Statistics

Jonas Rademacker at TESHEP 2018

New! With links to solutions to the problem sheet & the TESHEP-MC jupyter notebook on slide 2

Problems, Solutions and other links

Problem sheet: **https://tinyurl.com/TeshepProblems**

Solutions: **https://tinyurl.com/TeshepSolutions**

Jupyter Workbook for Monte Carlo à la TESHEP **https://tinyurl.com/TeshepMC**

Solutions: **https://tinyurl.com/TeshepMCSolved**

Additional Jupyter notebooks to play around with:

https://tinyurl.com/TeshepStatCode

Links for installing jupyter and anaconda:

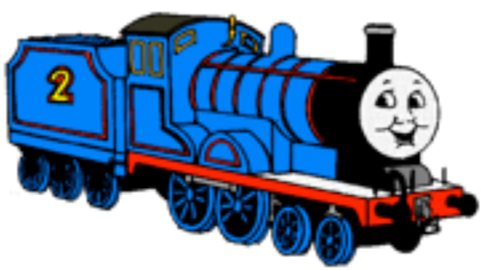
http://jupyter.readthedocs.io/en/latest/install.html

https://docs.anaconda.com/anaconda/

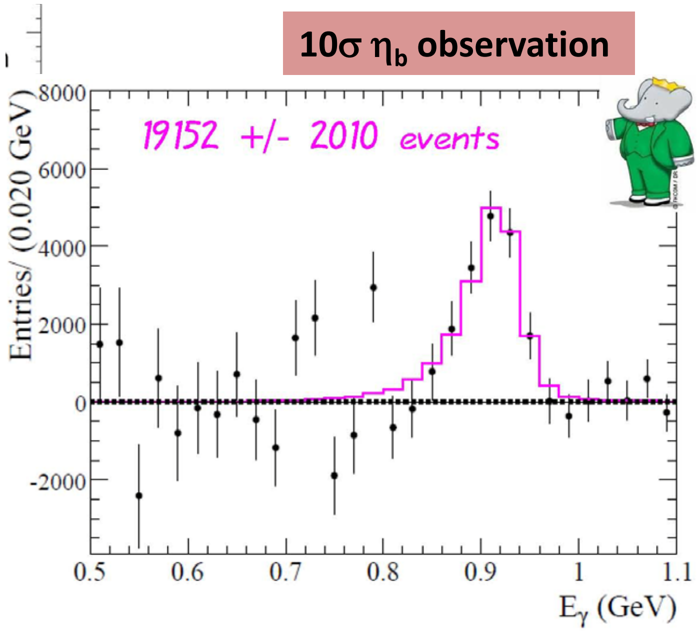
Statistics,Probability and Physics

$$i\hbar\frac{\partial}{\partial t}\psi = -\frac{\hbar^2}{2m}\nabla^2\psi + V\psi$$

Quantum Mechanics



Thermodynamics



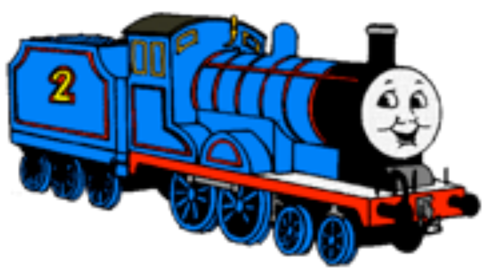
Interpretation of data

Statistics,Probability and Physics

$$i\hbar\frac{\partial}{\partial t}\psi = -\frac{\hbar^2}{2m}\nabla^2\psi + V\psi$$

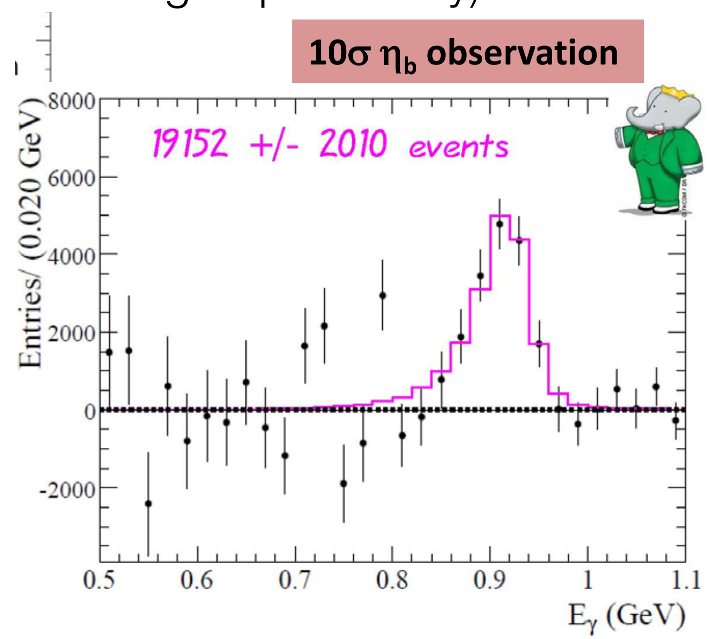
Quantum Mechanics

(a different, fundamental meaning to probability)



Thermodynamics

Probability, law of large numbers, combinatorics



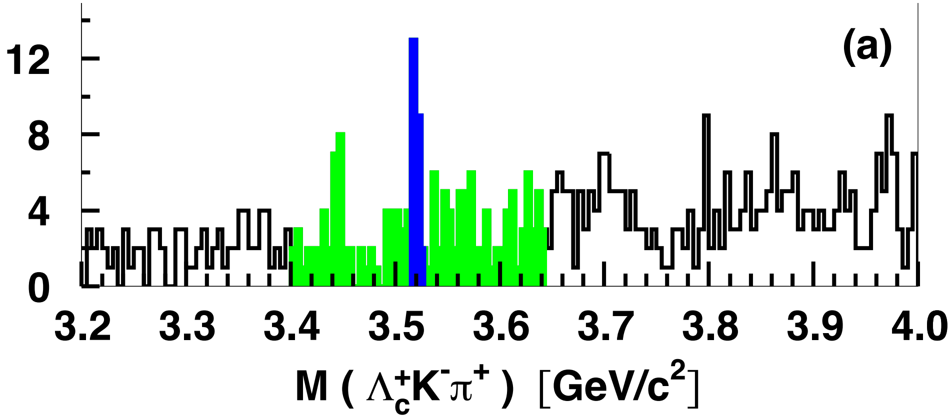
Interpretation of data

measurement errors, statistical fluctuations, Central Limit Theorem, confirming & rejecting theories, what constitutes a discovery?

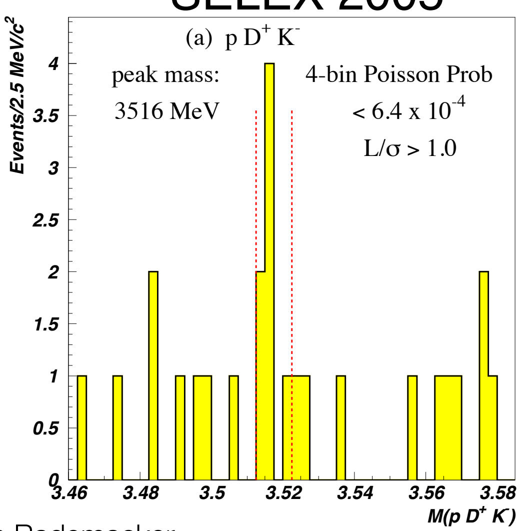
A Ξ_{cc} at 3.5 GeV?

SELEX see it twice

SELEX 2002



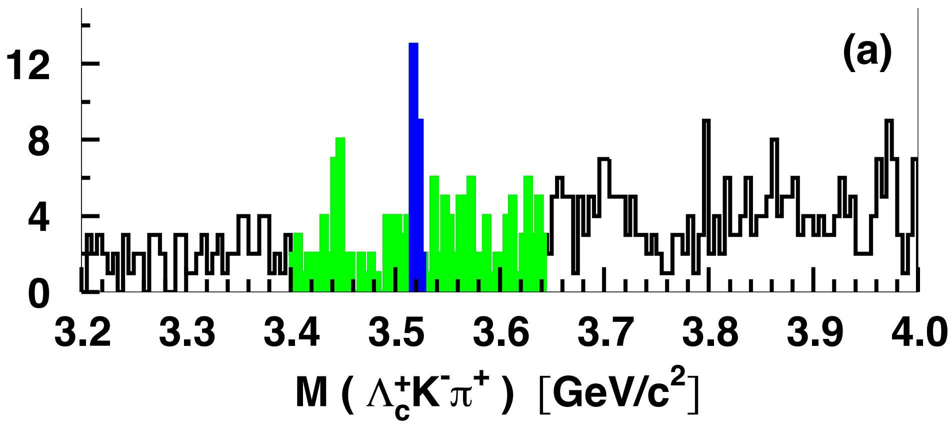
SELEX 2005



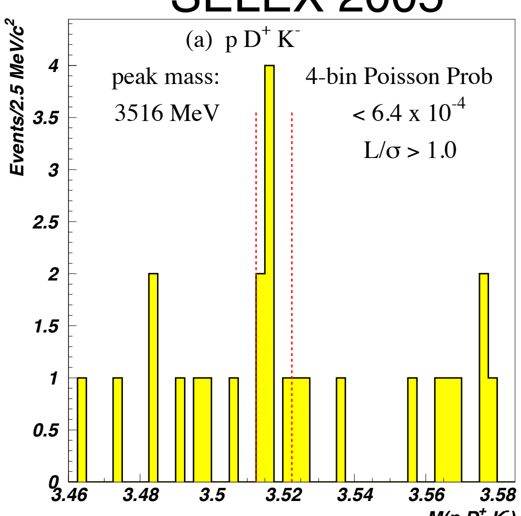
A Ξ_{cc} at 3.5 GeV?

SELEX see it twice

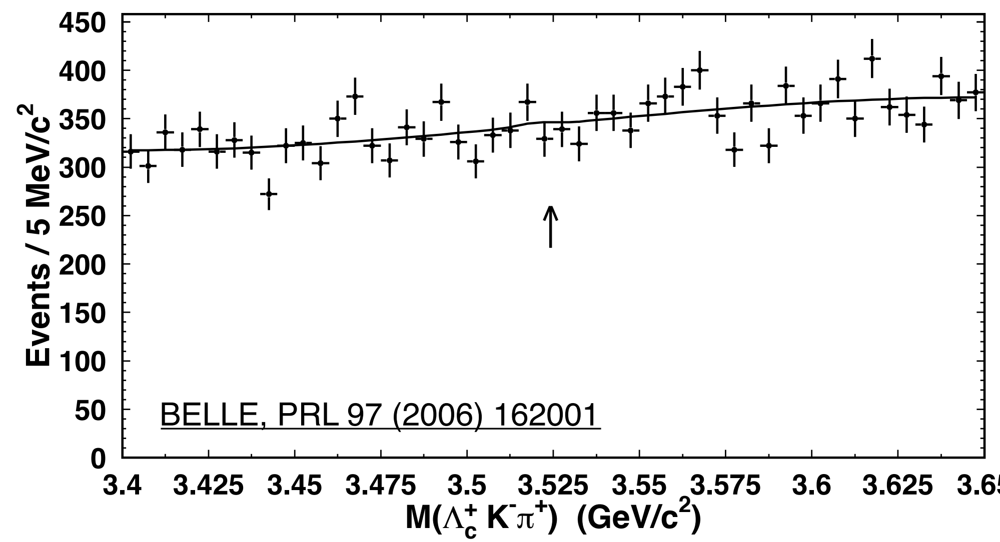
SELEX 2002



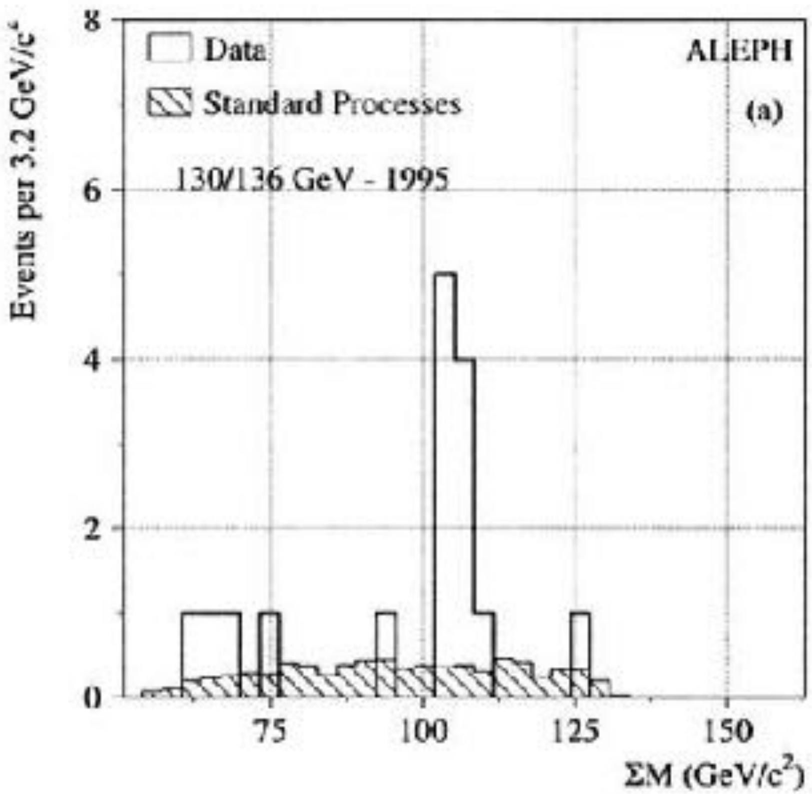
SELEX 2005



FOCUS, BaBar, BELLE, LHCb don't



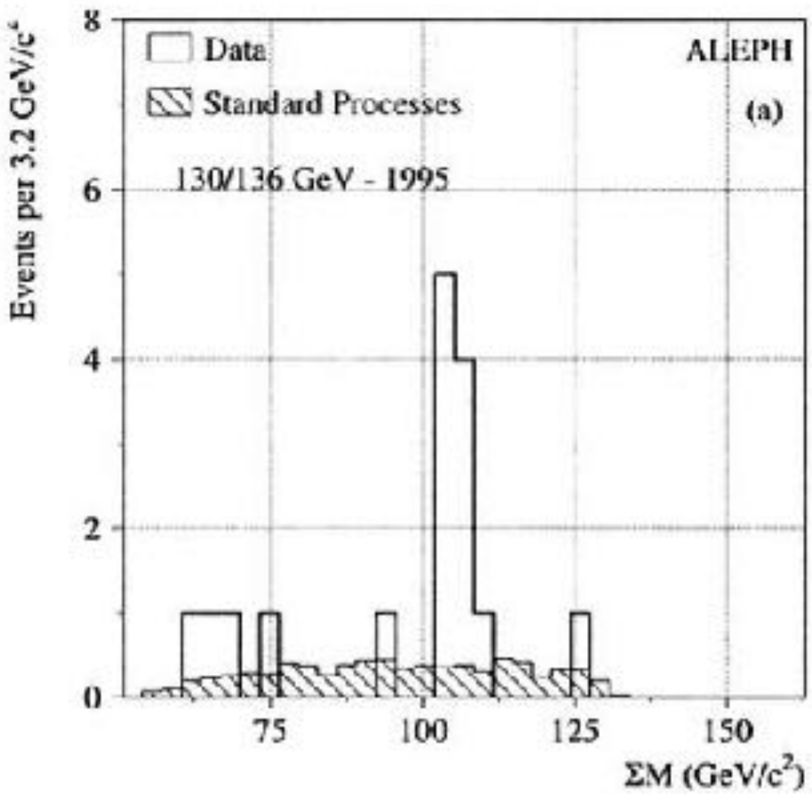
Higgs: true or false?



see: http://www.science20.com/a_quantum_diaries_survivor/true_and_false_discoveries_how_to_tell_them_apart-141024

Higgs: true or false?

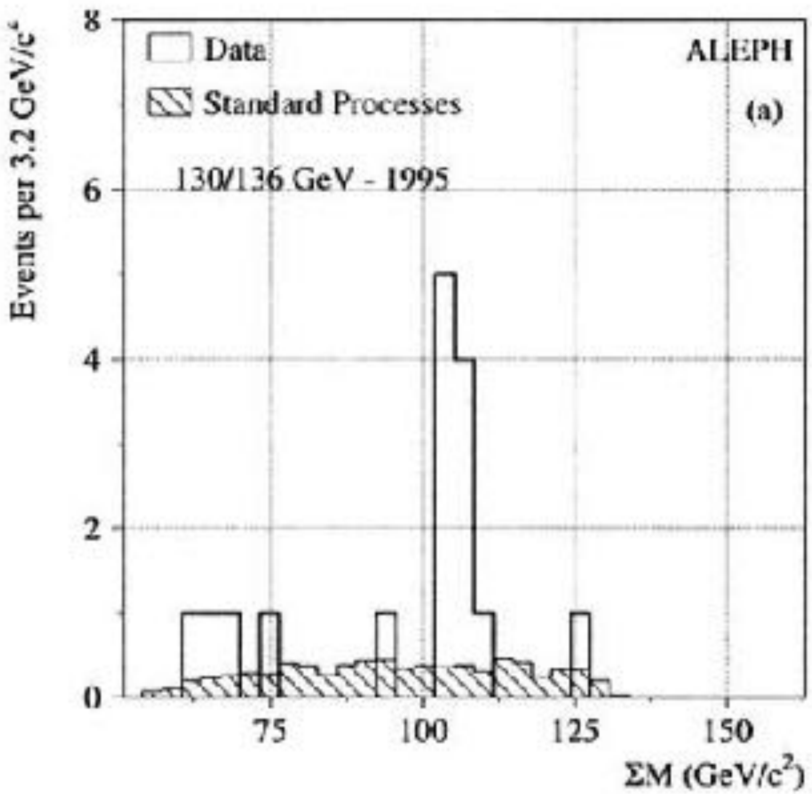
false Higgs (ALEPH/LEP 1996)



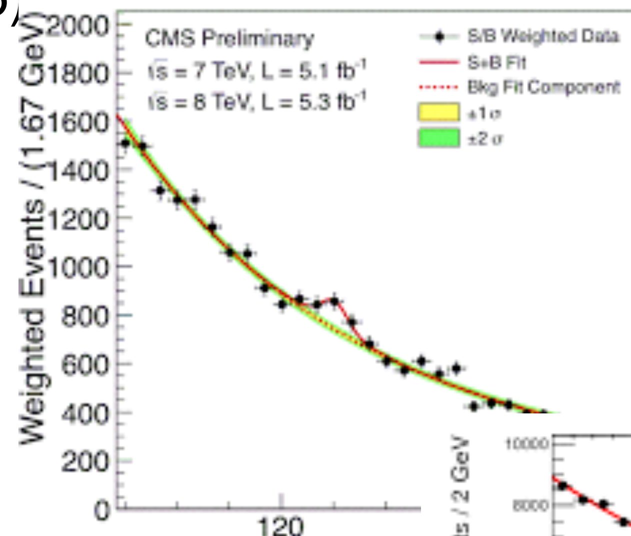
see: http://www.science20.com/a_quantum_diaries_survivor/true_and_false_discoveries_how_to_tell_them_apart-141024

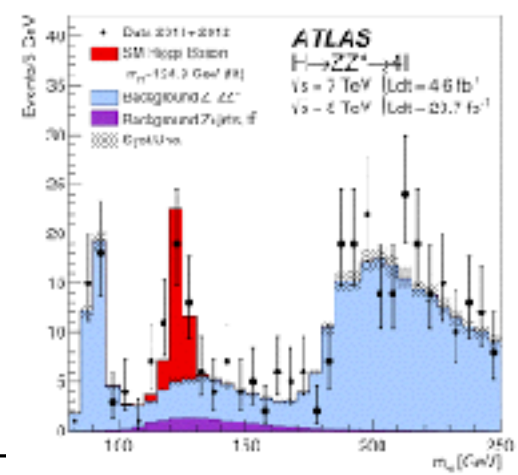
Higgs: true or false?

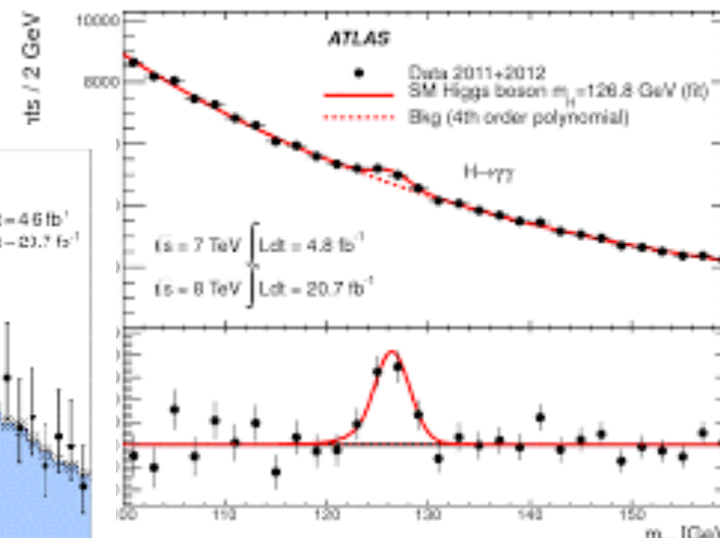
false Higgs (ALEPH/LEP 1996)



real Higgs (LHC 1996)



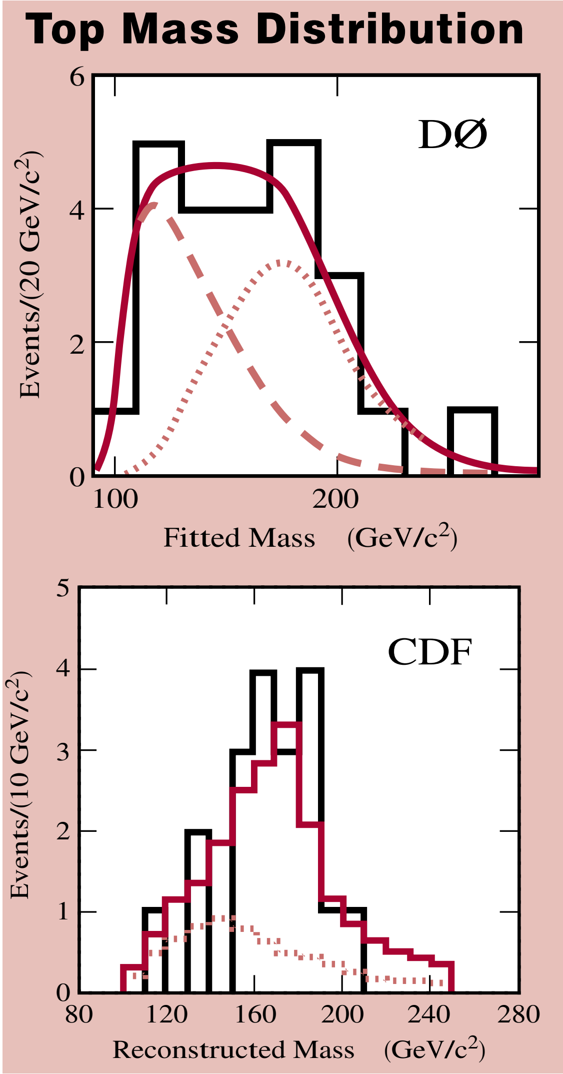




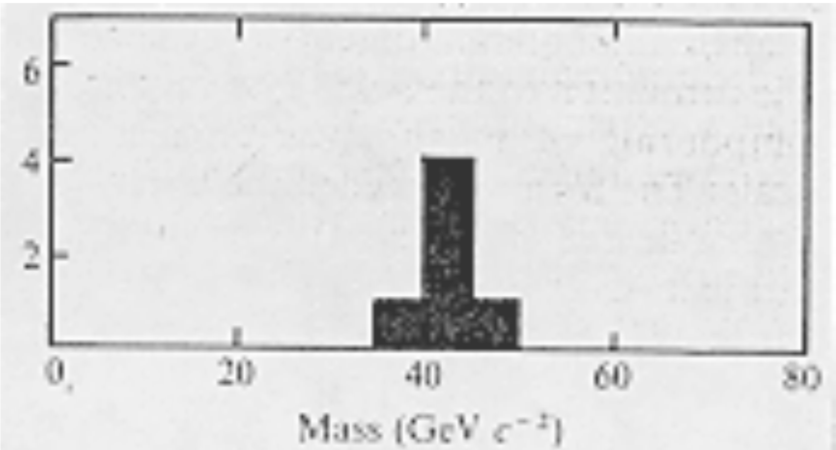
see: http://www.science20.com/a_quantum_diaries_ … _how_to_tell_them_apart-141024

True and False

True top (1996)



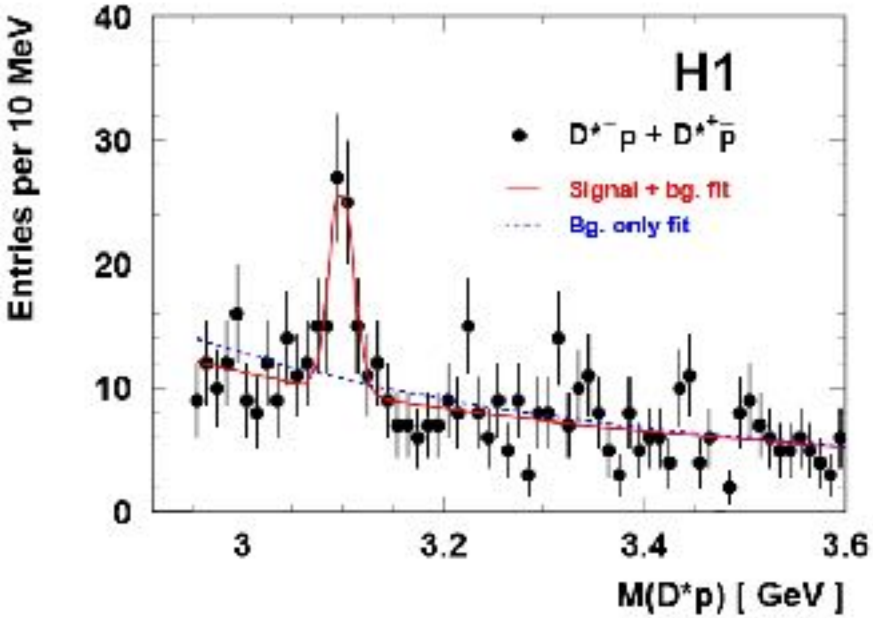
False top (1985)



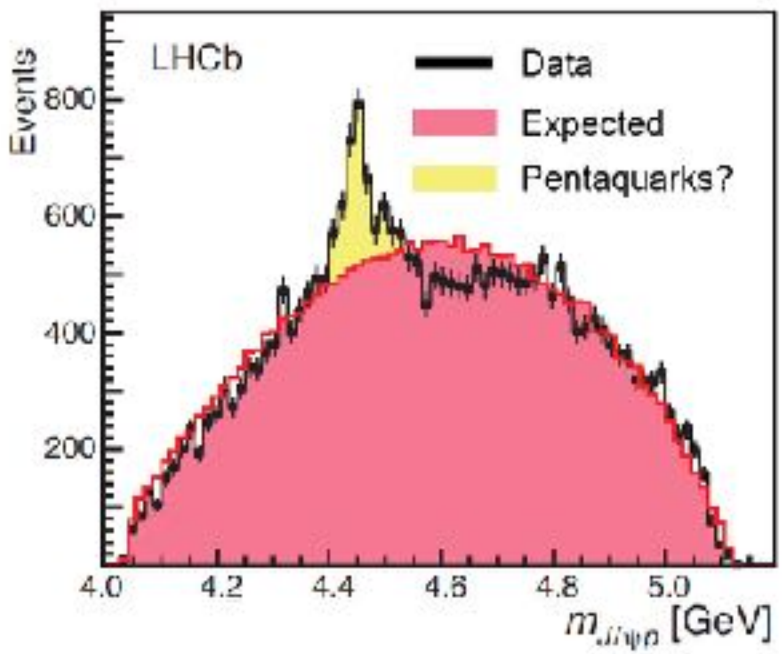
Distribution of measured top quark mass.

