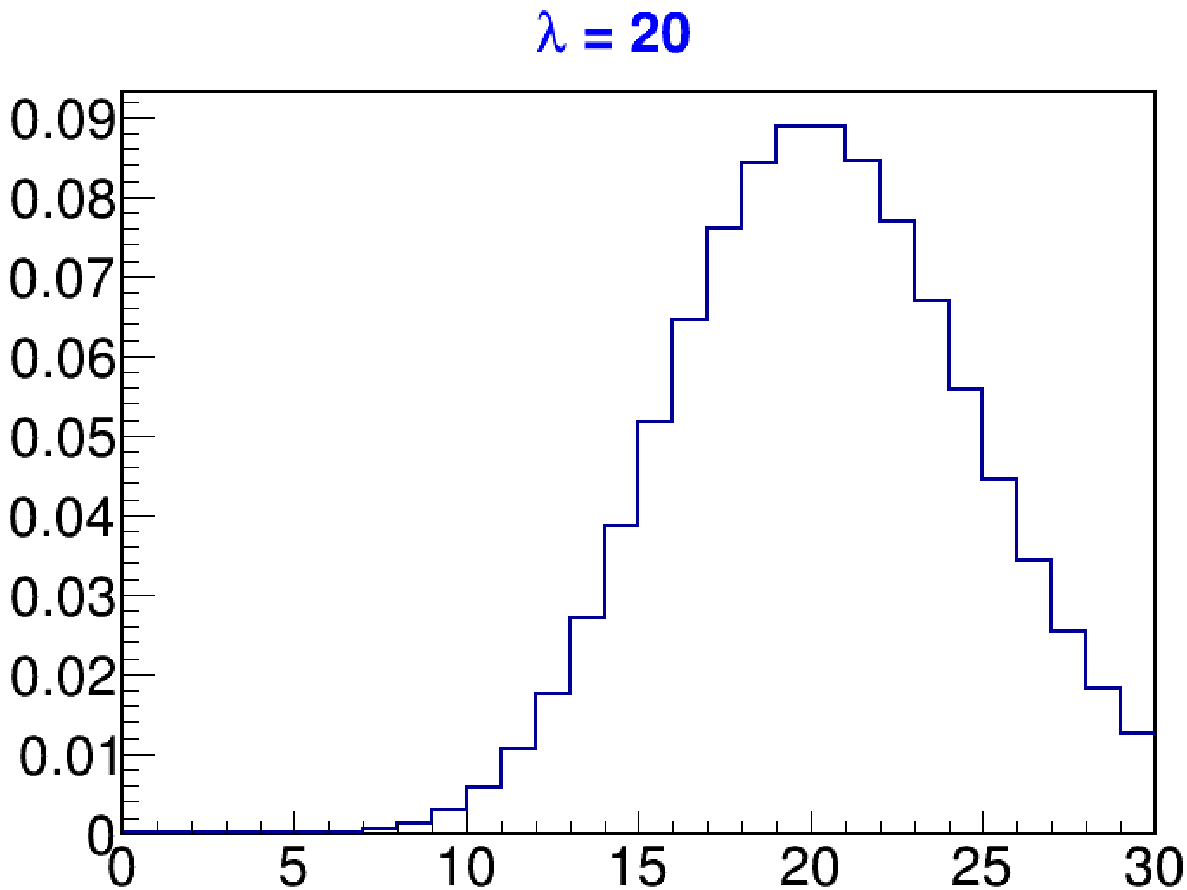
RMS = $\sqrt{\lambda}$

- **Discrete distribution** (integers only), **asymmetric for small λ**
- Central limit theorem : becomes **Gaussian for large λ**
- Typical uncertainty (RMS) on N events is **$\sqrt{N}$**, for large N

Poisson Distributions

$$P(k;\lambda) = e^{-\lambda} \frac{\lambda^k}{k!}$$

λ : expected number of events



Mean = λ

Variance = λ

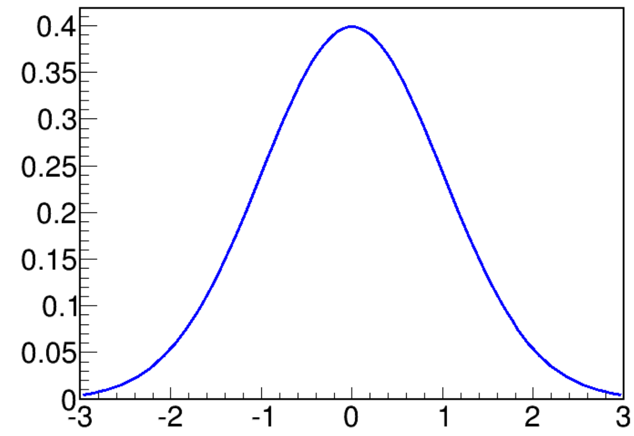
RMS = $\sqrt{\lambda}$

- **Discrete distribution** (integers only), **asymmetric for small λ**
- Central limit theorem : becomes **Gaussian for large λ**
- Typical uncertainty (RMS) on N events is **$\sqrt{N}$**, for large N

What we have learned so far (1)

- **PDFs**: give the probability to obtain each possible value in a random process
- **Examples**
  - **Gaussian**:

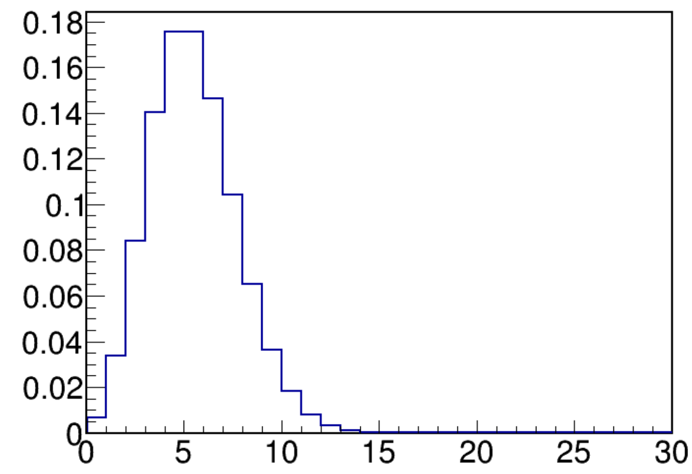
$$G(x;x_{0,}\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-x_0)^2}{2\sigma^2}}$$



    - To describe a continuous variable
    - For large numbers of events, processes become Gaussian

  - **Poisson** :

$$P(k;\lambda)=e^{-\lambda}\frac{\lambda^k}{k!}$$



    - generally used for counting events

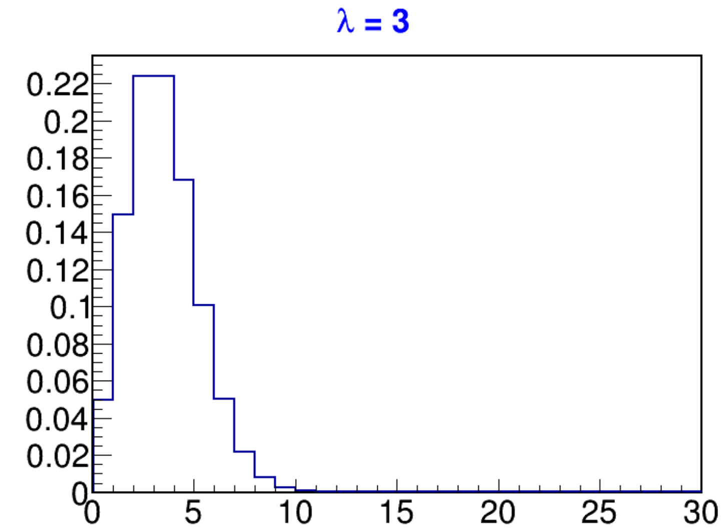
Building a Statistical Model

Statistical Model

- **Goal**: Quantify our knowledge using PDFs:
  Build a **Statistical Model**
- Includes
  - Assumptions about **what we know** (physics, etc.)
  - PDFs of random variables: statistical description **what we don't know.**
- The statements we can make have a **probabilistic** meaning:
  - **Not** m_H=125.5 GeV but **$124.95 < m_H < 125.77$ GeV with 68% confidence**
  - **Not** "there exists a Higgs boson" but **"exists with 99.9999% (5σ) confidence"**
- For these statements to be correct the **PDFs need to correctly describe the distributions of the random variables..**
  - Not always easy or possible...

Example: Model for Counting

- **Counting experiment:**
  - observe a **number of events n**
  - describe by a **Poisson distribution**



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

- With signal and bkg:

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

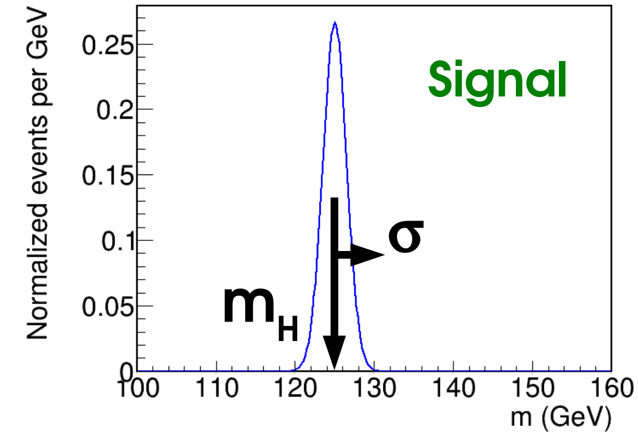
- We have **assumed** a Poisson distribution for n : This is our model, based on physics knowledge.
- Model has **parameters** (s,b), a priori unknown.
- For example, can **assume b is known**.
  ⇒ **Goal**: use the **measured n** to find out about the **parameter s.**

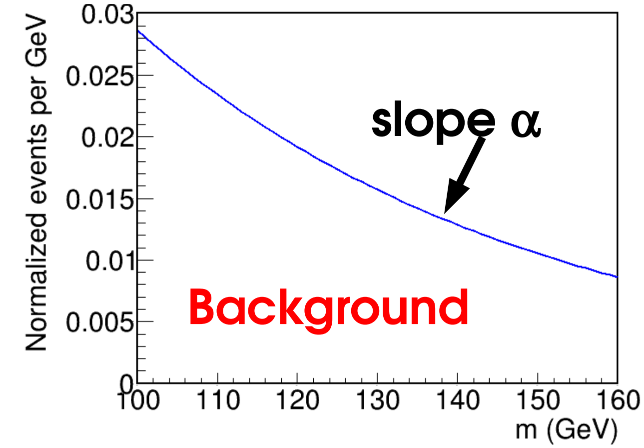
Example: Shape Analysis

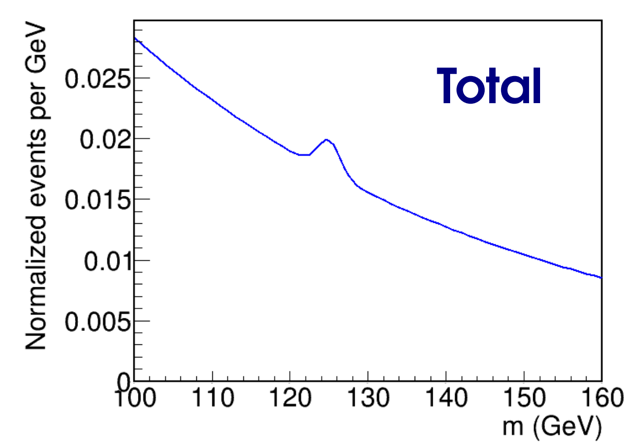
- Shape analysis experiment
  - **observe a set of masses m_1... m_n**
- Describe **shape** of m_i distribution using
  - **Gaussian signal** $P_{signal}(m) = G(m; m_0, \sigma)$
  - **Exp. background** $P_{bkg}(m) = \alpha \exp(-\alpha m)$
  - expected yields : **s, b**
- **Overall PDF:**

$$P_{Total}(m) = \frac{s}{s+b} G(m; m_H, \sigma) + \frac{b}{s+b} \alpha \exp(-\alpha m)$$

- We have **assumed**
  - A signal shape (detector response)
  - A background shape (physics)
- **Parameters s, b, m_H**... are **unknown**: measure using the **observed m_i**







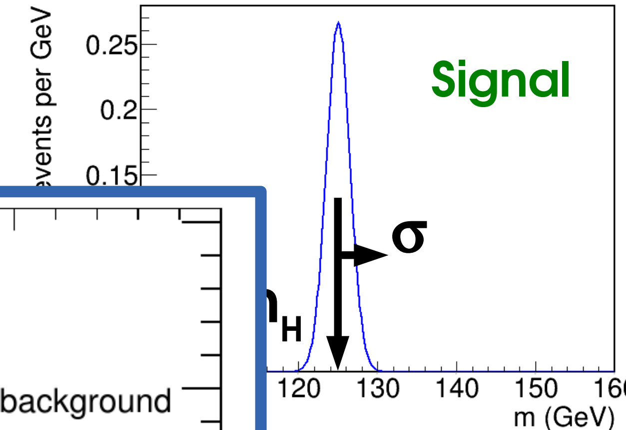
Example: Shape Analysis

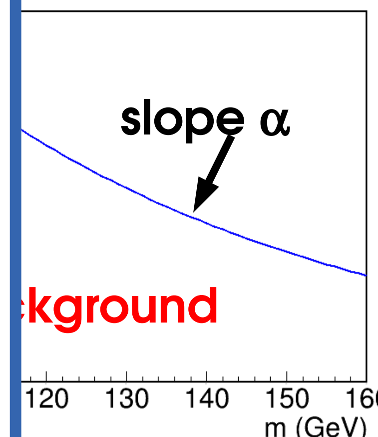
- Shape (
  – obse
- Describ
  – Gaus
  – Exp.
  – expe
- **Overall**

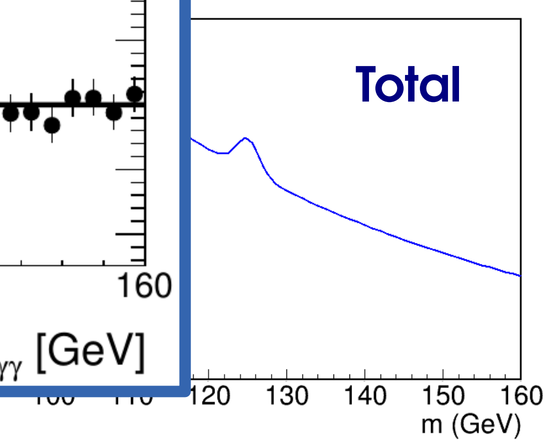
$(-\alpha\ m)$

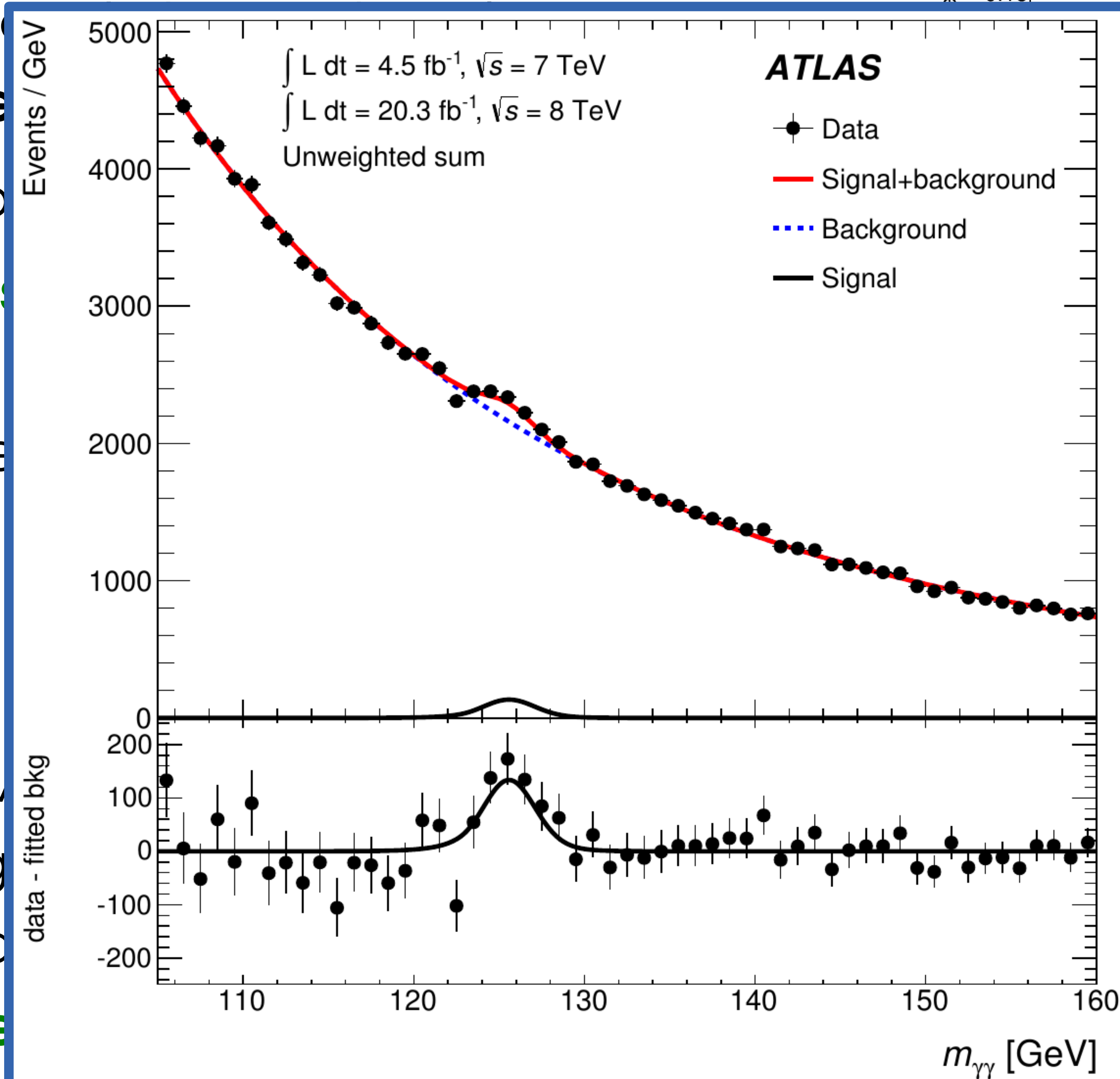
- We hav
  – A sig
  – A bo
- Parame

measure using the observed m_i



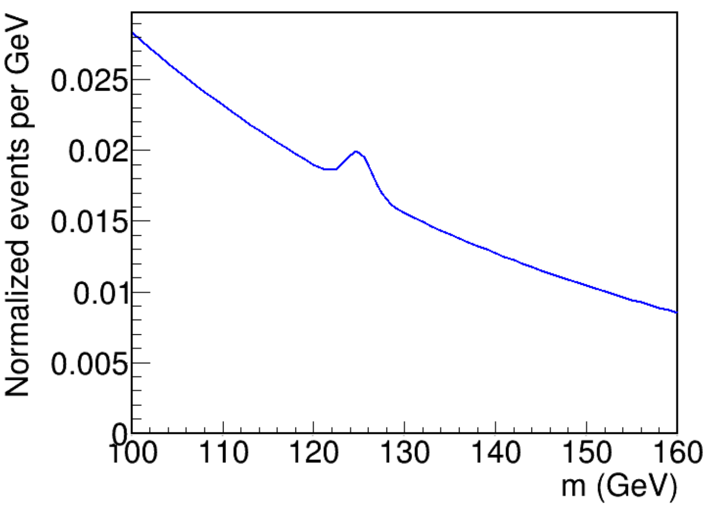




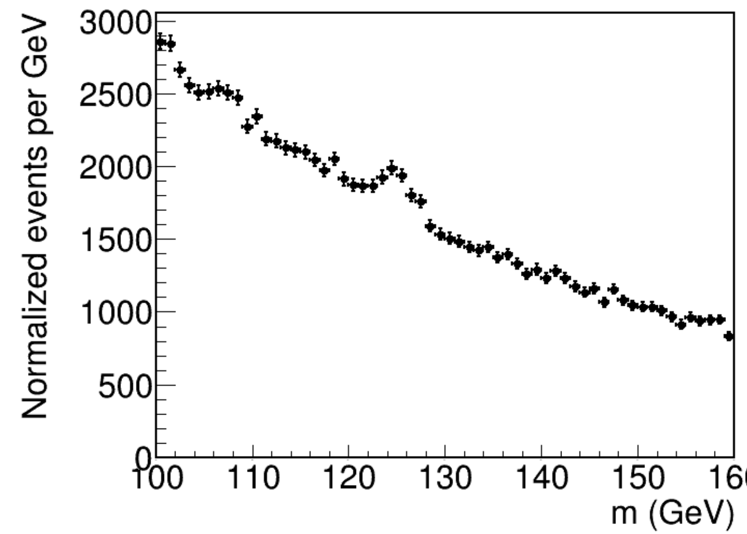


Monte-Carlo Generation

- Model describes the **distribution of the observable:**
  ⇒ Possible outcomes of the experiment,
  for given parameter values
- Can draw random events according to PDF
  "generate pseudo-data" (a.k.a. **"Monte Carlo"**)



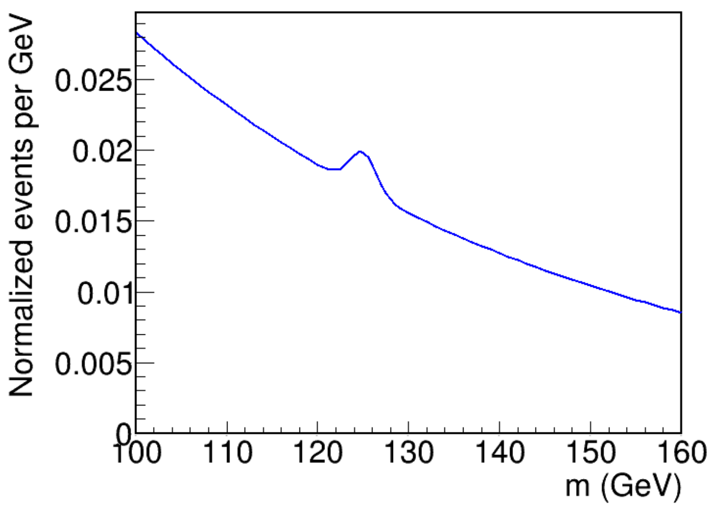
Generate
100,000 events



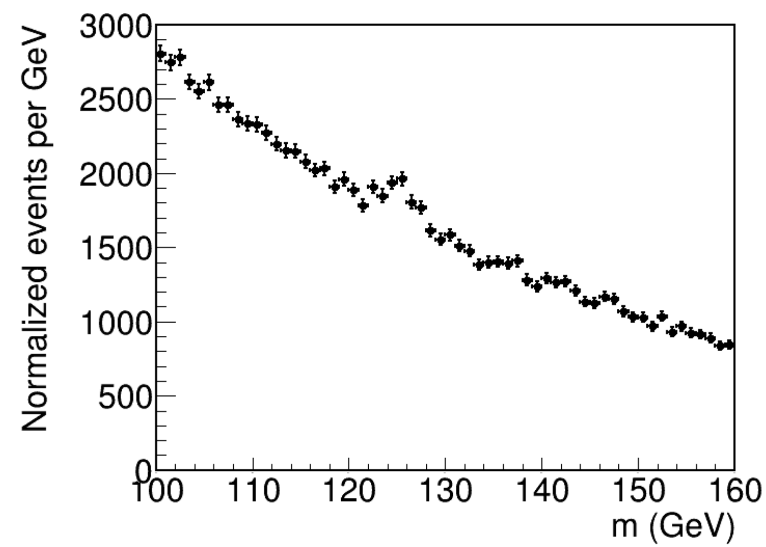
- Useful to design measurement, compute expected results
- Real MC involves realistic physics models, detector response, etc. this is **"Toy MC"**.