True & False: Pentaquark

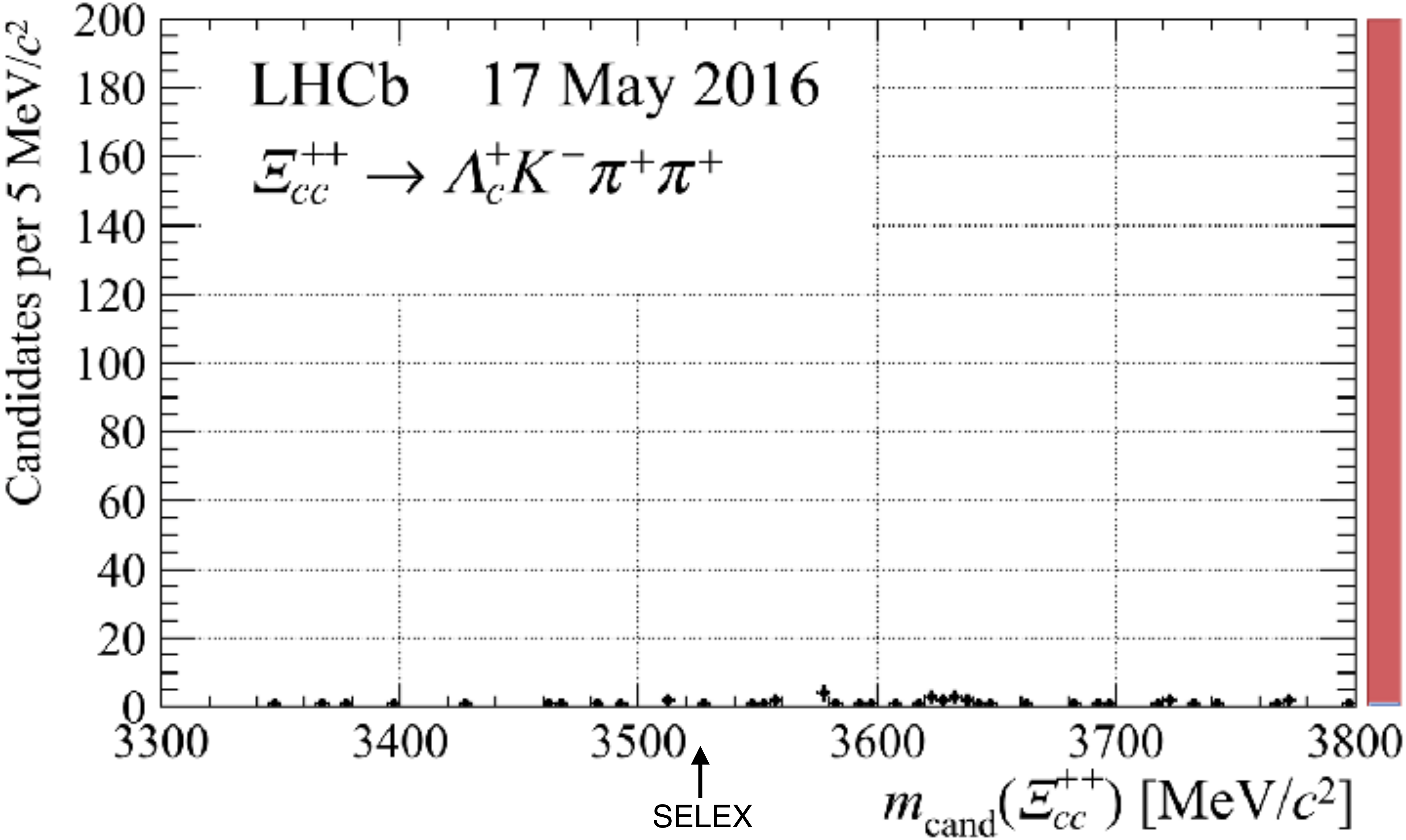
false, 2004, H1 (DESY)



true (LHCb, 2015)

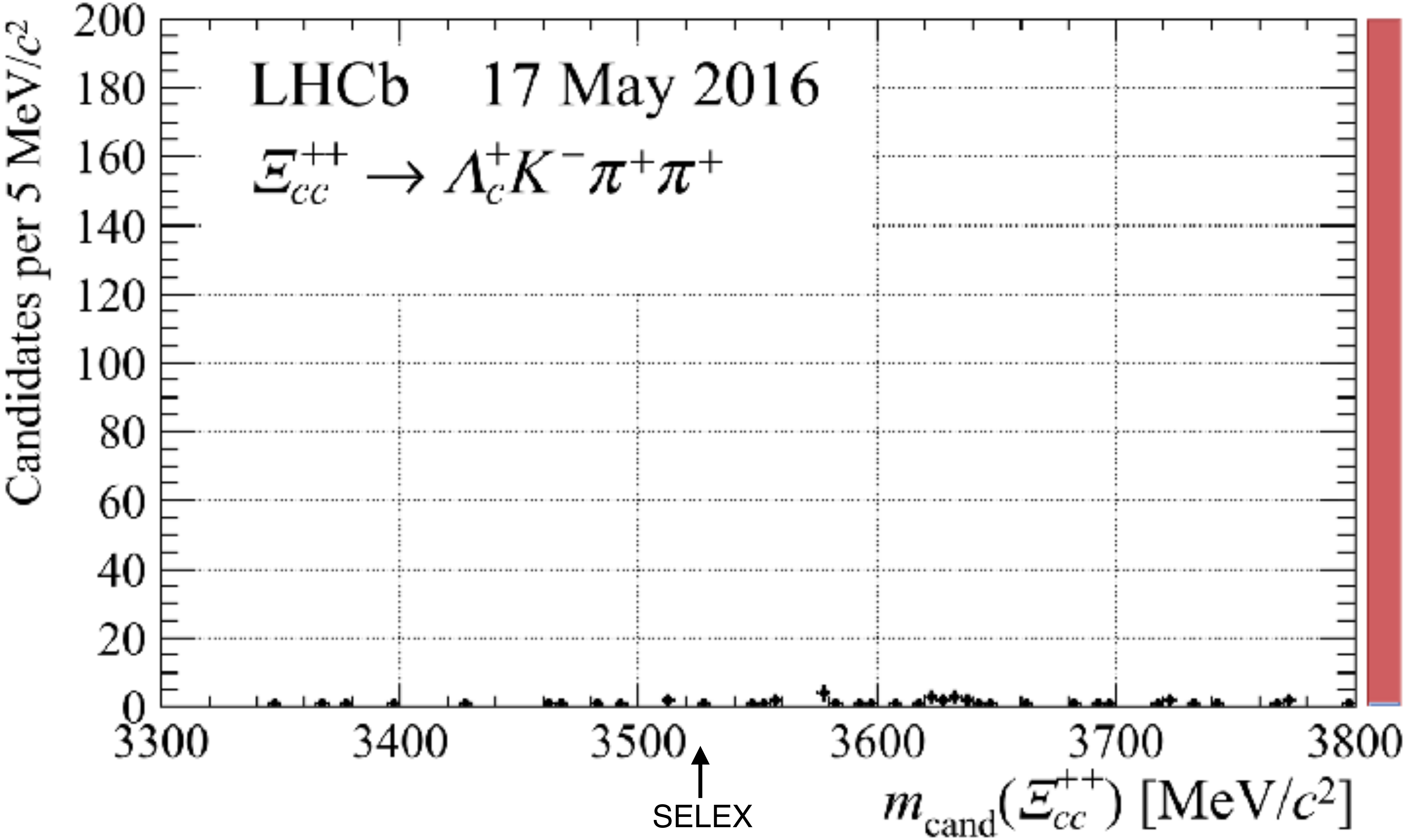


Ξ_{cc} at LHCb?



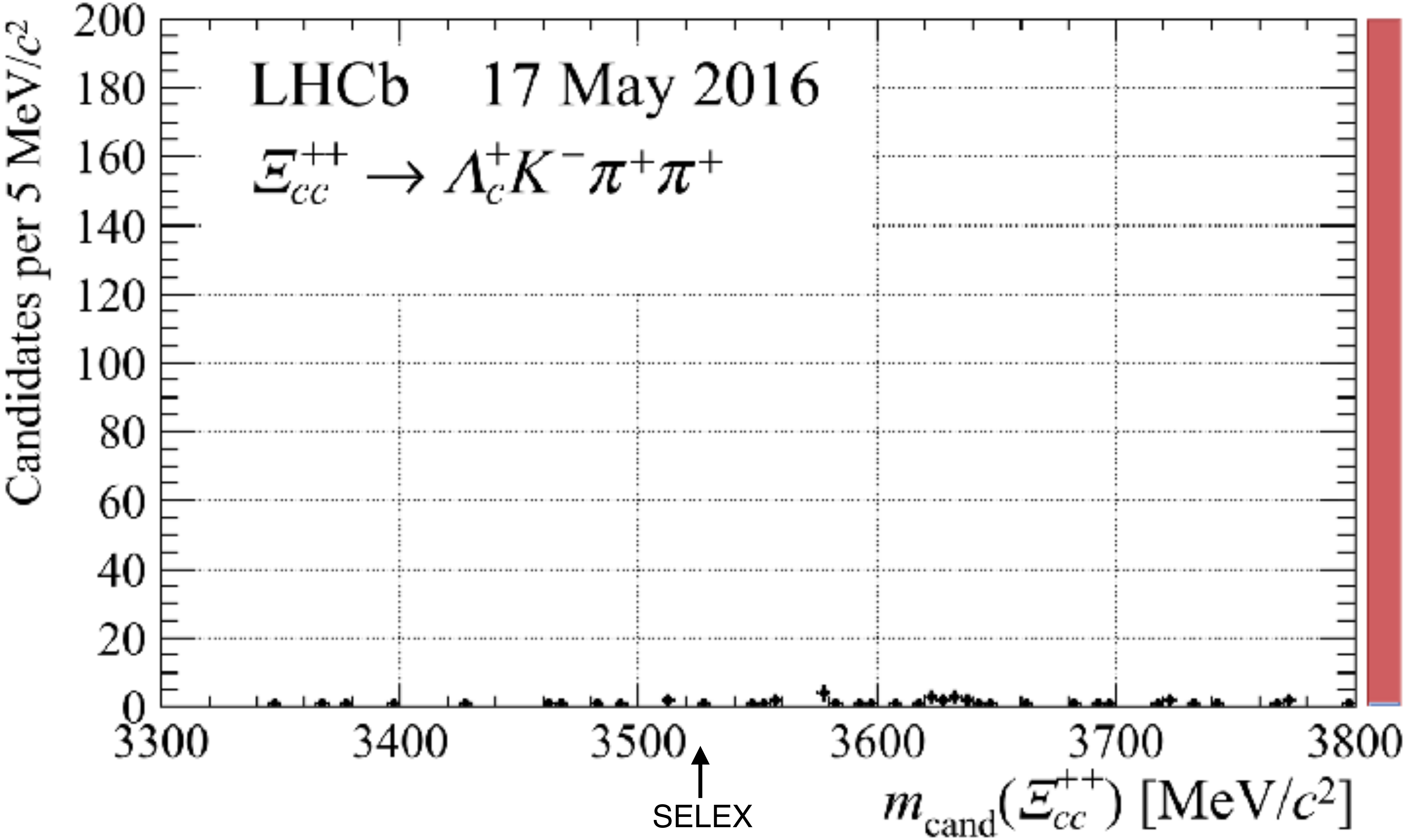
LHCb 17 May 2016
$\Xi_{cc}^{++} \to \Lambda_c^+ K^- \pi^+ \pi^+$
Candidates per 5 MeV/c^2
200
180
160
140
120
100
80
60
40
20
0
3300
3400
3500
3600
3700
3800
$m_{\text{cand}}(\Xi_{cc}^{++})$ [MeV/c^2]
SELEX

Ξ_{cc} at LHCb?



LHCb 17 May 2016
$\Xi_{cc}^{++} \to \Lambda_c^+ K^- \pi^+ \pi^+$
Candidates per 5 MeV/c^2
$m_{\text{cand}}(\Xi_{cc}^{++})$ [MeV/c^2]
0
20
40
60
80
100
120
140
160
180
200
3300
3400
3500
3600
3700
3800
SELEX

When did this become a discovery?



Discoveries...

Discoveries…

- **Particle physics is rife with false hints of discoveries - even the Higgs was seen and unseen at several energies before the LHC had its famous 5σ discovery.**

Discoveries…

- Particle physics is rife with false hints of discoveries - even the Higgs was seen and unseen at several energies before the LHC had its famous 5σ discovery.

- The problem: Nature does not allow us a direct view on its fundamental parameters.

What we want

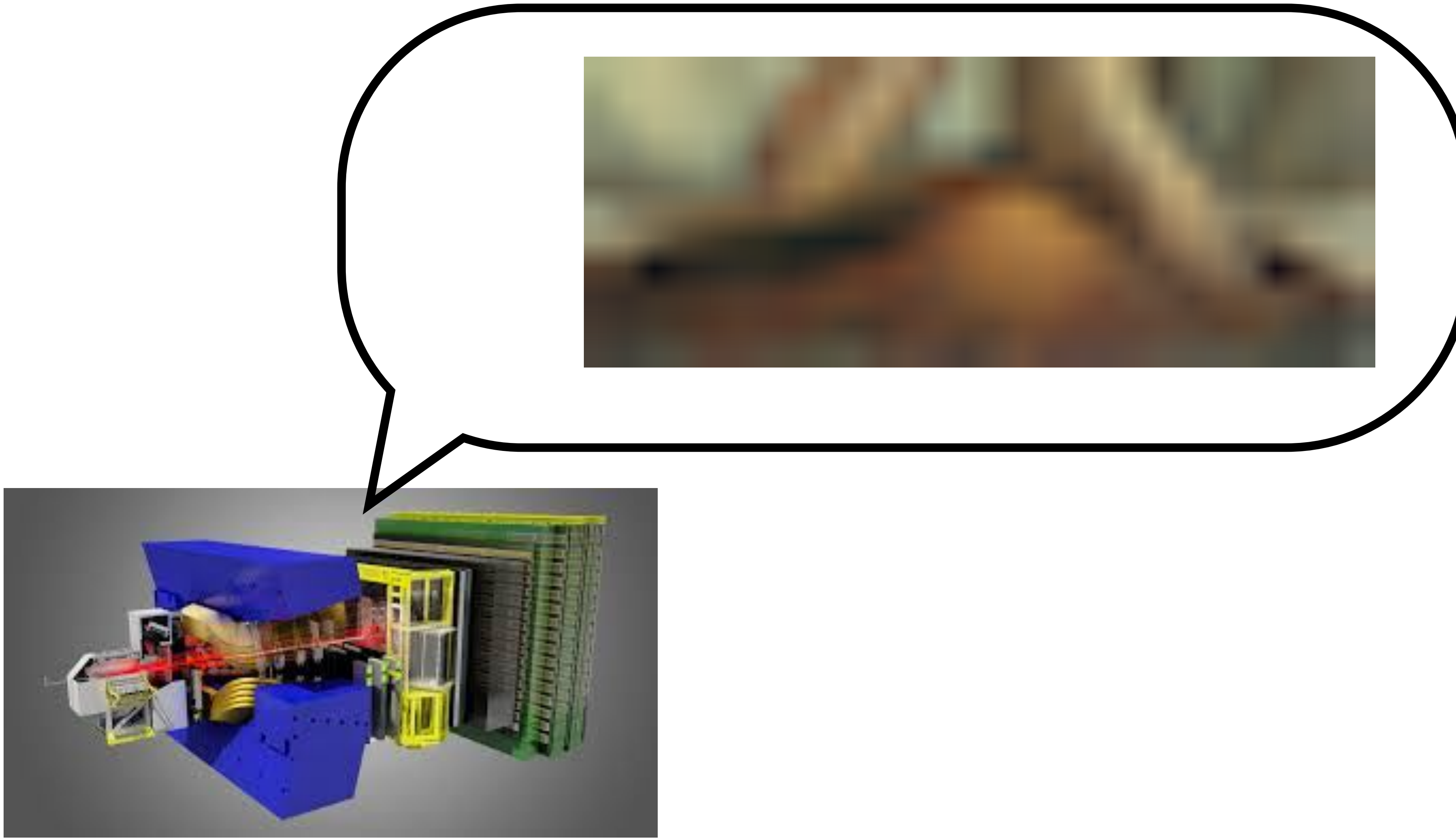
L =

The Birth of Venus by Sandro Botticelli, c. 1482–1486. tempera on canvas, 172.5 × 278.5 cm, Uffizi, Florence

What we get from Sébastien

$L_{\text{low Energy}} \approx$ 

What we get from experiments



Statistics and Measurements

Statistics and Measurements

- **Each measurement is messed up by millions of little perturbations that we cannot possibly all take into account, or even know about, individually.**

Statistics and Measurements

- Each measurement is messed up by millions of little perturbations that we cannot possibly all take into account, or even know about, individually.
- Statistics is the tool that allows us to separate the effect of those fluctuations from the underlying data. And it provides us with tools that tell us how confident we should be in our measurements.

Statistics and Measurements

- Each measurement is messed up by millions of little perturbations that we cannot possibly all take into account, or even know about, individually.
- Statistics is the tool that allows us to separate the effect of those fluctuations from the underlying data. And it provides us with tools that tell us how confident we should be in our measurements.
- After this lecture, you won't discover a false Ξ_{cc} (OK, it's too late for that anyway) or a false Z'. I hope. Discover something surprising, and real!

For a physics Masters/Ph.D….

For a physics Masters/Ph.D….

- **You'll be looking at and interpreting a lot of data.**

For a physics Masters/Ph.D….

- **You'll be looking at and interpreting a lot of data.**
- **You'll deal with a few basic distributions**

For a physics Masters/Ph.D….

- You'll be looking at and interpreting a lot of data.
- You'll deal with a few basic distributions
  - Gaussian, Poisson, binomial, ... (and possibly a few others that you'll pick up as you go along)

For a physics Masters/Ph.D….

- You'll be looking at and interpreting a lot of data.
- You'll deal with a few basic distributions
  - Gaussian, Poisson, binomial, ... (and possibly a few others that you'll pick up as you go along)
  - You'll deal with error estimates and error matrices

For a physics Masters/Ph.D….

- You'll be looking at and interpreting a lot of data.
- You'll deal with a few basic distributions
  - Gaussian, Poisson, binomial, ... (and possibly a few others that you'll pick up as you go along)
  - You'll deal with error estimates and error matrices
- You'll measure parameters doing likelihood and χ^2 fits

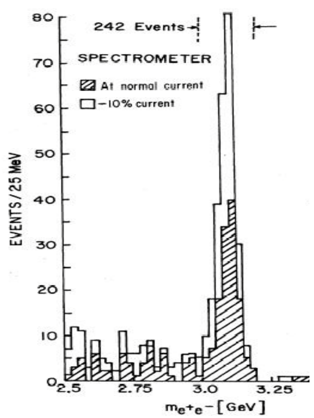
For a physics Masters/Ph.D….

- You'll be looking at and interpreting a lot of data.
- You'll deal with a few basic distributions
    - Gaussian, Poisson, binomial, ... (and possibly a few others that you'll pick up as you go along)
    - You'll deal with error estimates and error matrices
- You'll measure parameters doing likelihood and χ^2 fits
    - You'll need to translate physics into PDF's

For a physics Masters/Ph.D….

- You'll be looking at and interpreting a lot of data.
- You'll deal with a few basic distributions
  - Gaussian, Poisson, binomial, ... (and possibly a few others that you'll pick up as you go along)
  - You'll deal with error estimates and error matrices
- You'll measure parameters doing likelihood and χ^2 fits
  - You'll need to translate physics into PDF's
  - You'll interpret the fit result: what's the error? Is it a discovery? Are the data consistent with the PDF?

Roadmap



What do I see?

Describing Data

What do I expect?

Probability and probability distributions, Probability density functions

Central Limit Theorem

Is what I see compatible with what I expect?

Discoveries
Confidence Levels
Hypothesis testing
Fitting

Monte Carlo simulation

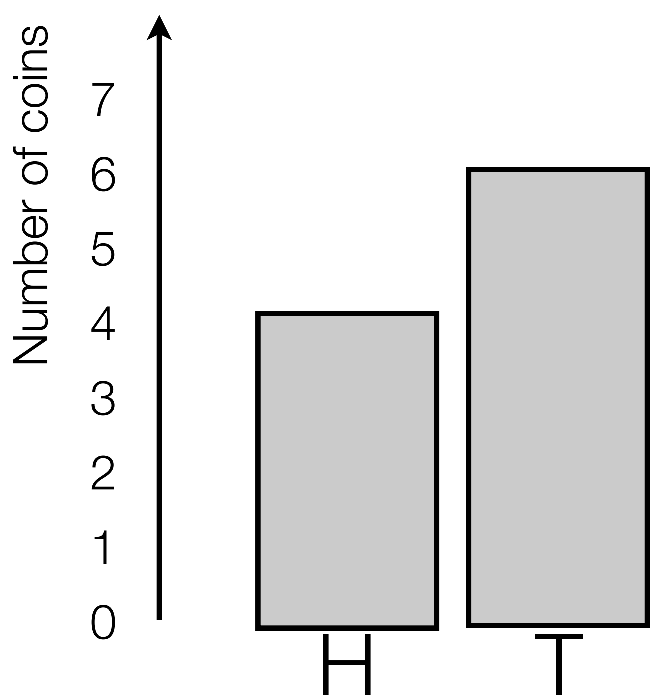
Today: Describing Data

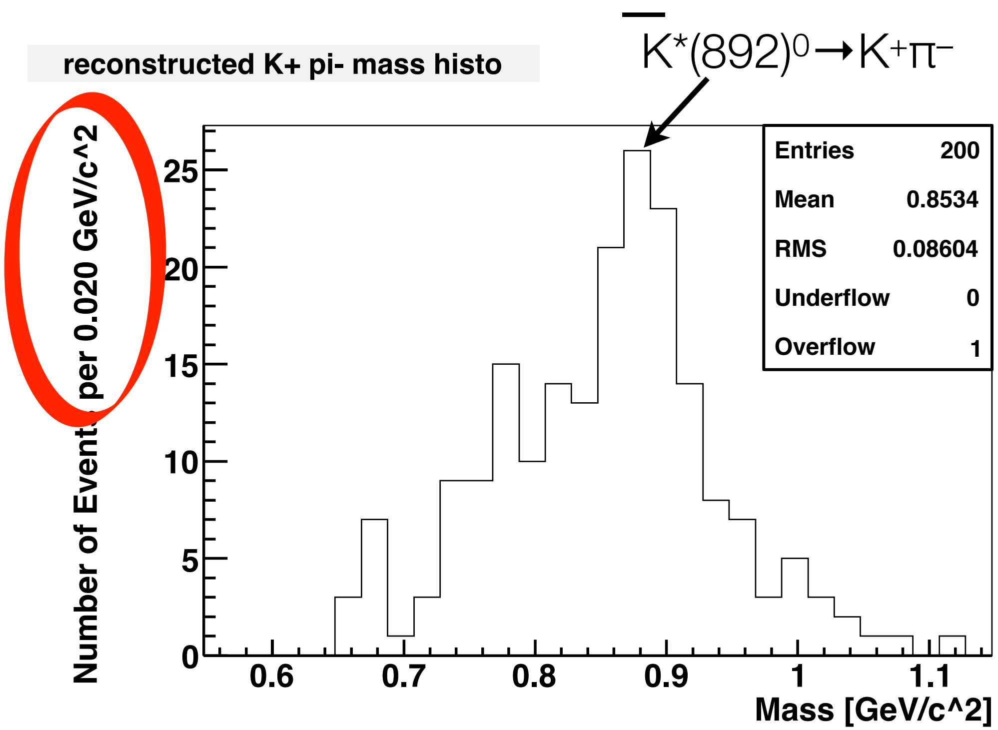
- **Describing real data.**
    - **Displaying them**
    - **Describing them with meaningful, “characteristic” numbers.**

Histograms vs Bar Charts

- Bar chart: length ∝ # events. Histogram: area ∝ # events.

Binwidth matters!





Books

- R. J. Barlow: "Statistics", John Wiley & Sons, ISBN 0-471-92295-1.
- Louis Lyons: "Statists for nuclear and particle physicists", Cambridge University Press, ISBN 0–521–37934–2
- Frederick James: "Statistical Methods in Experimental Physics", World Scientific, ISBN 981-270-527-9 (pbk).

Problems

Problem sheets:

https://tinyurl.com/TeshepProblems

Code (Jupyter Notebooks):

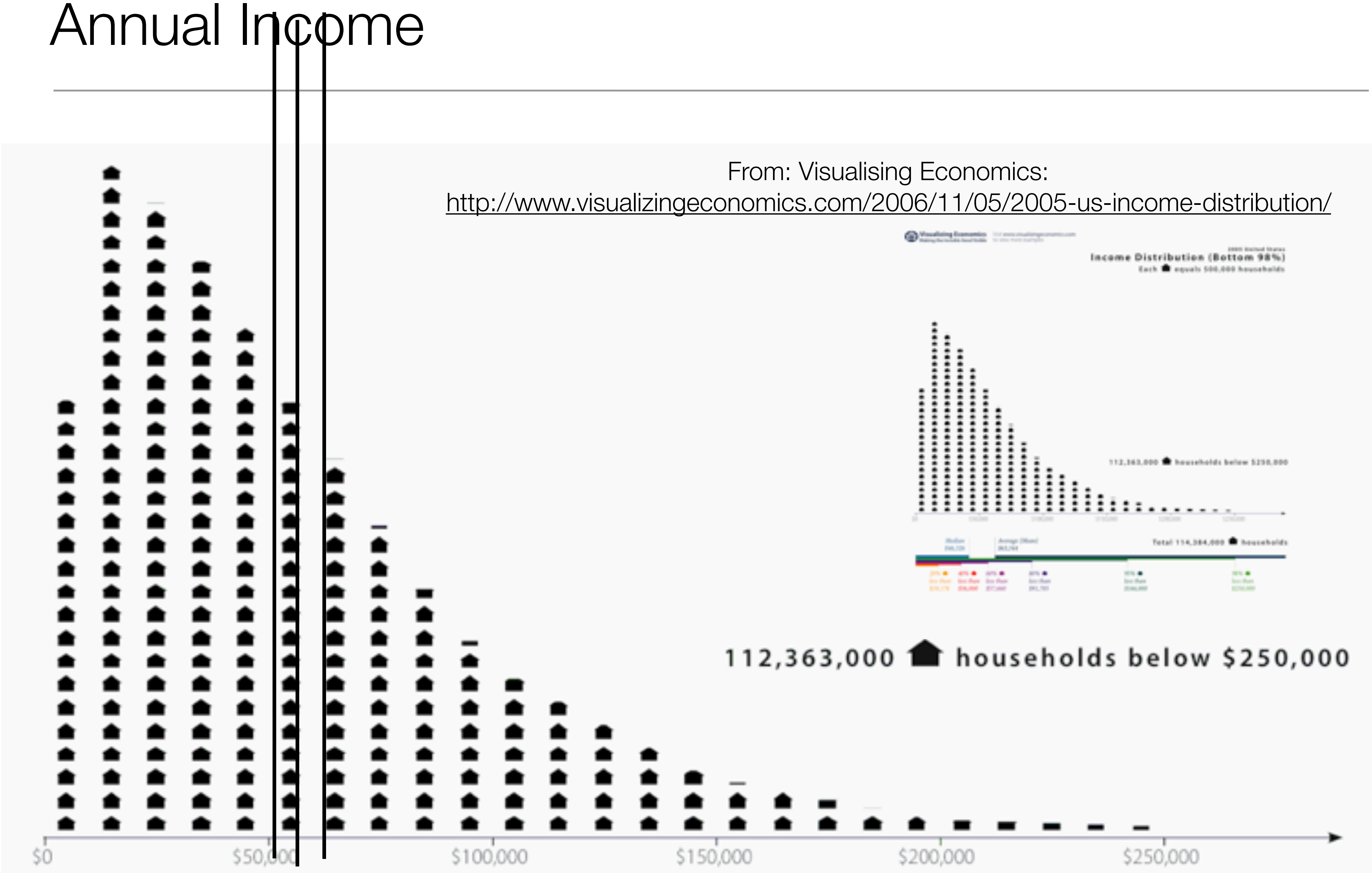
https://tinyurl.com/TeshepStatCode

Describing data with numbers

- **How do we describe a set of measurements with just a couple of characteristic, meaningful numbers?**