Monte-Carlo Generation

- Model describes the **distribution of the observable:**
  ⇒ **Possible outcomes of the experiment,**
  for given parameter values
- Can draw random events according to PDF
  “generate pseudo-data” (a.k.a. **“Monte Carlo”**)



Generate
100,000 events



- Useful to design measurement, compute expected results
- Real MC involves realistic physics models, detector response, etc. this is **“Toy MC”**.

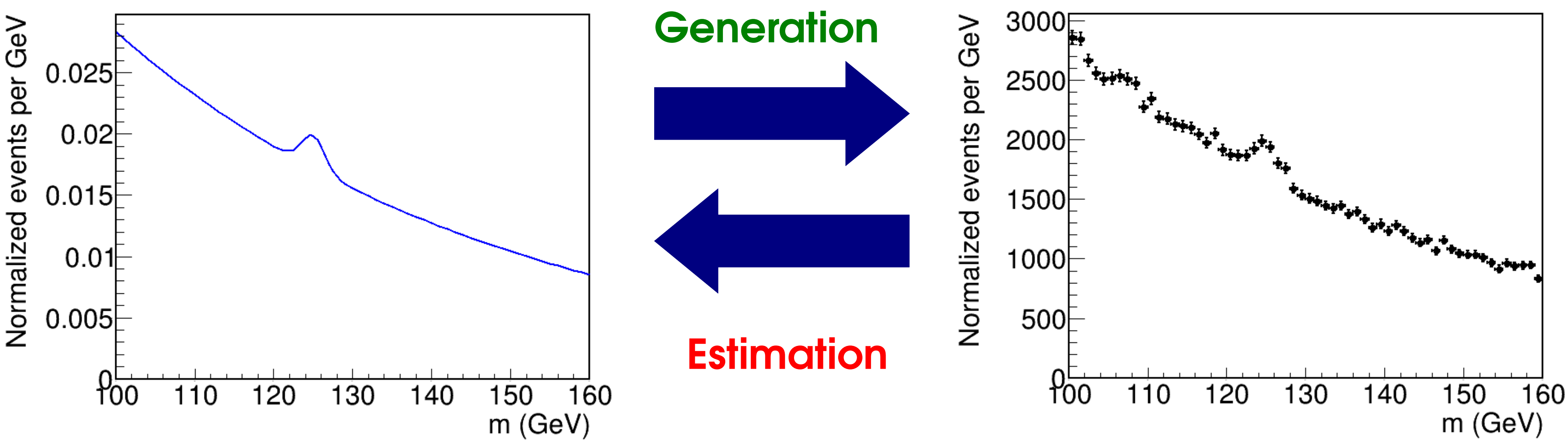
Inversion

- **MC generation**: parameter values (s, b, m_H) as **input:**

  model + **parameter values** ⇒ **pseudo-dataset**

- But what we really want is **the other direction**:

  model + **(real) dataset** ⇒ **parameter values**



⇒ **Parameter Estimation**

What we have learned so far (2)

- Need **probabilistic** description for some aspects of a measurement.
- **Use PDFs** as **building blocks** to construct a **model:**
  - **Event counting**: use Poisson distribution

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

  - **Shape analysis**: use PDF shapes that describe the distribution of signal and background.

$$P_{Total}(m;\theta_{signal},\theta_{bkg})=\frac{s}{s+b}P_{signal}(m;\theta_{signal})+\frac{b}{s+b}P_{bkg}(m;\theta_{bkg})$$

- Directly usable to **generate pseudo-data** for given parameter values.
- **Goal of the rest of these lectures: how to use data to measure the parameters**

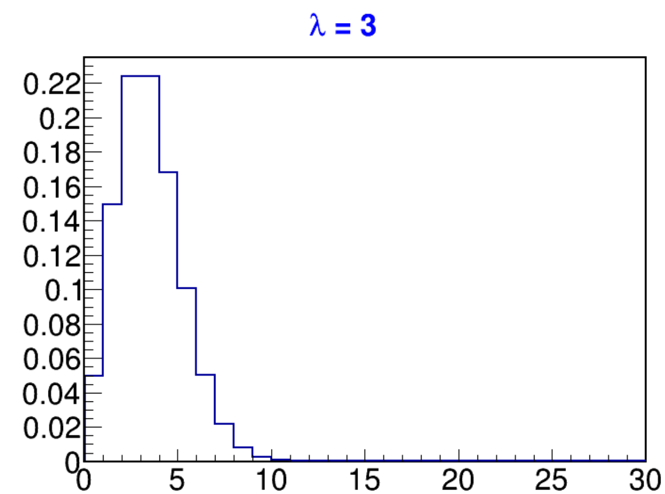
Parameter Estimation

Likelihood

- Likelihood function: same as PDF, but considered as **a function of model parameters, not the random variable**

Poisson PDF $$P(n;\lambda)=e^{-\lambda}\frac{\lambda^n}{n!} \rightarrow L(\lambda;n)=e^{-\lambda}\frac{\lambda^n}{n!}$$ Poisson Likelihood

- Purely a difference of interpretation!
- **PDF**: **given λ, how probable to observe n**
  - **Variable** : the observed data
  - **High values of PDF**: range of n where data is probable to appear



- **Likelihood**: **Given an observed n, how likely was this outcome for some λ value ?**
  - **Variable** : the model parameters.
  - **High value of the likelihood** : value of λ for which the data we observed was likely

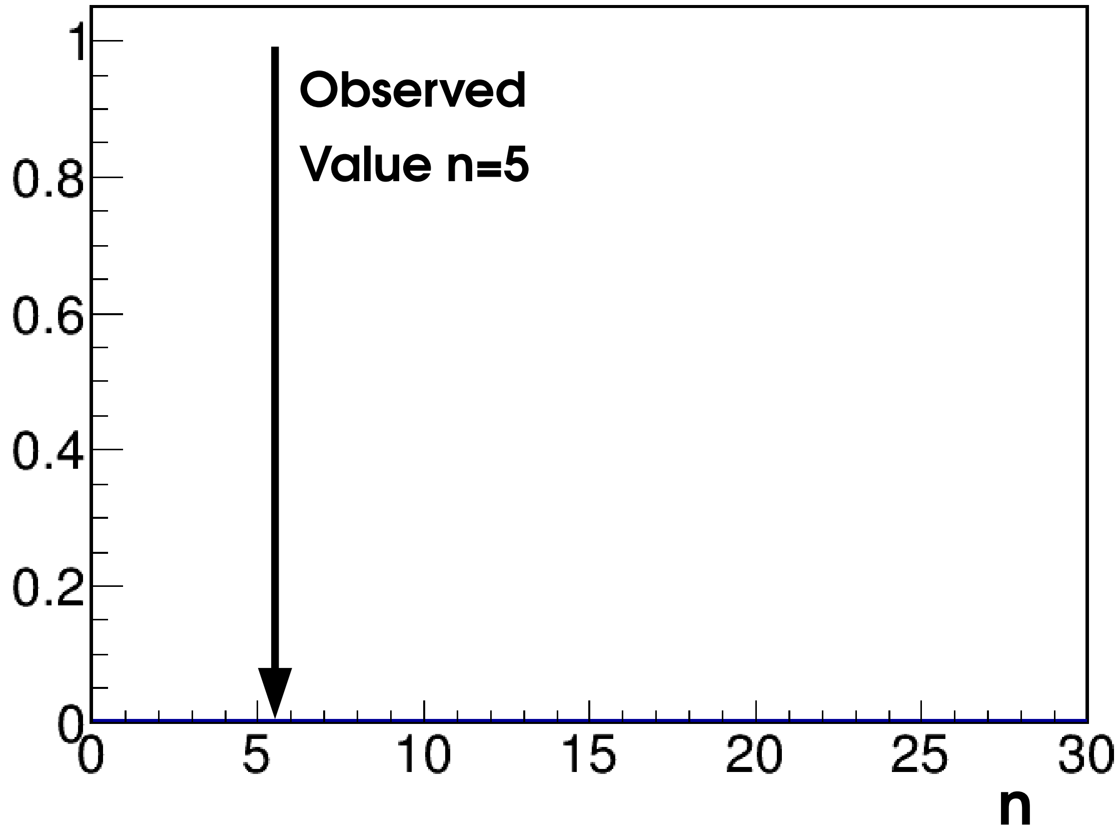
Poisson Example

- Assume **Poisson distribution** with **no background:**

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

- Say we observed **n=5**
- Data is **fixed**, **parameter s varies**

$$L(s;n=5)=e^{-s}\frac{s^5}{5!}$$



Poisson Example

- Assume **Poisson distribution** with **no background:**

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

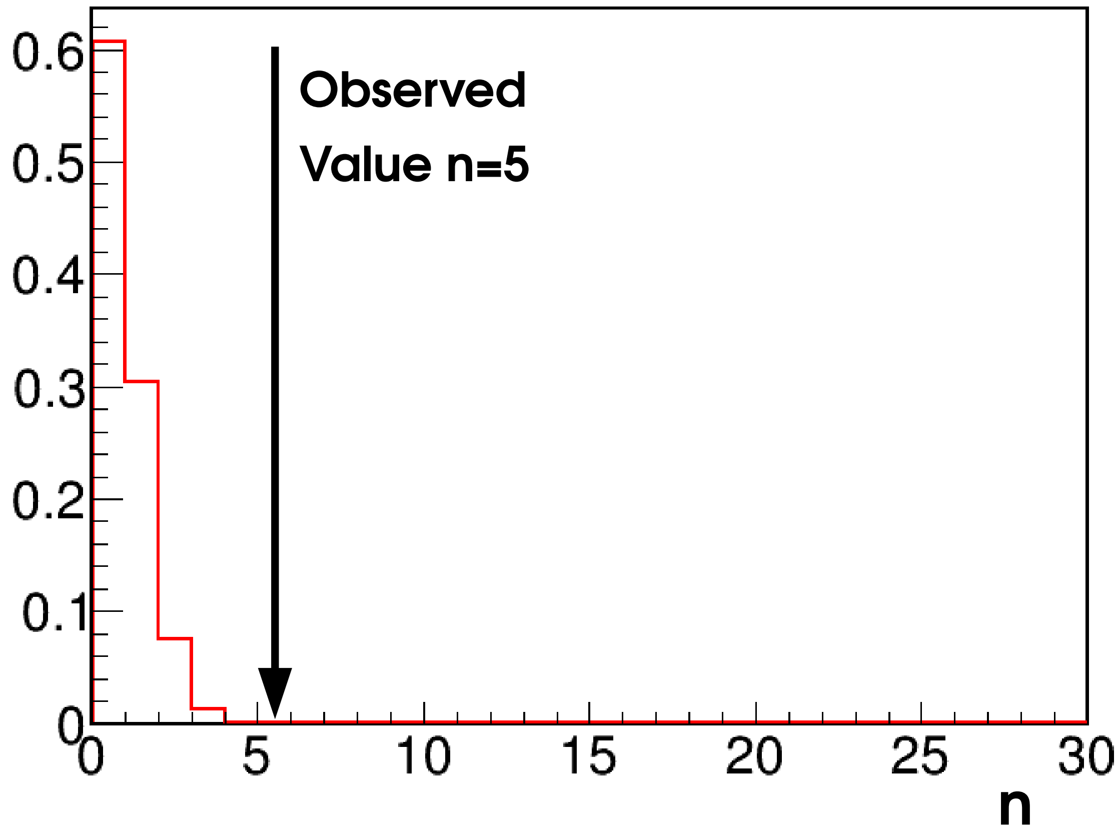
- Say we observed **n=5**
- Data is **fixed**, **parameter s varies**

$$L(s;n=5) = e^{-s}\frac{s^5}{5!}$$

s = 0.5

Low

likelihood



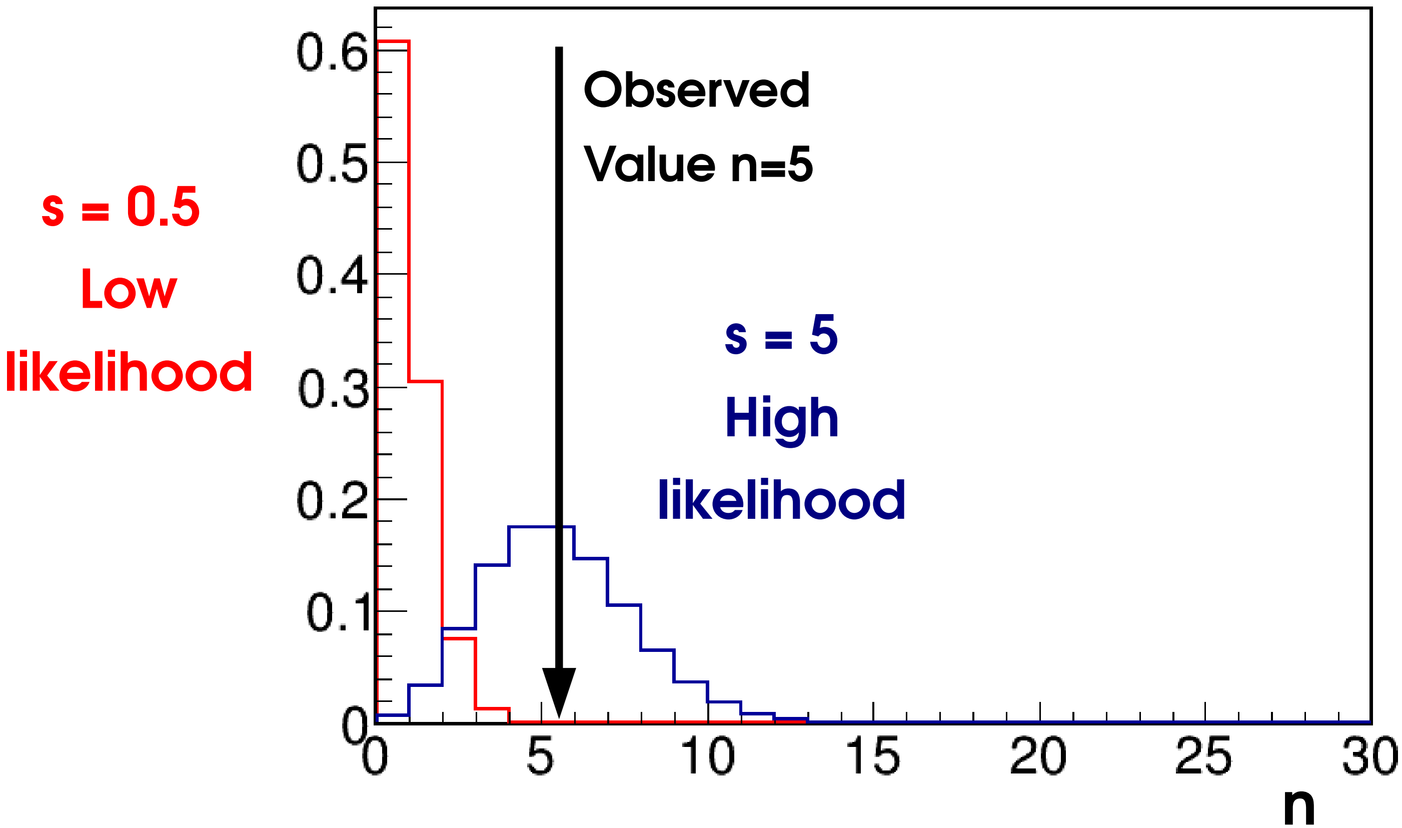
Poisson Example

- Assume **Poisson distribution** with **no background:**

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

- Say we observed **n=5**
- Data is **fixed**, **parameter s varies**

$$L(s;n=5)=e^{-s}\frac{s^5}{5!}$$



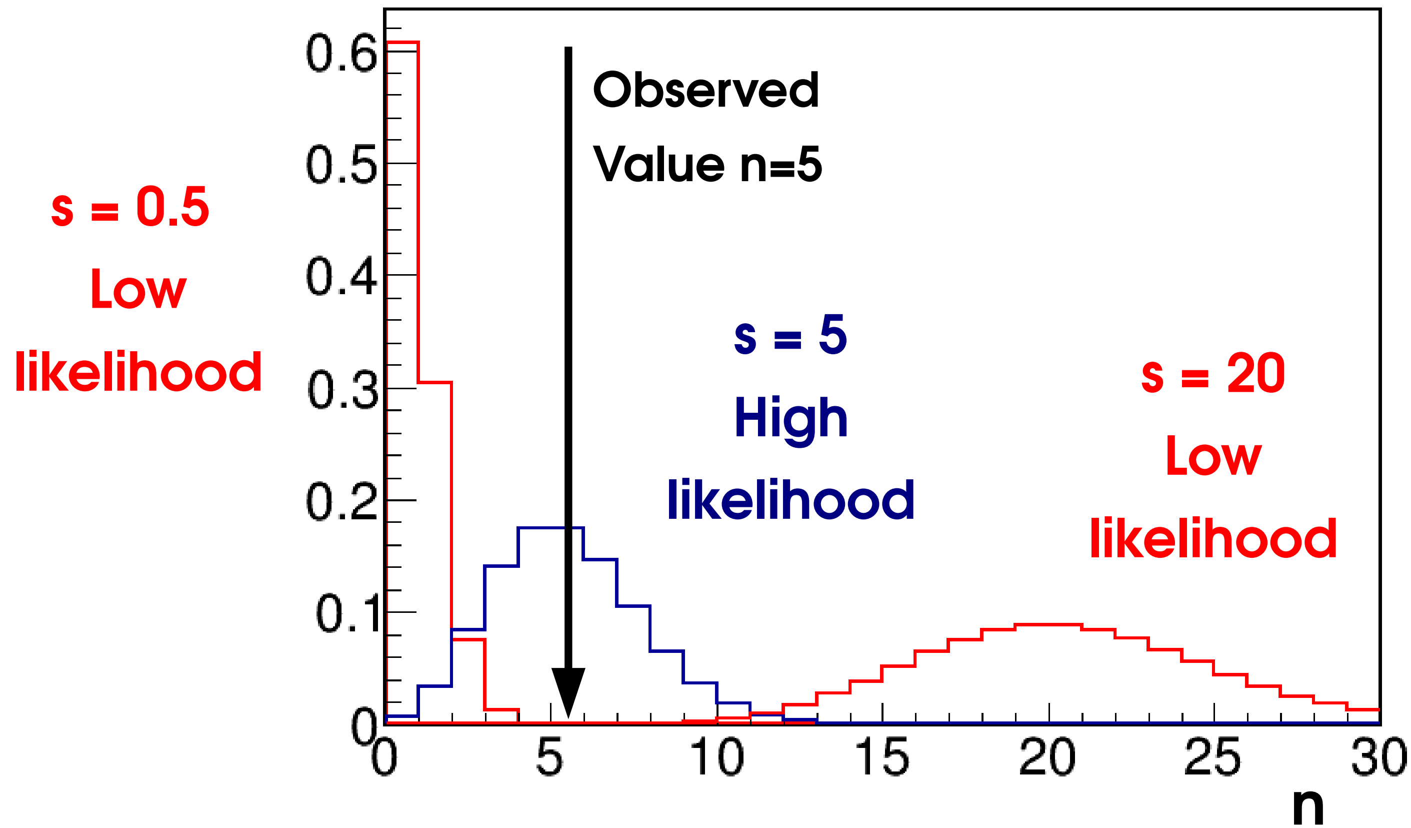
Poisson Example

- Assume **Poisson distribution** with **no background:**

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

- Say we observed **n=5**
- Data is **fixed**, **parameter s varies**

$$L(s;n=5)=e^{-s}\frac{s^5}{5!}$$



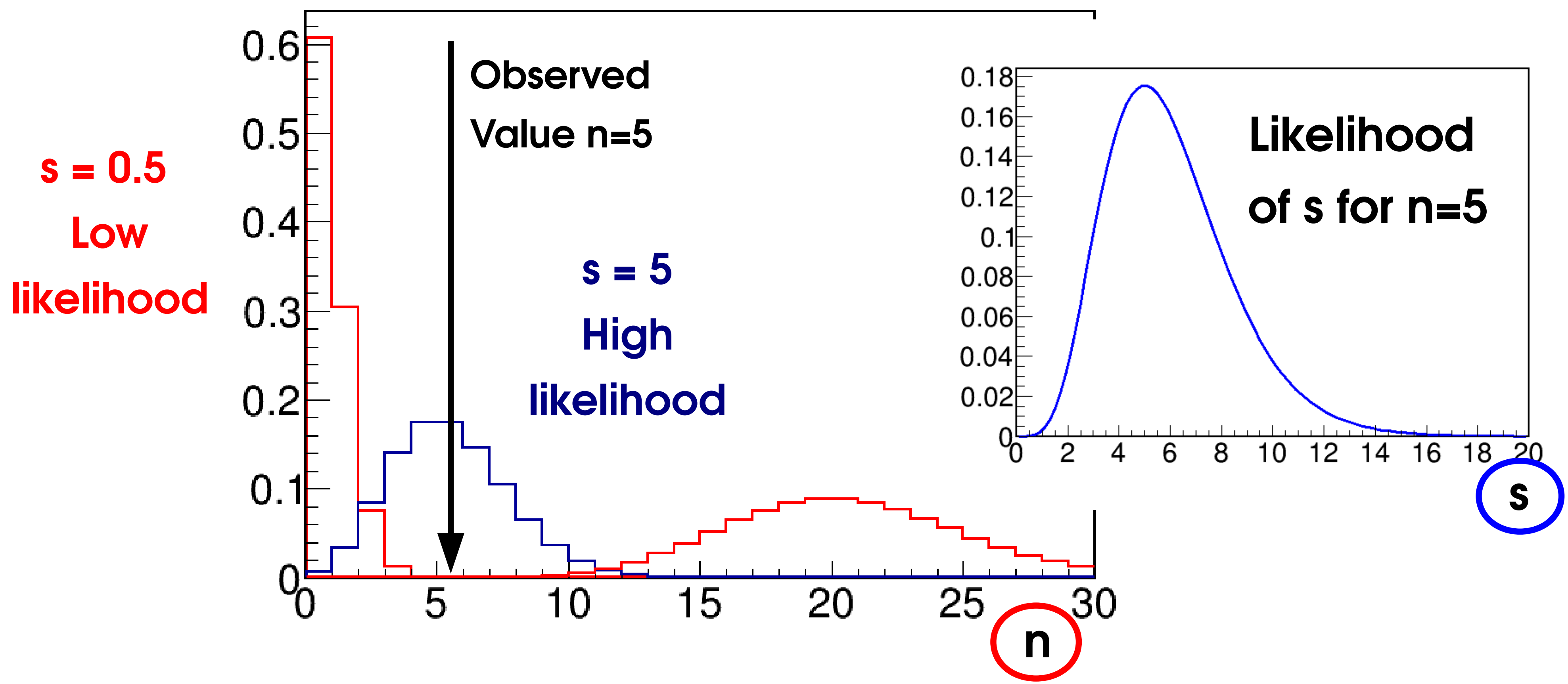
Poisson Example

- Assume **Poisson distribution** with **no background:**

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

- Say we observed **n=5**
- Data is **fixed**, **parameter s varies**

$$L(s;n=5)=e^{-s}\frac{s^5}{5!}$$

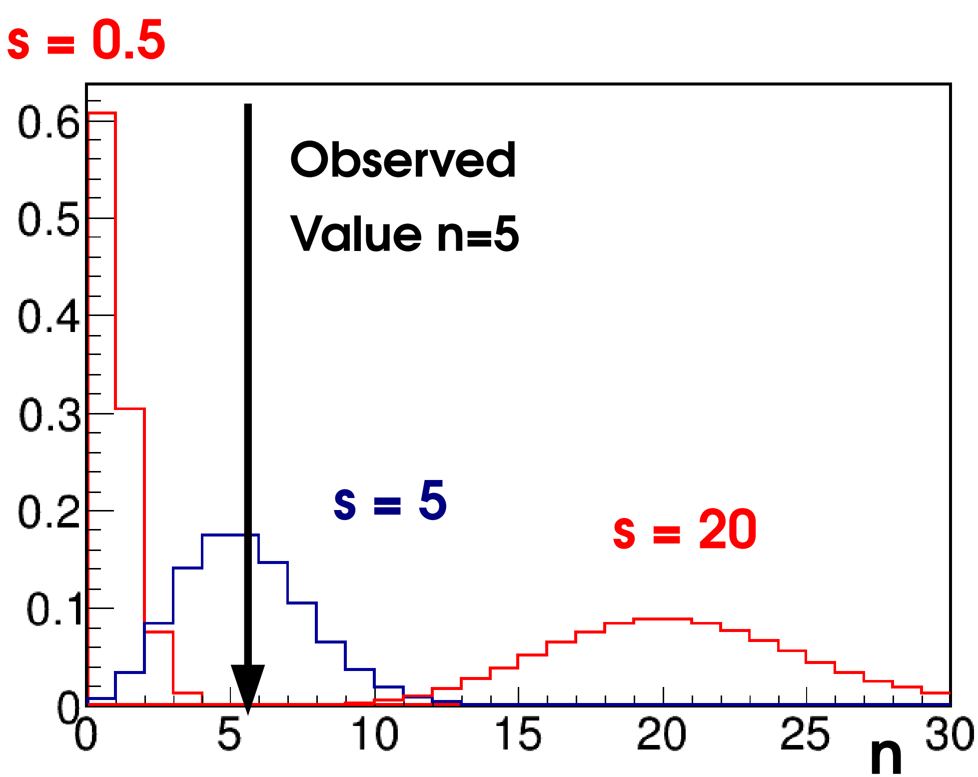


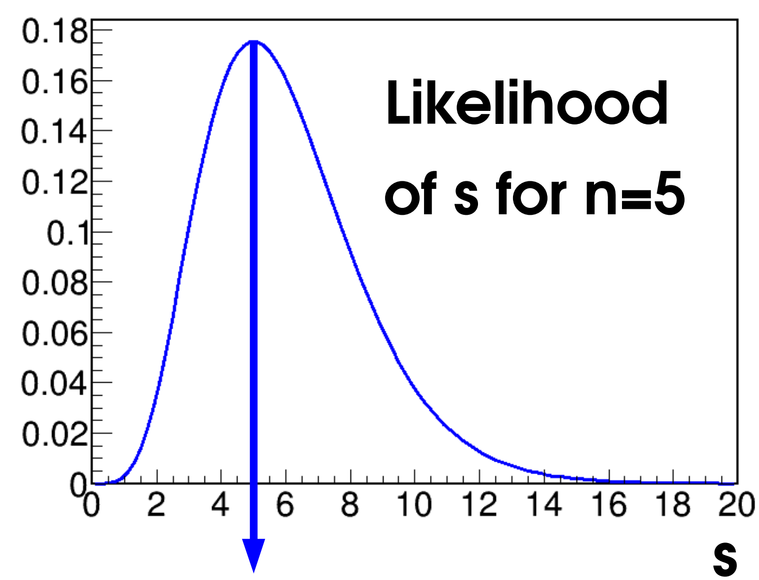
Maximum Likelihood

- To estimate a parameter θ : find the **value that maximizes $L(\theta)$**

  The value of θ for which the data was **most likely to occur**

  $\Rightarrow$ **Maximum Likelihood Estimator, $\hat{\theta}$**
  - A function of the data: $\hat{\theta}(n)$ or $\hat{\theta}(m_1,\ldots,m_n)$
  - Not guaranteed that $\hat{\theta}$ is the true value
    - sometimes the observed data is unlikely...





Maximum for s=5 : $\hat{s} = 5$

Maximum Likelihood Properties

- **Consistent**: $\hat{\theta}$ gives the true value **on average** $E(\hat{\theta})=\theta^*$
- **Asymptotically Gaussian** :

  for large datasets

$$P(\hat{\theta}) \sim \exp\left(-\frac{(\hat{\theta}-\theta^*)^2}{2\sigma_\theta^2}\right) \quad \textit{for } n \rightarrow \infty$$

- **Asymptotically Efficient** : σ_θ is the lowest possible value for an estimator for θ (in the limit $n \rightarrow \infty$)

- **Log-likelihood :**
  - Can also **minimize** λ **= -2 log L**
  - If L is Gaussian, λ is parabolic:

$$\lambda(\theta) = \left(\frac{\hat{\theta}-\theta}{\sigma_\theta}\right)^2$$

- Can drop multiplicative constants in L (additive constants in λ)

Poisson Example

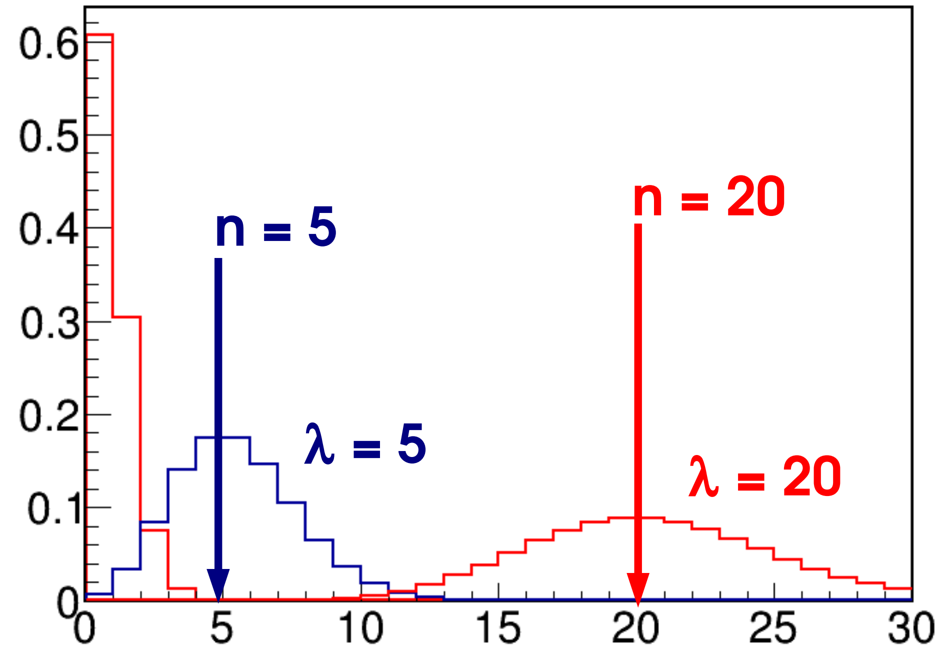
- Event counting with Poisson model, b=0

$$L(s;n)=e^{-s}s^n$$ ← dropped n!

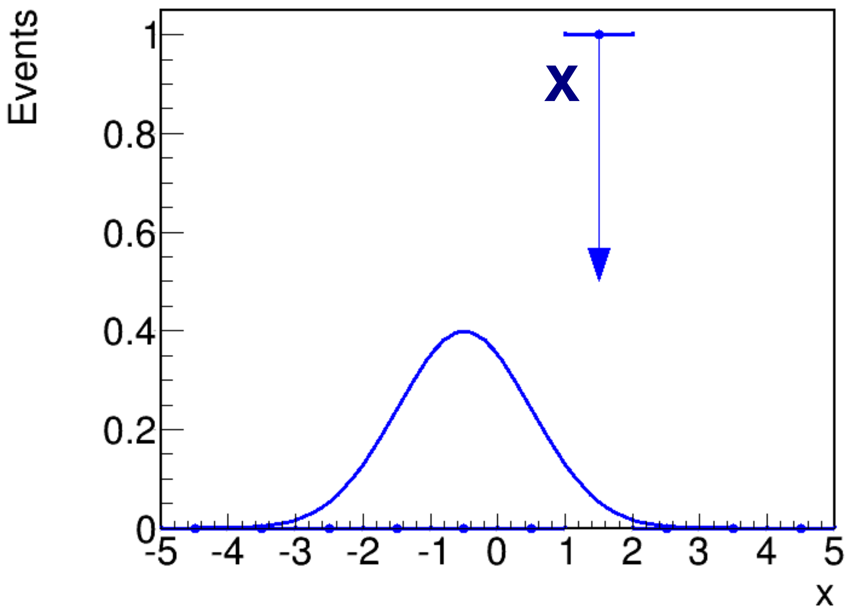
  - Peak of the poisson is always at n=s
  - ML estimate: **ŝ = n**
  - So ŝ=n is Poisson-distributed

- Properties:
  - **Consistent** E(ŝ)=E(n)=s*
  - **Gaussian** for large n

- Kind of trivial...



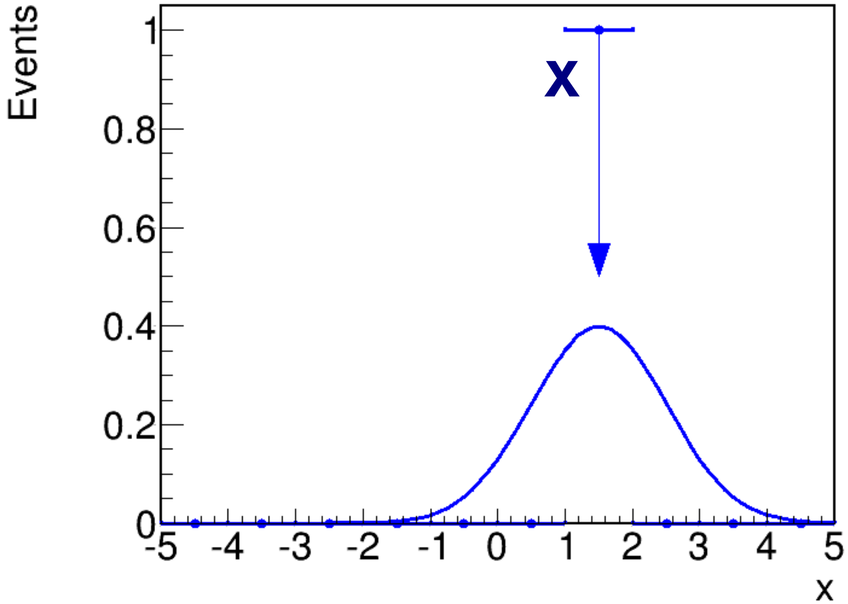
Gaussian Examples



- **Gaussian case, one measurement**
  - We measure x
  - Likelihood: $L(\theta; x) = G(x; \theta, \sigma)$
  - What is θ ? ML estimate : $\hat{\theta}$ = x.
- **Gaussian case, two measurements**
  - Measure the same parameter twice – how to **combine** ?
  - Both cases Gaussian, same mean, different resolutions
  - **Combined likelihood**: $L(\theta; x_{1,} x_2) = G(x_1; \theta, \sigma_1) G(x_2; \theta, \sigma_2)$

Log-likelihood: $\lambda(\theta; x_{1,} x_2) = \left(\frac{x_1 - \theta}{\sigma_1}\right)^2 + \left(\frac{x_2 - \theta}{\sigma_2}\right)^2$

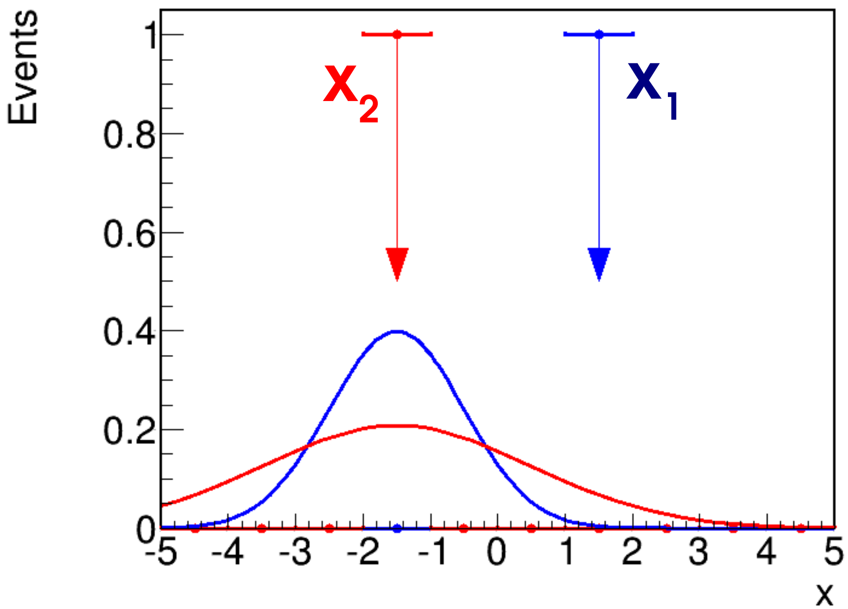
Gaussian Examples



- **Gaussian case, one measurement**
  - We measure x
  - Likelihood: $L(\theta;x) = G(x;\theta,\sigma)$
  - What is θ ? ML estimate : $\hat{\theta}$ = x.
- **Gaussian case, two measurements**
  - Measure the same parameter twice – how to **combine** ?
  - Both cases Gaussian, same mean, different resolutions
  - **Combined likelihood**: $L(\theta;x_{1,}x_2) = G(x_1;\theta,\sigma_1)G(x_2;\theta,\sigma_2)$

Log-likelihood: $\lambda(\theta;x_{1,}x_2) = \left(\frac{x_1-\theta}{\sigma_1}\right)^2 + \left(\frac{x_2-\theta}{\sigma_2}\right)^2$

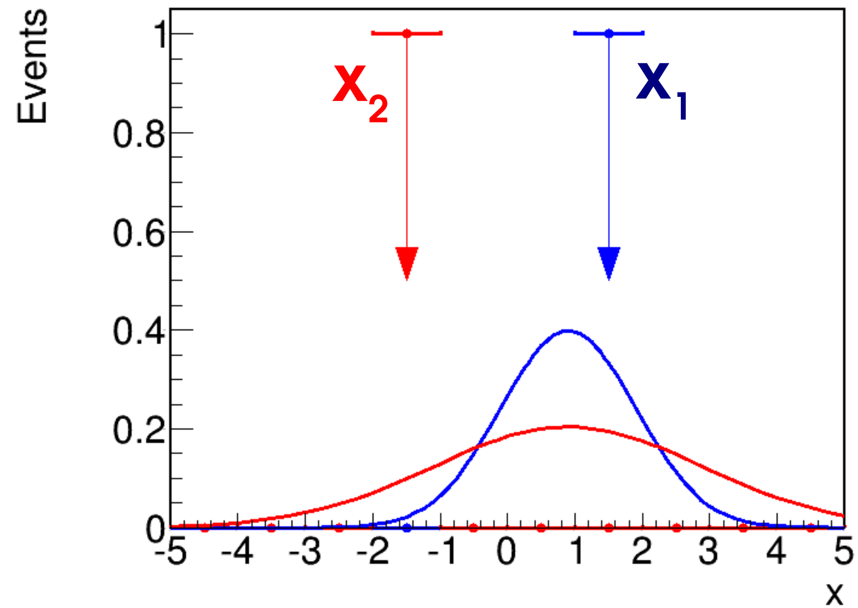
Gaussian Examples



- **Gaussian case, one measurement**
  - We measure x
  - Likelihood: $L(\theta;x)=G(x;\theta,\sigma)$
  - What is θ ? ML estimate : $\hat{\theta}$ = x.
- **Gaussian case, two measurements**
  - Measure the same parameter twice – how to **combine** ?
  - Both cases Gaussian, same mean, different resolutions
  - **Combined likelihood**: $L(\theta;x_{1,}x_2)=G(x_1;\theta,\sigma_1)G(x_2;\theta,\sigma_2)$

Log-likelihood: $\lambda(\theta;x_{1,}x_2)=\left(\frac{x_1-\theta}{\sigma_1}\right)^2+\left(\frac{x_2-\theta}{\sigma_2}\right)^2$

Gaussian Examples



- **Gaussian case, one measurement**
  - We measure x
  - Likelihood: $L(\theta; x) = G(x; \theta, \sigma)$
  - What is θ ? ML estimate : $\hat{\theta}$ = x.
- **Gaussian case, two measurements**
  - Measure the same parameter twice – how to **combine** ?
  - Both cases Gaussian, same mean, different resolutions
  - **Combined likelihood**: $L(\theta; x_1, x_2) = G(x_1; \theta, \sigma_1) G(x_2; \theta, \sigma_2)$

Log-likelihood: $$\lambda(\theta; x_1, x_2) = \left(\frac{x_1 - \theta}{\sigma_1}\right)^2 + \left(\frac{x_2 - \theta}{\sigma_2}\right)^2$$

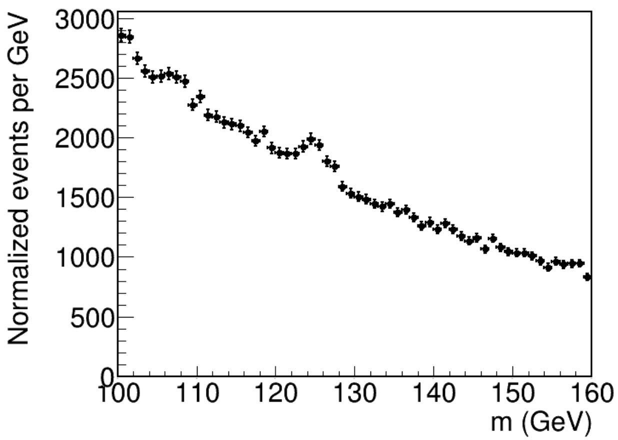
ML Estimate for θ: $$\hat{\theta} = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