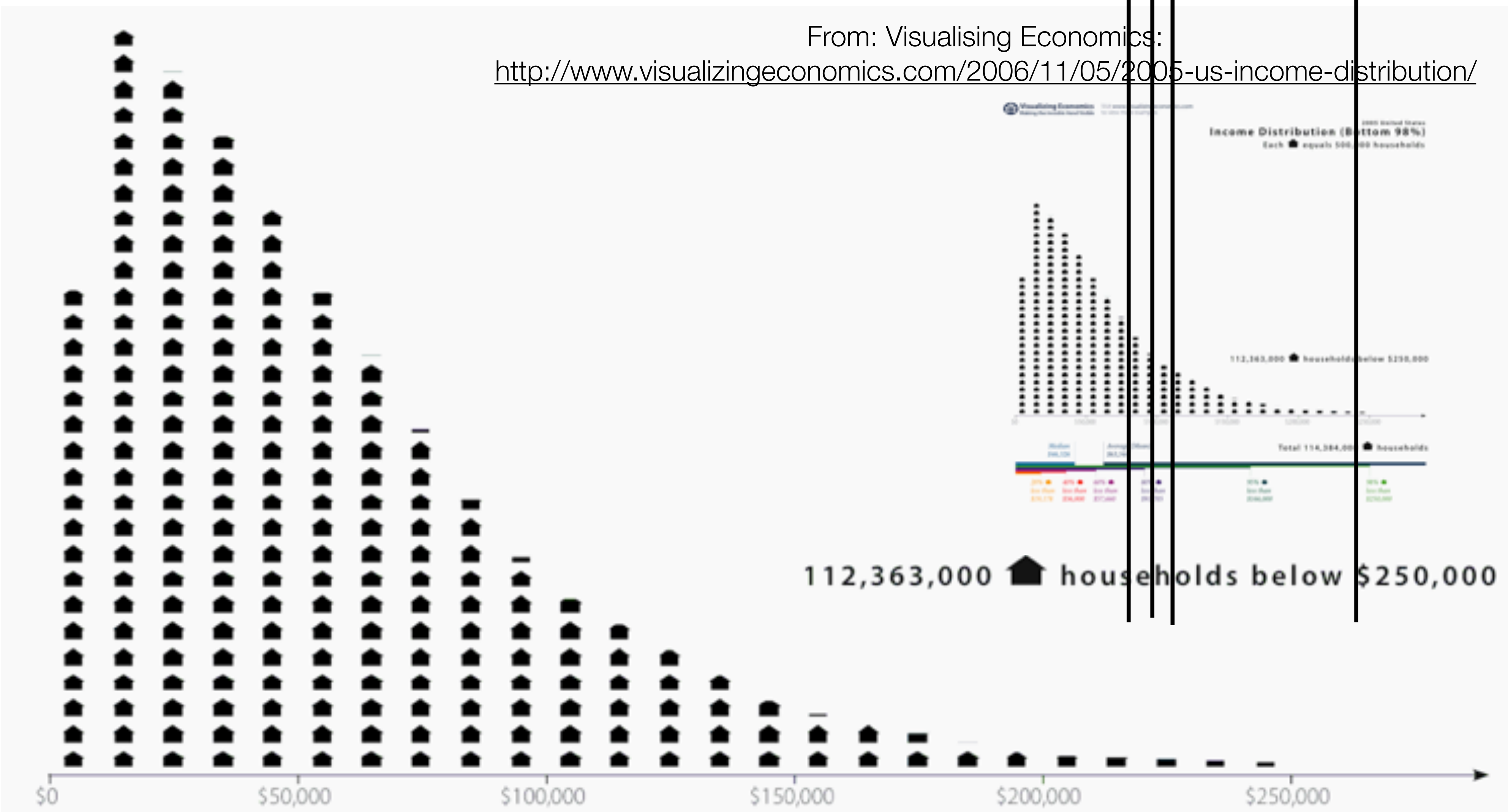
Annual Income

From: Visualising Economics:
http://www.visualizingeconomics.com/2006/11/05/2005-us-income-distribution/

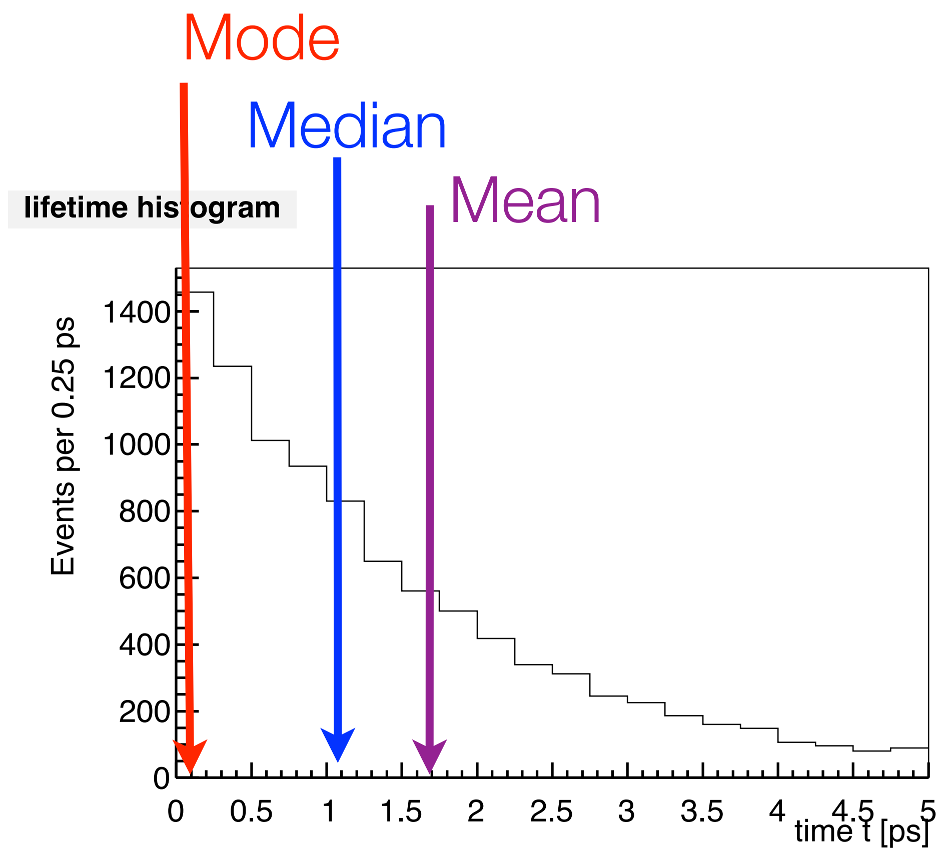


Annual Income

From: Visualising Economics:
http://www.visualizingeconomics.com/2006/11/05/2005-us-income-distribution/



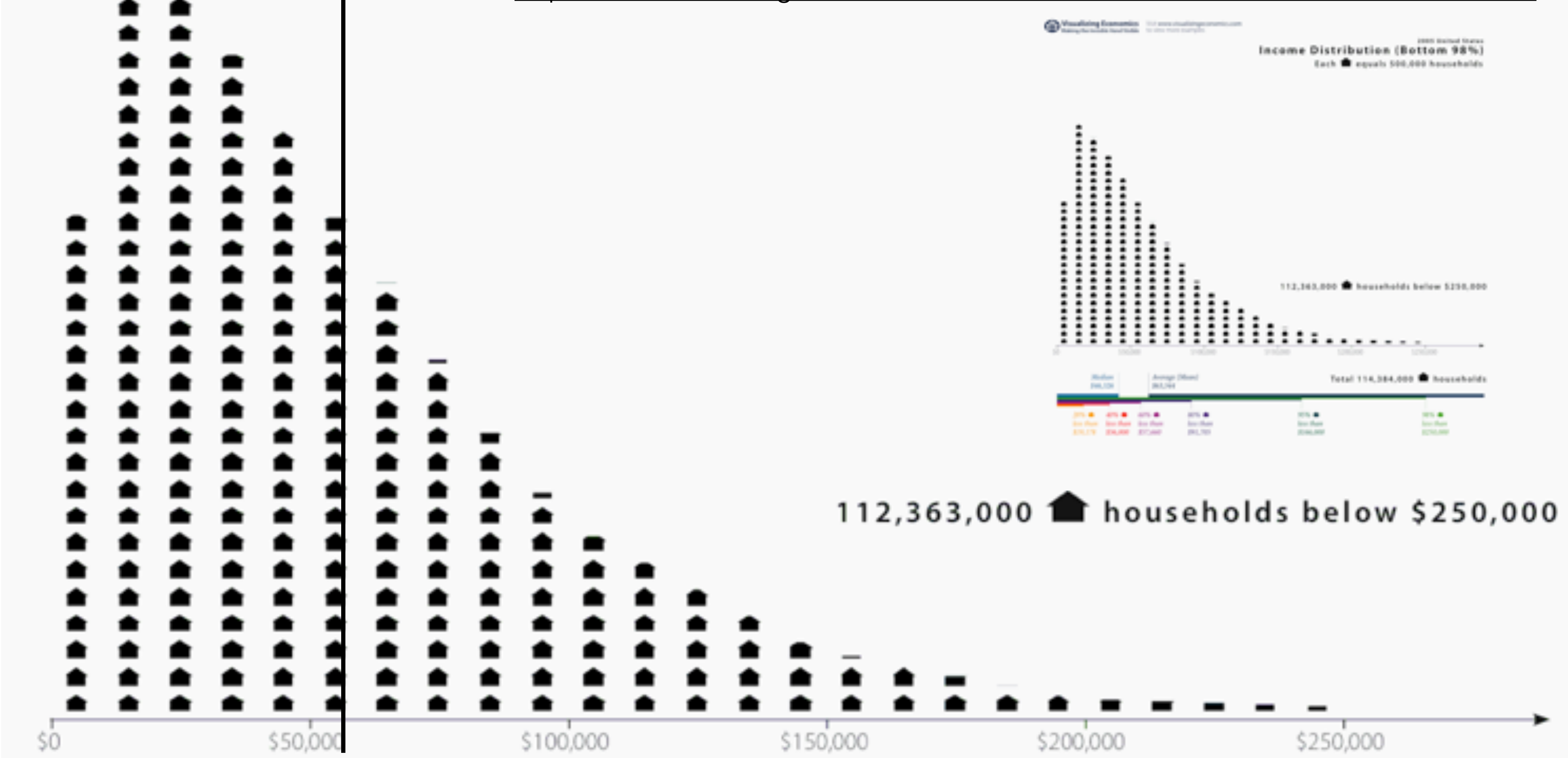
Central Values



- **Mode: highest population**
- **Median: As many events below as above.**
- **Arithmetic Mean: $(1/N)\ \Sigma_{i=1,N}\ x_i$**

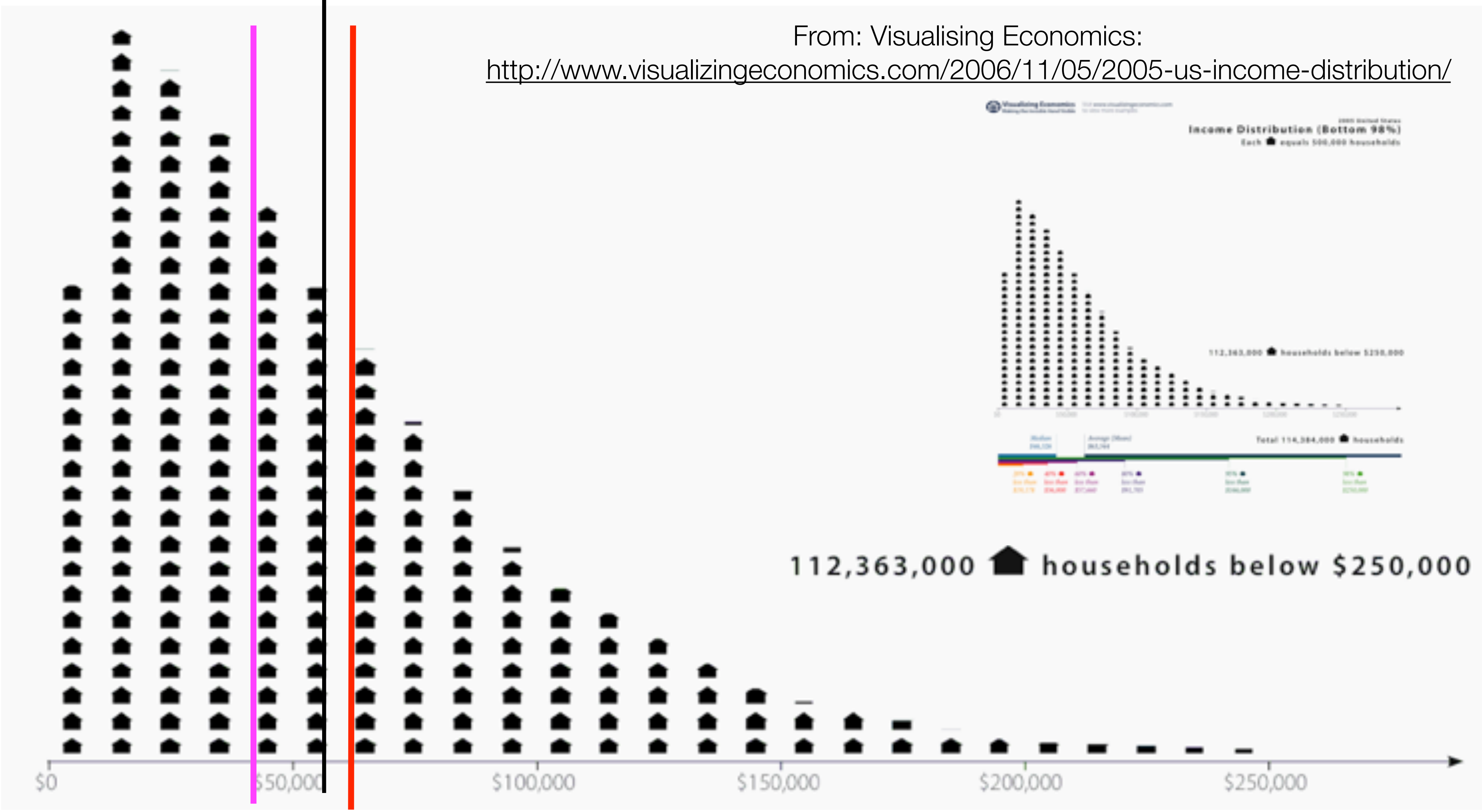
Annual Income

From: Visualising Economics:
http://www.visualizingeconomics.com/2006/11/05/2005-us-income-distribution/



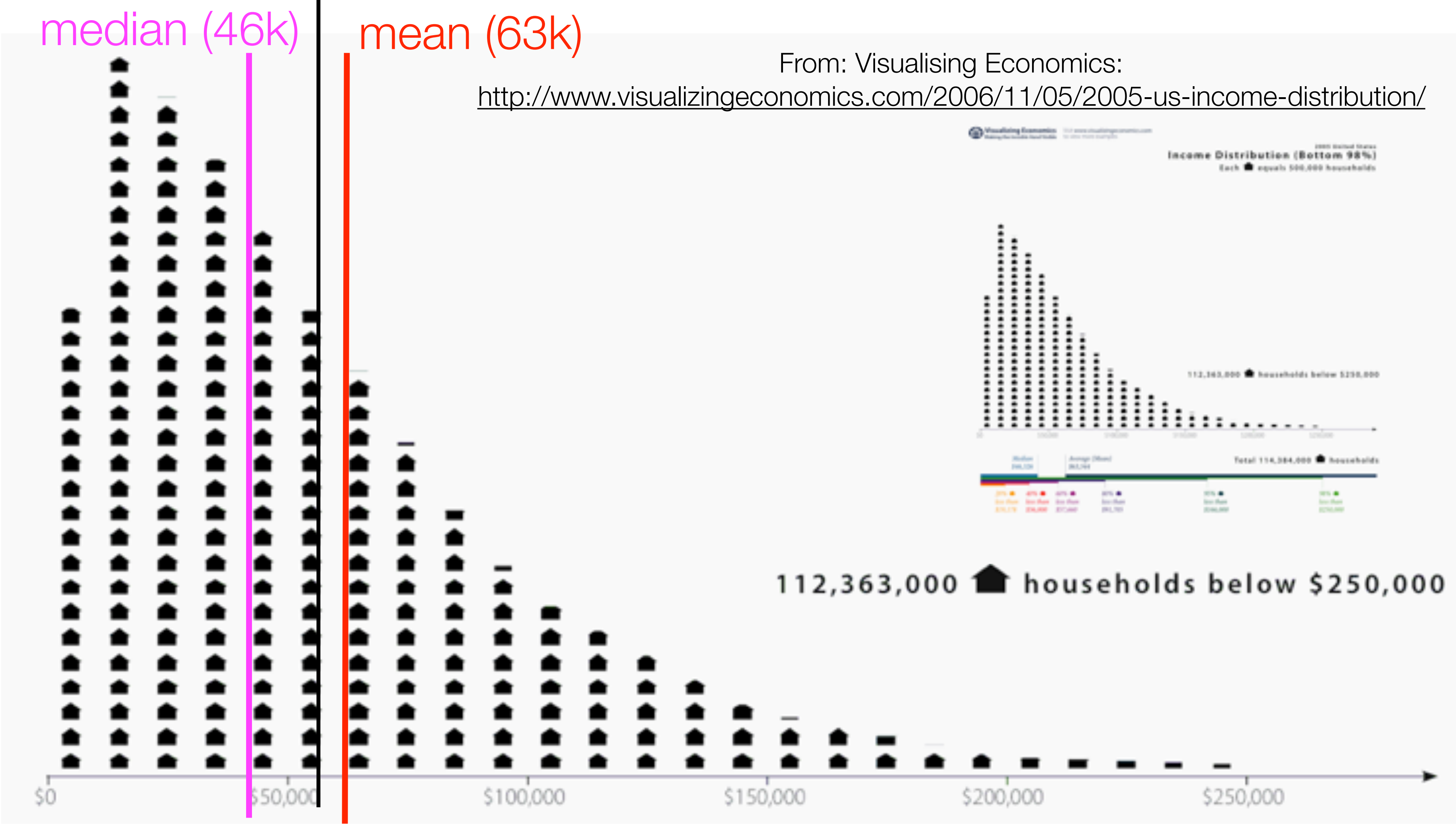
Annual Income

From: Visualising Economics:
http://www.visualizingeconomics.com/2006/11/05/2005-us-income-distribution/



Annual Income

Annual Income



From: Visualising Economics:
http://www.visualizingeconomics.com/2006/11/05/2005-us-income-distribution/

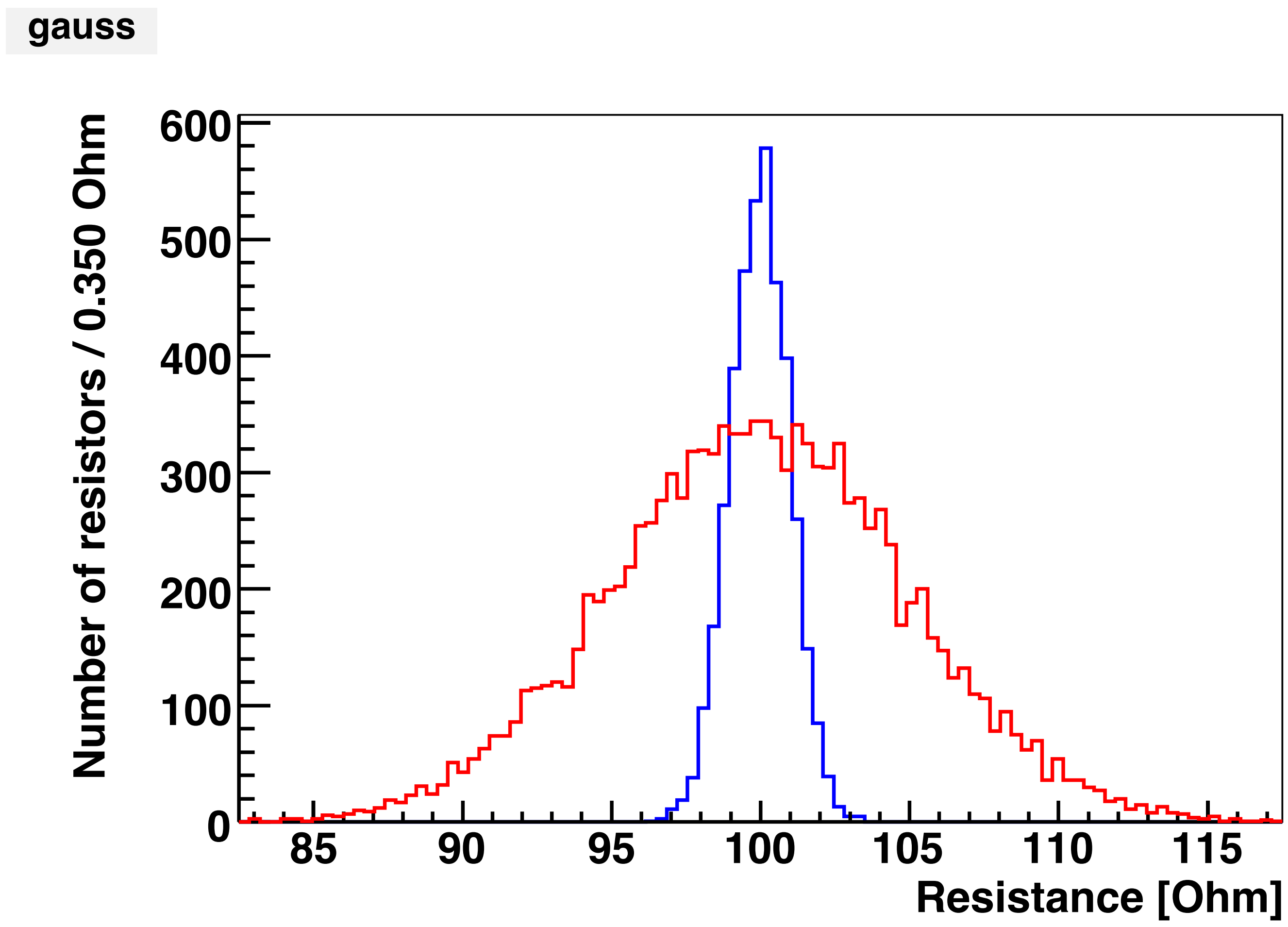
Mean

- For all practical purposes we will usually use the arithmetic mean: $(1/N)\ \Sigma_{i=1,N}\ x_i$

- Motivated to a large degree by its friendly mathematical properties.

- But other central values, other means exist (see also harmonic, geometric, etc) and they have their uses.

Width



gauss
Number of resistors / 0.350 Ohm
600
500
400
300
200
100
0
85
90
95
100
105
110
115
Resistance [Ohm]

Variance

- **We could calculate the total difference from the mean:**

  $d = \Sigma_{i=1,N} (x_i - \overline{x})$ but that's zero by the definition of the mean (check!)

- **The variance is the *average* (difference)^2 from the mean, the variance:**

- **$V \equiv \overline{(x - \overline{x})^2} = 1/N \, \Sigma_{i=1,N} (x_i - \overline{x})^2$**

Calculating the Variance

$$V = \overline{x^2} - \overline{x}^2$$

Home work: verify this

- In words: The variance is equal to

  THE MEAN OF THE SQUARES

  MINUS

  THE SQUARE OF THE MEAN

- You'll always get the order of the terms right if you imagine a wide distribution centered at zero. $\overline{x}^2$ would zero, $\overline{x^2}$ positive and large, and the overall variance must not be negative.

Standard Deviation

- The Standard Deviation is the square-root of the variance:

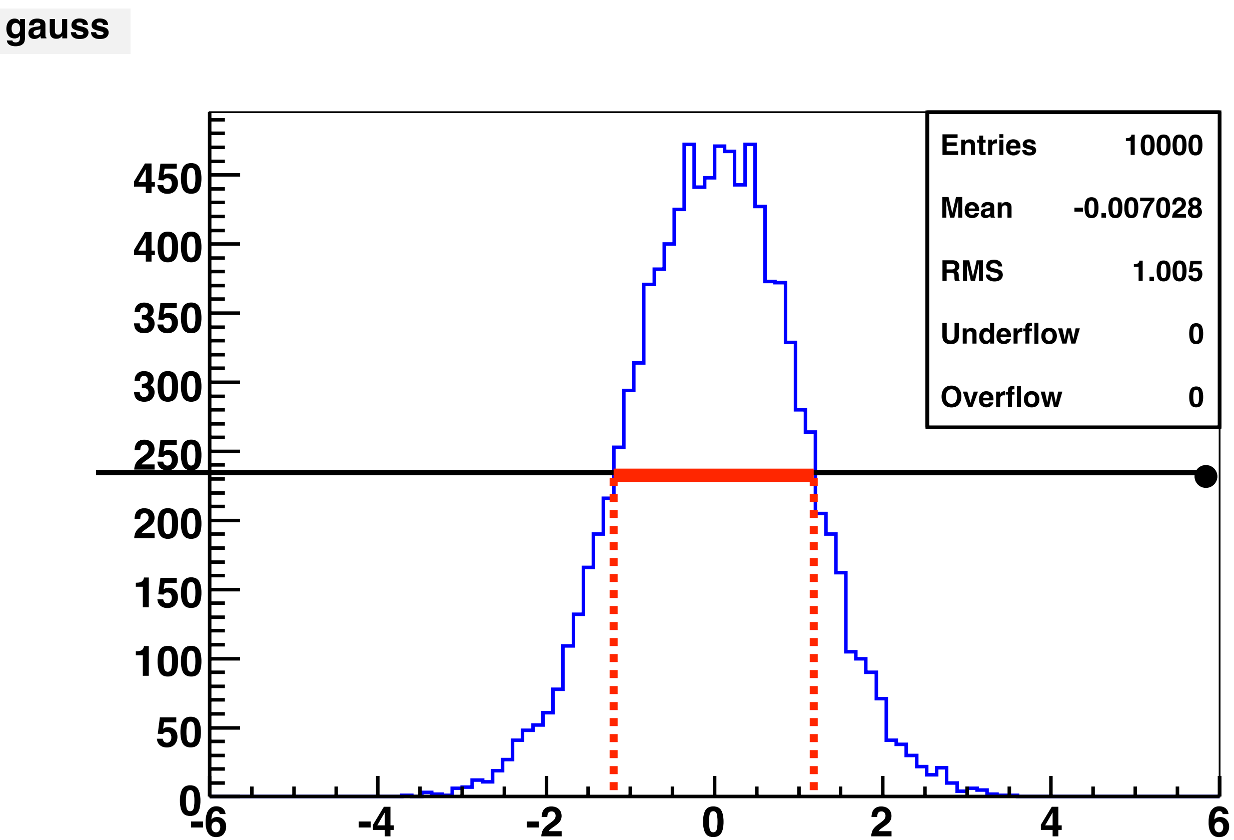
$$\sigma = \sqrt{V}$$

- The Standard Deviation has the same units as the data itself.

- It gives you a “typical” amount by which an individual measurement can be expected to deviate from the mean.

- Usually, a measurement that’s one or two σ away is fine, while 3 σ will raise a few eyebrows. We’ll quantify later what the probabilities for 1, 2, 3 σ deviations are under certain (common) circumstances.

FWHM and standard deviation



- For Gaussian distributions (why these are so important, later):

FWHM ≈ 2.35σ

- Check histogram on the left:

σ =RMS = 1.0,

FWHM= 1.2 – (–1.2) = 2.4

Close enough.

Covariance

- Consider a data sample where each measurement consists of a pair of numbers: *{(x_1, y_1), (x_2, y_2), ...}*

- The *covariance between x and y* is defined as:

$$\mathrm{cov}(x, y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})$$

- The covariance between two parameters is a quantity that has units; its value depends on the units you chose, difficult to interpret.

Covariance

- Consider a data sample where each measurement consists of a pair of numbers: *{(x_1, y_1), (x_2, y_2), ...}*

- The *covariance between x and y* is defined as:

$$\begin{aligned} \mathrm{cov}(x,y) &= \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right) \\ &= \overline{xy} - \overline{x}\cdot\overline{y} \end{aligned}$$

- The covariance between two parameters is a quantity that has units; its value depends on the units you chose, difficult to interpret.