Just **average** the measurements **using 1/σ² as weight.**

Likelihood for a Shape Analysis

- For a **single** measurement, **$L(\theta; x) = P(x; \theta)$**
- For a distribution of n_{obs} events, **product over events:**

$$L(\theta; x_1 \ldots x_{n_{obs}}) = \prod_{i=1}^{n_{obs}} P(x_i; \theta)$$



- **Also variations for n_{obs}**: include **Poisson term**

$$L(N_{exp}, \theta; x_1 \ldots x_{n_{obs}}) = e^{-N_{exp}} \frac{N_{exp}^{n_{obs}}}{n_{obs}!} \prod_{i=1}^{n_{obs}} P(x_i; \theta)$$

N_{exp} = total number of events expected, model parameter

- If we use $P_{Total}(m; s, b, \theta) = \frac{s}{s+b} P_{signal}(m; \theta) + \frac{b}{s+b} P_{bkg}(m, \theta)$

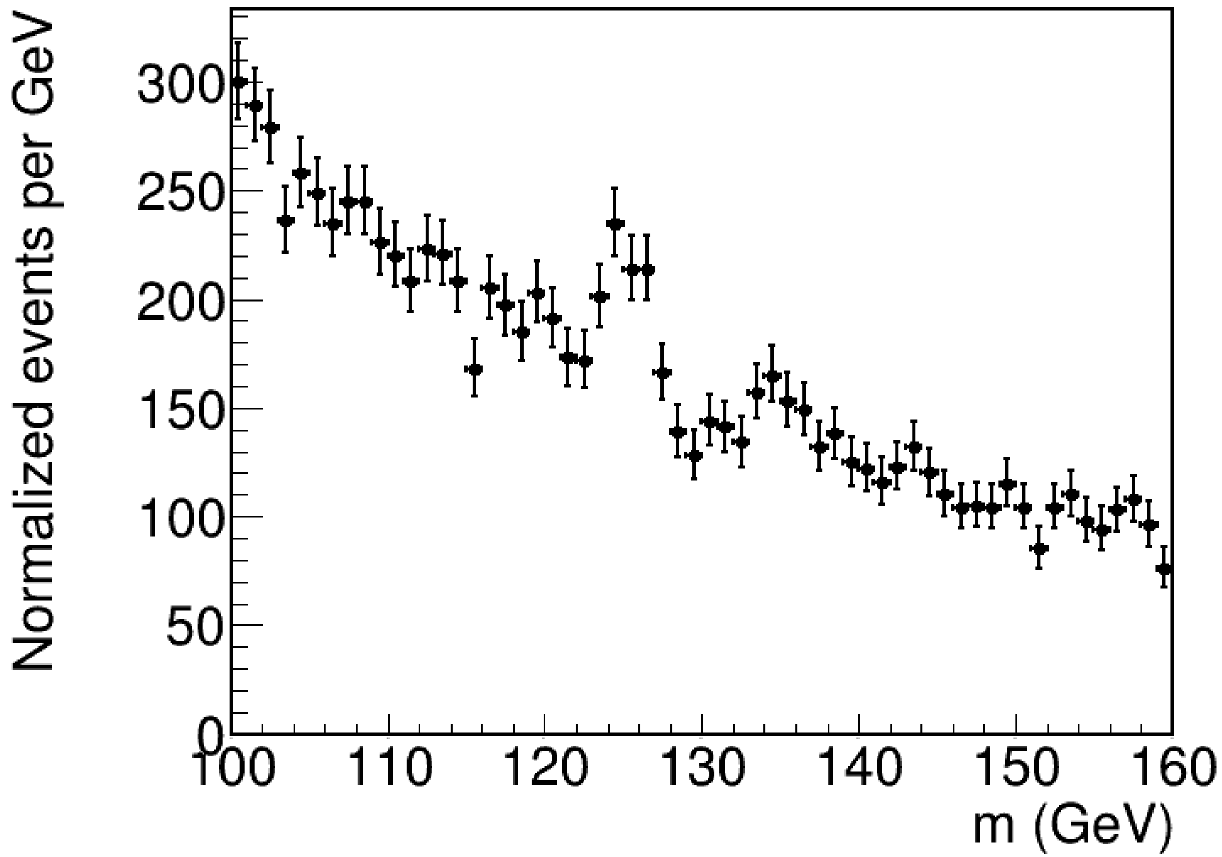
then **N_{exp} = s+b** and

$$L(s, b, \theta; m_1 \ldots m_{n_{obs}}) = e^{-(s+b)} \prod_{i=1}^{n_{obs}} s P_{signal}(m_i; \theta) + b P_{bkg}(m_i, \theta)$$

"Extended Likelihood"

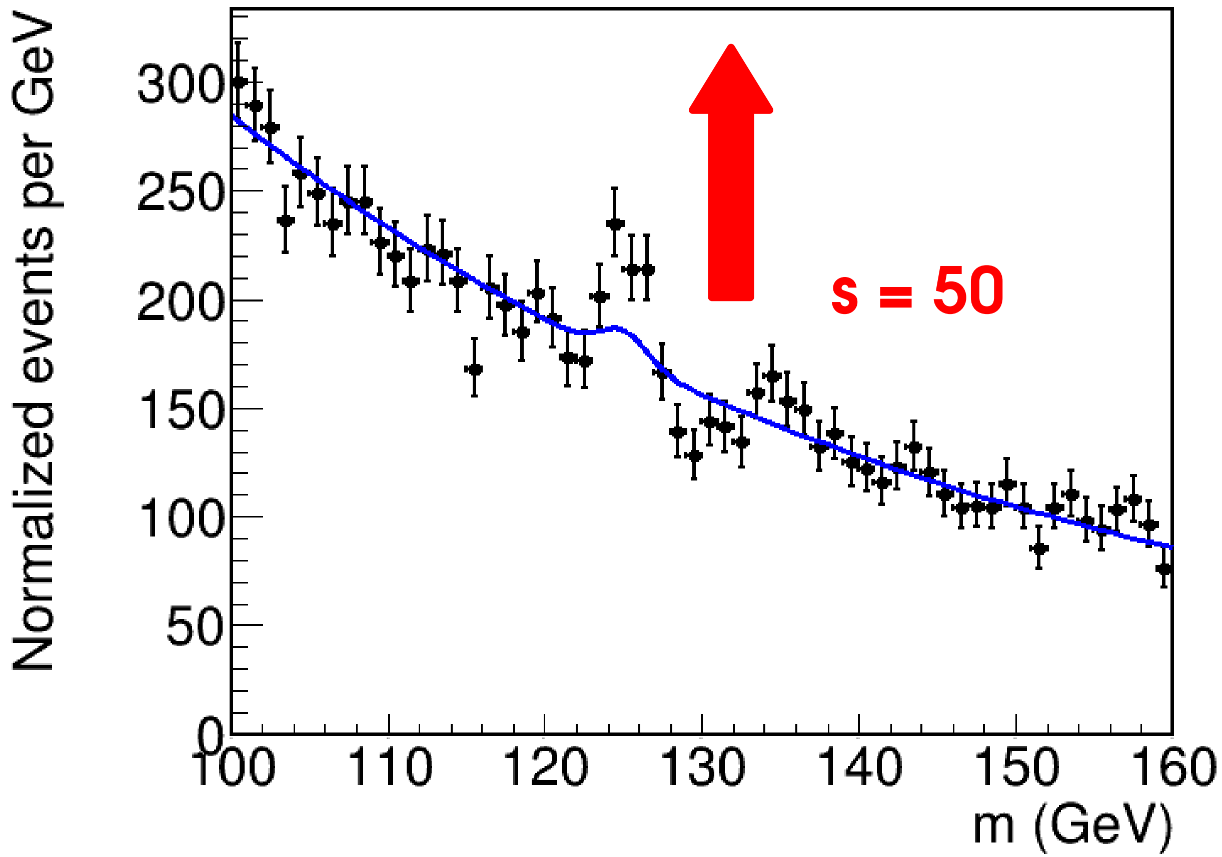
H→γγ Example

- Use the H→γγ-inspired model from before
- Generate 10k events of pseudo-data with **s=200, m_H=125 GeV**
- Evaluate **$\hat{s}$, $\hat{m}_H$** from the pseudo-data



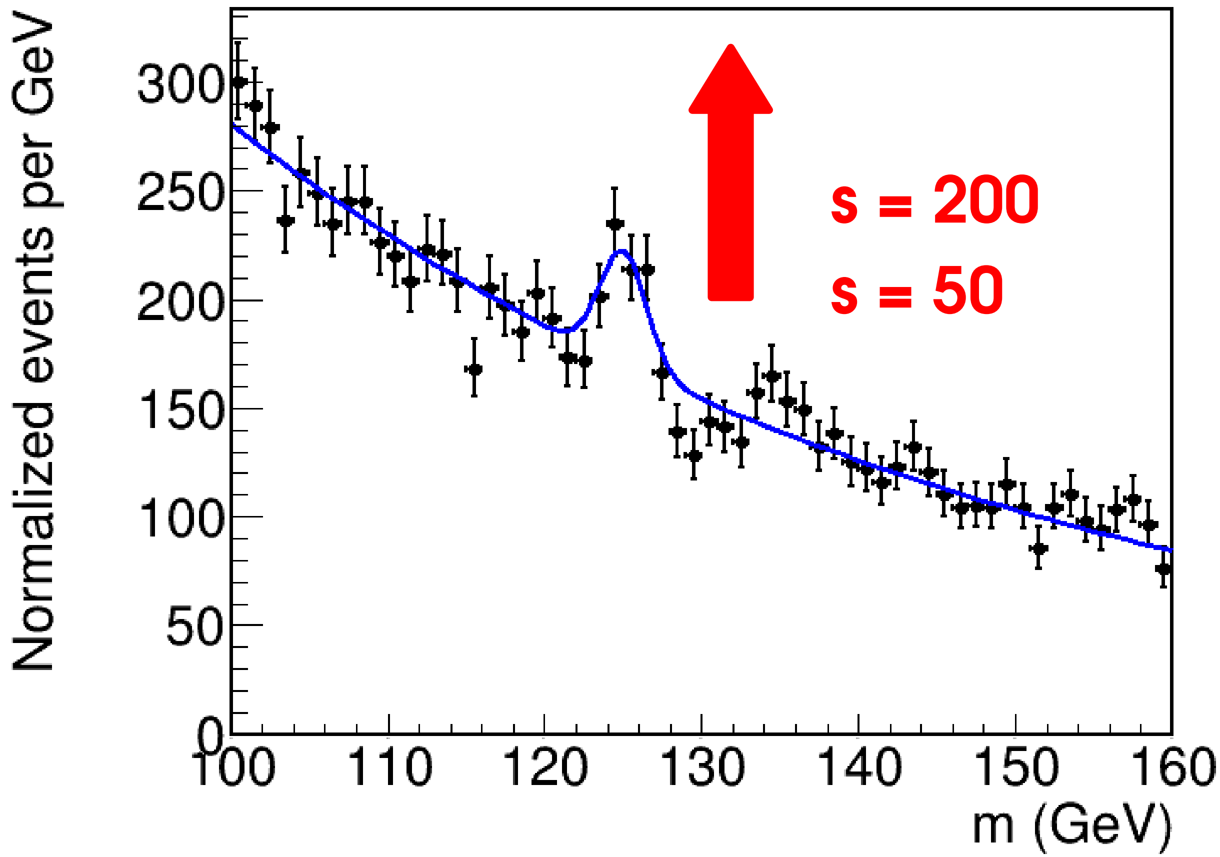
H→γγ Example

- Use the H→γγ-inspired model from before
- Generate 10k events of pseudo-data with **s=200, m_H=125 GeV**
- Evaluate **$\hat{s}$, $\hat{m}_H$** from the pseudo-data



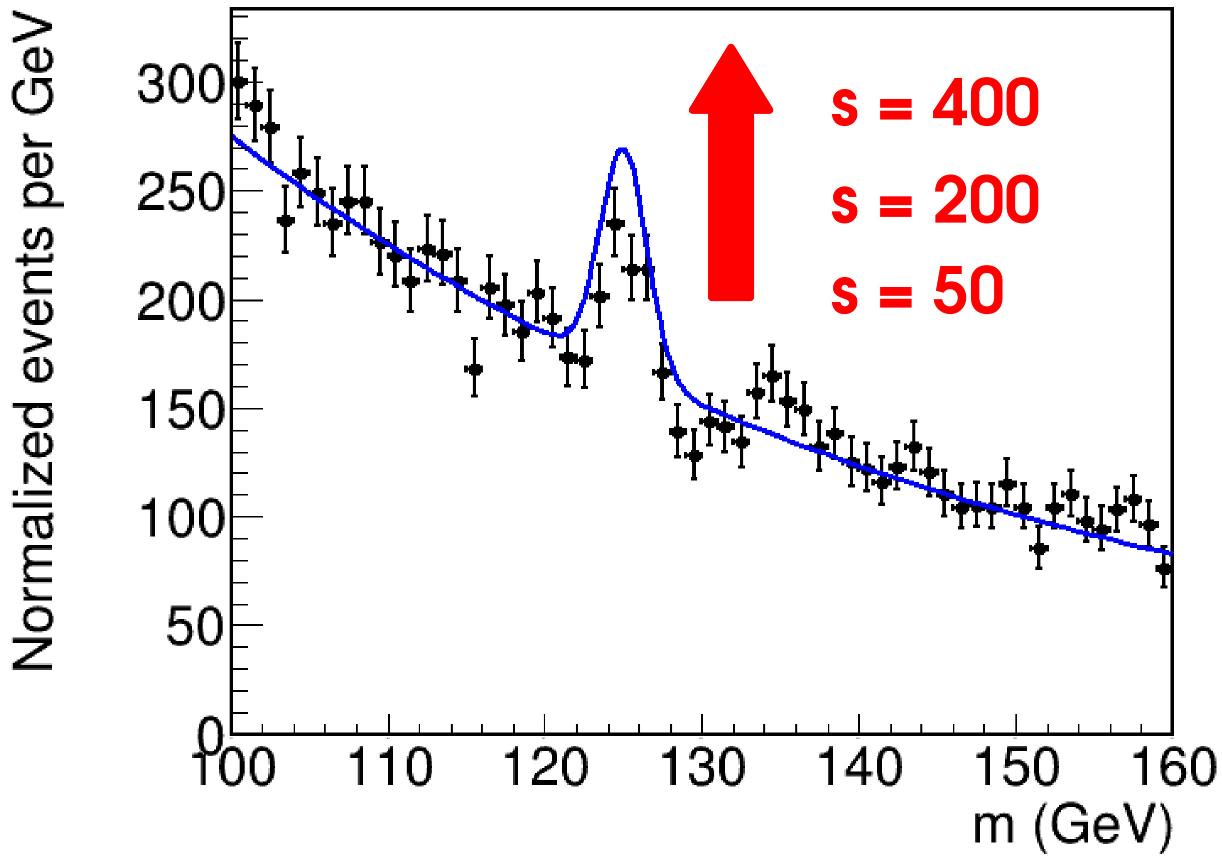
H→γγ Example

- Use the H→γγ-inspired model from before
- Generate 10k events of pseudo-data with **s=200, m_H=125 GeV**
- Evaluate **$\hat{s}$, $\hat{m}_H$** from the pseudo-data



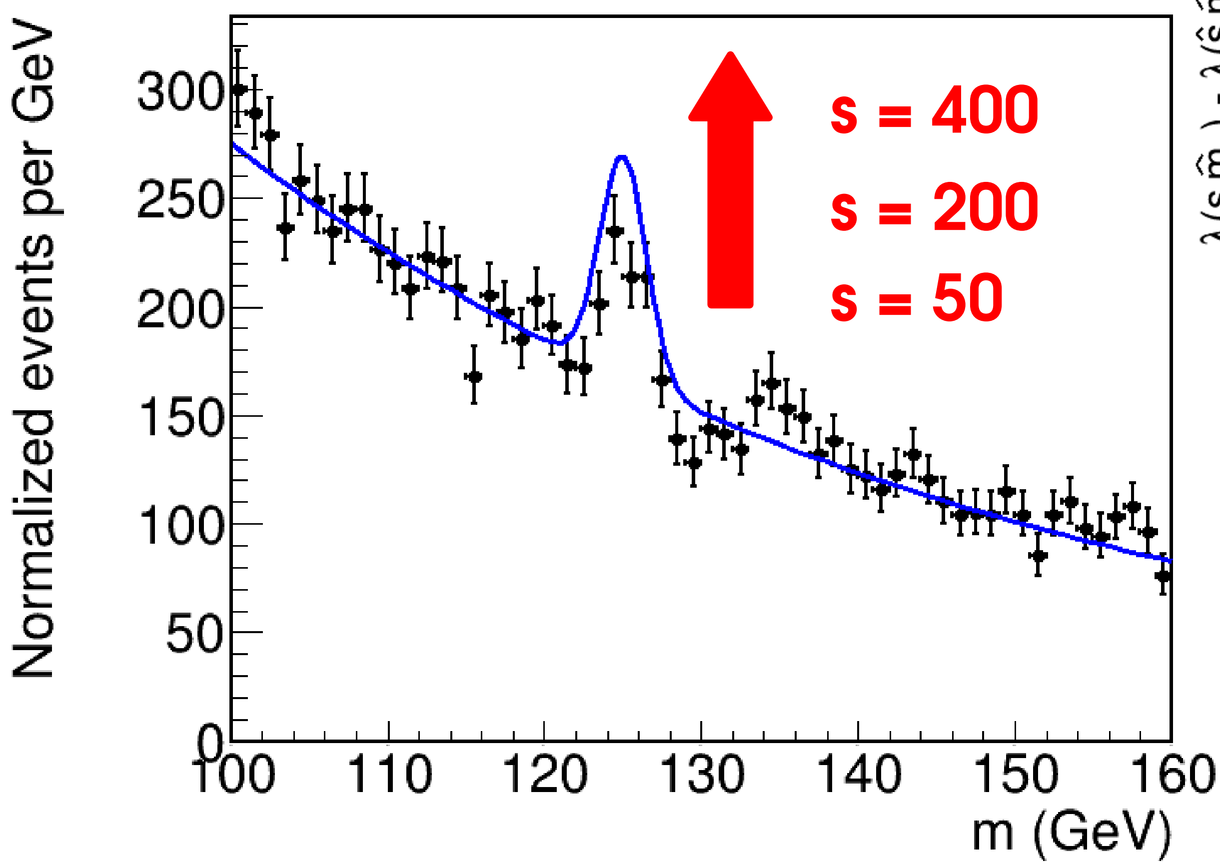
$H \to \gamma\gamma$ Example

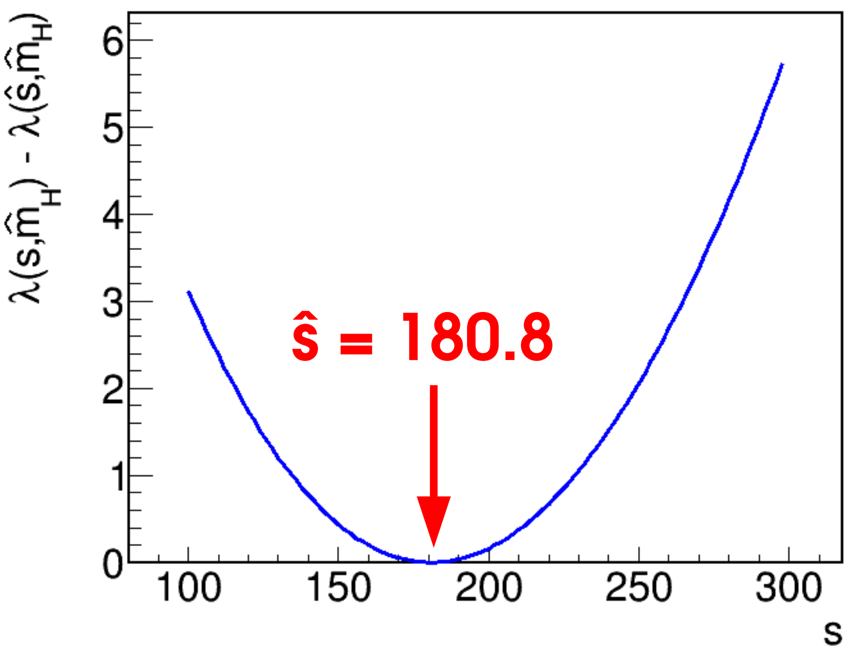
- Use the $H \to \gamma\gamma$-inspired model from before
- Generate 10k events of pseudo-data with **s=200, m_H=125 GeV**
- Evaluate **$\hat{s}$, $\hat{m}_H$** from the pseudo-data



$H \to \gamma\gamma$ Example

- Use the $H \to \gamma\gamma$-inspired model from before
- Generate 10k events of pseudo-data with **s=200, m_H=125 GeV**
- Evaluate **$\hat{s}$, $\hat{m}_H$** from the pseudo-data