Correlation Coefficient

- The correlation coefficient is defined as:

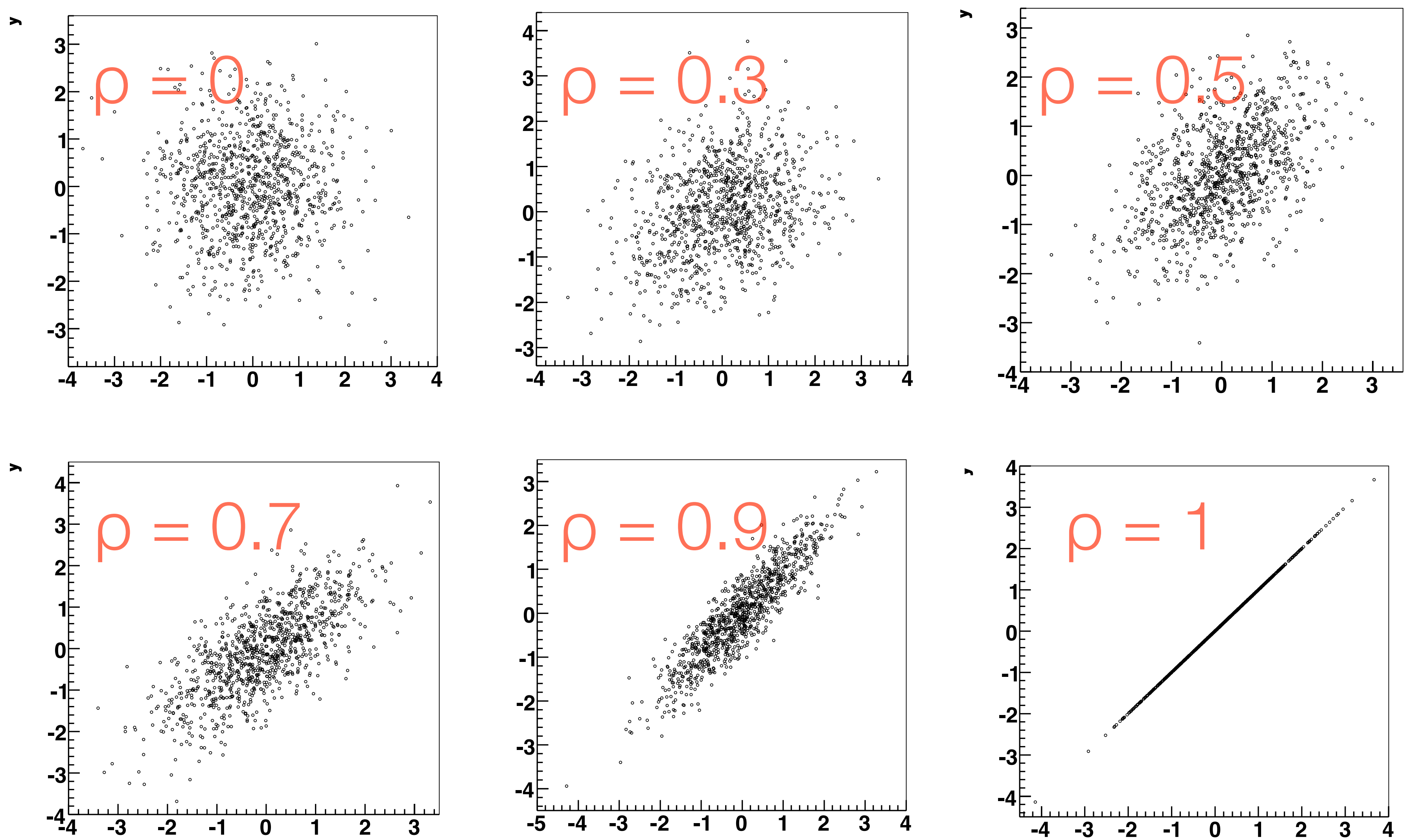
$$\rho_{xy} = \frac{\mathrm{cov}(x,y)}{\sigma_x \cdot \sigma_y}$$

- It has no units and varies between -1 and 1. This provides a measure of how related to quantities are.

- For independent variables, ρ=0 while the correlation coefficient of a parameter with itself (can't get more correlated) is:

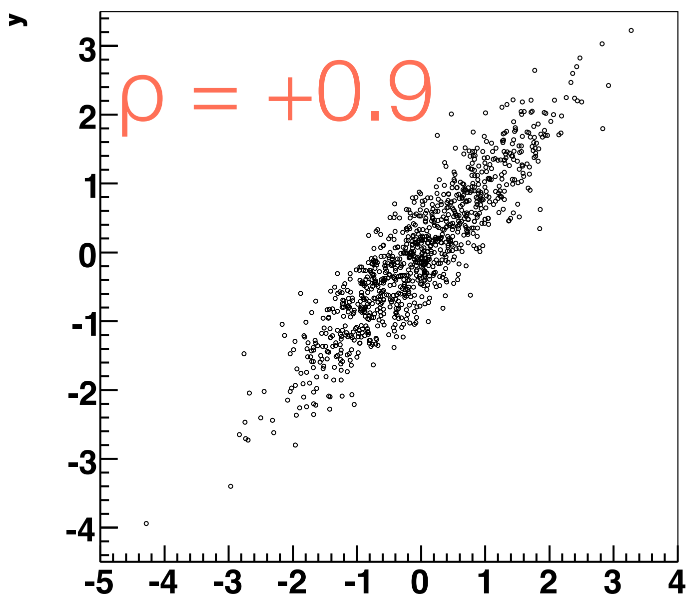
$$\begin{aligned}\rho_{xx} &= \frac{\mathrm{cov}(x,x)}{\sigma_x \cdot \sigma_x} \\ &= \frac{\mathrm{Var}(x)}{\sigma_x^2} = \frac{\sigma_x^2}{\sigma_x^2} = 1\end{aligned}$$

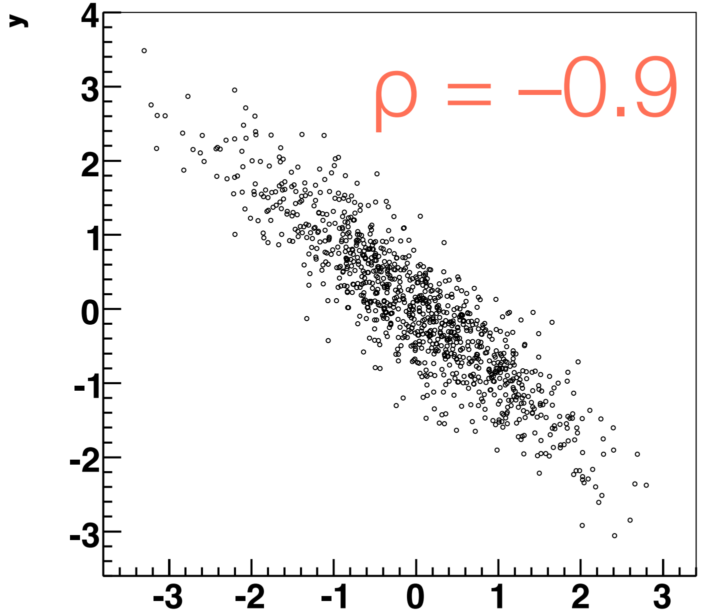
Correlation Coefficient Examples



Correlation Coefficients Examples

- **Correlation coefficients can be positive or negative:**





The Covariance/Error Matrix

- **For N variables, named $x^{(1)}$, ..., $x^{(N)}$**

$$V_{ij} \equiv \text{cov}\left(x^{(i)}, x^{(j)}\right)$$

$$V \equiv \begin{pmatrix} \text{cov}\left(x^{(1)}, x^{(1)}\right) & \text{cov}\left(x^{(1)}, x^{(2)}\right) & \cdots & \text{cov}\left(x^{(1)}, x^{(N)}\right) \\ \text{cov}\left(x^{(2)}, x^{(1)}\right) & \text{cov}\left(x^{(2)}, x^{(2)}\right) & \cdots & \text{cov}\left(x^{(2)}, x^{(N)}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}\left(x^{(N)}, x^{(1)}\right) & \text{cov}\left(x^{(N)}, x^{(2)}\right) & \cdots & \text{cov}\left(x^{(N)}, x^{(N)}\right) \end{pmatrix}$$

- **Symmetric. Diagonal = variances. Off-diagonal: covariances.**

- **Will become very important when we discuss errors and multidimensional parameter transformations.**

The Correlation Matrix

- **Defined equivalently, for N variables $x^{(1)}, \ldots, x^{(N)}$**

$$\rho_{ij} \equiv \frac{\mathrm{cov}\left(x^{(i)}, x^{(j)}\right)}{\sigma_i \sigma_j}$$

$$\rho \equiv \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1N} \\ \rho_{21} & 1 & \cdots & \rho_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N1} & \rho_{N2} & \cdots & 1 \end{pmatrix}$$

- **symmetric**
- **diagonal = 1**
- **Related to covariance matrix by:**

$$V_{ij} = \rho_{ij}\, \sigma_i \sigma_j$$

Correlation and Causality

- Statistics does not tell us if two correlated variables are also connected by causality, i.e. if one causes the other.
- For example there is a strong correlation between rain and wet roads. It is clear that rain causes roads to be wet, and that wet roads do not cause rain. But the statistics won't tell you that.
- There is also a clear correlation between wet roads and the the number of people running around with wet hair. Here neither causes the other, but both are correlated because they have a common cause.

Correlation and Causality

- **Among my favourite correlations is this one:**
- **During doctors' strikes the death-rate tends to go down - in Israel the death-rate went down by 39% in a recent doctors' strike. So there is a positive correlation between life-expectancy and the number of doctors on strike (this phenomenon has been observed in other countries, too). Does this mean that fewer doctors would be good for the nation's health?**
- **Listen to this BBC programme if you like this sort of thing:**

http://news.bbc.co.uk/2/hi/programmes/more_or_less/7408337.stm

Homework

- **Write down 100 times:**

  “Correlation is not causation”

Summary: Representing Data

- Histograms: Area proportional to number of events
  - Label y-axis as Number of Events/(bin size) as in N/(4 GeV)
- Central value: Usually use arithmetic mean. Nice: Means add up. (i.e. <x + y> = <x> + <y>)
- Width: Use standard deviation. Standard deviations do not add up. Variances do, i.e. V(x+y) = V(x) + V(y) (if variable uncorrelated).
- Multiparameter distributions: Covariance, Correlation.

Correlation is not causation.
Correlation is not causation.

Summary: Representing Data

- **Histograms: Area proportional to number of events**
    - **Label y-axis as Number of Events/(bin size) as in N/(4 GeV)**
- **Central value: Usually use arithmetic mean. Nice: Means add up. (i.e. <x + y> = <x> + <y>)**
- **Width: Use standard deviation. Standard deviations do not add up. Variances do, i.e. V(x+y) = V(x) + V(y) (if variable uncorrelated).**
- **Multiparameter distributions: Covariance, Correlation.**

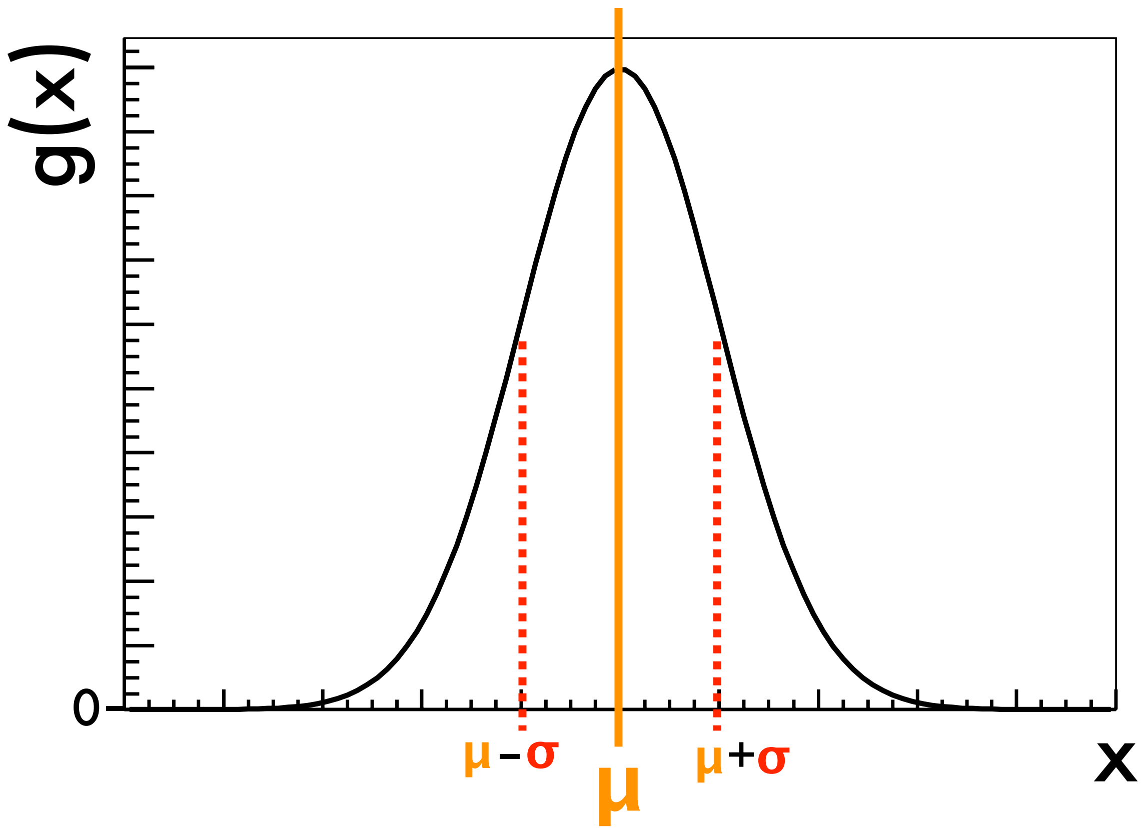
Blur

https://www.youtube.com/watch?v=SSbBvKaM6sk

https://www.youtube.com/watch?v=WDswiT87oo8

We only ever see a slightly blurred picture of nature

Why the blur is Gaussian



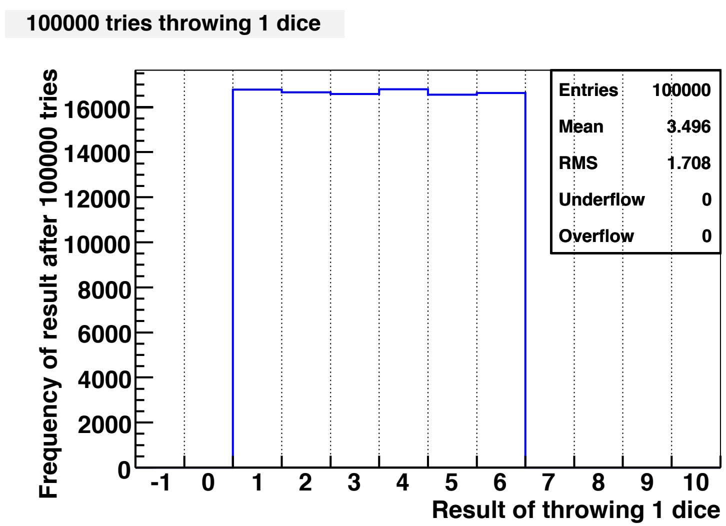
Rolling Dice (macro)

localhost:8888/notebooks/CentralLimitTheoremWithDice.ipynb#

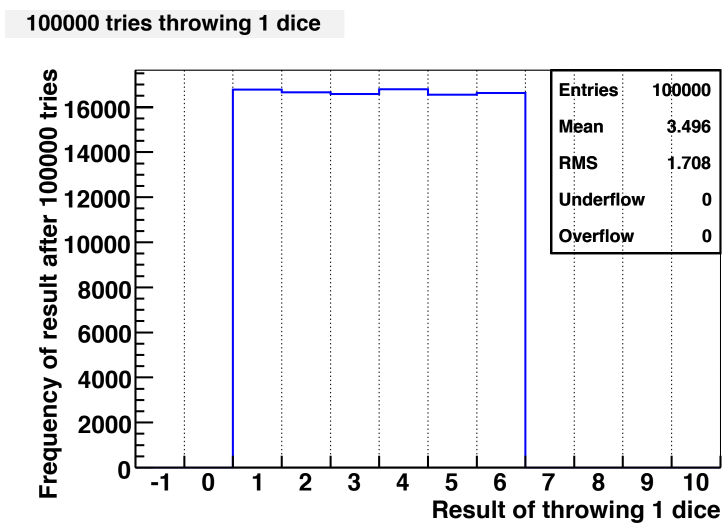
https://tinyurl.com/TeshepStatCode

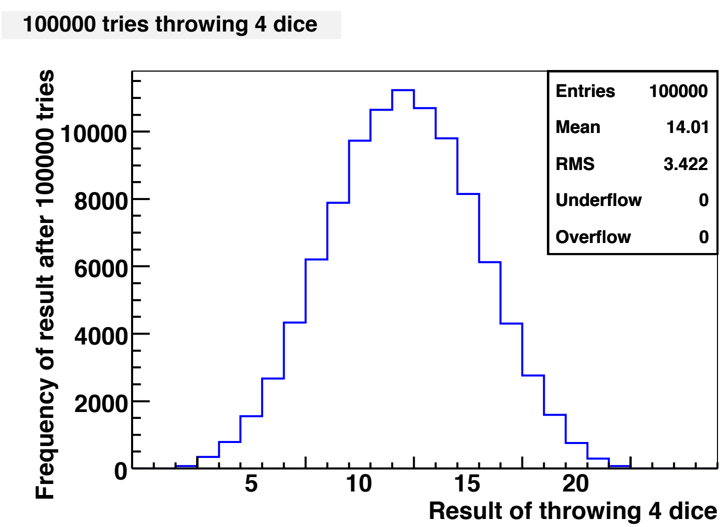
Rolling more and more dice

Rolling more and more dice

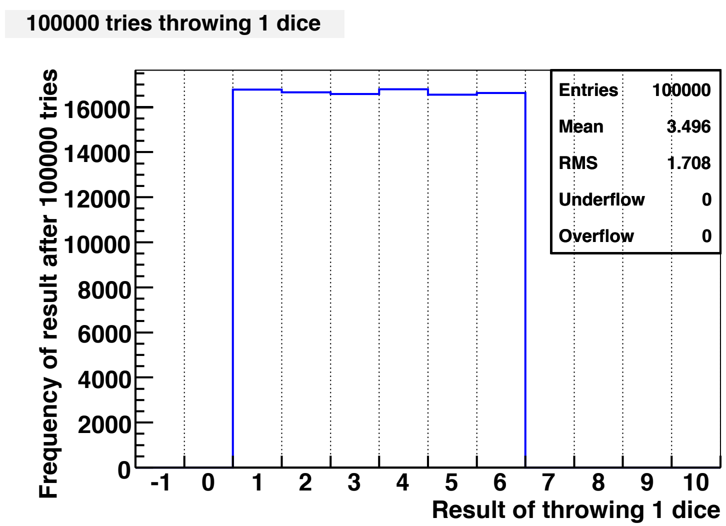


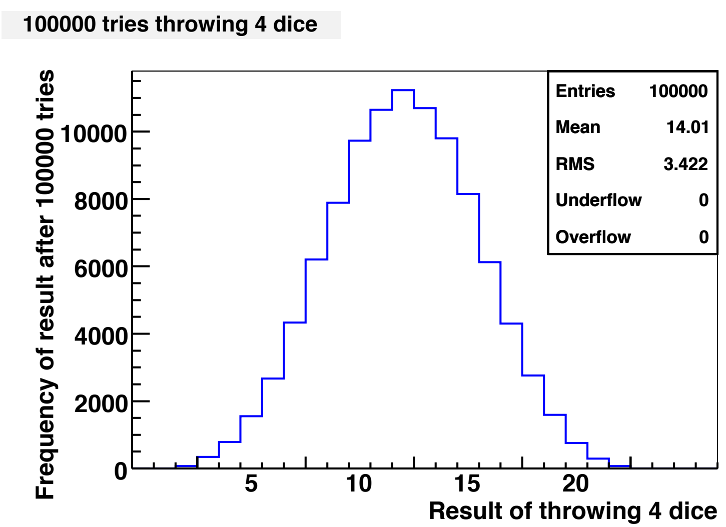
Rolling more and more dice

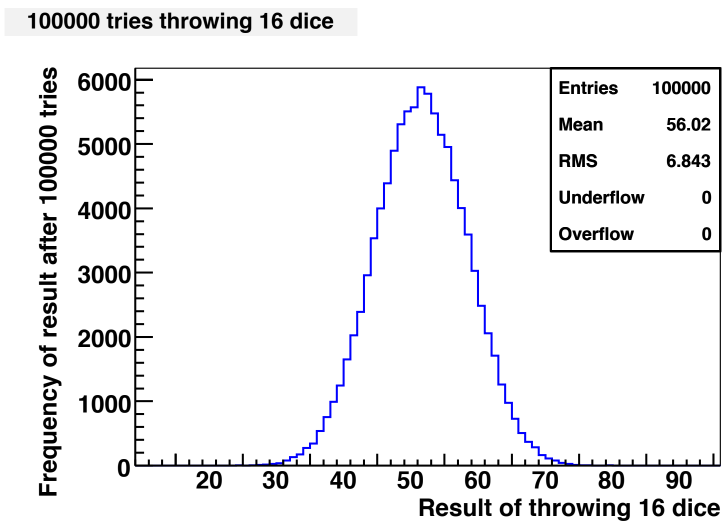




Rolling more and more dice

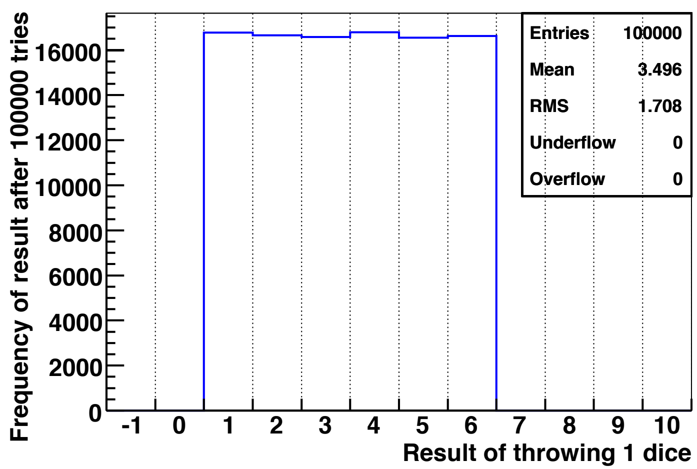




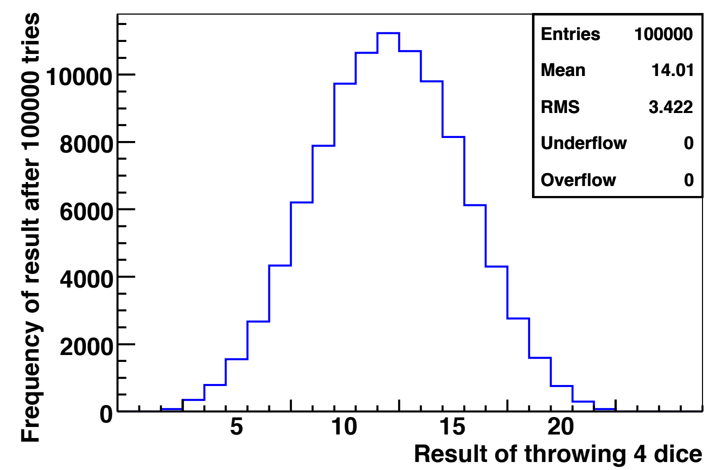


Rolling more and more dice

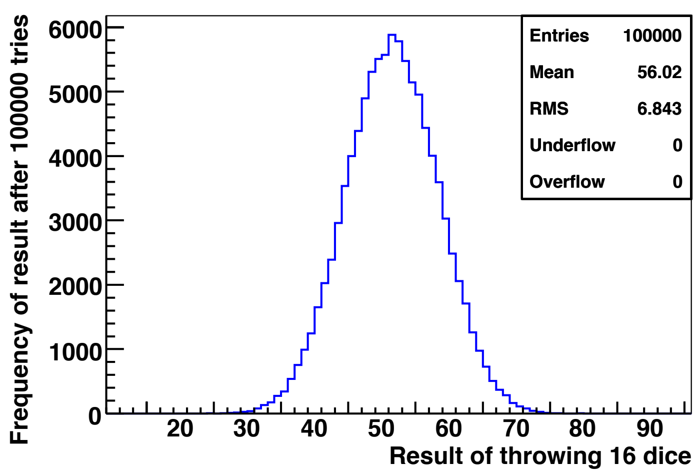
100000 tries throwing 1 dice



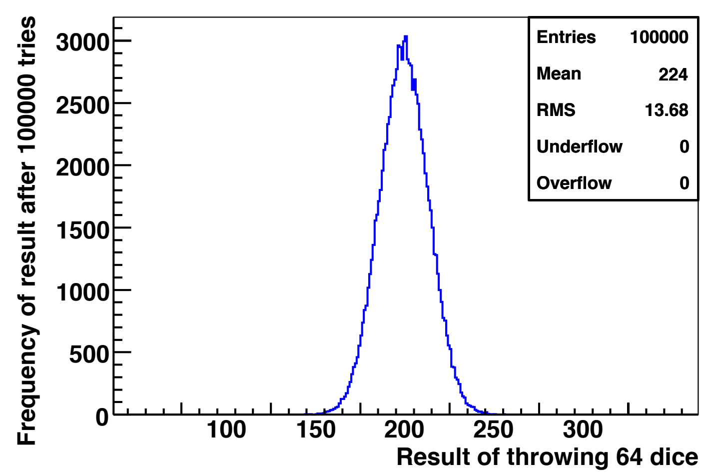
100000 tries throwing 4 dice



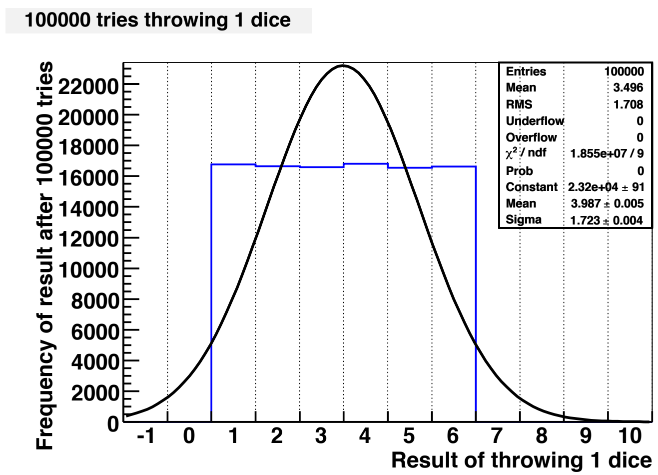
100000 tries throwing 16 dice

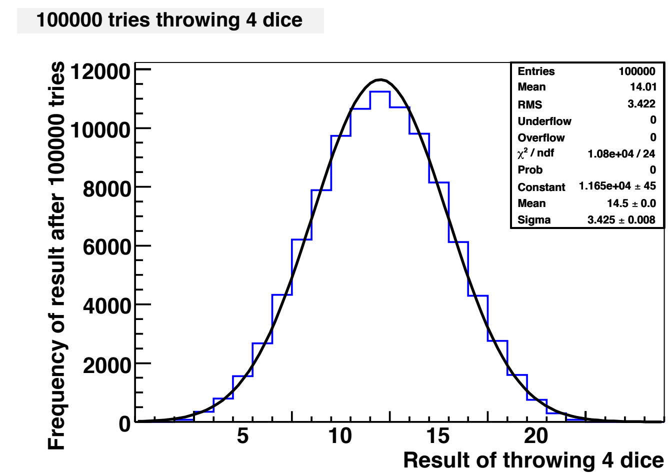


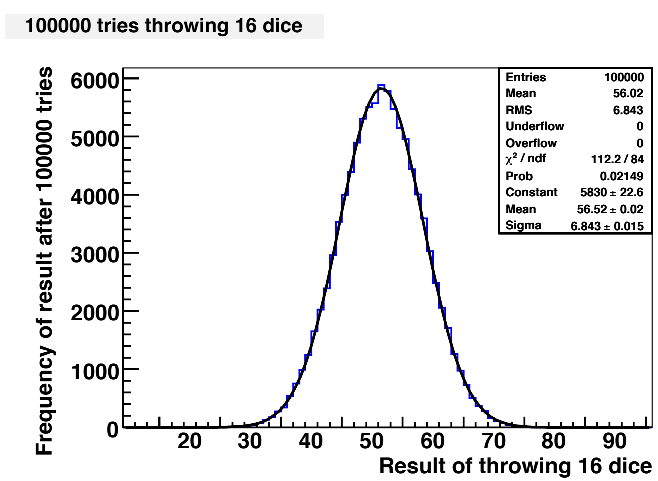
100000 tries throwing 64 dice

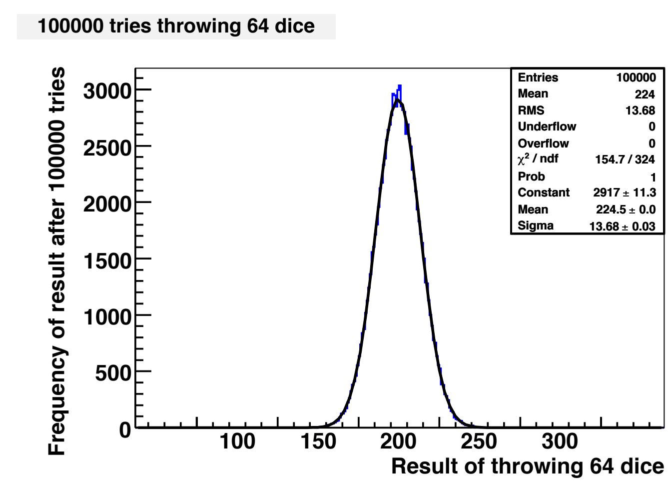


Comparing Gaussians to 1, 4, 16, 64-dice distributions

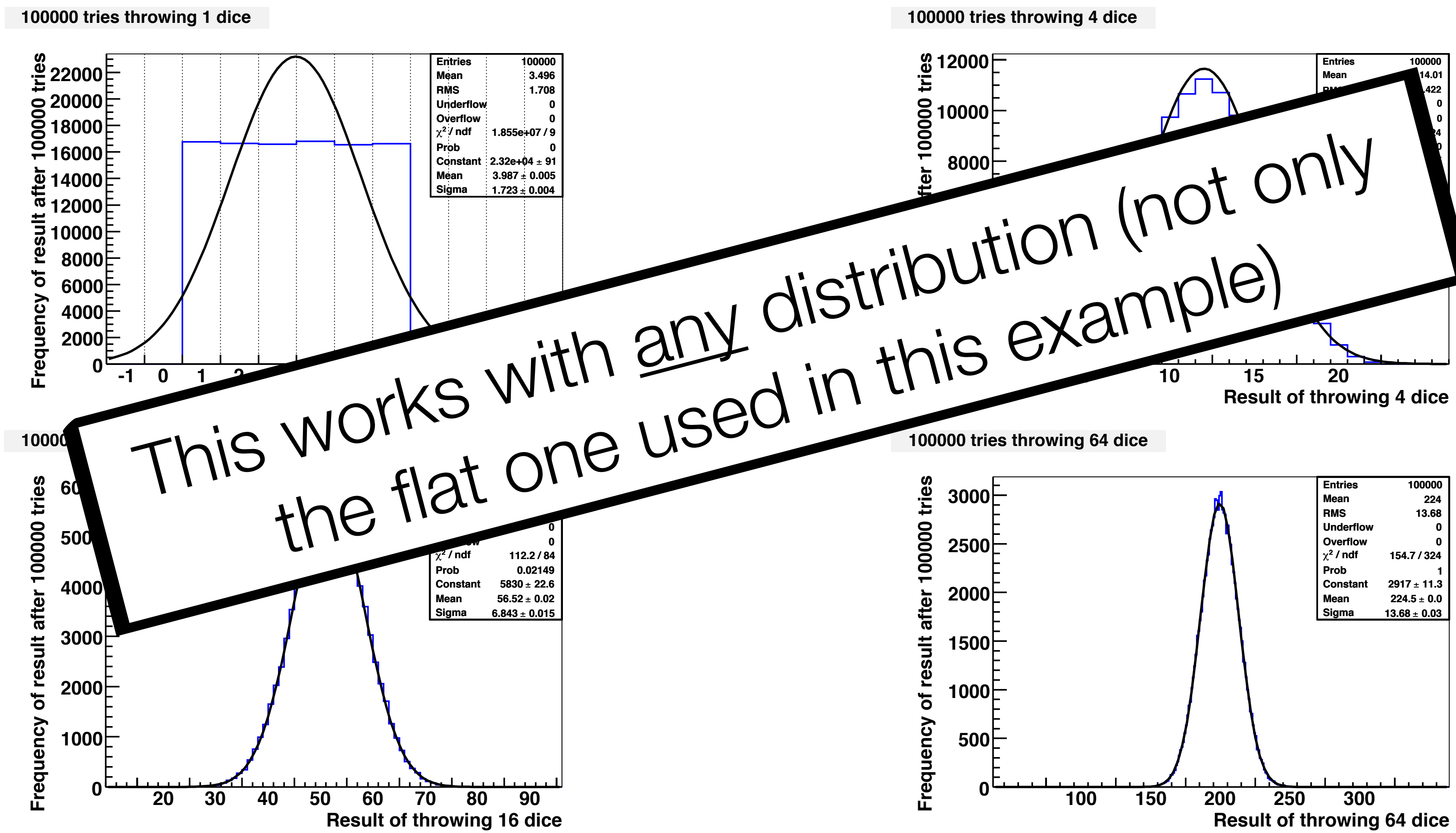








Comparing Gaussians to 1, 4, 16, 64-dice distributions



This works with <u>any</u> distribution (not only the flat one used in this example)

Gauss & me hanging out in Göttingen

Gauss on old money

bokeh serve jonas_singletoy.py

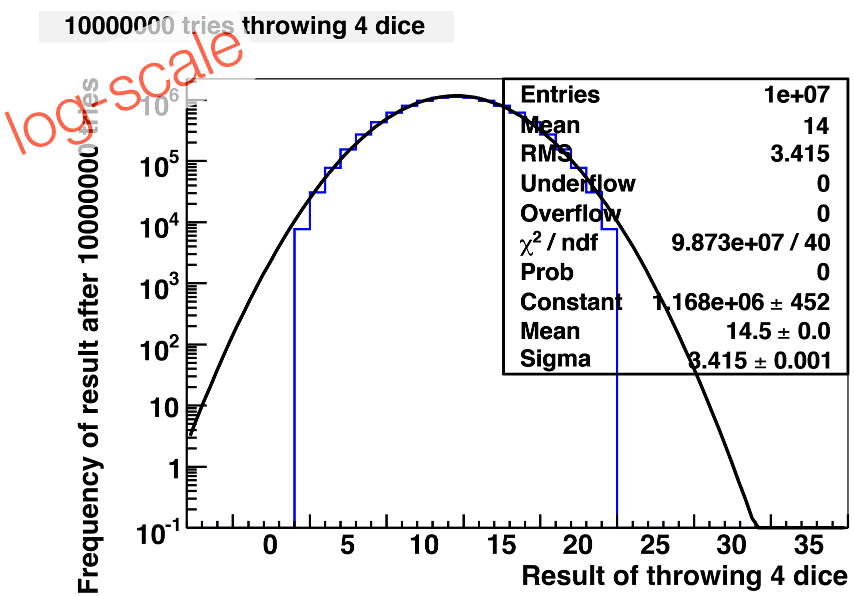
localhost:5006/jonas_singletoy

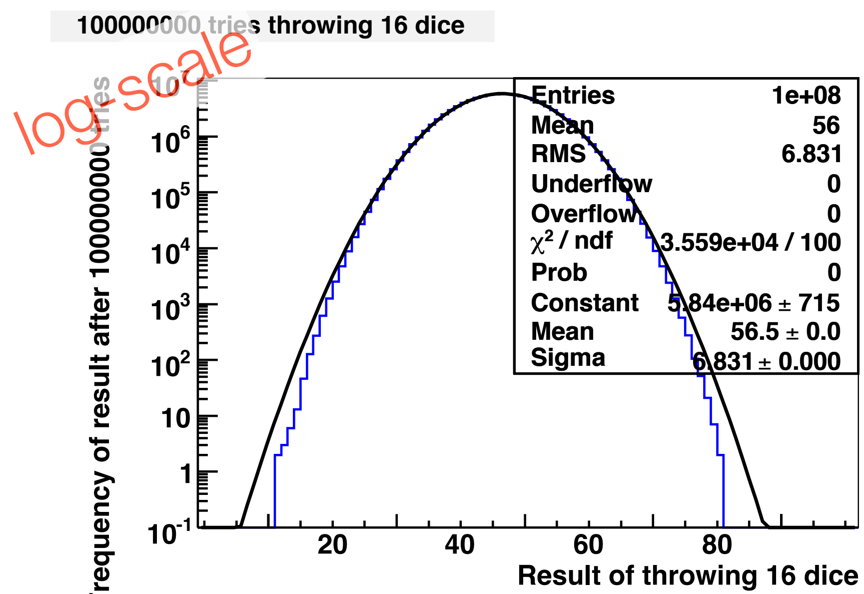
The Central Limit Theorem

- Take the sum X of N independent variables x_i
- Each x_i is taken from a distribution with mean $\langle x_i \rangle$ and variance $V_i = \sigma_i^2$.
- Then
    - X has an expectation value $\langle X \rangle = \Sigma \langle x_i \rangle$
    - X has a variance (the square of the standard deviation, $V=\sigma^2$) $V(X) = \Sigma V_i$.
    - The distribution of X becomes Gaussian as $N \rightarrow \infty$.

Variances add up! (Standard deviations don't)

Central Limit Theorem holds in the centre, not in the tails(!)





- **Central limit theorem ensures that within a few sigma of the mean, we get a good approximation to a Gaussian.**
- **Differences remain in the tails of the distribution (doesn't have to be fewer events, such as here, can also be more).**

Gaussians, errors, confidence

- **Within ±1σ: "1σ Confidence Level", or "68.27% Confidence level"**

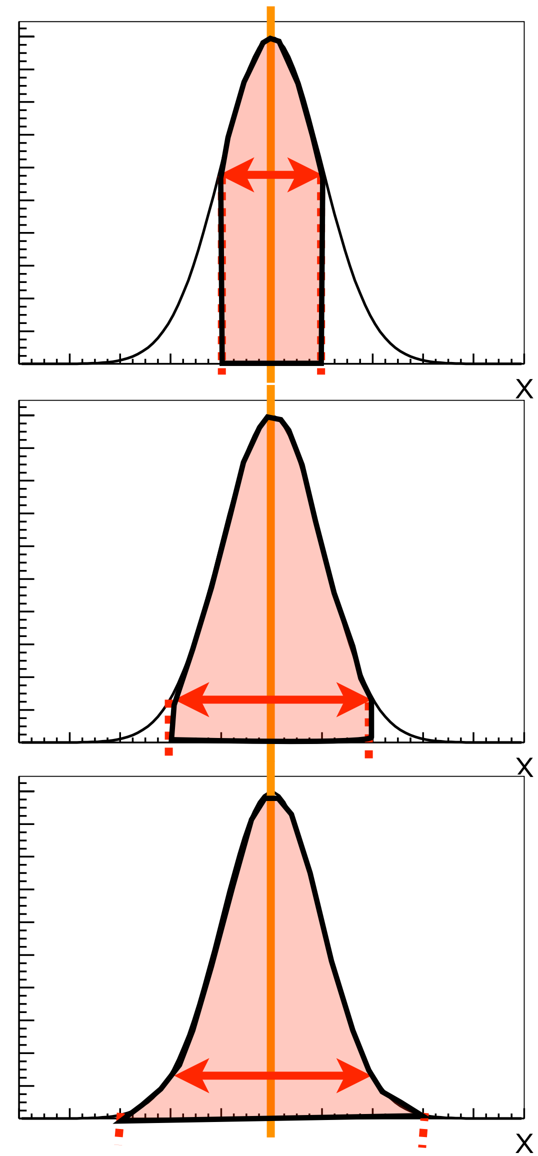
$$\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 68.27\%$$

- **Within ±2σ: "2σ CL" or "95.45% CL"**

$$\int_{-2}^{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 95.45\%$$

- **Within ±3σ: "3σ" or "99.73% CL"**

$$\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 99.73\%$$



Talking to Engineers

- **Physicists quote their errors as 1σ (Gaussian) confidence intervals.**
- **The probability that a result is outside the quoted error is 32%. About 1/3 of measurements should be outside the error bars. Results outside error bars are OK - it just shouldn't happen too often. And it shouldn't be too far: *P(outside μ±2σ) ~5%, P(outside μ±3σ) ~0.3%*)**
- **Engineers *guarantee* that the actual value is within mean ± tolerance.**

Cool Hand Luke (1967)

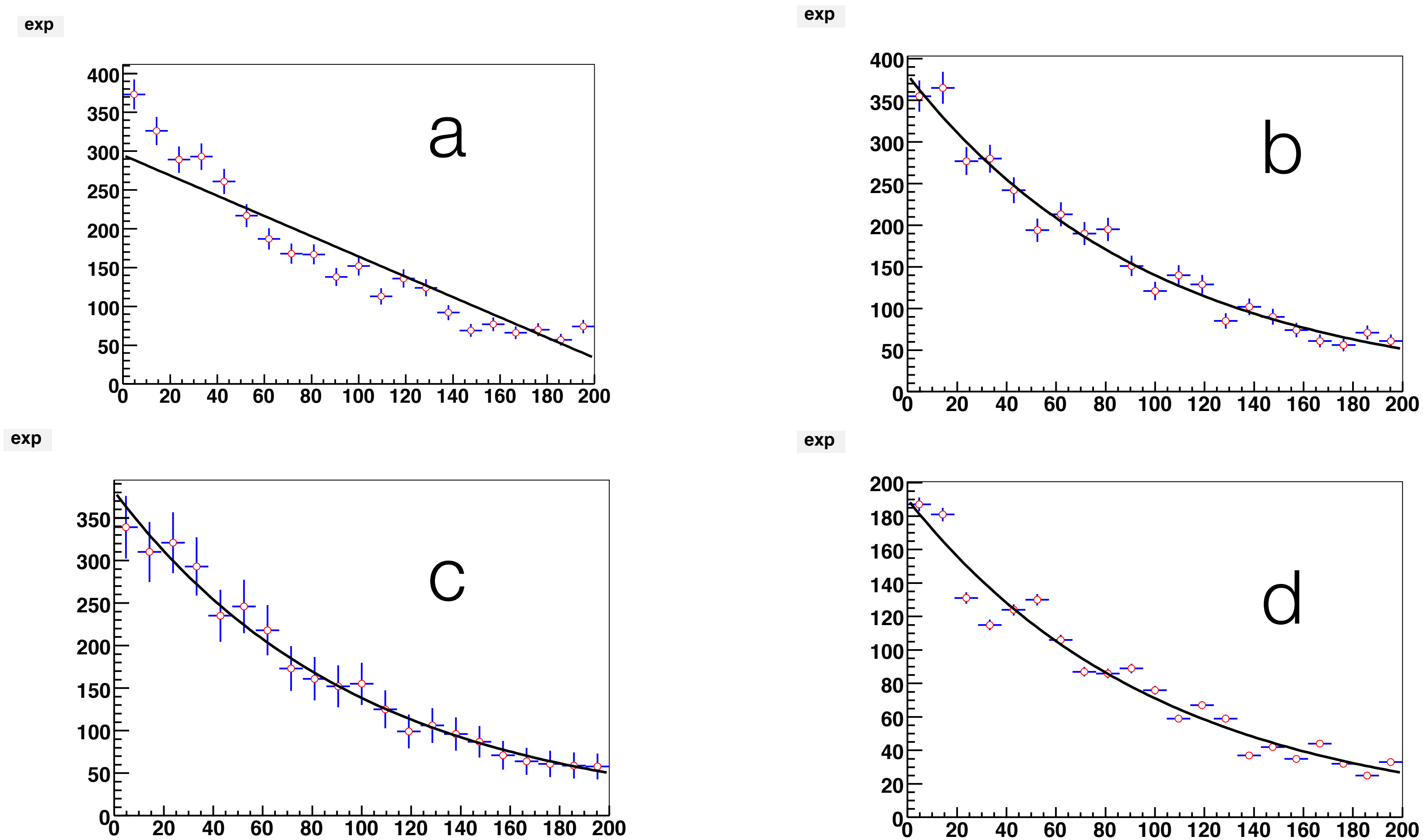
"What we've got here is...failure to communicate.

Some men you just can't reach."

https://tinyurl.com/TeshepProblems

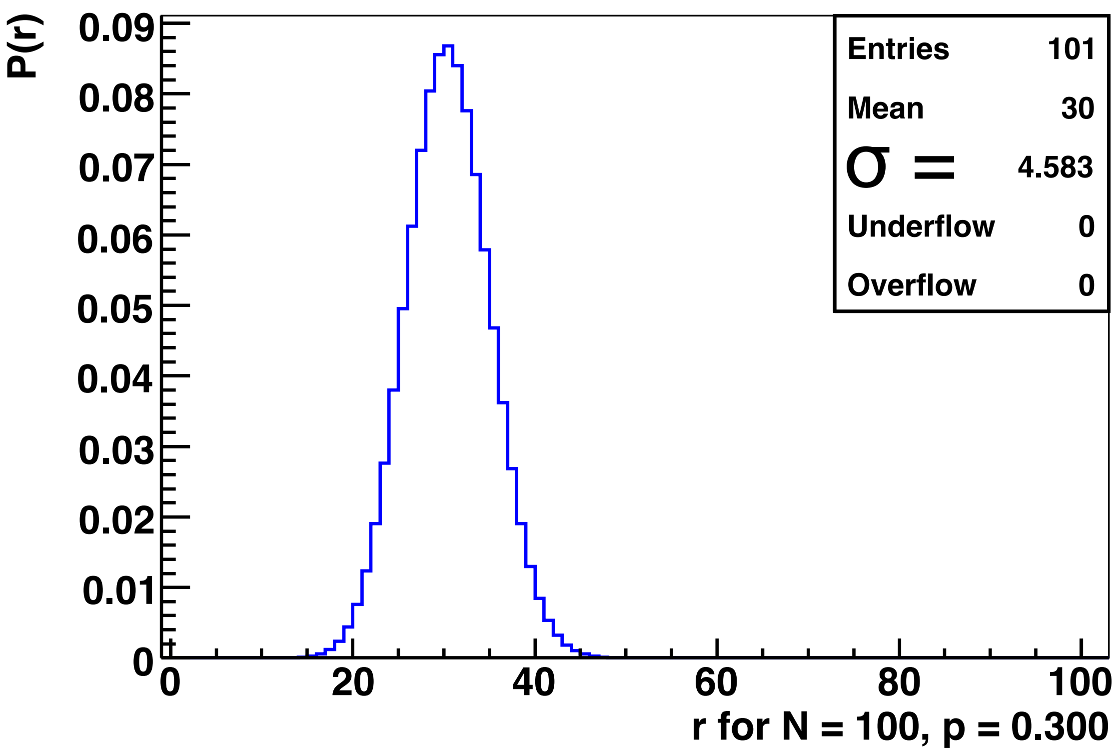
Which plot makes most sense?

What is the most plausible plot if the line represents theory, dots data distributed according to that theory, and the vertical lines are 1σ error bars.

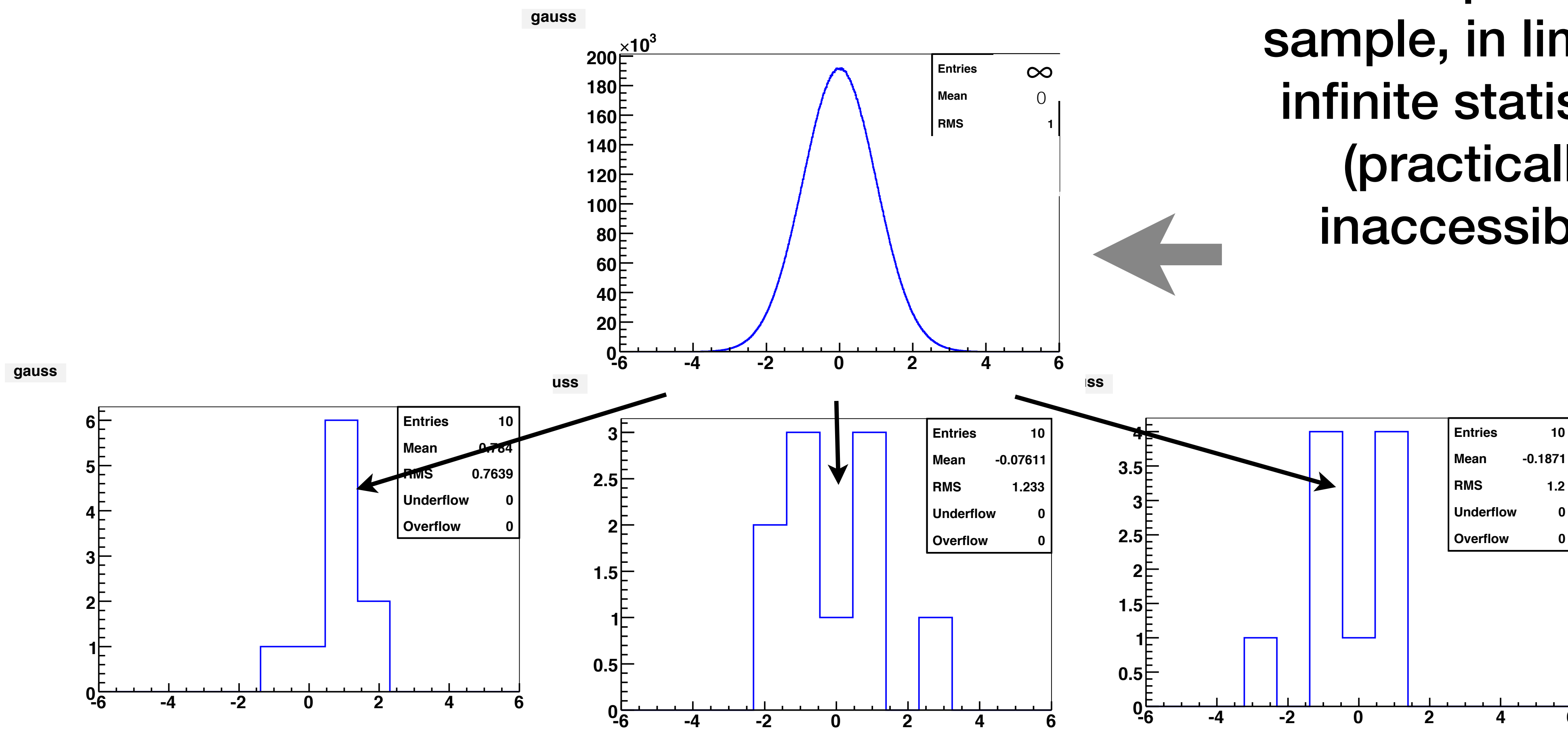


What's the uncertainty on the mean?

Theory with N = 100, p = 0.300



Idea of "ideal" parent sample



Ideal parent sample, in limit of infinite statistics (practically inaccessible)

Uncertainty on the mean: if I repeat the measurement with N data points again and again, and record each time the mean, what is the width/standard deviation of that distribution?

Central Limit theorem

- Take the sum X of N independent variables x_i [as in the case of the radioactive cats].
  - $\langle X \rangle = \Sigma \langle x_i \rangle$
  - Variance $V(X) = \Sigma V_i$.
  - Std dev. $\sigma_{\Sigma xi} = (\Sigma V_i)^{½}$
  - Gaussian as $N \to \infty$.

the 1st miracle of √N

Central Limit theorem

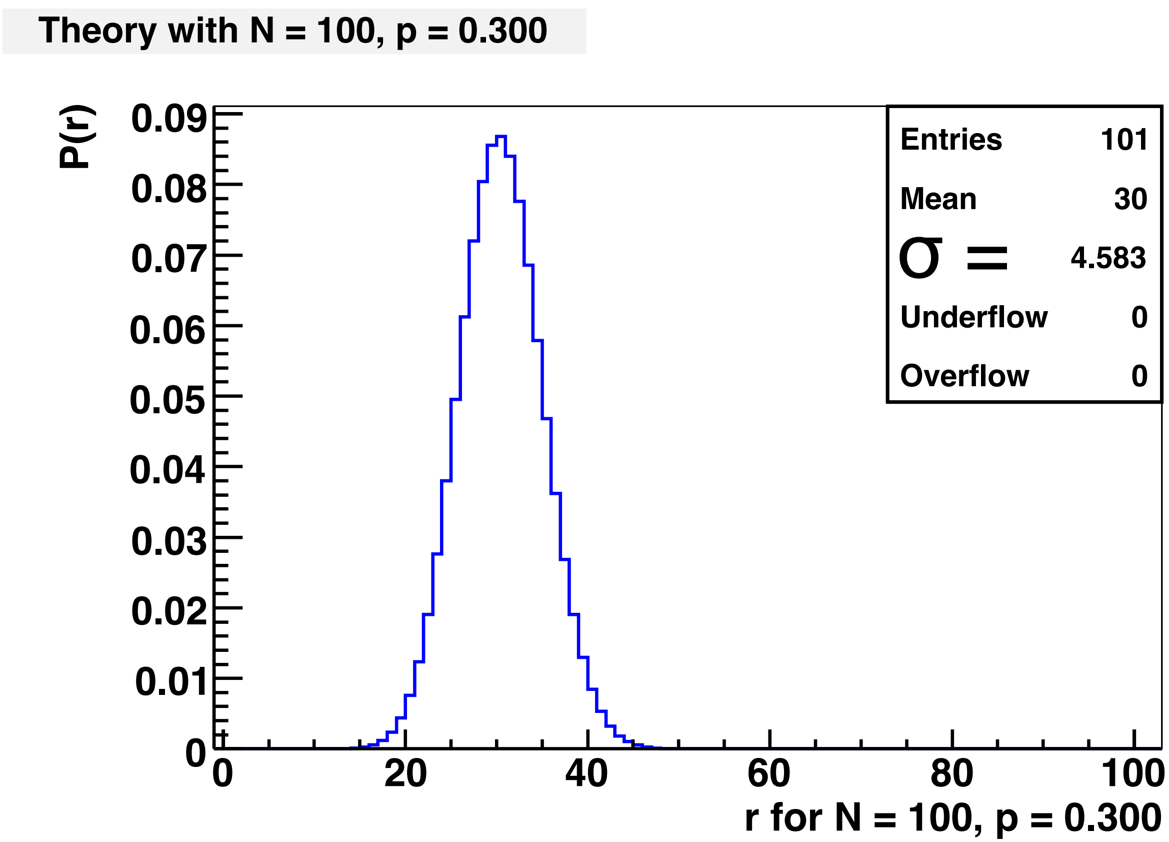
- Take the sum X of N independent variables x_i [as in the case of the radioactive cats].
  - $\langle X \rangle = \Sigma \langle x_i \rangle$
  - Variance $V(X) = \Sigma V_i$.
  - Std dev. $\sigma_{\Sigma xi} = (\Sigma V_i)^{½}$
  - Gaussian as $N \rightarrow \infty$.

- Take the average Y of N independent variables x_i: $Y = \Sigma x_i / N$.
  - $\langle Y \rangle = \Sigma \langle x_i \rangle / N$
  - $V(Y) = (\Sigma V_i)/N^2$
    if all V_i the same: $= V_i/N$
  - $\sigma_y = (\Sigma \sigma^2_i)^{½}/N$
    if all σ_i the same: $= \sigma_i/\sqrt{N}$

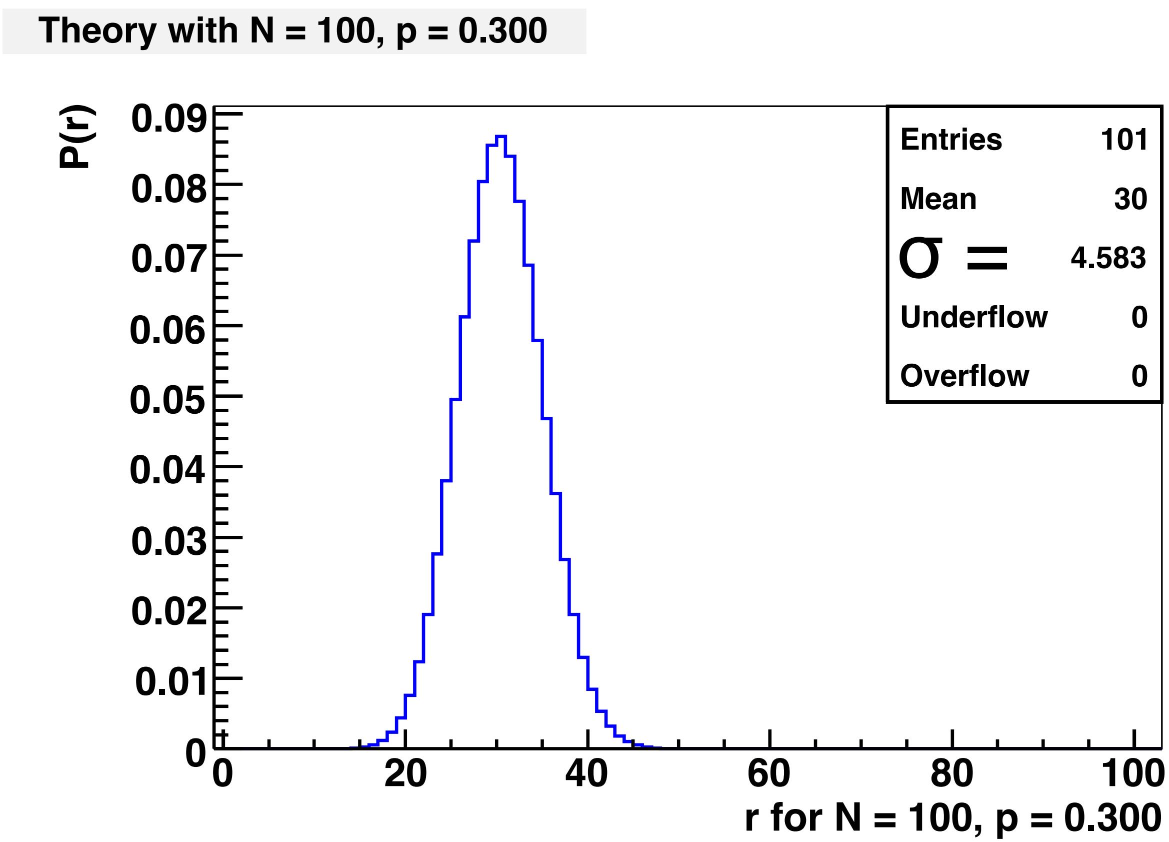
  the 1st miracle of $\sqrt{N}$

  - Gaussian as $N \rightarrow \infty$.