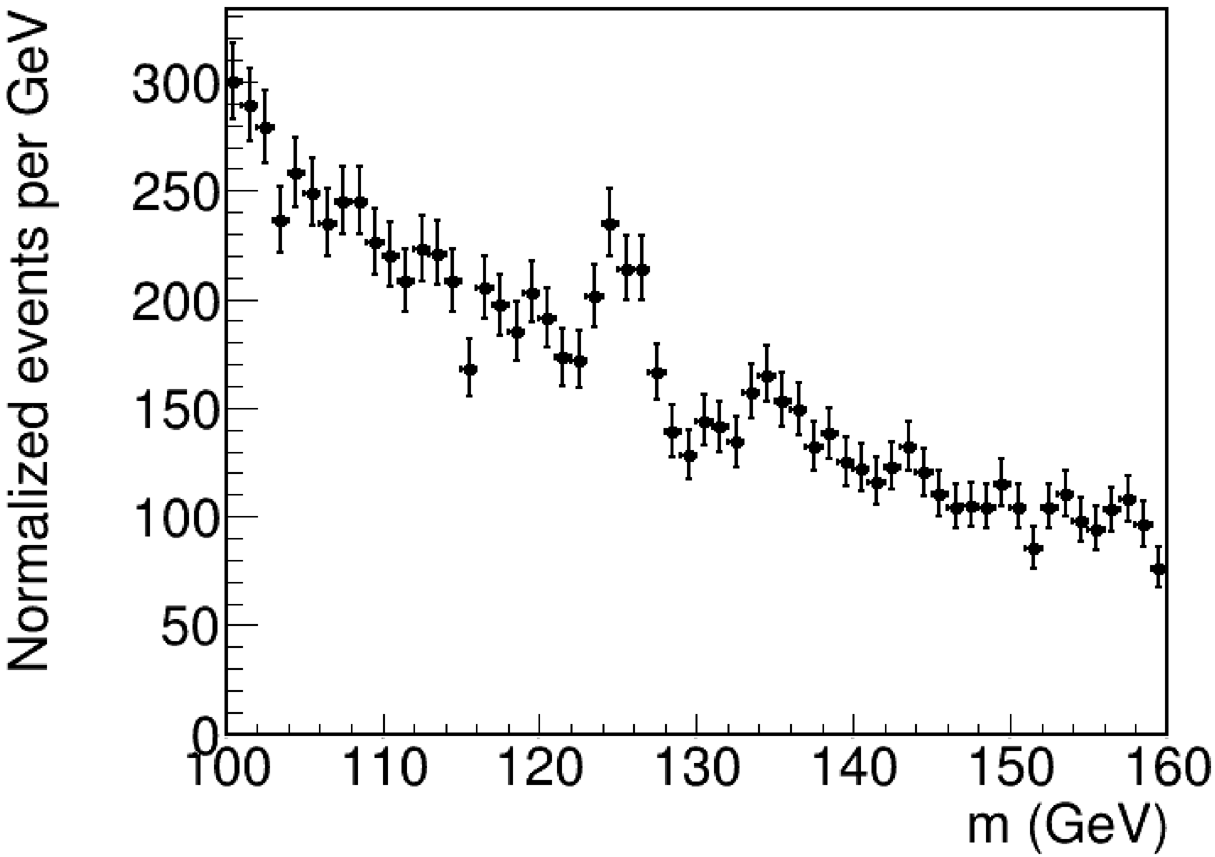




- λ is parabolic ($\Rightarrow$Gaussian)
- ML estimate $\hat{s} \neq$ true value, but close

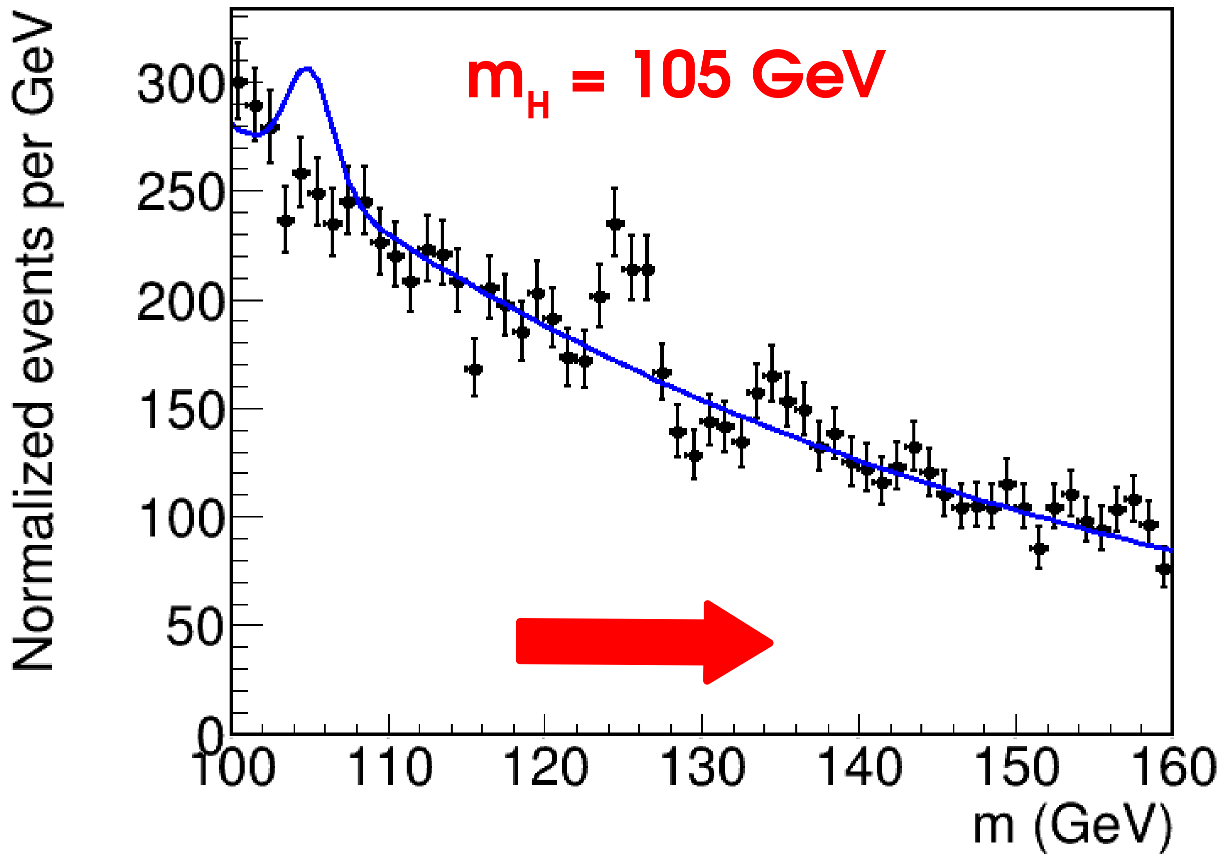
H→γγ Example

- Use the H→γγ-inspired model from before
- Generate pseudo-data with **s=200, m_H=125 GeV**
- Evaluate $\hat{s}$, $\hat{m}_H$ from the pseudo-data



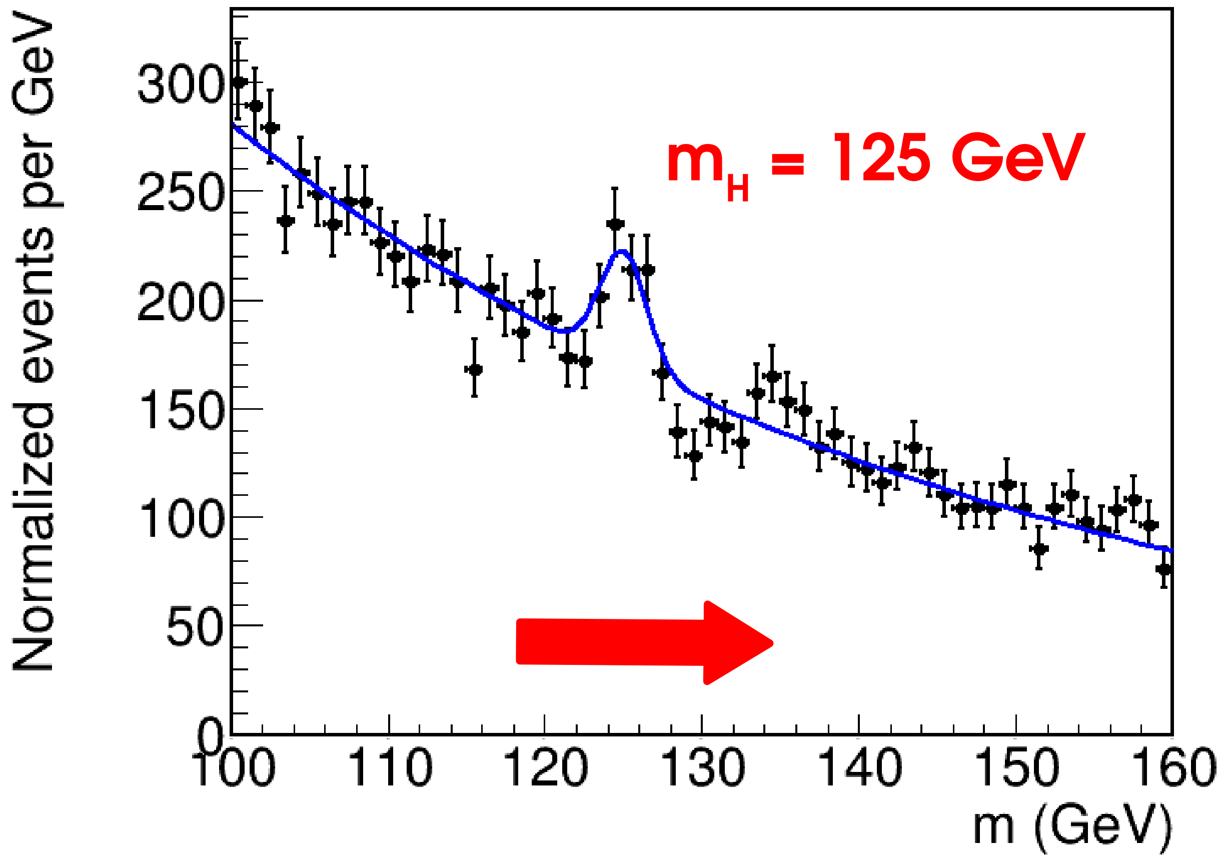
$H\to\gamma\gamma$ Example

- Use the $H\to\gamma\gamma$-inspired model from before
- Generate pseudo-data with **s=200, m_H=125 GeV**
- Evaluate **$\hat{s}$, $\hat{m}_H$** from the pseudo-data



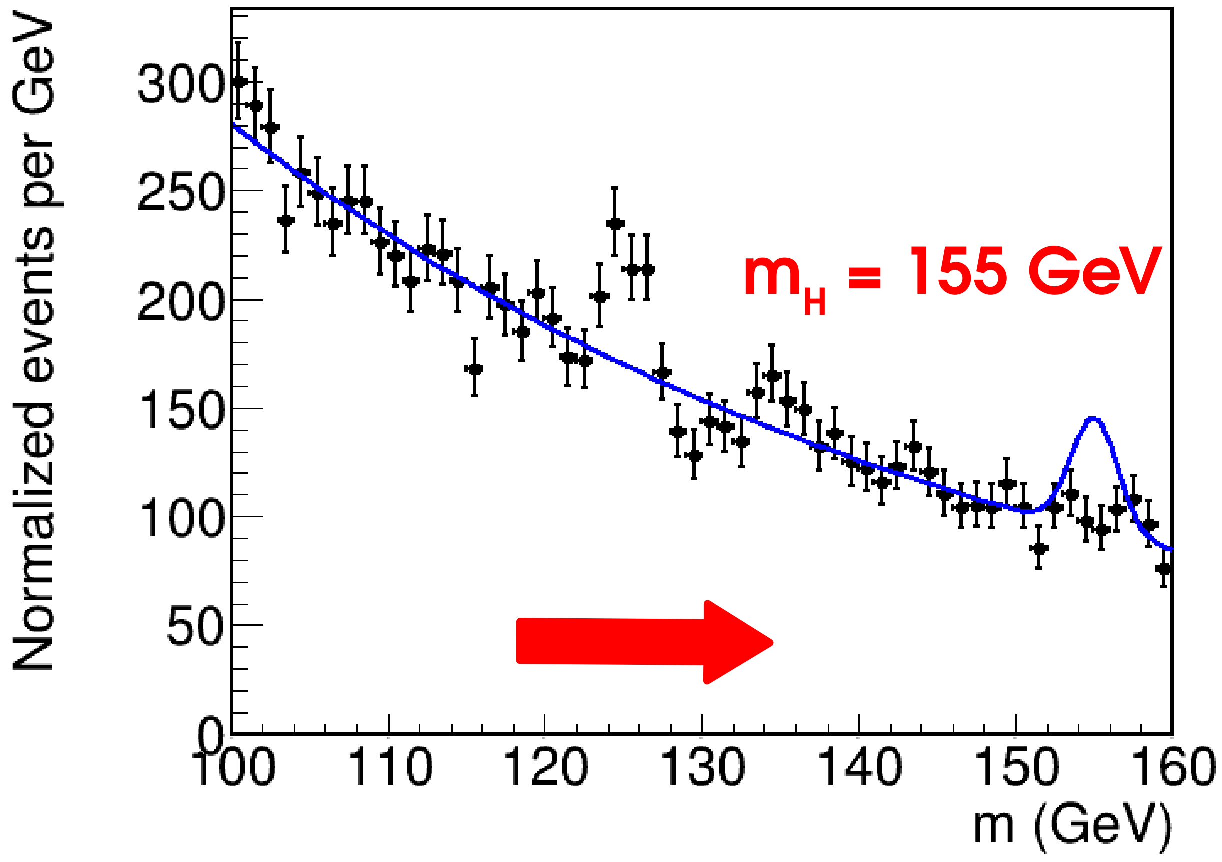
H→γγ Example

- Use the H→γγ-inspired model from before
- Generate pseudo-data with **s=200, m_H=125 GeV**
- Evaluate **$\hat{s}$, $\hat{m}_H$** from the pseudo-data



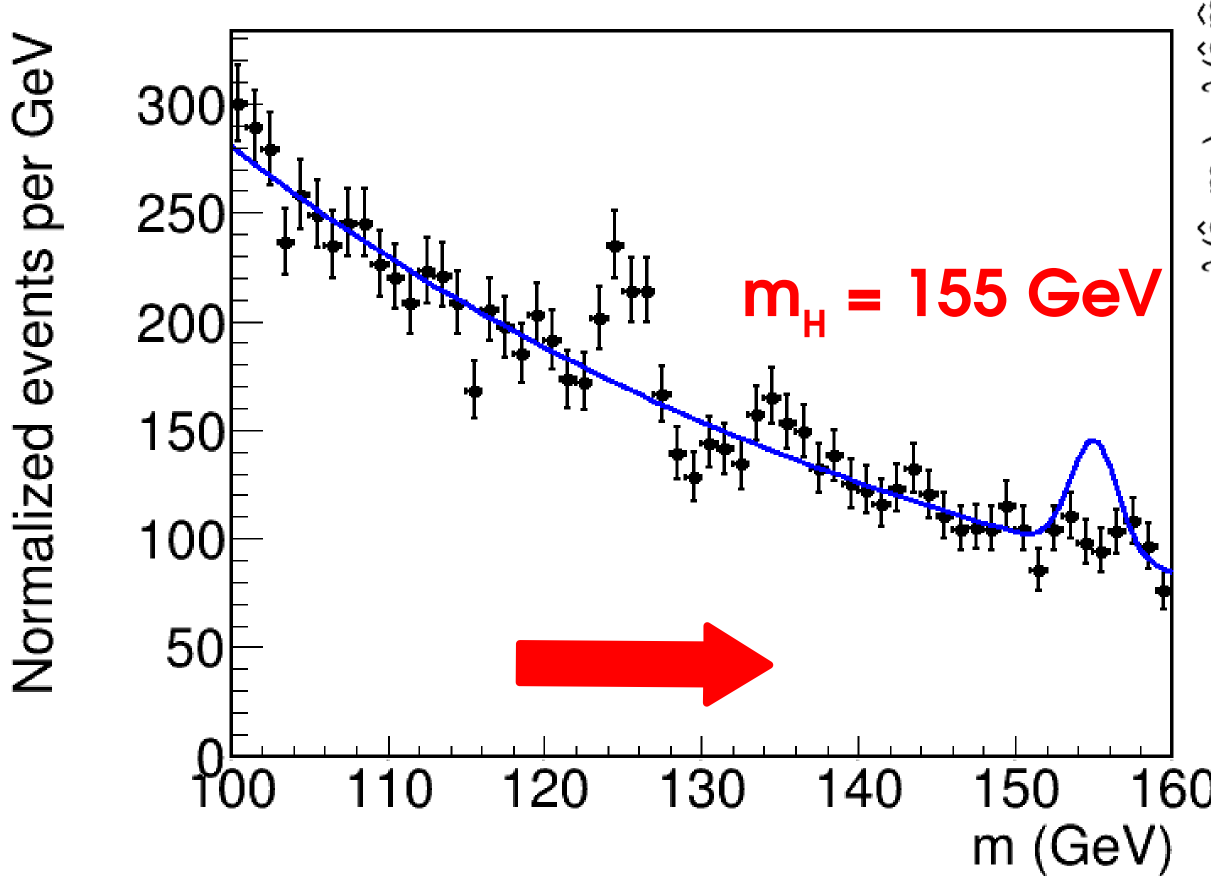
H→γγ Example

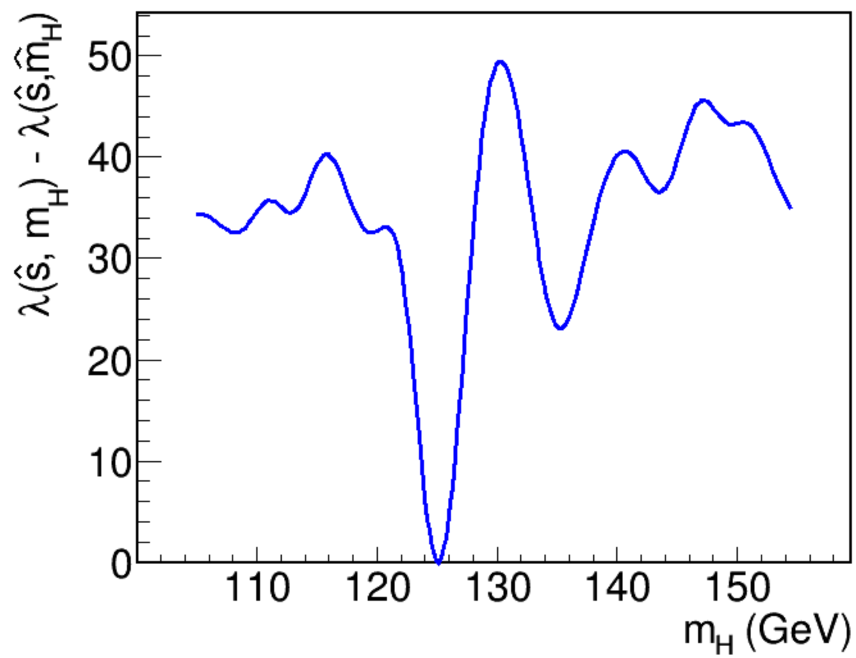
- Use the H→γγ-inspired model from before
- Generate pseudo-data with **s=200, m_H=125 GeV**
- Evaluate **$\hat{s}$, $\hat{m}_H$** from the pseudo-data



$H\to\gamma\gamma$ Example

- Use the $H\to\gamma\gamma$-inspired model from before
- Generate pseudo-data with **s=200, m_H=125 GeV**
- Evaluate **$\hat{s}$, $\hat{m}_H$** from the pseudo-data