What's the uncertainty on the mean?

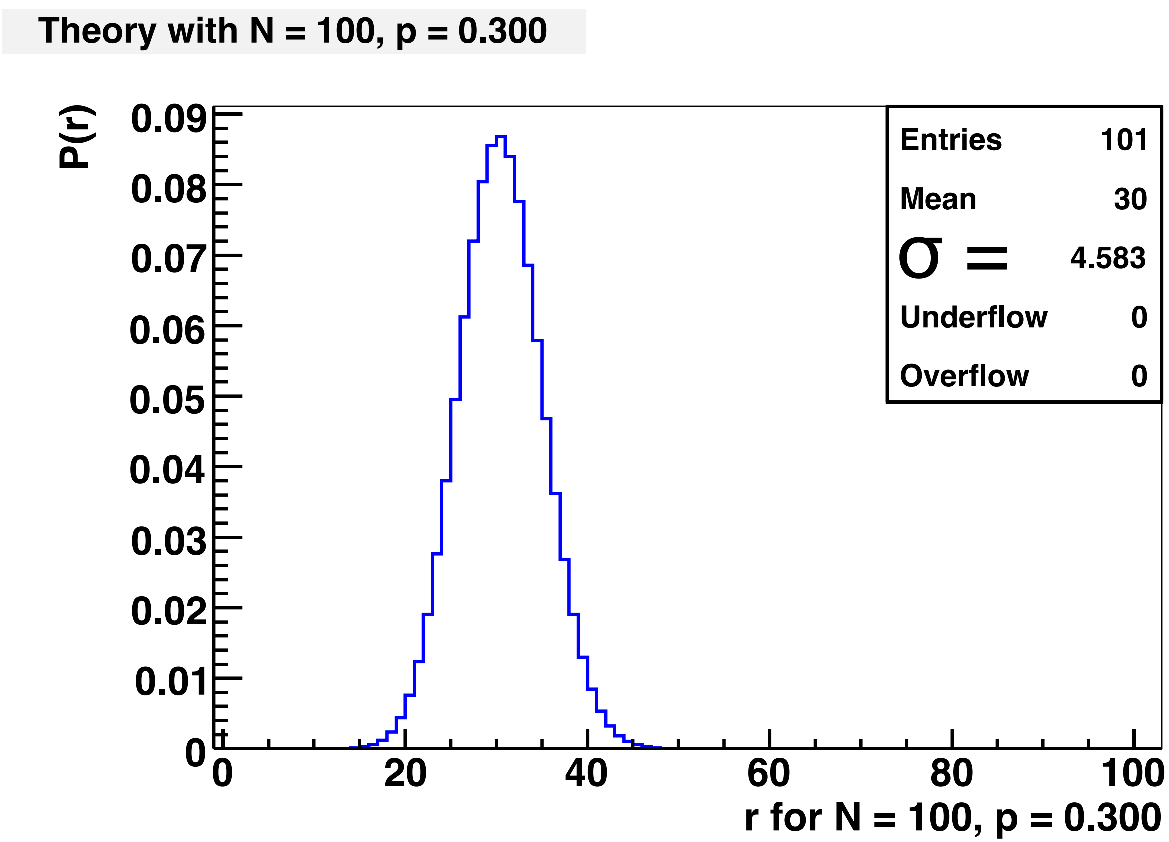


What's the uncertainty on the mean?



$$\sigma_{mean} = \sigma/\sqrt{N}$$

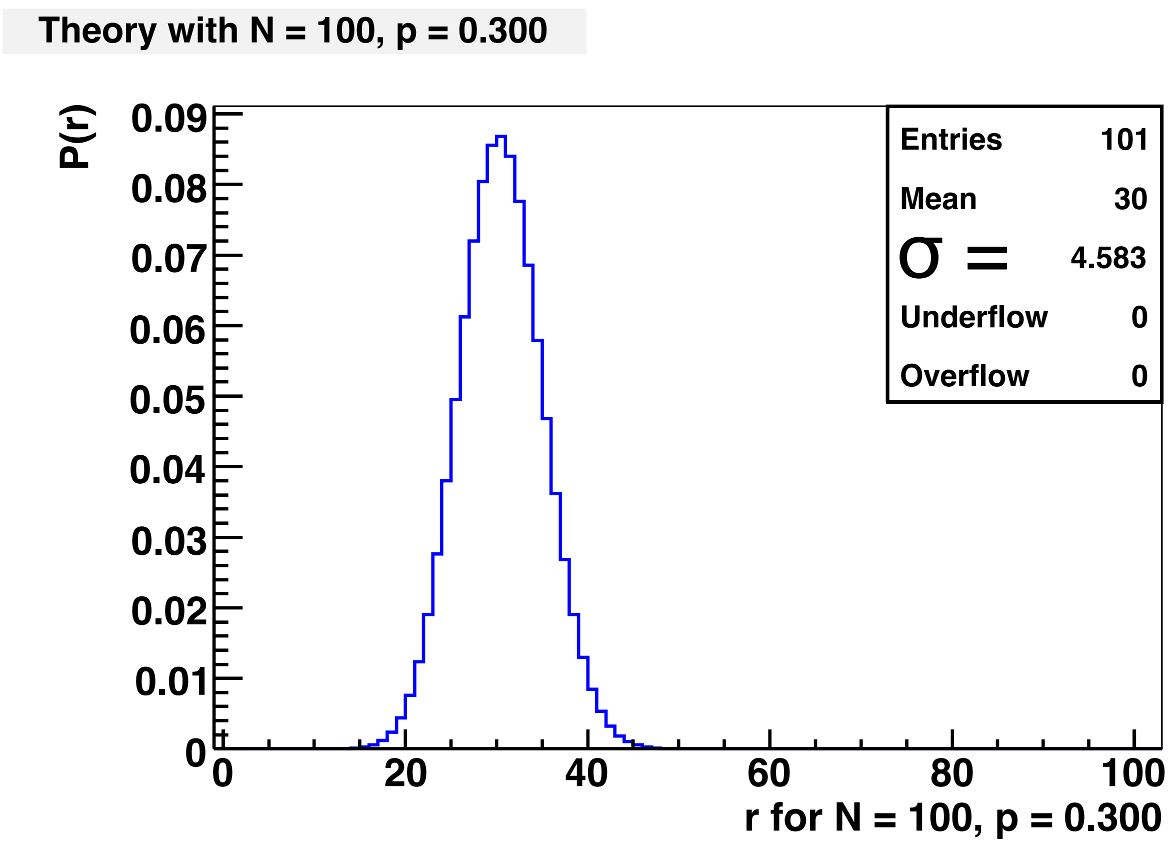
What's the uncertainty on the mean?



$\sigma_{mean} = \sigma/\sqrt{N}$

N=101

What's the uncertainty on the mean?



$\sigma_{mean} = \sigma/\sqrt{N}$

N=101

$\sigma_{mean} = 0.46$

Further important theoretical distributions...

- **In the next few slides I'll introduce the binomial and the Poisson distribution - you will meet them a lot in your particle physics research!**

The Binomial Distribution

- **Fixed number of "trials" (measurements), *N***
- **Two possible outcomes, usually termed "*Success*" and "*Failure*" (but can be *green* and *orange*, or *>5* and *<=5*, or anything else mutually exclusive).**
- **The probability for a success in a single trial is *p*.**
- **Question: What is the probability to get *r* successes and *(N–r)* failures in *N* trials:**

(whiteboard)

$P(r; N, p) = ?$

The Binomial Distribution

number of "successes"

probability of failure in single trial

probability of success in single trial

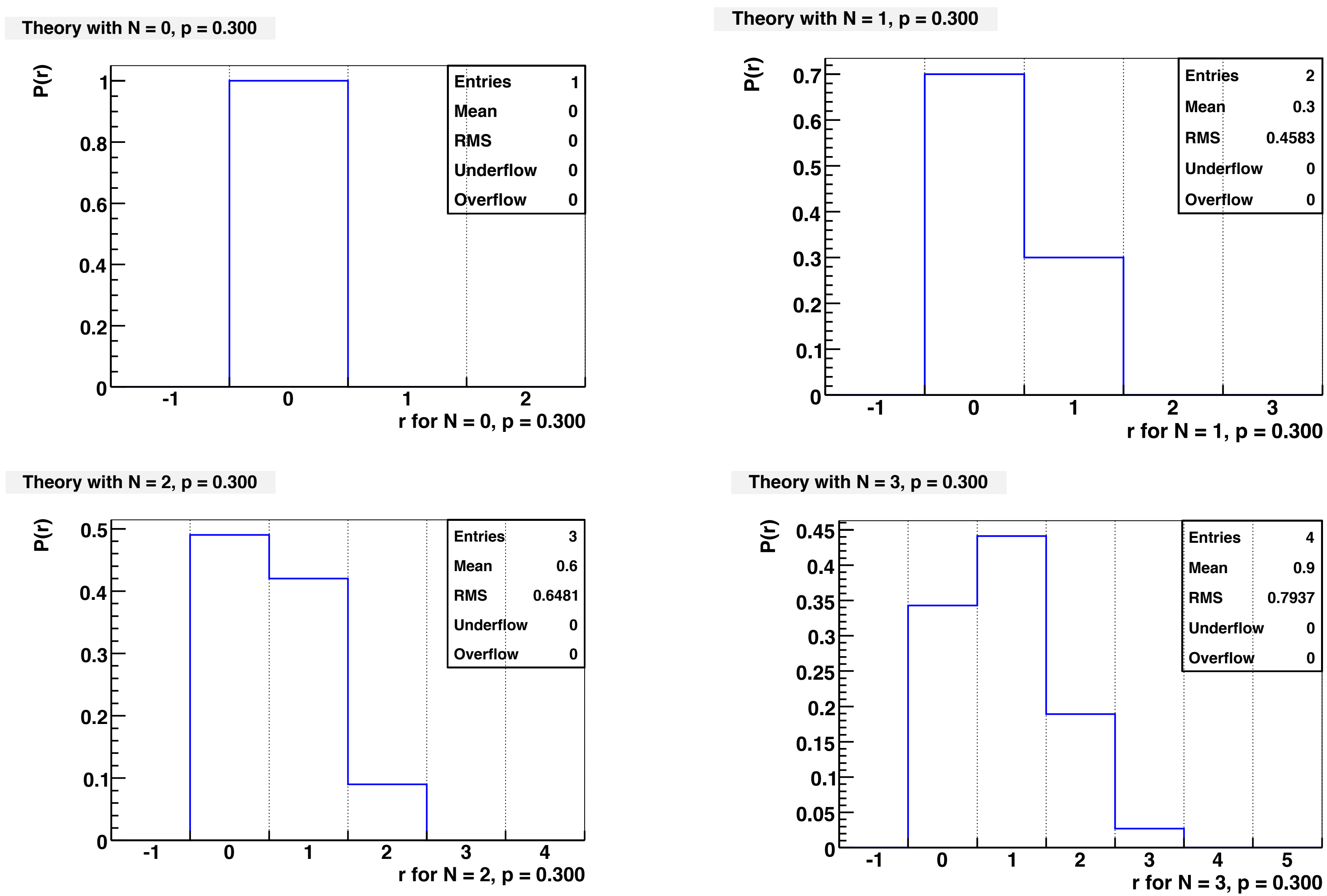
number of "failures"

$$P(r; N, p) = p^r (1-p)^{N-r} \binom{N}{r}$$

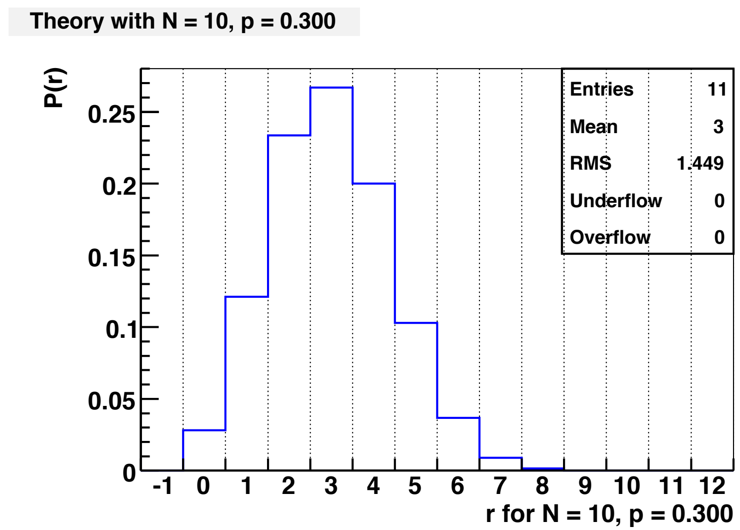
$$= p^r (1-p)^{N-r} \frac{N!}{r!\,(N-r)!}$$

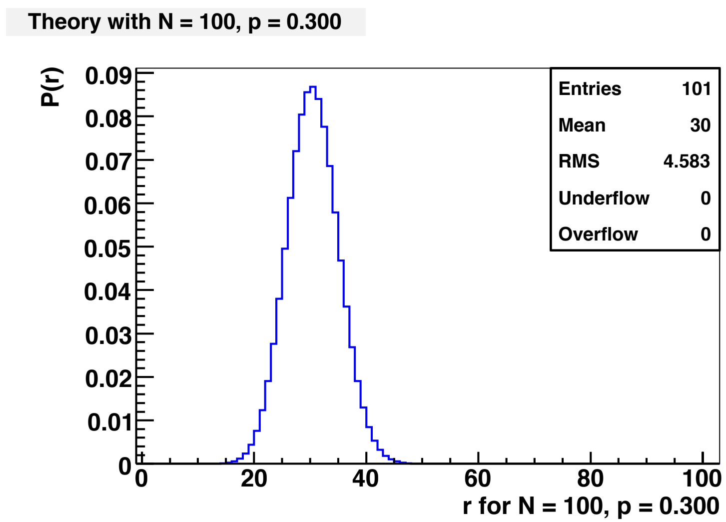
number of different sequences in which one can have r successes and N – r failures

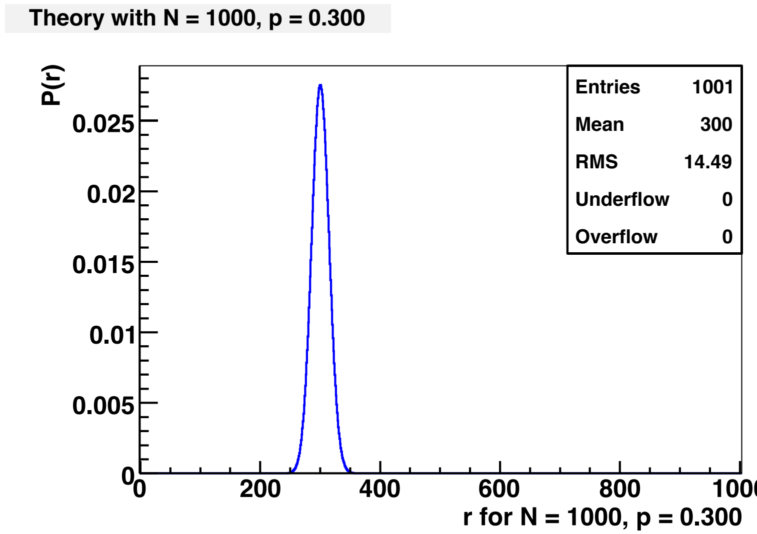
Binomi Examples

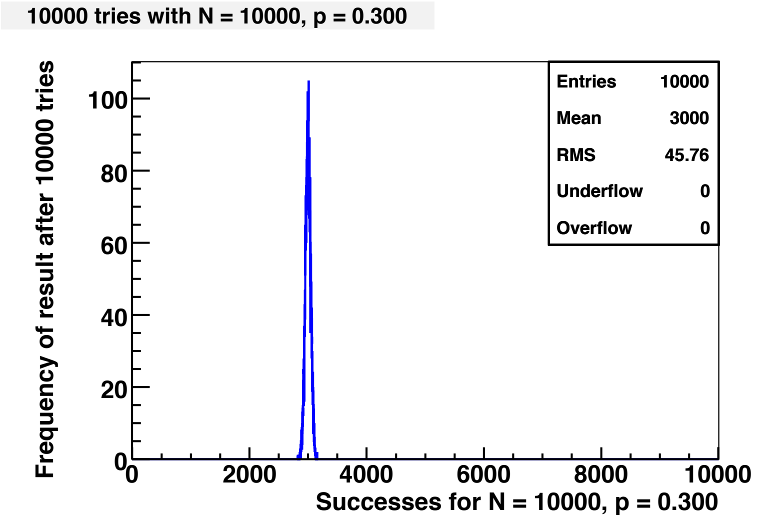


Binomi Examples

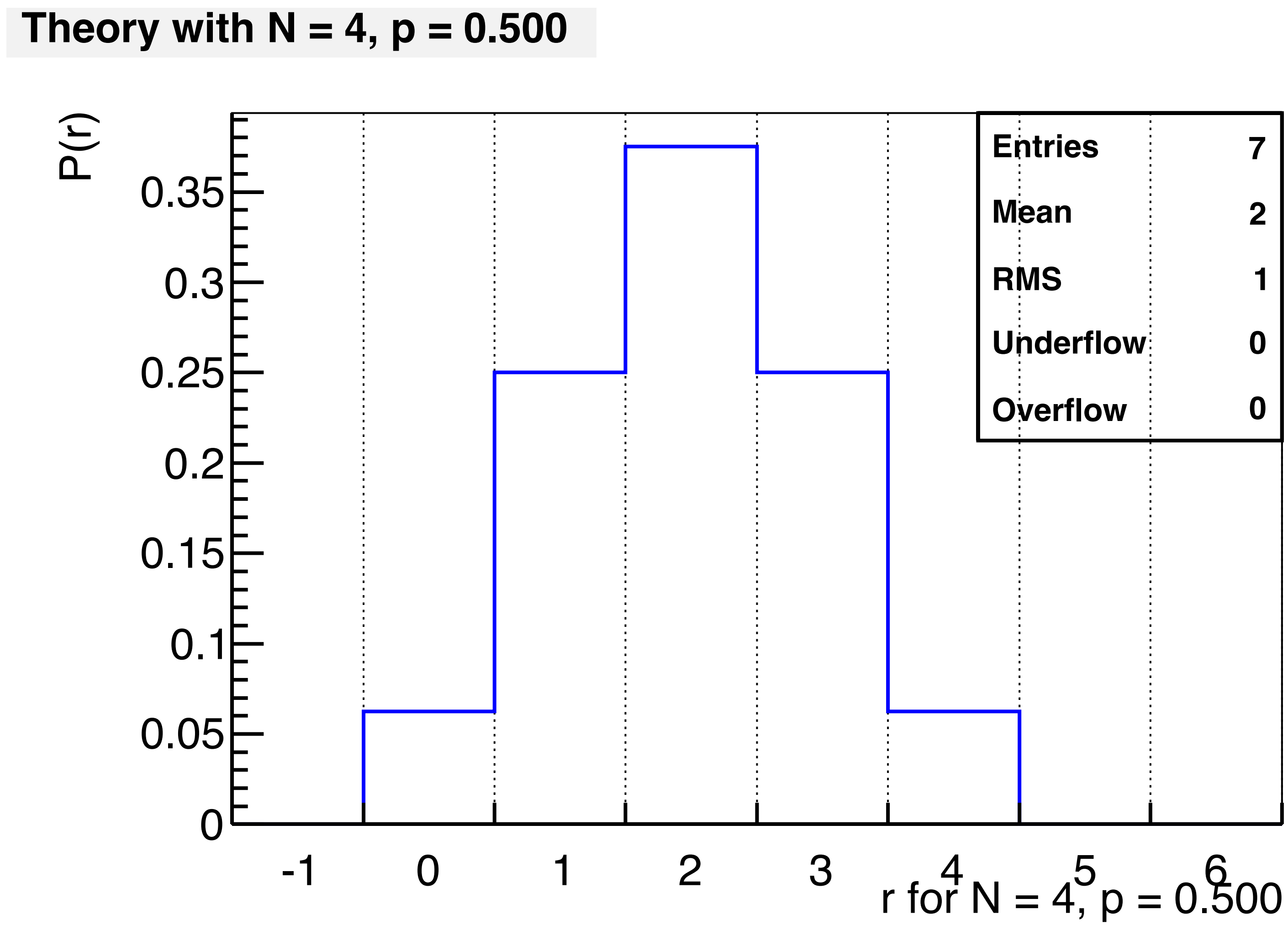








Binomi Examples



Example: Lightning

- The Poisson distribution describes sharp events in a continuum.
- There is still a fixed outcome (flash), but not a fixed number of trials. It doesn't make sense to ask how many non-flashes we saw.
- But we can ask how many flashes we expect to see in a given time interval. Or clicks in a Geiger counter.

Lightning striking the Eiffel Tower, June 3, 1902, at 9:20 P.M. This is one of the earliest photographs of lightning in an urban setting In:"Thunder and Lightning", Camille Flammarion, translated by Walter Mostyn Published in 1906.

Binomial → Poisson

- **We'll start with our trusted Binomial Distribution.**

$$P(r; N, p) = p^r (1-p)^{N-r} \binom{N}{r}$$
$$= p^r (1-p)^{N-r} \frac{N!}{r!\ (N-r)!}$$

- **How can we modify it such that it describes the number of flashes in a continuum?**

Binomial → Poisson

- **Strategy:**
  - Divide the time over which we observe the sky and count flashes into small intervals.
  - If the intervals are small enough, we do have a binomial distribution - each interval is a trial and can have two outcomes, success (flash) or failure (no flash).
  - Important: The intervals must be *so small that we can get at most one flash* - otherwise we would have more than two possible outcomes (0, 1, 2, ,... flashes), and the binomial distribution would not work.

- ...derivation on whiteboard, if time permits

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

$$
\begin{aligned}
P(r;N,p) &= p^r(1-p)^{N-r}\frac{N!}{r!(N-r)!} \\
P(r;N,\lambda) &= \frac{\lambda^r}{N^r}\left(1-\frac{\lambda}{p}\right)^{N-r}\frac{N!}{r!(N-r)!} \\
&= \frac{\lambda^r}{r!}\left(1-\frac{\lambda}{N}\right)^{N-r}\frac{N!}{N^r(N-r)!} \\
&= \frac{\lambda^r}{r!}\left(1-\frac{\lambda}{N}\right)^{N-r}\frac{N(N-1)(N-2)\cdots(N-r+1)}{N^r} \\
&= \frac{\lambda^r}{r!}\left(1-\frac{\lambda}{N}\right)^{N}\left(1-\frac{\lambda}{N}\right)^{-r}\frac{N^r+\alpha_1 N^{r-1}+\alpha_2 N^{r-2}\cdots}{N^r} \\
\lim_{N\to\infty} P(r;N,\lambda) &= \frac{\lambda^r}{r!}e^{\lambda}\,(1)^{-r}\left(1+\alpha\frac{1}{N}+\alpha_2\frac{1}{N^2}+\ldots\right) \\
&= \frac{\lambda^r}{r!}e^{\lambda}\,(1)^{-r}
\end{aligned}
$$

```
P(r; N, p) &= p^r (1-p)^{N-r} \frac{N!}{r! (N-r)!}
\\
P(r; N, \lambda) &=
\frac{\lambda^r}{N^r} \left(1-\frac{\lambda}{p}\right)^{N-r} \frac{N!}{r! (N-r)!}
\\
&= \frac{\lambda^r}{r!}
\left(1-\frac{\lambda}{N}\right)^{N-r}
\frac{N!}{N^r (N-r)!}
\\
&= \frac{\lambda^r}{r!}
\left(1-\frac{\lambda}{N}\right)^{N-r}
\frac{N(N-1)(N-2)\cdots (N-r+1)}{N^r}
\\
&= \frac{\lambda^r}{r!}
\left(1-\frac{\lambda}{N}\right)^{N}
\left(1-\frac{\lambda}{N}\right)^{-r}
\frac{N^r + \alpha_1 N^{r-1} + \alpha_2 N^{r-2} \cdots}{N^r}
\\
\lim_{N\to\infty} P(r; N, \lambda)
&= \frac{\lambda^r}{r!}
     e^{\lambda} \left( 1 \right)^{-r}
 \left( 1 + \alpha \frac{1}{N} + \alpha_2 \frac{1}{N^2} + \ldots \right)
\\
& = \frac{\lambda^r}{r!}
     e^{\lambda} \left( 1 \right)^{-r}
```

Poisson Summary $P(r;\lambda) = e^{-\lambda}\frac{\lambda^r}{r!}$

- Describes cases where we do not have a fixed number of trials, but discrete events in a continuum.

- It has only one single parameter - the expected mean number of events, λ.

$\langle r \rangle = \lambda$

$\sigma = \sqrt{\lambda}$

- The probability to see r events, given an expected mean of λ, is:

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

Poisson Summary $P(r;\lambda) = e^{-\lambda}\frac{\lambda^r}{r!}$

- **Describes cases where we do not have a fixed number of trials, but discrete events in a continuum.**
- **It has only one single parameter - the expected mean number of events, λ.**

$\langle r \rangle = \lambda$

$\sigma = \sqrt{\lambda}$ the 2nd miracle of √N.

If I expect N events, the uncertainty on this is √N, and the relative uncertainty is √N/N = 1/√N.

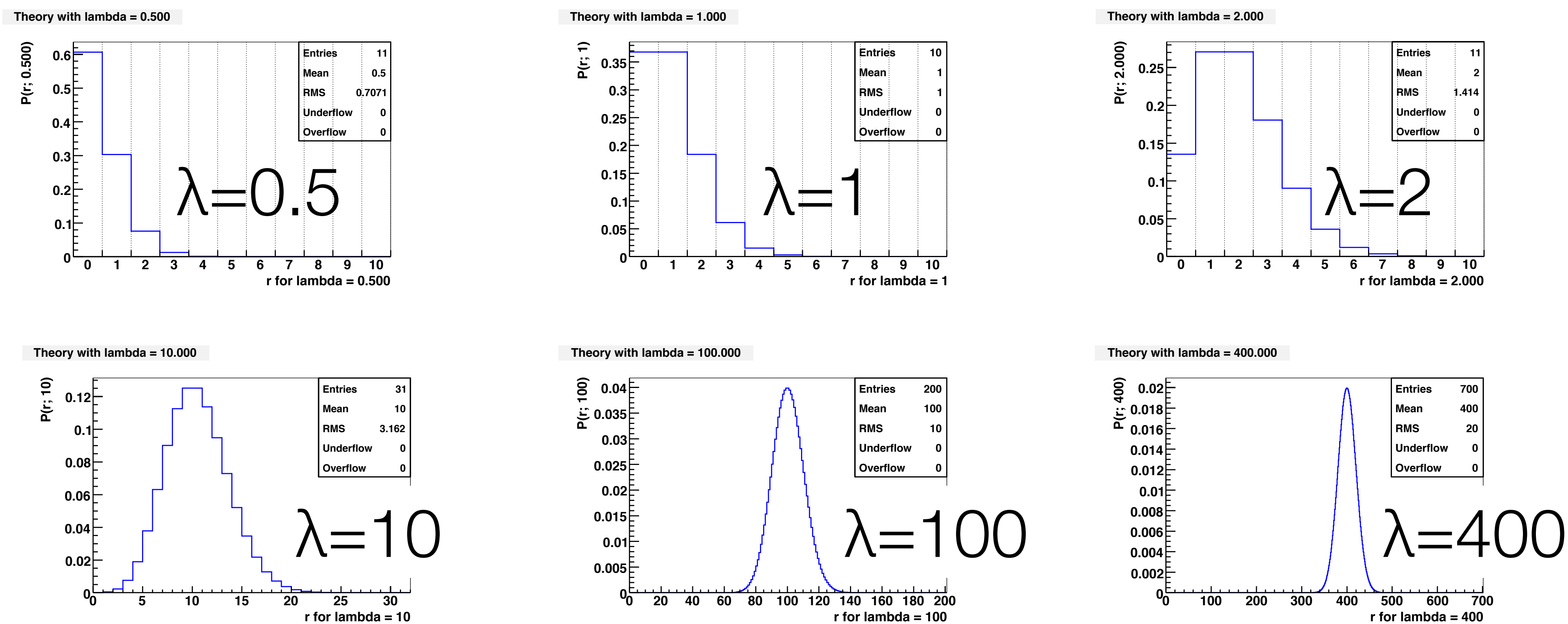
- **The probability to see *r* events, given an expected mean of *λ*, is:**

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

Binomial → Poisson

- ... our derivation (if we did it) implies that the Poisson distribution with $\lambda = Np$ is a decent approximation of the Binomial distribution in cases where p is small and N is large.