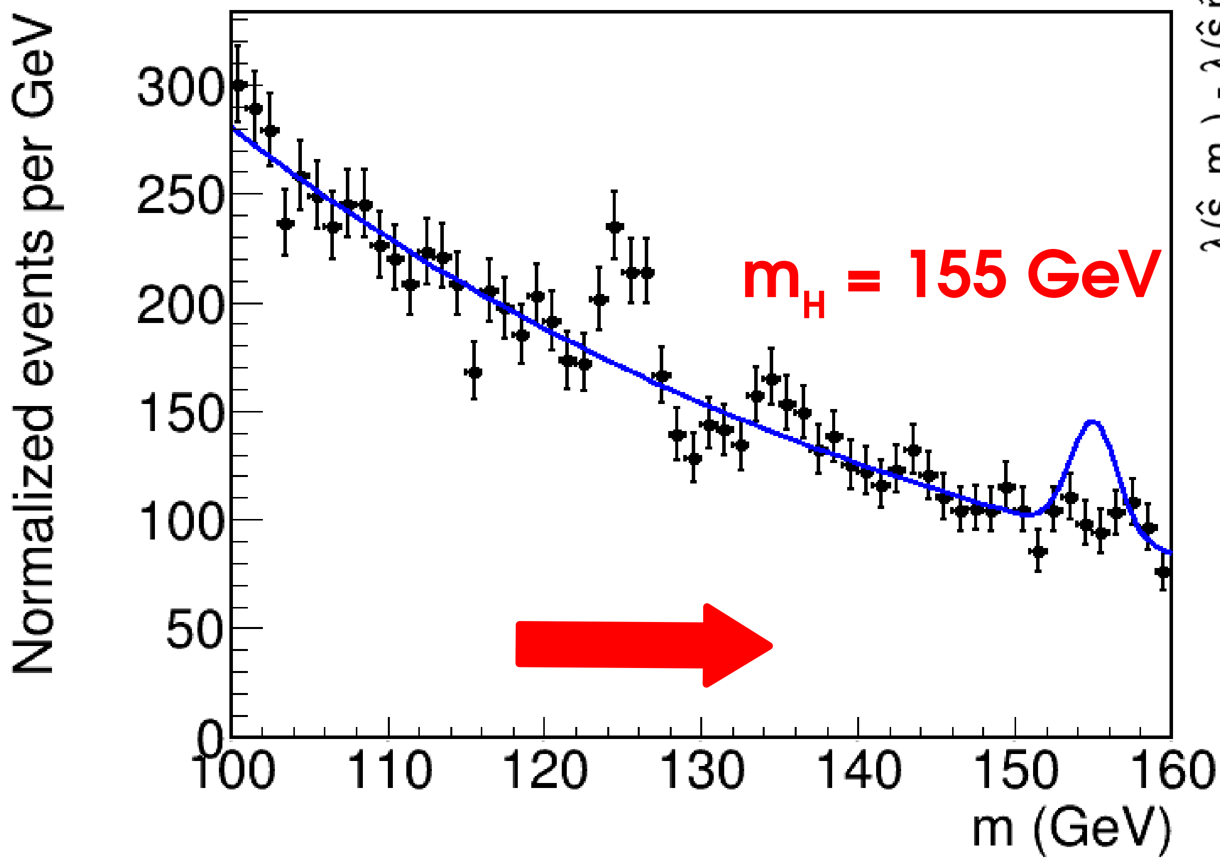


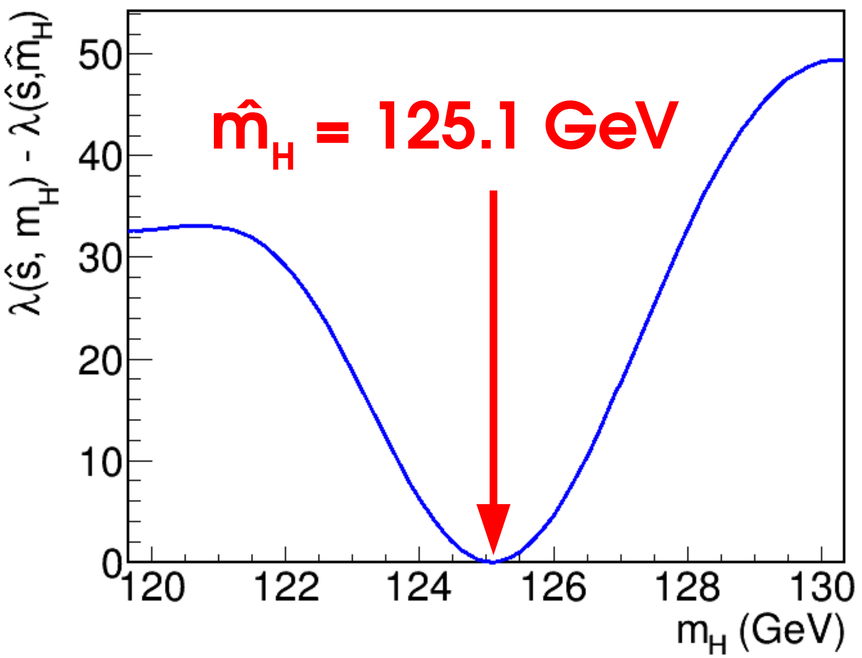


- λ parabolic only for m_H~125 GeV

H→γγ Example

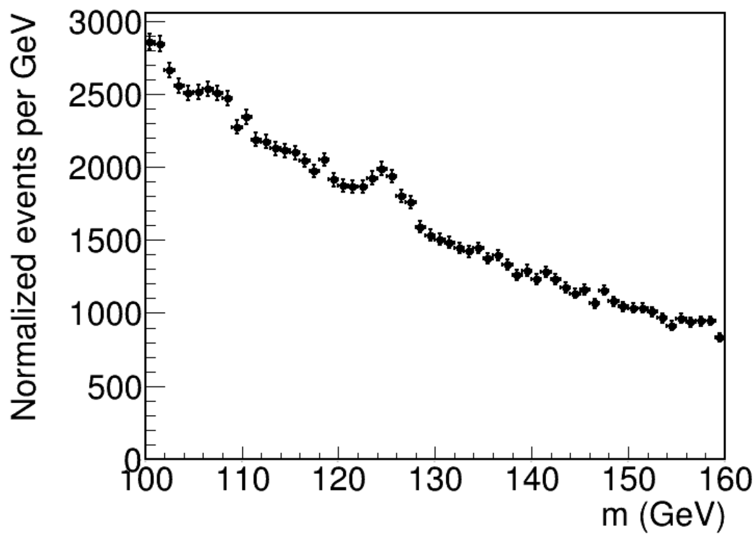
- Use the H→γγ-inspired model from before
- Generate pseudo-data with **s=200, m_H=125 GeV**
- Evaluate **$\hat{s}$, $\hat{m}_H$** from the pseudo-data





- λ parabolic only for m_H~125 GeV

Continuous case, binned



- Previous slide: consider individual **events.**
- Another option:
  - Define a binning
  - **Consider each bin as a counting experiment**

$$L\left(s_1 \ldots s_{n_{bins}}, b_1 \ldots b_{n_{bins}}; n_1 \ldots n_{n_{bins}}\right) = \prod_{i=1}^{n_{bins}} e^{-(s_i+b_i)} \frac{(s_i+b_i)^{n_i}}{n_i!}$$

  - **s_i , b_i = expected signal and bkg yields in bin i.**
- **For fine enough binning, equivalent to unbinned case**
- ⊖ depends on binning, can influence the result if not fine enough
- ⊕ Binned computations can be much faster for large numbers of events (H→γγ : 100k events, but ~1000 bins enough)

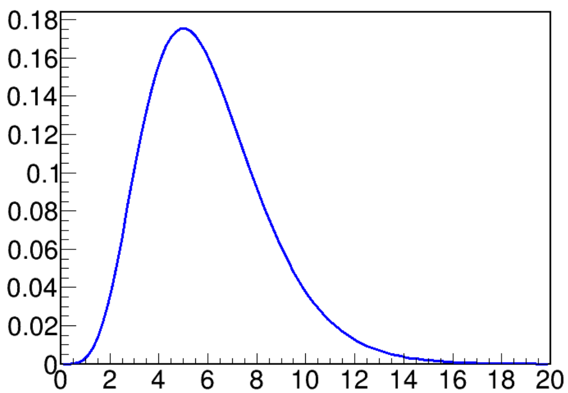
Summary of Likelihood Definitions

• Method	• Observable	• Likelihood
• Counting	• **n** : measured number of events	• Poisson $L(s,b;n_i)=e^{-(s+b)}\frac{(s+b)^{n_{obs}}}{n_{obs}!}$ • b : expected background
• Binned shape analysis	• n_i, $i=1..n_{bins}$: measured events in each bin.	• Poisson product $L(s_i,b_i;n_i)=\prod_{i=1}^{n_{bins}} e^{-(s_i+b_i)}\frac{(s_i+b_i)^{n_{obs}}}{n_{obs}!}$ f_i : fraction of signal in each bin b_i : expected background in each bin
• Unbinned shape analysis	• m_i, $i=1..n_{events}$: observable value for each event	• Extended Likelihood $L(s,b;m_i)=e^{-(s+b)}\prod_{i=1}^{n_{obs}} s\,P_{signal}(m)+b\,P_{bkg}(m)$ • P_S, P_B : PDFs for x in signal and background

What we have learned so far (3)

Estimating a parameter value

- **Build a likelihood for the measurement**
  - see previous page
  - Usually the hard part of the problem!



- **Compute the likelihood of the data** L_{data}, or $\lambda = -2 \log L_{data}$

- **Adjust the parameter of the likelihood to maximize $L_{data}(\theta)$**

$\Rightarrow$ Maximum is reached for $\hat{\theta}$.

Confidence Intervals

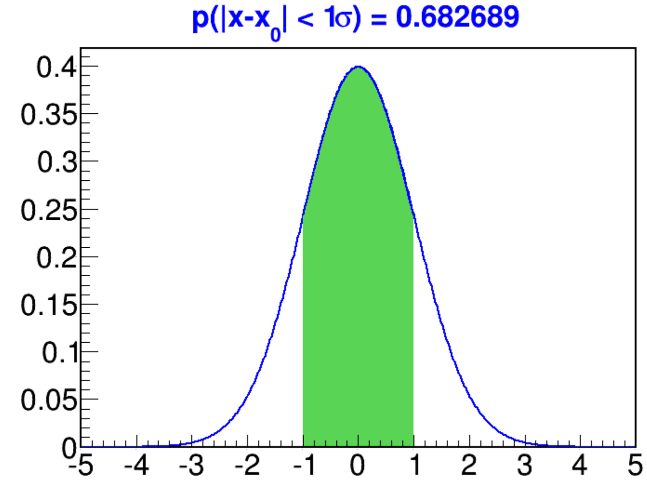
Definition

- What we want : $\theta^* = \theta_0$
- OK, so what about : $\theta^* = \theta_0 \pm \sigma$, i.e. $\theta_0 - \sigma < \theta^* < \theta_0 + \sigma$
  - Large fluctuations can happen, although unlikely

- But we **can** have $P(\theta_0 - \sigma < \theta^* < \theta_0 + \sigma) \geq 1\text{-}\alpha$ for a **small** α.
  Confidence Interval: a region where θ is **very likely** to be

- Usually, use "1σ uncertainties", i.e. $1\text{-}\alpha = 68\%$

N_{sigmas}	$1\text{-}\alpha$	α
1	**0.68**	**0.32**
1.645	**0.90**	**0.10**
1.96	**0.95**	**0.05**
2	**0.955**	**0.045**



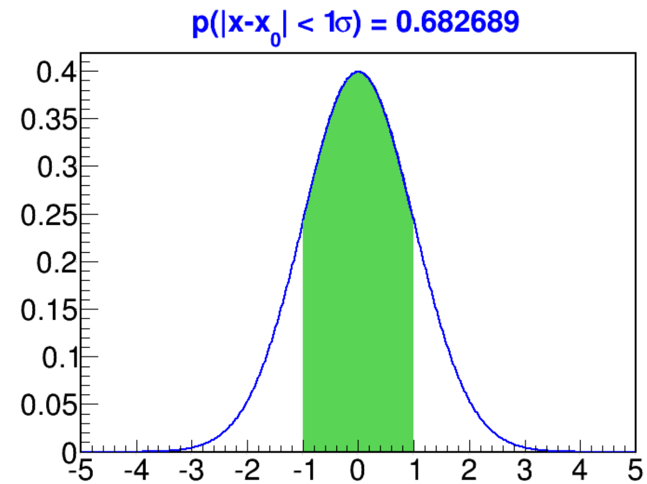
Definition

- What we want : ~~$\theta^* = \theta_0$~~ **Impossible (usually)**
- OK, so what about : $\theta^* = \theta_0 \pm \sigma$, i.e. $\theta_0 - \sigma < \theta^* < \theta_0 + \sigma$
  – Large fluctuations can happen, although unlikely

- But we **can** have $P(\theta_0 - \sigma < \theta^* < \theta_0 + \sigma) \geq 1-\alpha$ for a **small** α.
  Confidence Interval: a region where θ is **very likely** to be

- Usually, use "1σ uncertainties", i.e. $1-\alpha = 68\%$

N_{sigmas}	$1-\alpha$	α
1	**0.68**	**0.32**
1.645	**0.90**	**0.10**
1.96	**0.95**	**0.05**
2	**0.955**	**0.045**



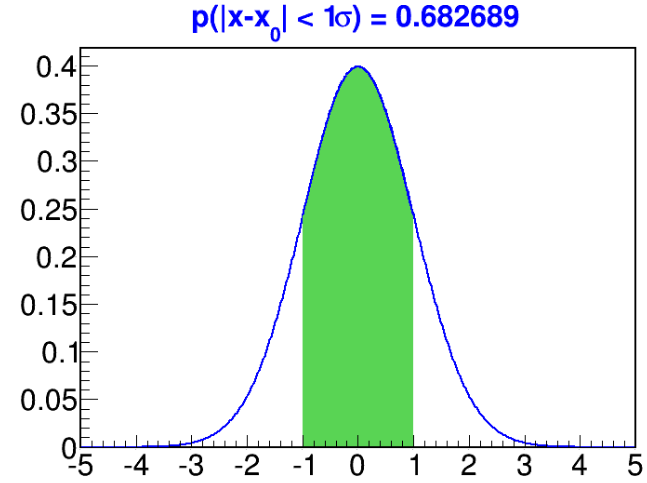
Definition

- What we want : ~~$\theta^* = \theta_0$~~ **Impossible (usually)**
- OK, so what about : ~~$\theta^* = \theta_0 \pm \sigma$, i.e. $\theta_0 - \sigma < \theta^* < \theta_0 + \sigma$~~
  - Large fluctuations can happen, although unlikely

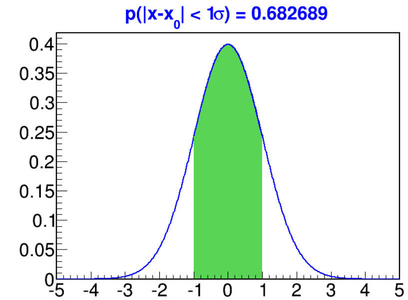
- But we **can** have $\mathbf{P(\theta_0 - \sigma < \theta^* < \theta_0 + \sigma) \geq 1-\alpha}$ for a **small** α.
  Confidence Interval: a region where θ is **very likely** to be

- Usually, use "1σ uncertainties", i.e. $1-\alpha = 68\%$

N_{sigmas}	$1-\alpha$	α
1	**0.68**	**0.32**
1.645	**0.90**	**0.10**
1.96	**0.95**	**0.05**
2	**0.955**	**0.045**



Gaussian case



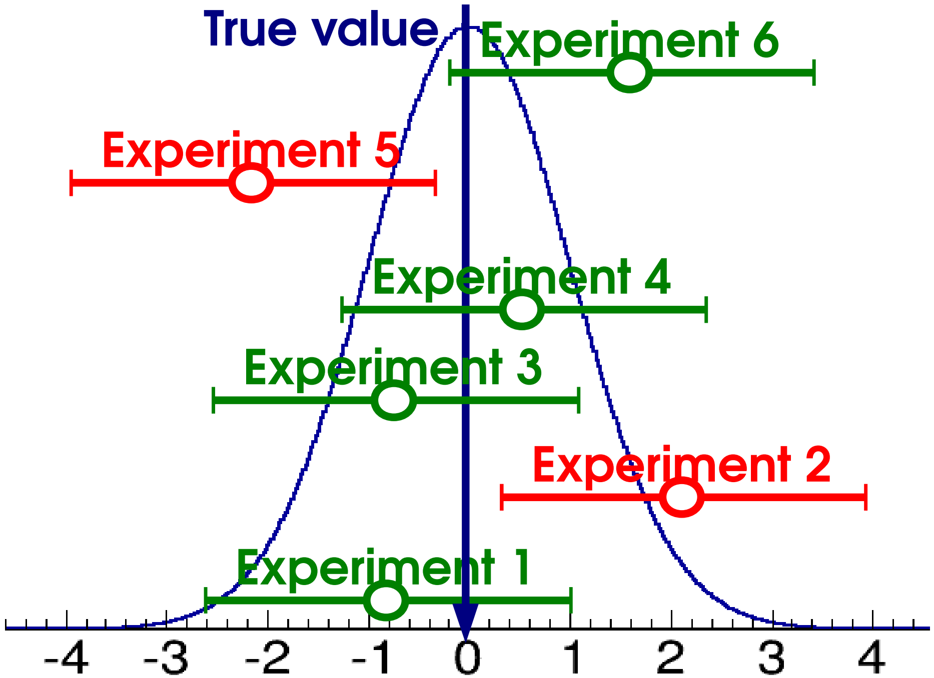
- If $\hat{\theta}$ is Gaussian, known quantiles :

  $P(\theta^* - \sigma < \hat{\theta} < \theta^* + \sigma) = 68\%$
- This is a probability **for $\hat{\theta}$, not θ^***
- But we can invert the relation:

  $P(\theta^* - \sigma < \hat{\theta} < \theta^* - \sigma) = 68\%$

  $P(|\hat{\theta} - \theta^*| < \sigma) = 68\%$

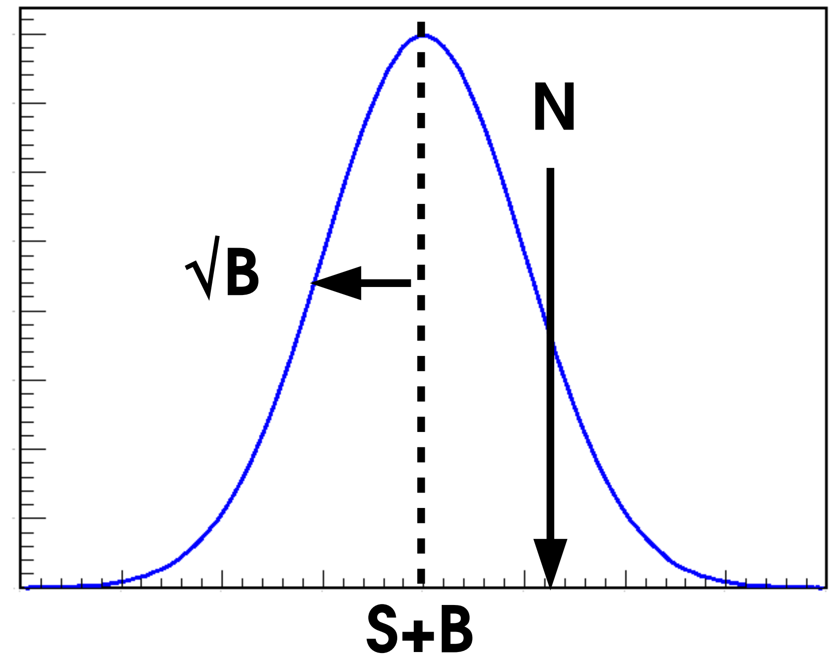
  $P(\hat{\theta} - \sigma < \theta^* < \hat{\theta} + \sigma) = 68\%$



- This gives the statement on θ^* we wanted: "if we repeat the experiment many times, **$[\hat{\theta} - \sigma, \hat{\theta} + \sigma]$** will contain the true value 68% of the time"
- **θ_0 is fixed** -- actually a statement on the interval **$[\hat{\theta} - \sigma, \hat{\theta} + \sigma]$**, obtained for each experiment
- Can adjust the probability : 95% → **$[\hat{\theta} - 1.96\sigma, \hat{\theta} + 196\sigma]$** etc.

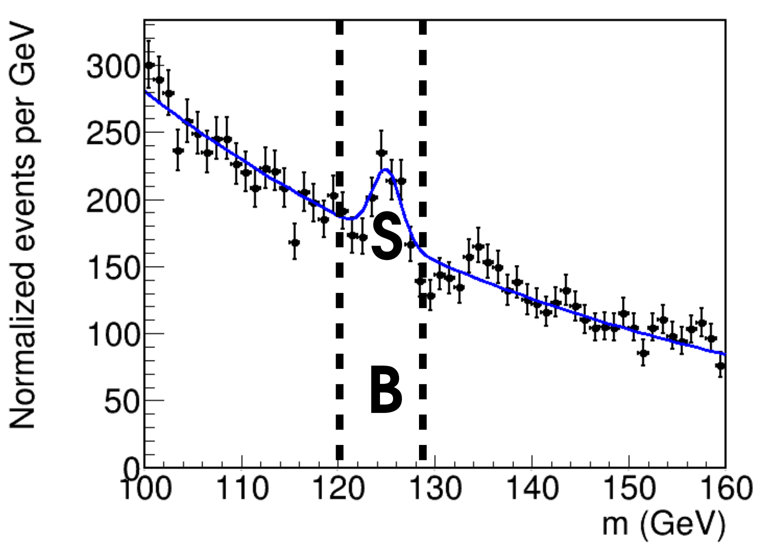
Trivial Application: Gaussian counting

- Suppose a counting experiment measuring N=S+B, with
  - B is known
  - B ≫ 1 so N is ~ Gaussian
  - B ≫ S so $\sigma = \sqrt{(S+B)} \sim \sqrt{B}$



- Then **L(S, B; N) = G(N; S+B, √B)**

- **Results**:
  - Best fit signal : $\hat{S}$ = N-B
  - 68% confidence interval : **[$\hat{S}-\sqrt{B}$, $\hat{S}+\sqrt{B}$]**
  - Finally : **S = (N-B) ± √B**

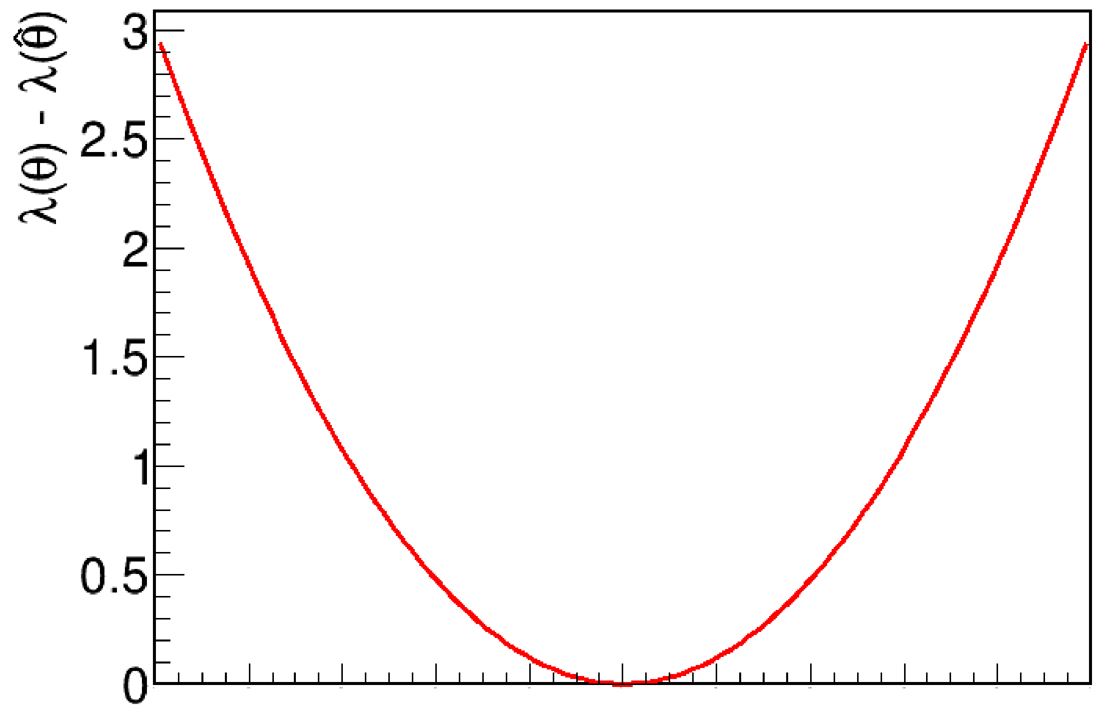


General Case: Likelihood intervals

- Gaussian case:
  $\lambda(\theta) - \lambda(\hat{\theta}) = (\theta - \hat{\theta})^2/\sigma^2$
- 68% interval : **$[\hat{\theta}-\sigma, \hat{\theta}+\sigma]$**
- So at the interval endpoints
  $\lambda(\hat{\theta} \pm \sigma) - \lambda(\hat{\theta}) = 1$

$\Rightarrow$ Find the endpoints by solving:

$$\lambda(\theta) - \lambda(\hat{\theta}) = 1$$



- Also good approximation for non-Gaussian case
- Very easy to apply
- Other interval sizes:

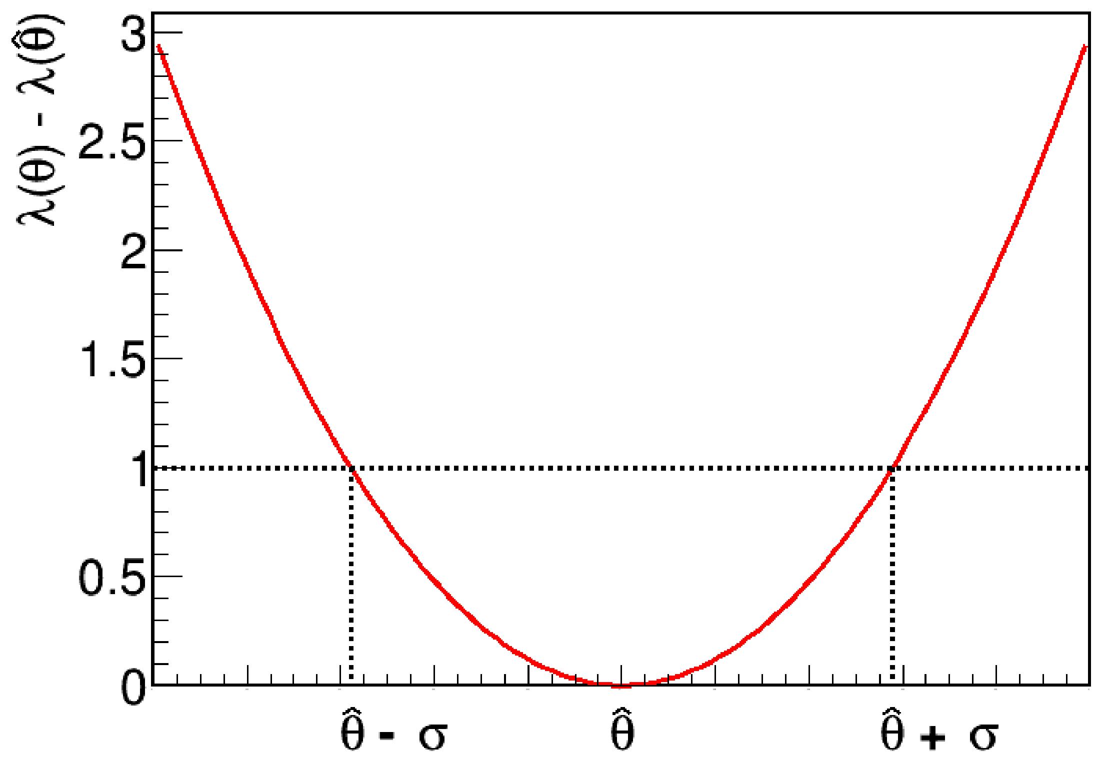
N_{sigmas}		$1-\alpha$
1	**$\lambda(\theta) < 1$**	**0.68**
1.645	**$\lambda(\theta) < 2.71$**	**0.90**
1.96	**$\lambda(\theta) < 3.84$**	**0.95**

General Case: Likelihood intervals

- Gaussian case:
  $\lambda(\theta) - \lambda(\hat{\theta}) = (\theta - \hat{\theta})^2/\sigma^2$
- 68% interval : $[\hat{\theta}-\sigma, \hat{\theta}+\sigma]$
- So at the interval endpoints
  $\lambda(\hat{\theta} \pm \sigma) - \lambda(\hat{\theta}) = 1$

$\Rightarrow$ Find the endpoints by solving:

$$\lambda(\theta) - \lambda(\hat{\theta}) = 1$$



- Also good approximation for non-Gaussian case
- Very easy to apply
- Other interval sizes:

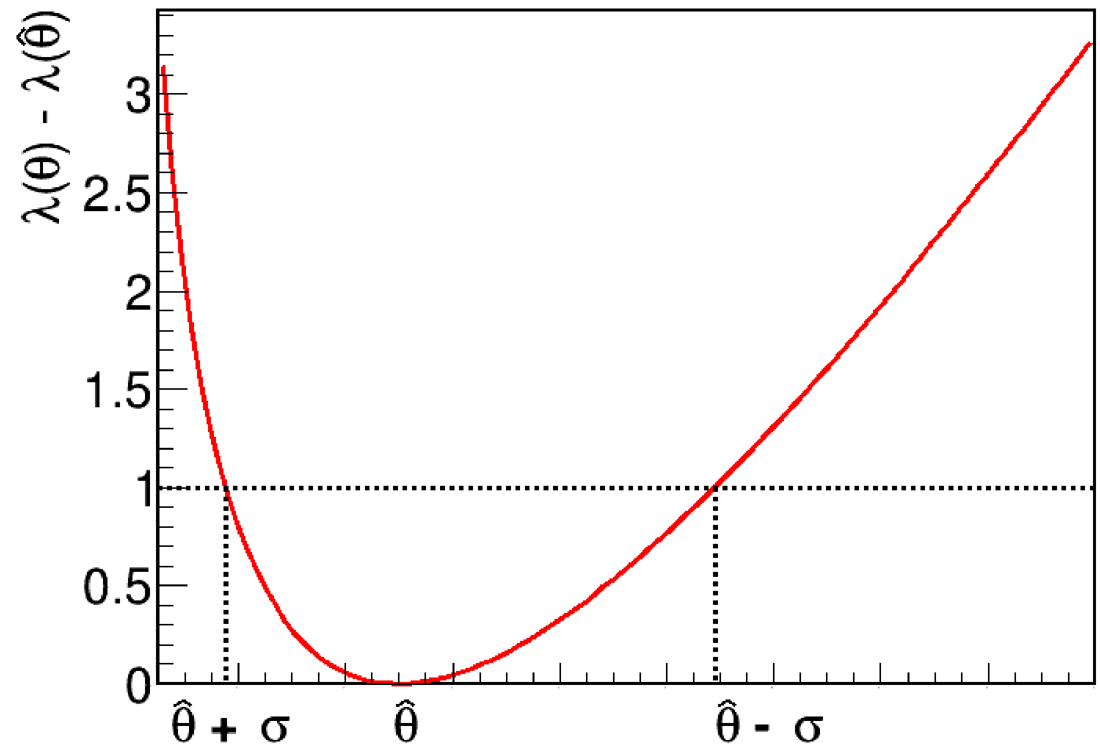
N_{sigmas}		$1-\alpha$
1	$\lambda(\theta) < 1$	0.68
1.645	$\lambda(\theta) < 2.71$	0.90
1.96	$\lambda(\theta) < 3.84$	0.95

General Case: Likelihood intervals

- Gaussian case:
  $\lambda(\theta) - \lambda(\hat{\theta}) = (\theta - \hat{\theta})^2/\sigma^2$
- 68% interval : $[\hat{\theta}-\sigma, \hat{\theta}+\sigma]$
- So at the interval endpoints
  $\lambda(\hat{\theta} \pm \sigma) - \lambda(\hat{\theta}) = 1$

$\Rightarrow$ Find the endpoints by solving:

$$\lambda(\theta) - \lambda(\hat{\theta}) = 1$$

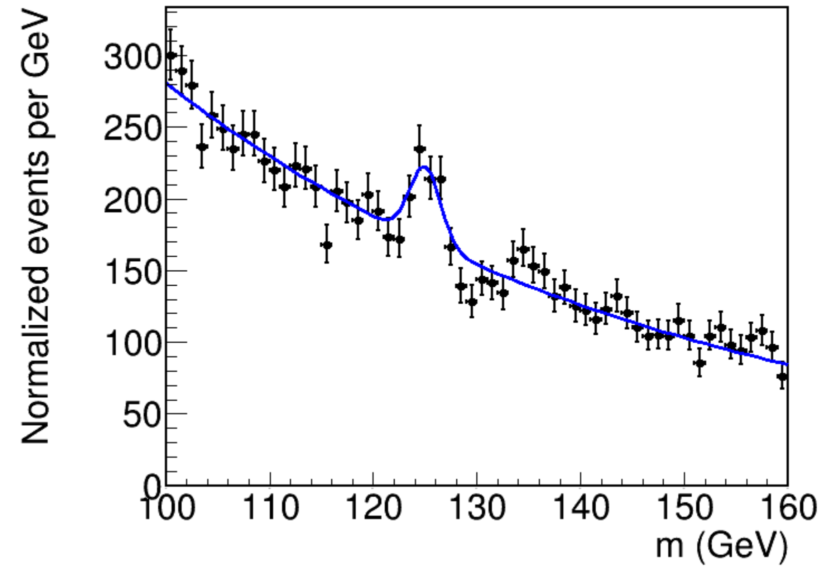


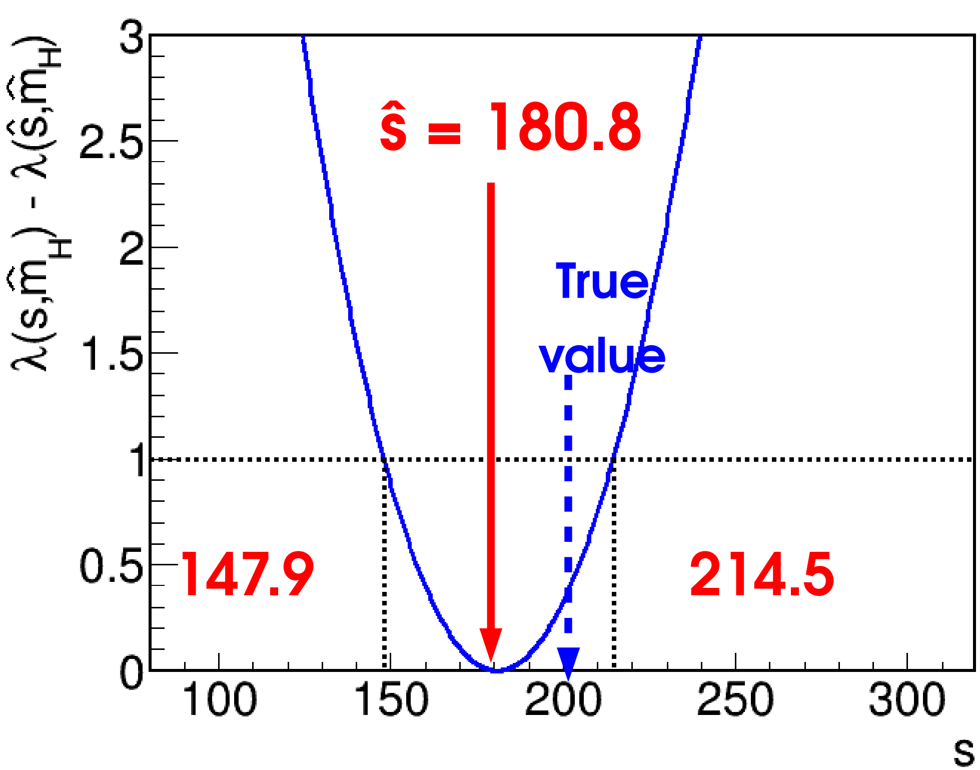
- Also good approximation for non-Gaussian case
- Very easy to apply
- Other interval sizes:

N_{sigmas}		$1-\alpha$
1	$\lambda(\theta) < 1$	**0.68**
1.645	$\lambda(\theta) < 2.71$	**0.90**
1.96	$\lambda(\theta) < 3.84$	**0.95**

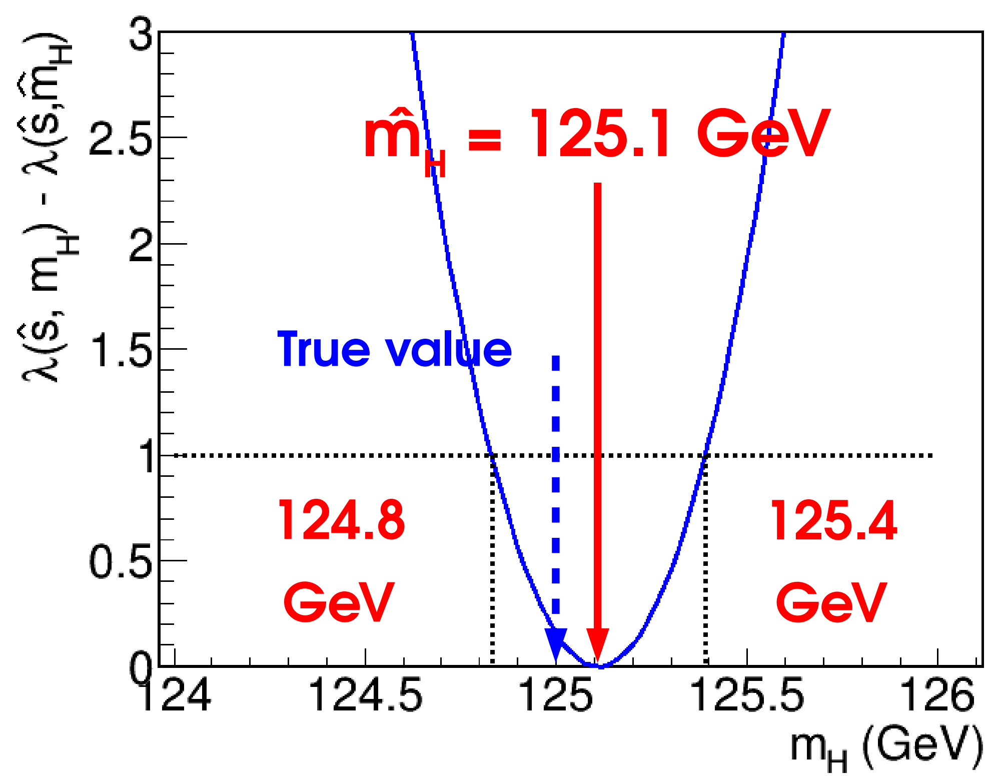
Example : Back to H→γγ

- Generate pseudo-data with **s=200, m_H=125 GeV**
- Measure **s, m_H** in the pseudo-data





$$s = 181^{+34}_{-33}$$



$$\hat{m}_H = 125.1 \pm 0.3 \text{ GeV}$$

Coverage & Toys

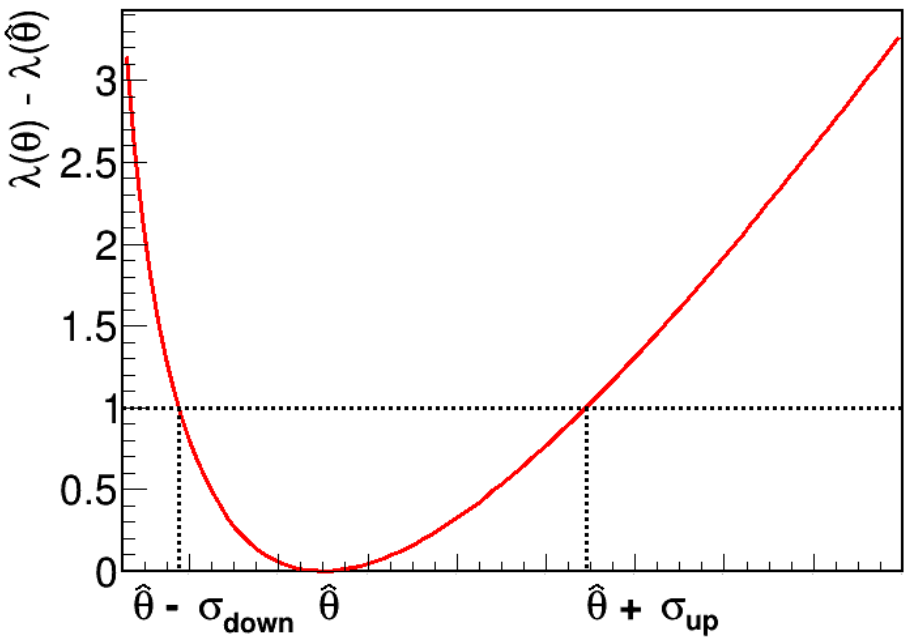
- We claim to have computed (θ_1, θ_2) so that $\mathbf{P(\theta_1 < \theta_0 < \theta_2) = 68\%}$
- We can check whether this is OK ("good coverage") :
  - Generate pseudo-data with $\theta=\theta_0$.
  - Compute the interval
  - Repeat many times, check fraction of cases when we do get $\theta_1 < \theta_0 < \theta_2$.
- Example on previous slide: run 5k toys with s=200, m_H=125 GeV
  - **134.5 < s < 228.7** : true 3530/5k = **70.6% of the time**
  - **124.7 < m_H < 125.5 GeV** : true 3414/5k = **68.3% of the time**
- Can also be used to compute (θ_1, θ_2) :
  - Choose some values, compute coverage
  - Adjust (θ_1, θ_2) until coverage is OK.

  ⊕ No approximations involved **⊖ Can be very slow**,

What we have learned so far (4)

Estimating a parameter

- Build a likelihood $\mathbf{L(\theta)}$ for the measurement



- Compute $\boldsymbol{\lambda(\theta) = -2 \log L_{data}(\theta)}$, as a function of θ.

- Find the minimum of $\lambda(\theta)$

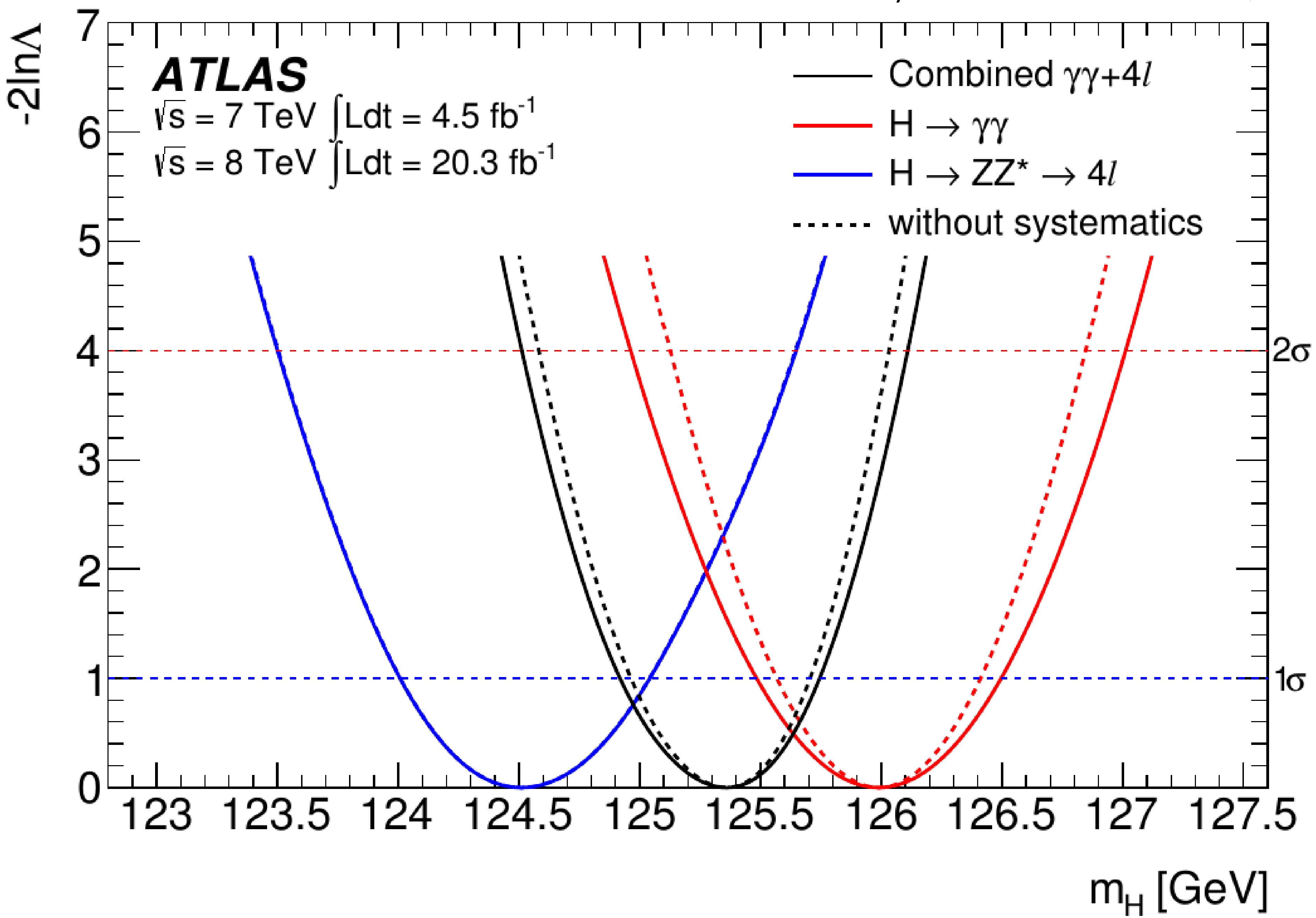
  ⇒ **Minimum is reached for $\hat{\theta}$.**

- Move the parameter up and down to get $\lambda(\hat{\theta}+\sigma_{up}) = \lambda(\hat{\theta}) + 1$ and $\lambda(\hat{\theta}-\sigma_{down}) = \lambda(\hat{\theta}) + 1$.

  Then $[\theta-\sigma_{down}, \theta+\sigma_{up}]$ is a **68% confidence interval** for θ : $\theta = \hat{\theta}^{+\sigma_{up}}_{-\sigma_{down}}$

Real-Life Case: ATLAS Higgs Mass Measurement

Phys. Rev. D. **90**, 052004 (2014)



Fisher Information

- Gaussian case: $\lambda(\theta) - \lambda(\hat{\theta}) = (\theta - \hat{\theta})^2/\sigma^2$ so $d^2\lambda/d\theta^2 = 2/\sigma^2$
- Define the **Fisher Information** as

$$I = -E\left(\frac{\partial^2}{\partial\theta^2}\log L(\theta)\right)$$

- Measure of the quantity of information in the measurement of θ
- Gaussian case, $I=1/\sigma^2$: more information => smaller uncertainty.

- In general, for any estimator $\hat{\theta}$,

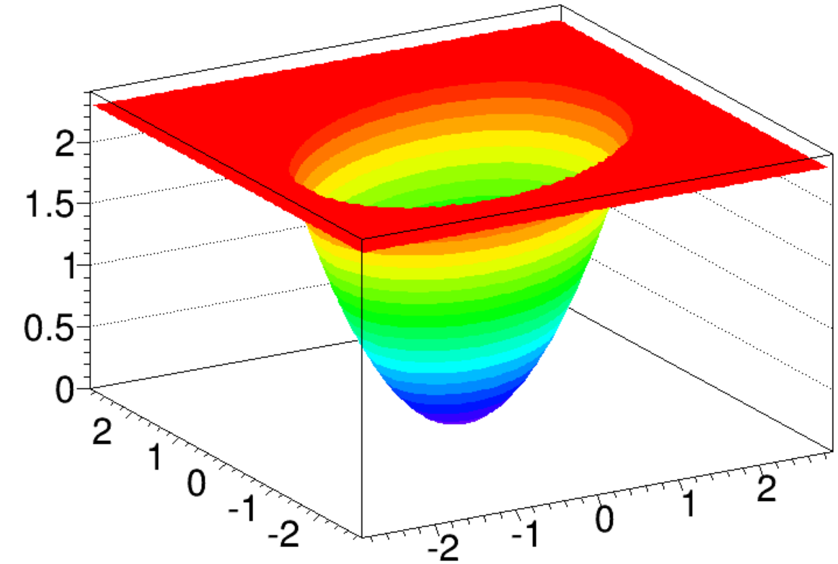
  $Var(\hat{\theta}) \geq 1/I$ (Cramer-Rao inequality)

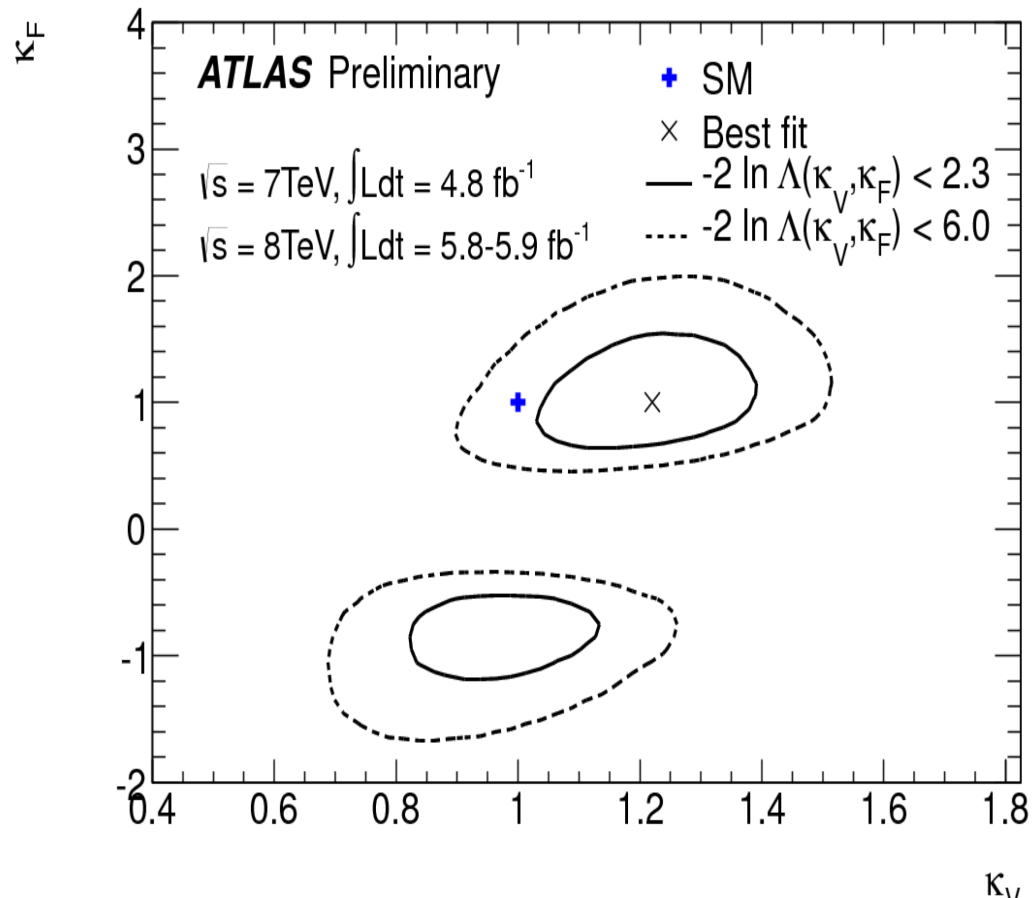
  (cannot be more precise than information allows.)
- Estimators which reach the bound are **efficient** – e.g. MLE in the large n limit.

2D Contours

- Two correlated parameters:
  - Now $\lambda(\theta_1,\theta_2)$, Gaussian Likelihood ⇒ Paraboloid
  - Find 2D maximum
  - Find 2D contour : $\lambda(\theta_1,\theta_2) = \lambda(\hat{\theta}_1, \hat{\theta}_2) + 2.30$
  - **Contour values are differen**
    ($\chi^2(n=1)$ vs. $\chi^2(n=2)$)

N_{sigmas}	For 2 degrees of freedom	1-α
1	$\lambda(\theta) < 2.30$	**0.68**
1.645	$\lambda(\theta) < 4.61$	**0.90**
1.96	$\lambda(\theta) < 6.00$	**0.95**





Relation with χ^2

- χ^2: say you measure $\hat{\theta}_1 ... \hat{\theta}_n$ with means $\theta^*_1 ... \theta^*_1$, uncertainty σ. Then

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{\hat{\theta}_i - \theta_i^*}{\sigma} \right)^2$$

- If good agreement : $\chi^2 \sim n$.
- If $\hat{\theta}_i$ are Gaussian (with same θ^*_i and σ as in the χ^2 expression), then χ^2 follows a **χ^2 distribution with n degrees of freedom**, $\boldsymbol{\chi^2_n}$
- Now go back to the likelihood picture, assume Gaussian measurements:

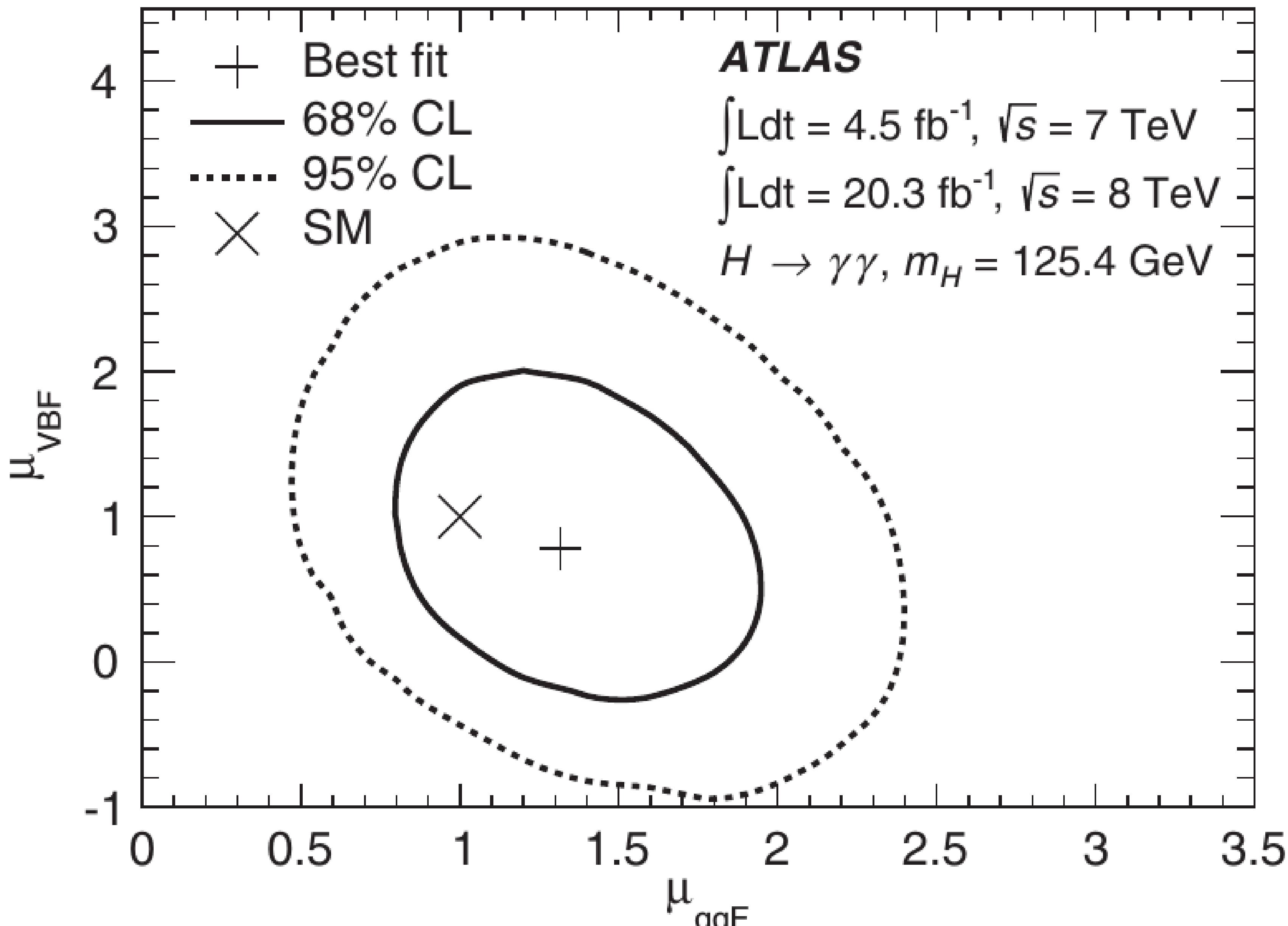
$$L = \prod_{i=1}^{n} e^{-\frac{1}{2}\left(\frac{\hat{\theta}_i - \theta_i^*}{\sigma}\right)^2} \qquad \lambda = -2 \log L = \sum_{i=1}^{n} \left(\frac{\hat{\theta}_i - \theta_0}{\sigma} \right)^2$$

- So
  - **λ is like a χ^2**
  - L is exp($-\chi^2/2$)
  - λ is $\sim \chi^2_n$. Quantiles :
    - for **n=1**, **same as Gaussian**
    - For n>1, look up the values...

N_{sigmas}	χ^2_1	χ^2_2	$1-\alpha$
1	**1**	**2.30**	**0.68**
1.645	**2.71**	**4.61**	**0.90**
1.96	**3.84**	**6.00**	**0.95**

Real-Life: (μ_{ggF}, μ_{VBF}) from $H \rightarrow \gamma\gamma$

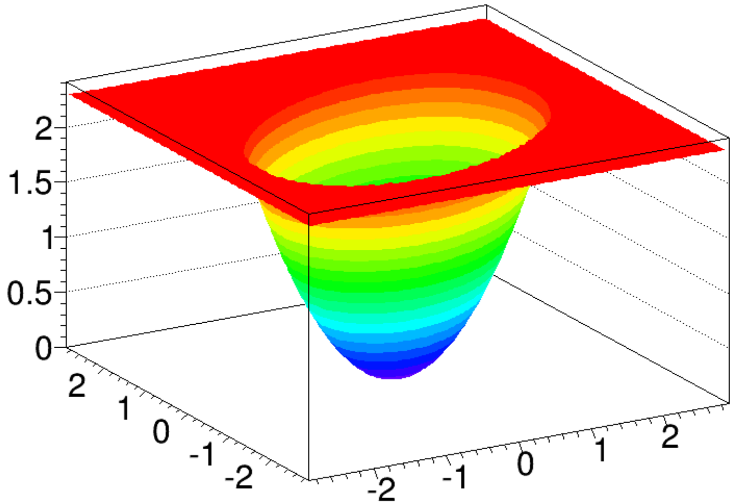
Physics Letters B 740 (2015) 222-242

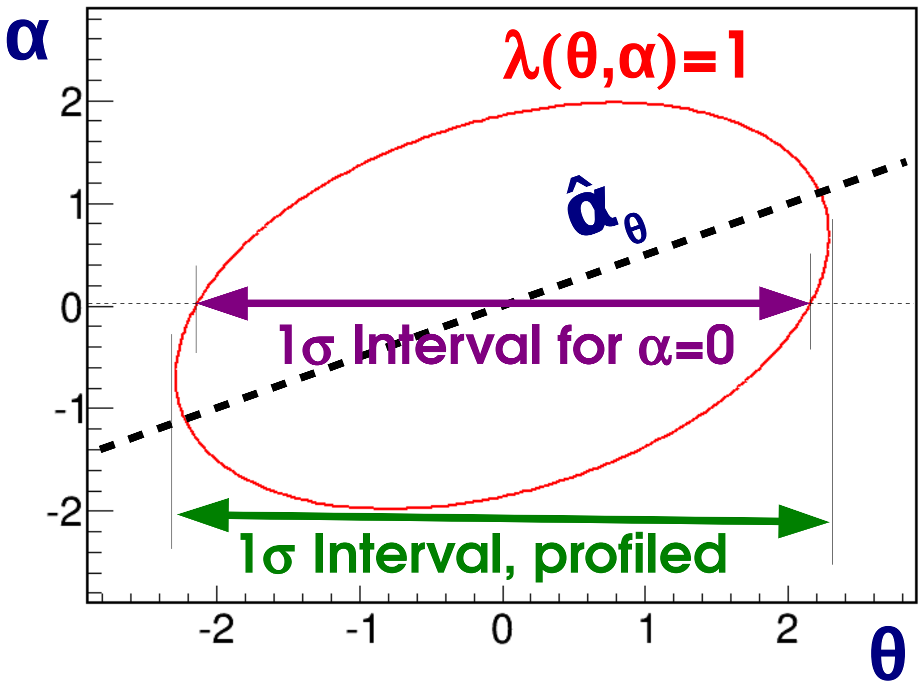


1D Contours with Multiple Parameters: Profiling

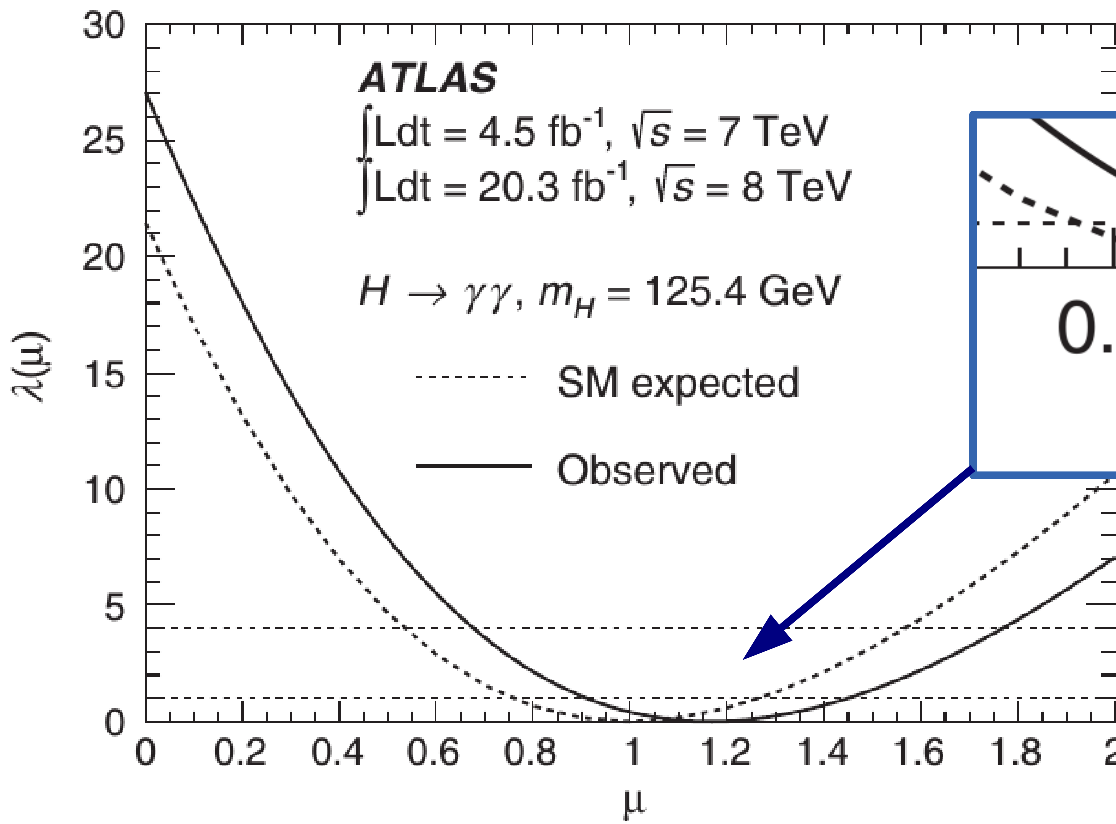
- What about 1D contours, when several parameters are present ? e.g. $\lambda(\theta,\alpha)$, , and we want an interval on θ only.
- Define the **profile likelihood** $\lambda(\theta) = \lambda(\theta,\hat{\hat{\alpha}}_\theta)$
  where $\hat{\hat{\alpha}}_\theta$ is the ML estimate of α for a **fixed value** of θ.
- Compute intervals as before with

$$\lambda(\theta) - \lambda(\hat{\theta}) = 1 \quad \text{i.e.} \quad \lambda(\theta,\hat{\hat{\alpha}}_\theta) - \lambda(\hat{\theta},\hat{\alpha}) = 1$$

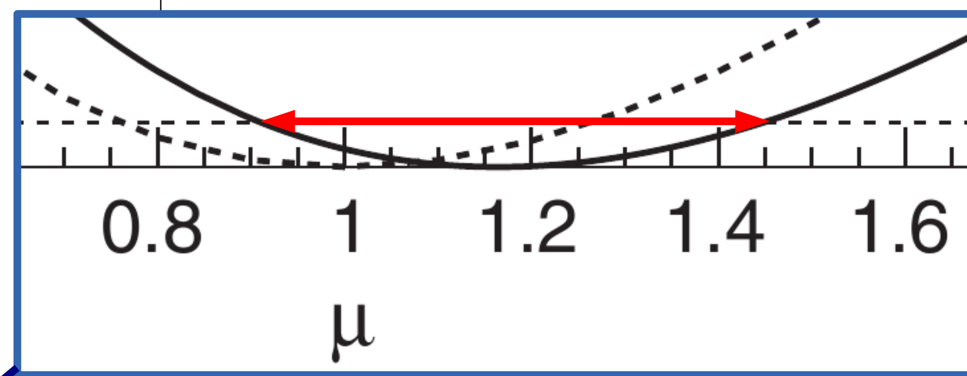




Real-Life: μ from H→γγ



Physics Letters B 740 (2015) 222-242



$\mu = 1.17 \pm 0.27$

FIG. 15. The profile of the negative log-likelihood ratio $\lambda(\mu)$ of the combined signal strength μ for $m_H = 125.4$ GeV. The observed result is shown by the solid curve, the expectation for the SM by the dashed curve. The intersections of the solid and dashed curves with the horizontal dashed line at $\lambda(\mu) = 1$ indicate the 68% confidence intervals of the observed and expected results, respectively.

Conclusion

- Seen today
  - Likelihoods
  - Point Estimation
  - Interval Estimation
- Tomorrow
  - Discovery significance
  - Upper Limits
  - Further topics

Exercises

- We perform a counting experiment where b=400. We observe 410 events. These counts are large enough so that result is Gaussian
  - Write the likelihood
  - Compute the best-fit value for s
  - Compute the 68% confidence interval for s.

- Combining two Gaussian measurements
  - Recall $L(\theta; x_{1,} x_2) = G(x_1; \theta, \sigma_1) G(x_2; \theta, \sigma_2)$

$$\hat{\theta} = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

  - Compute the (68%) "Combined error" for this estimate