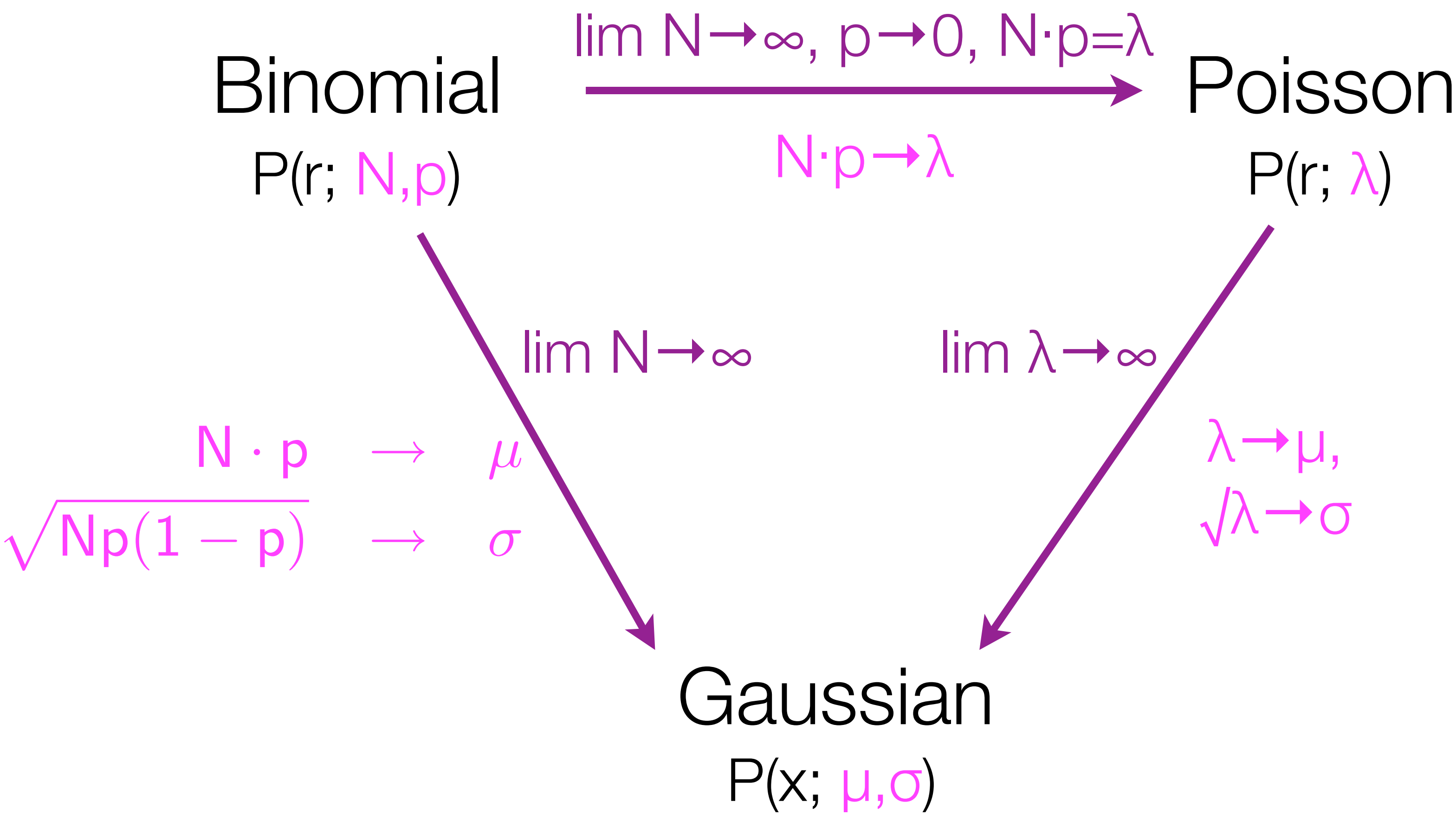
Poisson → Gaussian



Trinity

$$P\left(r; N, p\right) = p^{r}\left(1-p\right)^{N-r}\binom{N}{r}$$

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



$$g(x;\mu,\sigma = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Homework: Which distribution?

a) The number of flashes of lightening within on hour of a thunderstorm.
b) The number of Higgs events at the LHC in a year of running.
c) The number of students per hundred carrying the H1F1*virus.
d) Weight of individual A4 pieces of paper in a notebook
e) The number of sand grains in 1kg of sand.

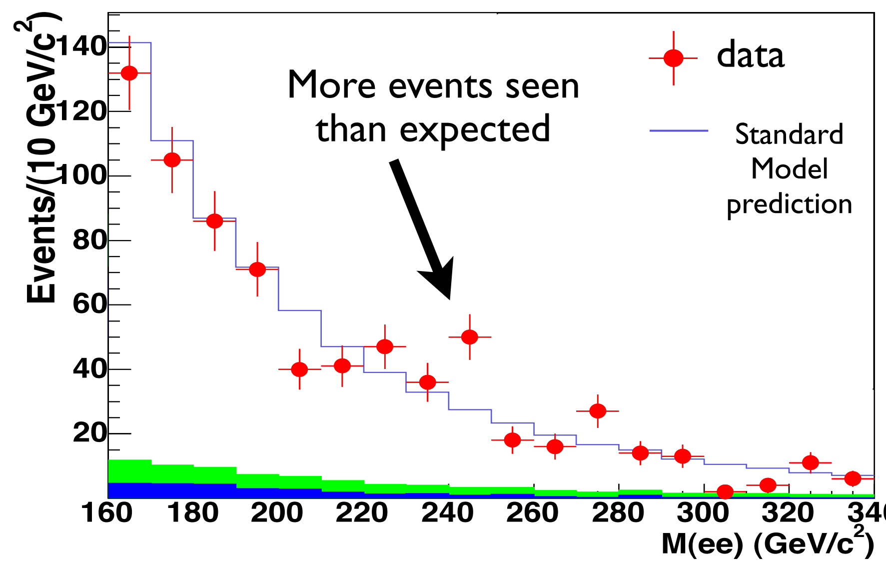
* H1F1 gives you bird flue

https://tinyurl.com/TeshepProblems

More Homework - calculate significances

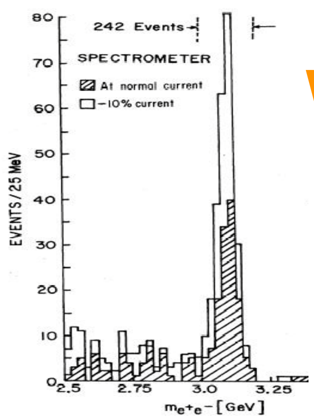
- Estimate the significance of this observation:
  - Step 1: calculate the probability so see an upward fluctuation this big or bigger in the Standard Model, in this one bin
  - Step 2: take into account that they looked in 84 bins (tricky!)
- You should get a fairly small number. Why, do you think, have you not read in the news about the discovery of the Z' at CDF?

Z' search at CDF



- In the bin with the arrow, we expect 28 events without the Z'
- See 48 events.

Roadmap



What do I see?

Describing Data

What do I expect?

Probability and probability distributions, Probability density functions

Central Limit Theorem

Is what I see compatible with what I expect?

Discoveries
Confidence Levels
Hypothesis testing
Fitting

Monte Carlo simulation

Fitting

Lifetime fit

- I have a decay time distribution that I want to describe with an exponential decay distribution:

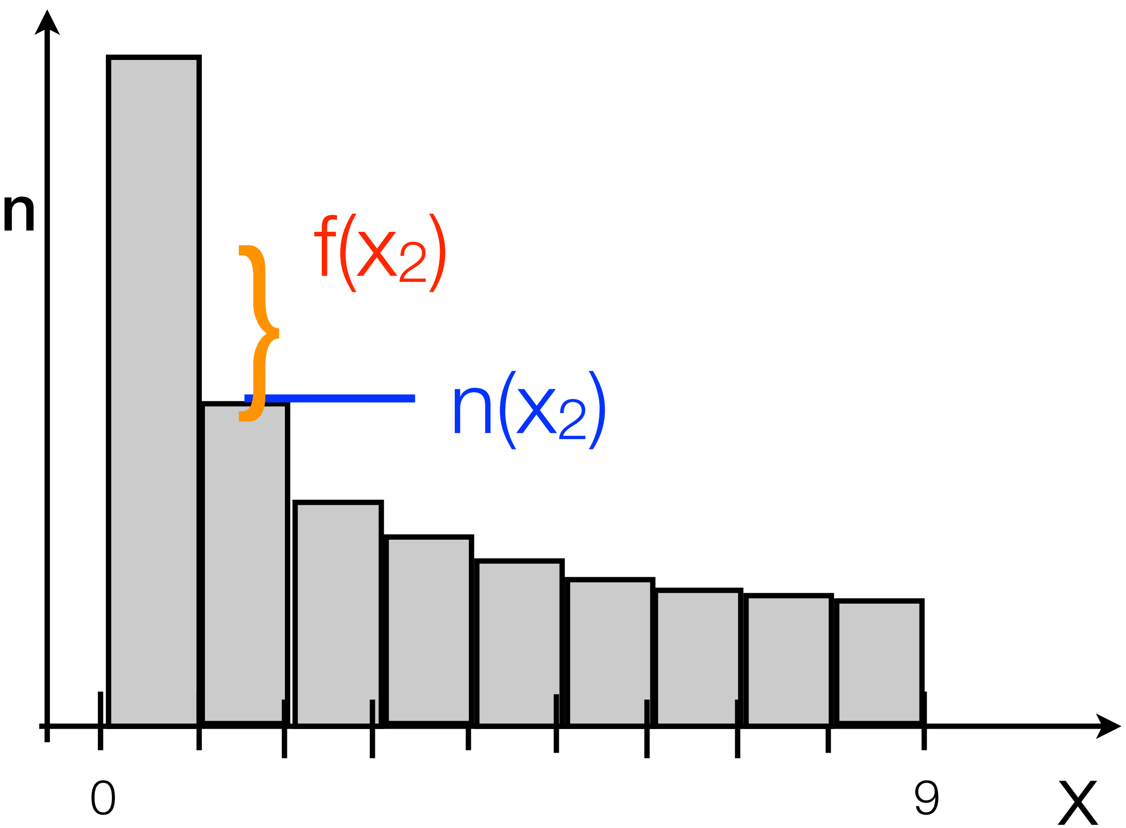
$$P(t) = \frac{1}{\tau} e^{-t/\tau}$$

- Question 1: What is the mean lifetime τ?

- Question 2: Did I pick the right function - are my data really described by an exponential decay?

χ^2 Fitting

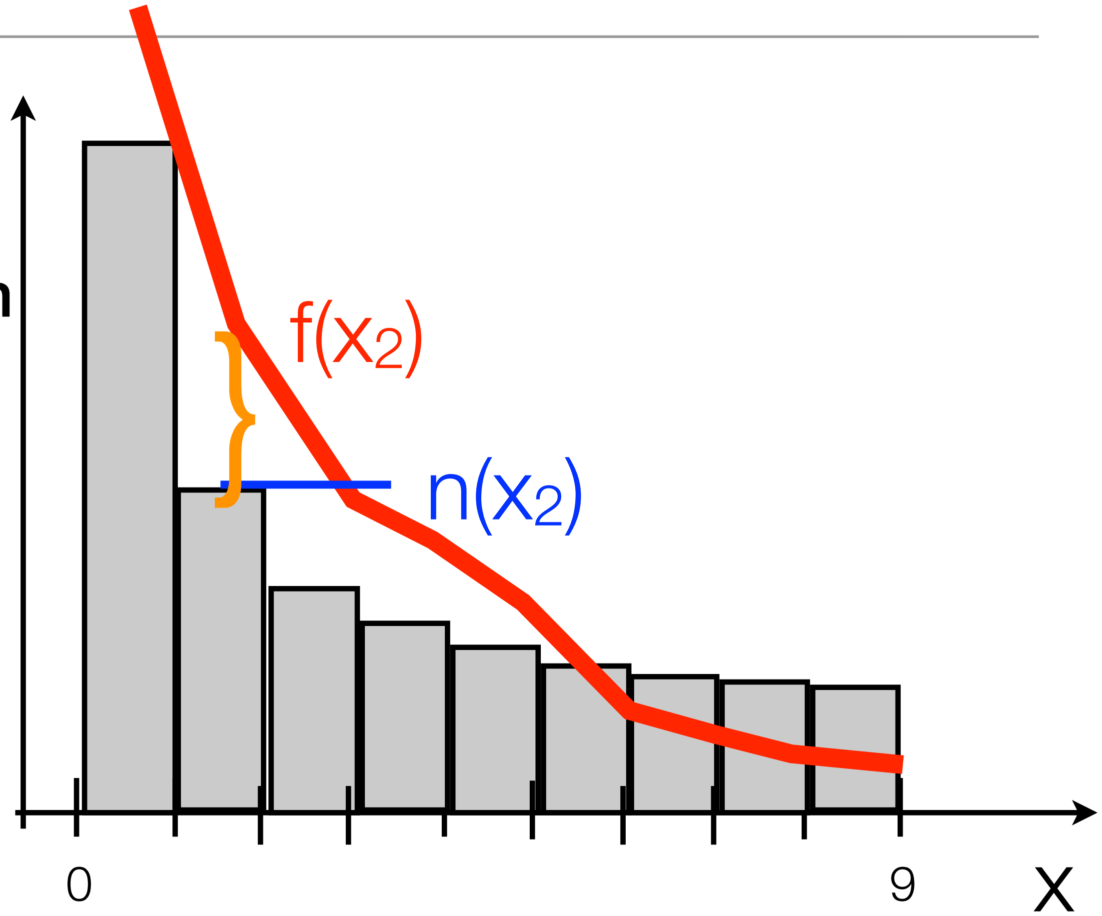
- **Use for binned data**
- **Minimise distance between data and function that describes data.**



usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

χ^2 Fitting

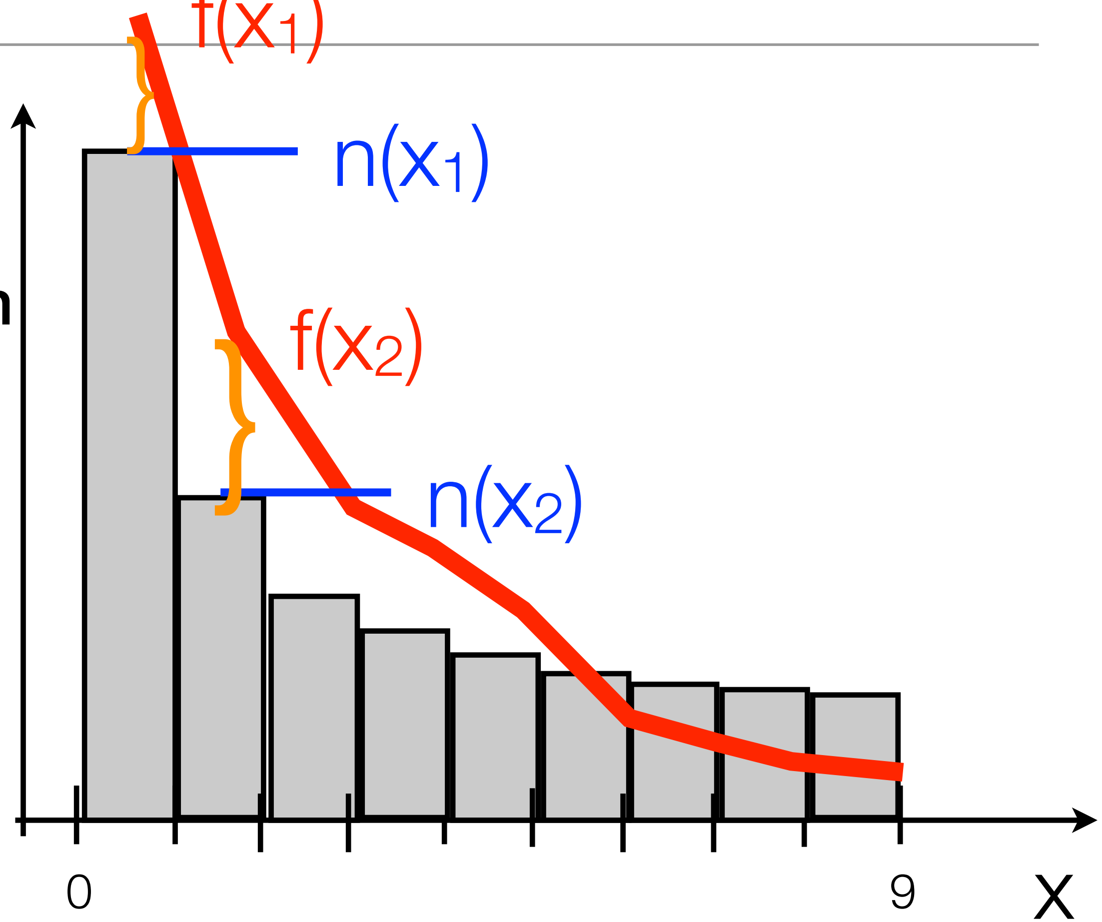
- **Use for binned data**
- **Minimise distance between data and function that describes data.**



usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

χ^2 Fitting

- **Use for binned data**
- **Minimise distance between data and function that describes data.**

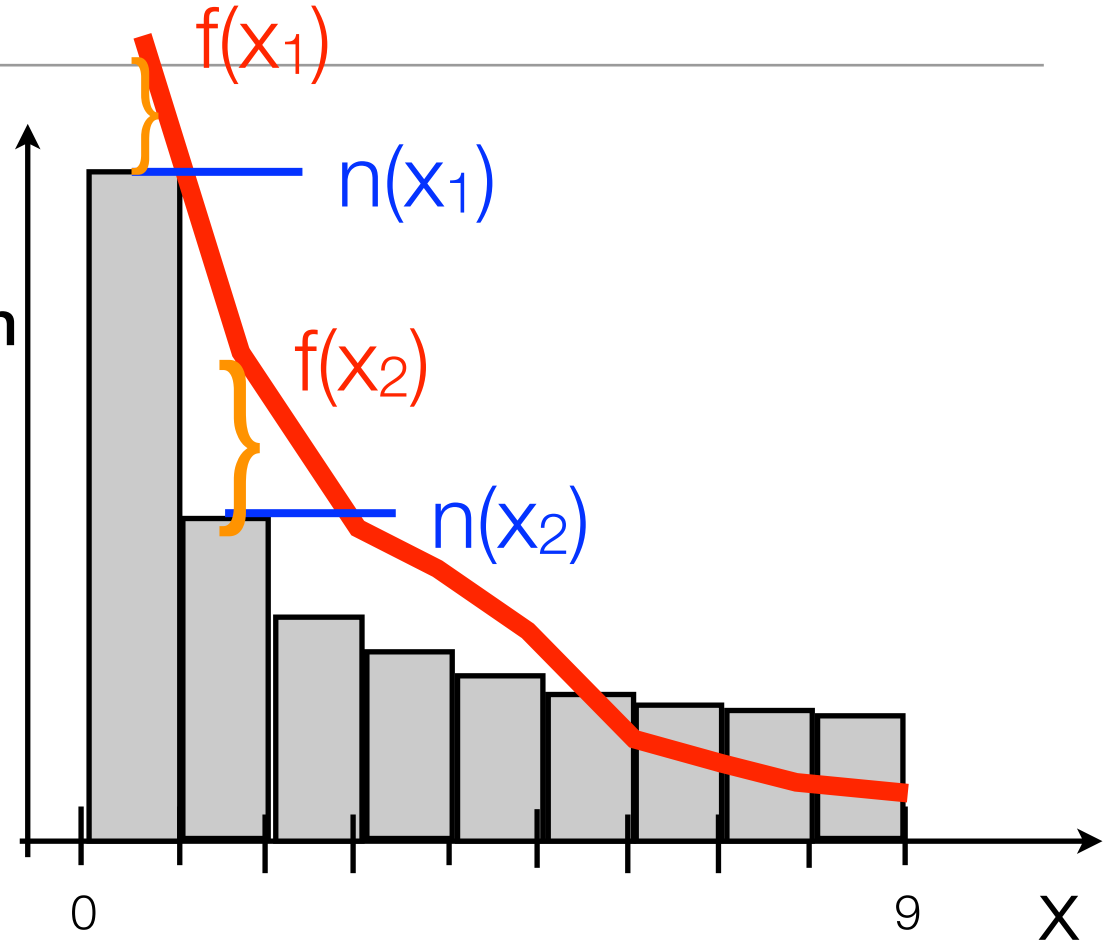


usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

χ^2 Fitting

- **Use for binned data**
- **Minimise distance between data and function that describes data.**
- **Possible definition:**

$d^2 = \Sigma(n(x_i) - f(x_i))^2$



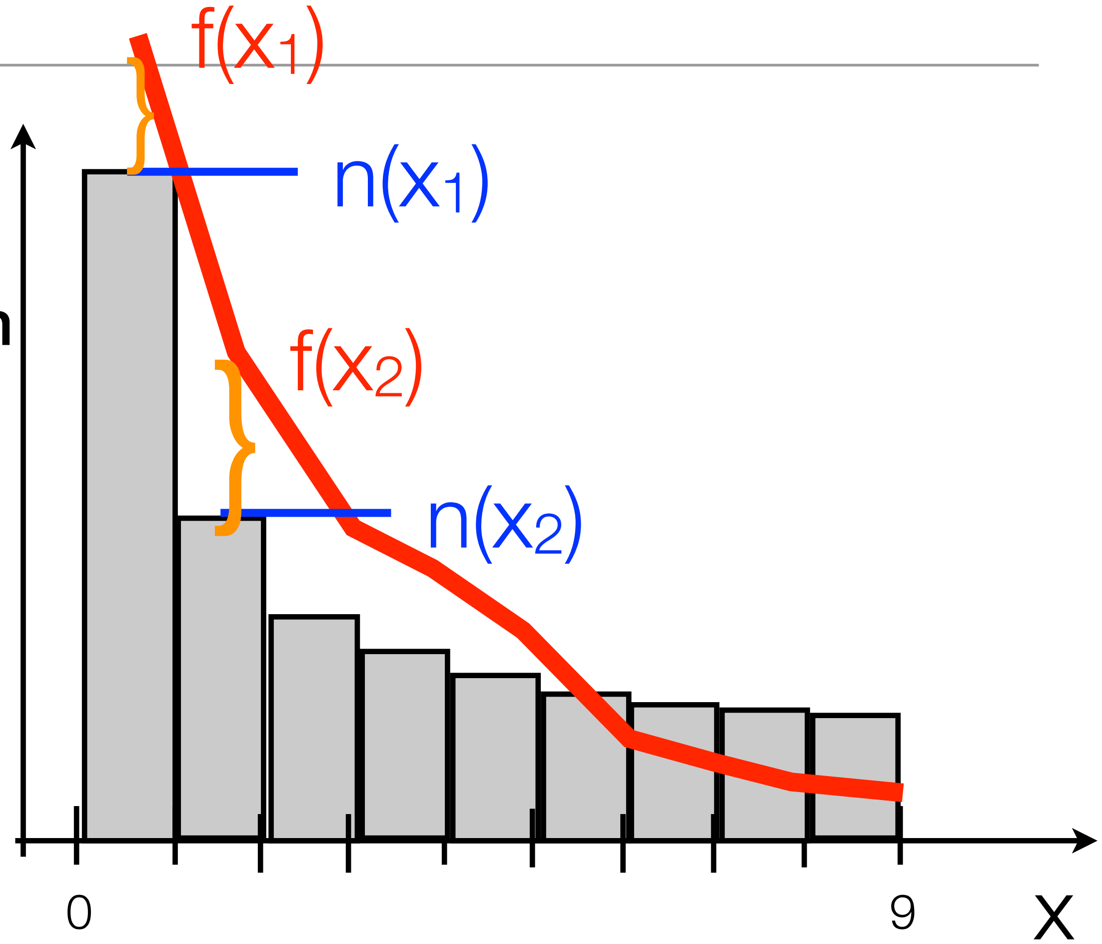
usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

χ^2 Fitting

- **Use for binned data**
- **Minimise distance between data and function that describes data.**
- **Possible definition:**

  $d^2 = \Sigma(n(x_i) - f(x_i))^2$

- **Better: Weight by error**



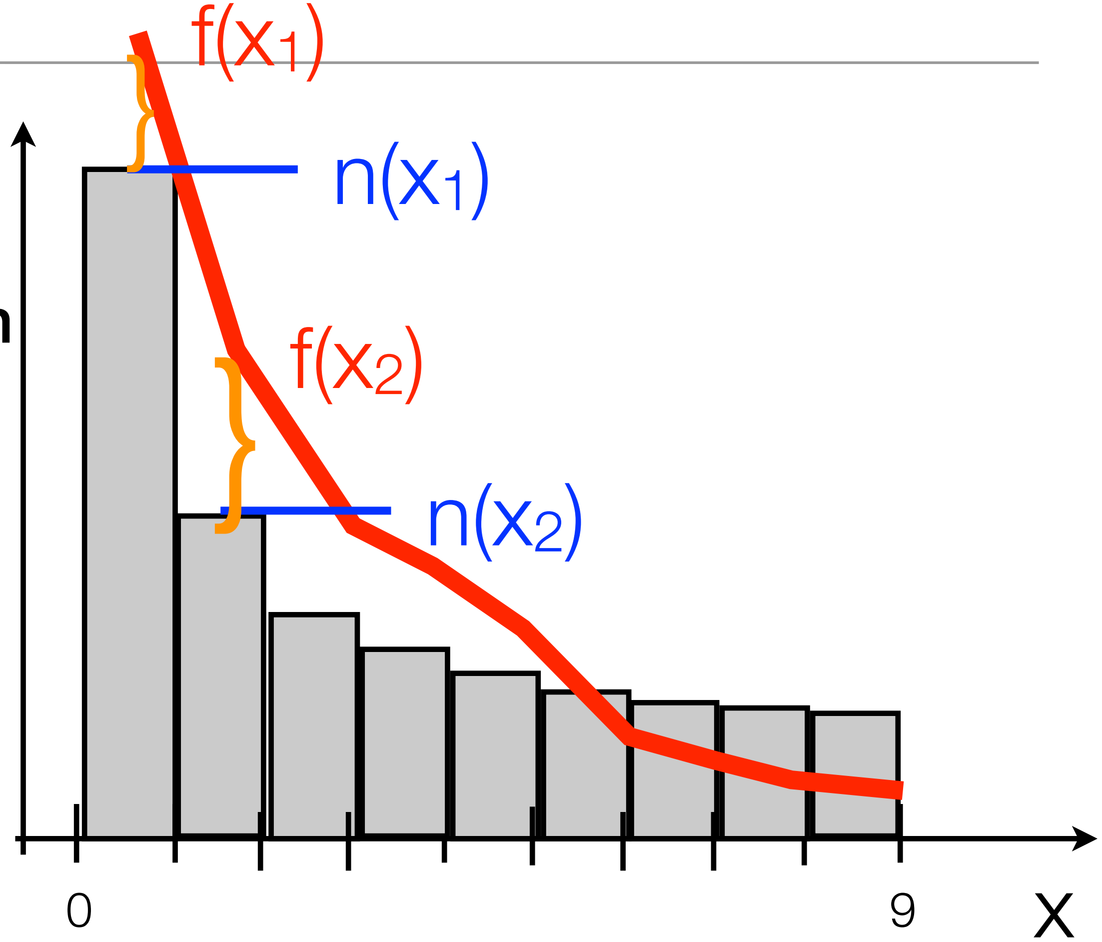
usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

χ^2 Fitting

- Use for binned data
- Minimise distance between data and function that describes data.
- Possible definition:

  $d^2 = \Sigma(n(x_i) - f(x_i))^2$

- Better: Weight by error



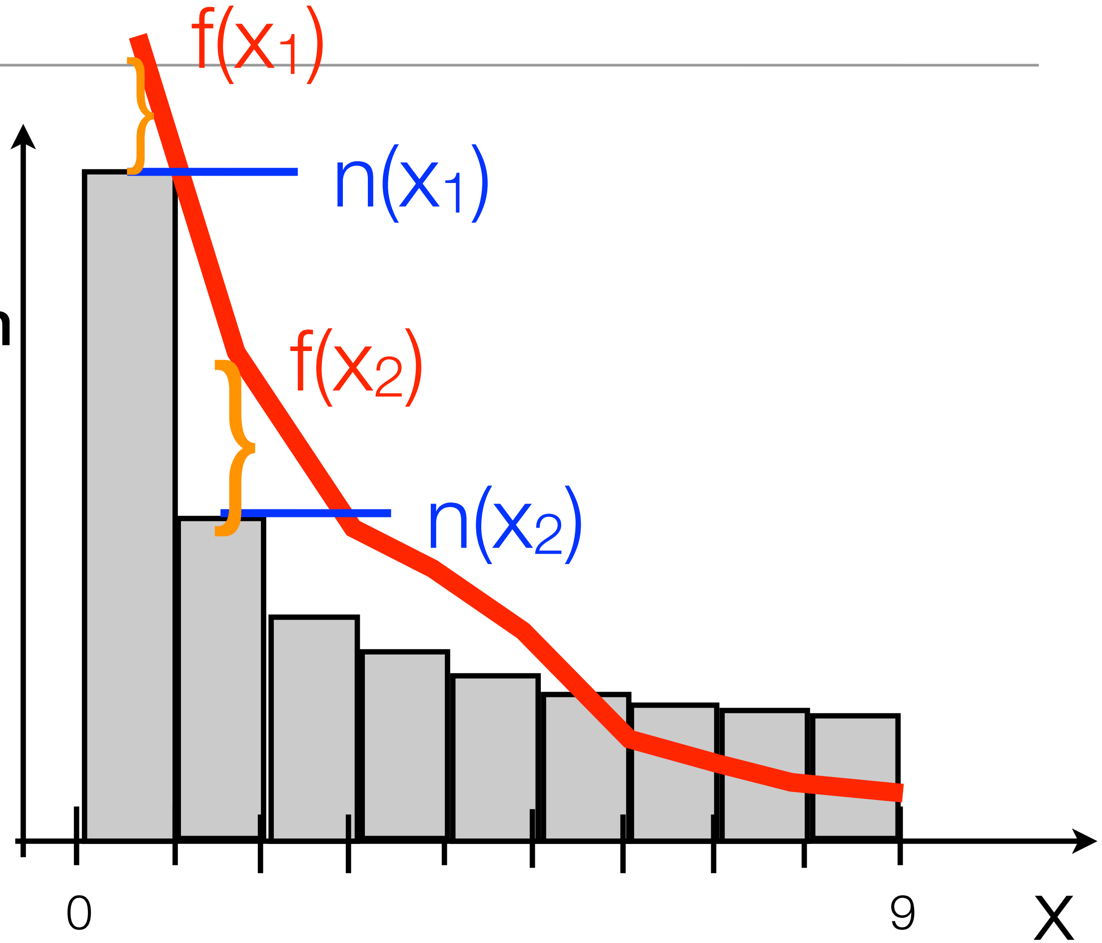
usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

χ^2 Fitting

- **Use for binned data**
- **Minimise distance between data and function that describes data.**
- **Possible definition:**

  $d^2 = \Sigma(n(x_i) - f(x_i))^2$

- **Better: Weight by error**

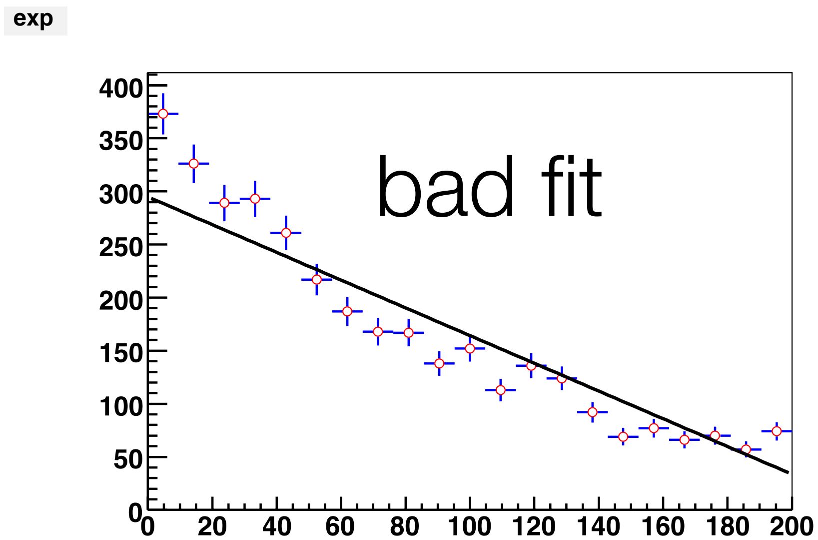


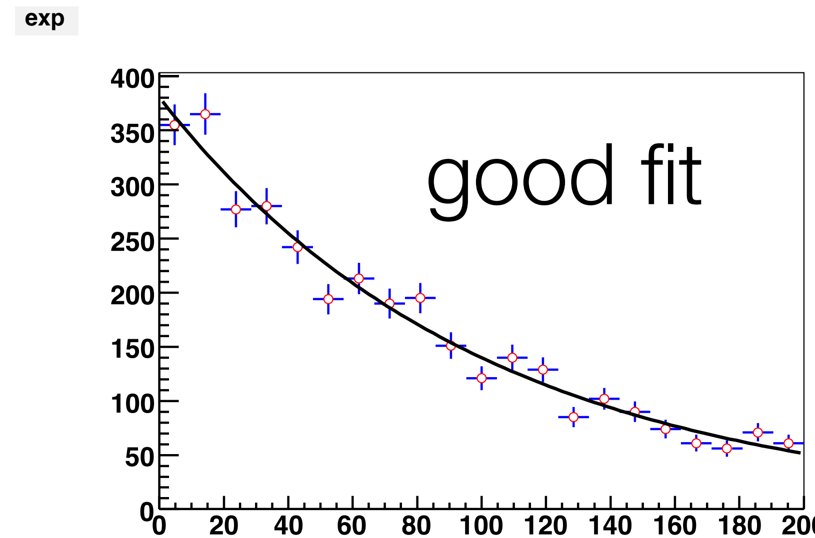
$$\chi^2 \equiv \sum_{\text{all bins}} \frac{\left(n_{\text{meas}}(x_i) - f(x_i)\right)^2}{\sigma^2}$$

usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

- root macros go here

Do I trust my fit?





- **Your fit programme will probably converge even if you use the wrong function. Need a way to pick this up - we want to the quantify badness of our fit.**

Goodness of fit and χ^2 distribution

- **Given this definition:**

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - f_i)^2}{\sigma_i^2}$$

what value for χ^2 would you expect?

Goodness of fit and χ^2 distribution

- **Given this definition:**

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - f_i)^2}{\sigma_i^2}$$

what value for χ^2 would you expect?

- **If we got our error estimates right, we'd expect a typical difference between model and data in each bin of 1σ.**

- **So we'd expect, for N bins:**

$$\chi^2 \approx N, \qquad \frac{\chi^2}{N} \approx 1$$

Goodness of fit and χ^2 distribution

- **χ^2 definition:**

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - f_i)^2}{\sigma_i^2}$$

- **However, we are not just comparing a model and data. We are allowed to adjust the model.**

- **To account for the extra wiggle-room each fit parameter provides, we define the number of degrees of freedom as**

$$\mathrm{ndf} \equiv N_{\mathrm{bins}} - N_{\mathrm{fit\ parameters}}$$

- **We expect** $$\frac{\chi^2}{\mathrm{ndf}} \approx 1$$

Fit quality as a probability: How likely am I to get a fit that bad or worse if my model is correct?

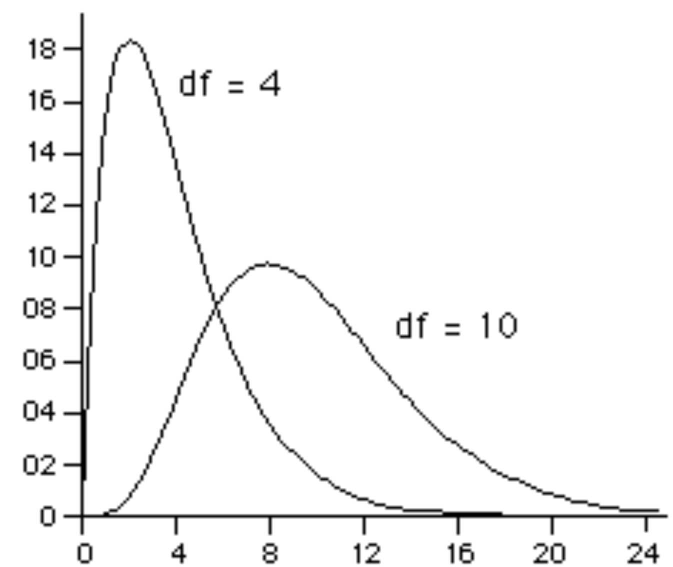
- **The probability density to get a certain χ^2 for a given number of degrees of freedom:**

$$P(\chi^2; \mathrm{ndf}) = \frac{1}{2^{\mathrm{ndf}/2}\Gamma(\mathrm{ndf}/2)} \chi^{\mathrm{ndf}-2} e^{-\chi^2/2}$$

- **Calculate the probability, p, to get a χ^2 this bad or worse***

$$p = \int_{\chi^2}^{\infty} P(\chi'^2; \mathrm{ndf})\, d(\chi'^2)$$

- **If *p* is smaller than a few %, it gets a bit worrying.**



***) root does it for you, with the stupidly named function TMath::Prob**

Resources for tomorrow’s problem classes:

This course:

https://tinyurl.com/TeshepProblems

https://tinyurl.com/TeshepStatCode

https://tinyurl.com/TeshepMC

External:

http://jupyter.readthedocs.io/en/latest/install.html

https://docs.anaconda.com/anaconda/

Probabilities, PDFs and likelihood fitting

Probability

Probability

- **As an average UK citizen, at the age of 20, the probability that you die within a year is 0.048%.**

Probability

- **As an average UK citizen, at the age of 20, the probability that you die within a year is 0.048%.**

- **But who is average?**

Probability

- **As an average UK citizen, at the age of 20, the probability that you die within a year is 0.048%.**

- **But who is average?**

- **If you are female, it is only 0.026% (male: 0.069%)**

Probability

- As an average UK citizen, at the age of 20, the probability that you die within a year is 0.048%.
- But who is average?
- If you are female, it is only 0.026% (male: 0.069%)
- If you are a male in Scotland, it is 0.1%

Probability

- As an average UK citizen, at the age of 20, the probability that you die within a year is 0.048%.
- But who is average?
- If you are female, it is only 0.026% (male: 0.069%)
- If you are a male in Scotland, it is 0.1%
- But what if you smoke? If you don't? If you are a heroin-addicted bomb-disposal expert?

What is Probability?

- Mathematically: Defines basic properties such as $0 \leq P \leq 1$ and calculation rules; all other definitions must satisfy also this one. But: No meaning.

- Frequentist: How many times n_E does something (event E) happen if I try N times? $P(E) = n_E/N$ for $N \to \infty$
Problem: What if I can try only once?

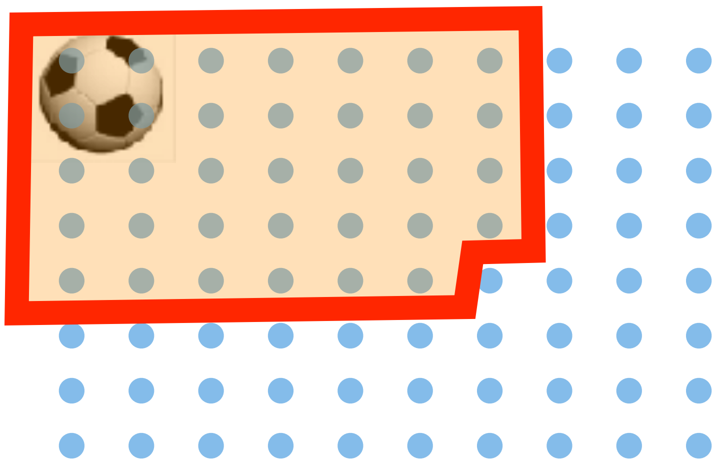
- Bayesian: Probability is a measure for the "degree of belief" that event E happens. One possible definition: I'd bet up to € n_E that E happens, if I get € N if I win: $P(E) = (£\ n_E)/(£\ N)$.
Problem: Subjective (not good for science, but occasionally unavoidable, e.g. for systematics.)

Probabilities nomenclatura

- P(A) = probability that A happens
- P(A or B) = probability that A happens, or B happens, or both.
- P(A & B) = P(A and B) probability that both A and B happen.
- P(A|B) = “P of A given B”, the probability that A happens given that B happens.
  - Note: while P(A & B) = P(B & A), P(A or B) = P(B or A), P(A|B) ≠ P(B|A), for example:
    P(pregnant | woman) ≈ a few %
    P(woman | pregnant) ≈ 100%

Probabilities

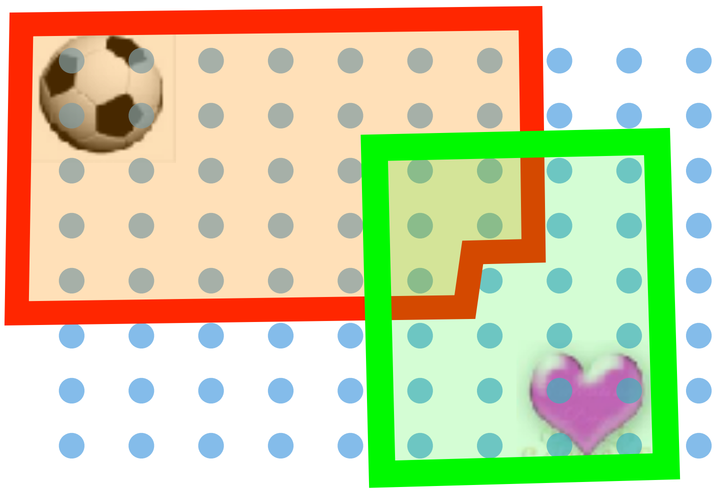
- **Inside the red box everyone who likes football.**



Adding non-exclusive Probabilities

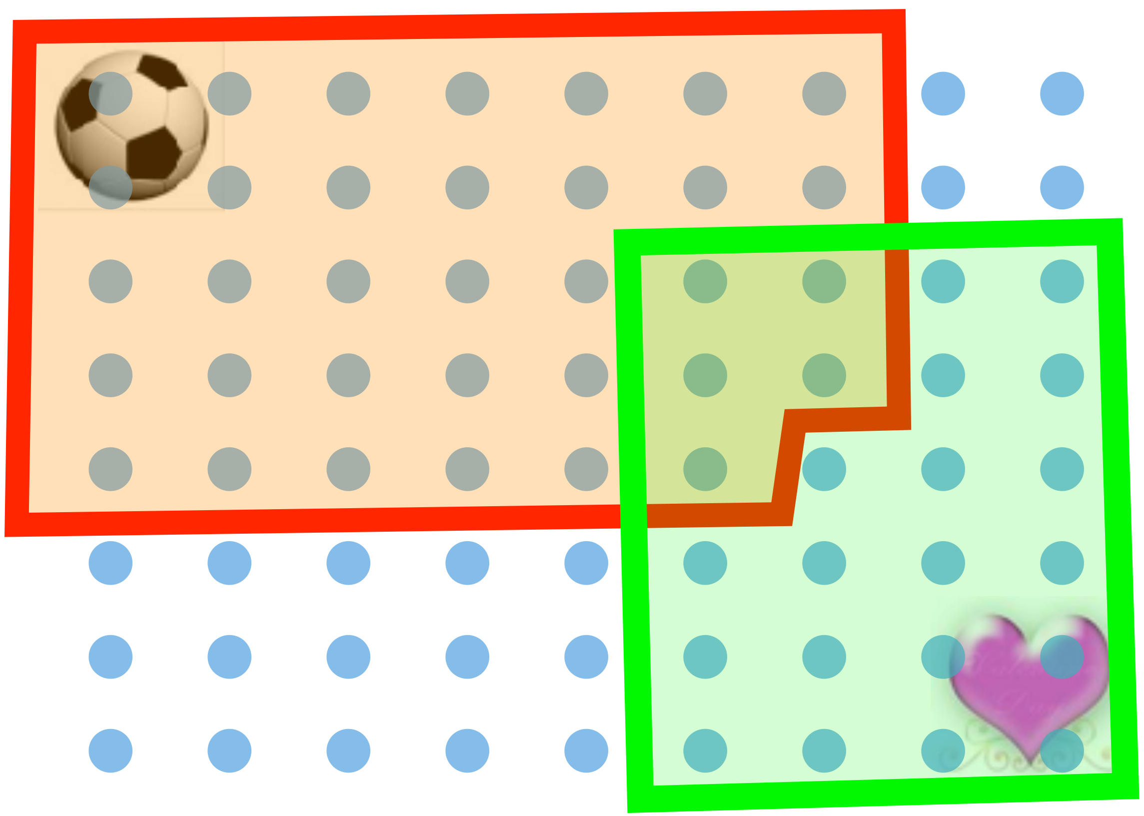
- What is the probability to pick somebody who likes football (outcome A) or the colour pink (outcome B)?

- Not ~~P(A or B) = P(A) + P(B),~~ wrong
  because we would be double-counting those who like football and the colour pink.



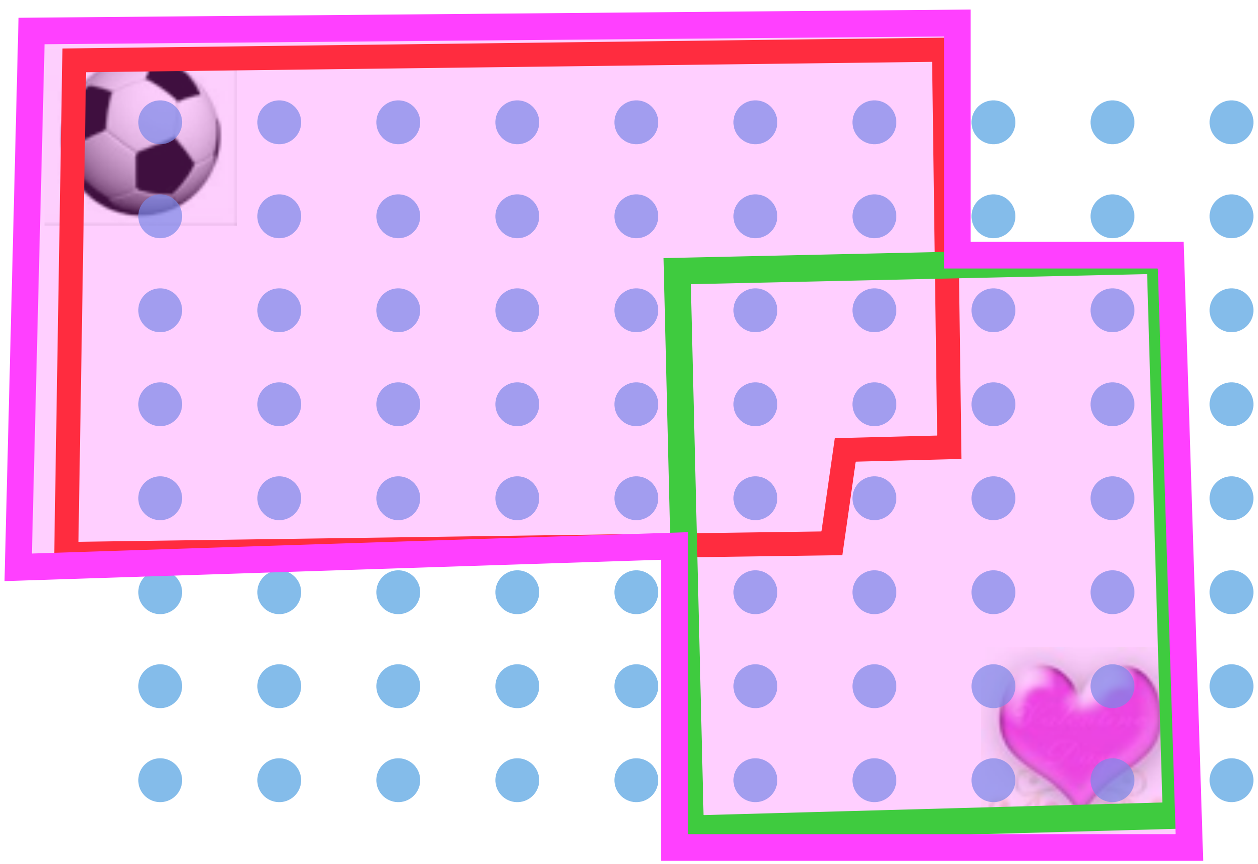
Adding Non-Exclusive Probabilities

- **P(A or B)**



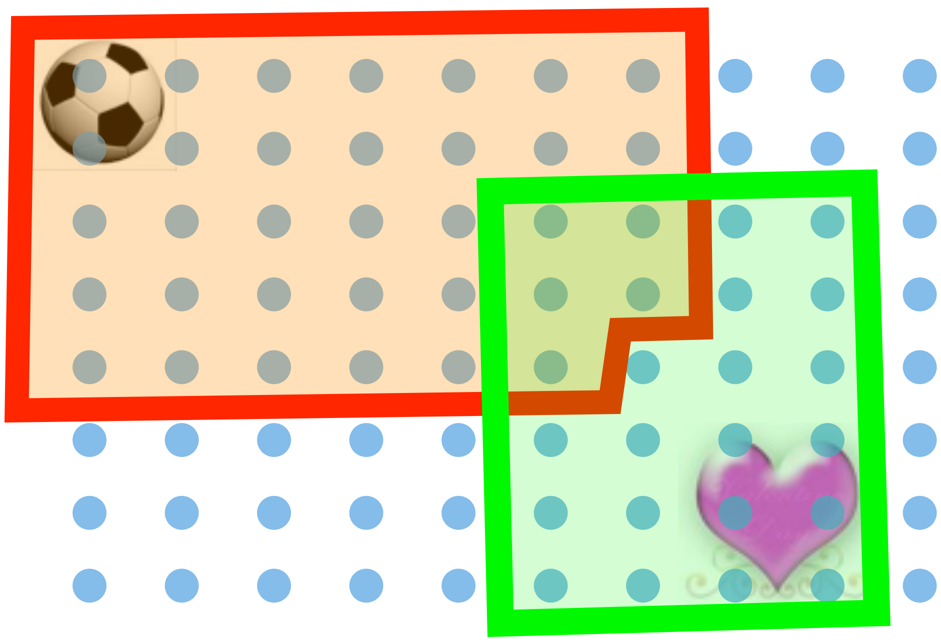
Adding Non-Exclusive Probabilities

- **P(A or B) = P(A) + P(B) – P(A and B)**



Conditional Probabilities

- P(A given B) = P(A|B) = P(A and B)/P(B)
- P(B given A) = P(B|A) = P(A and B)/P(A)
- P(A and B) = P(A) · P(B|A) = P(B) · P(A|B)



Bayes' Theorem

- P(A and B) = P(A) P(B|A) = P(B) P(A|B)

- From this follows Bayes' theorem:

  P(A|B) = P(B|A) P(A)/P(B)

Bayes' Theorem

- P(A and B) = P(A) P(B|A) = P(B) P(A|B)

- **From this follows Bayes' theorem:**

  P(A|B) = P(B|A) P(A)/P(B)

Very important theorem.
Also worth noting:
This is not Bayesian statistics (every frequentist will happily use Bayes' theorem)

Problem

- 0.01% of the population is infected with a nasty, contagious virus

  A test for this virus is developed. This test identifies correctly 100% of those carrying the virus. Amongst those that do not carry the virus, it gives the correct result in 99.8% of the cases.

- If you test positive, how worried should you be? Are you likely to be infected?

Problem

- 0.01% of the population is infected with a nasty, contagious virus

  A test for this virus is developed. This test identifies correctly 100% of those carrying the virus. Amongst those that do not carry the virus, it gives the correct result in 99.8% of the cases.

- If you test positive, how worried should you be? Are you likely to be infected?

- Task: calculate how likely you are infected if the test is positive

Probabilities for Continuous Distributions

Probabilities for Continuous Distributions

- **Say you have a 100 strings between 10cm and 12cm long and measure their length.**

Probabilities for Continuous Distributions

- **Say you have a 100 strings between 10cm and 12cm long and measure their length.**
- **How many are 11 cm?**

Probabilities for Continuous Distributions

- **Say you have a 100 strings between 10cm and 12cm long and measure their length.**
- **How many are 11 cm?**
- **But how do we describe a probability distribution where the probability of each event is zero?**

Probabilities for continuous variables

- *P(x)* = probability density function (PDF)
- PDFs are not probabilities. But we can use them to calculate probabilities that we find a value between **a** and **b**

$$P(x \in [a,b]) = \int_a^b P(x')\, dx'$$

- This integral is a probability. If you integrate over a small range, such as a histogram bin of width Δx, the probability to find an event in that bin is

P(find event in bin centered at x) ≈ P(x)Δx

Expected number of events in that bin ≈ N_{total} P(x)Δx

- BTW, the Gaussian discussed earlier is a PDF.

PDFs for real variables

- Frequent student mistake: decide which of the three great distributions applies (Binomial, Poisson, Gauss) based on whether a variable is continuous or not.
- But: You can use Probability Density Functions (and Gaussians) for discrete variables. It’s an approximation, but often a useful one.
- It’s the same as approximating discrete people with a population density or discrete atoms with a mass density.

PDFs: important properties

- **Normalisation - the probability that something happens is 1:**

$$\int_{-\infty}^{+\infty} P(x')\,dx' = 1$$

- **Expectation value of x, or any function of x, gives the average expected outcome for x (function of x)**

$$\langle x \rangle = \int x'\,P(x')\,dx' \qquad \langle f(x) \rangle = \int f(x')P(x')\,dx'$$

- **Variance** $V = \langle x^2 \rangle - \langle x \rangle^2$

PDFs and change of variables

- **Let $P(x)$ be a PDF. Then $P(x)\,dx$ is a probability.**

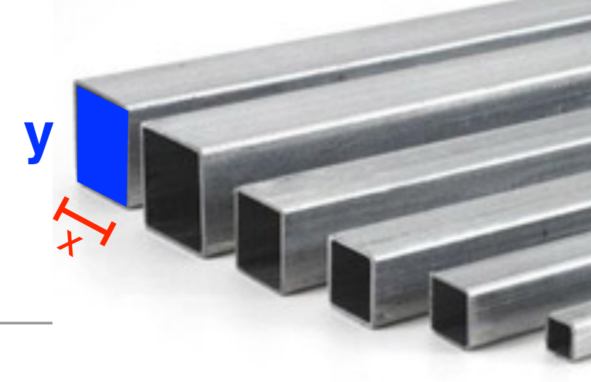
- **Let y be a function of x (suitable for co-ordinate transformations, i.e. bijective [one-to-one], and also differentiable).**

- **Then $P(y)\,dy = P(x)\,dx \Rightarrow P(y) = P(x)\,dx/dy$.**

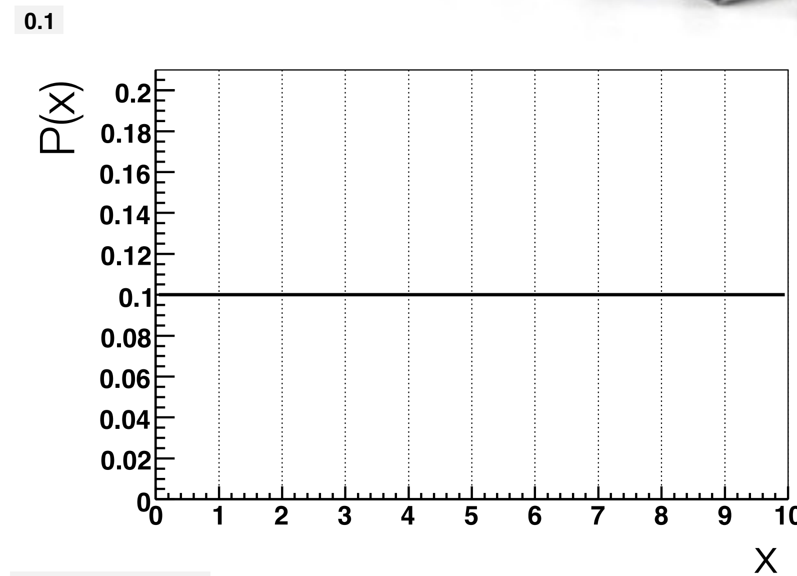
- This can give negative $P(y)$ because the derivative can be negative. This would be handled by the corresponding swap in integration limits, giving positive integrals. We'd rather have positive PDF's and decide that integration limits for PDFs will always be from the lower to the higher value.

- **Hence $P(y) = P(x)\,\left|dx/dy\right|$.**

Example: Variable Transformation



$$P(x) = \left\{ \begin{array}{lr} \frac{1}{10} & \text{between 0 and 10} \\ 0 & \text{otherwise} \end{array} \right\}$$



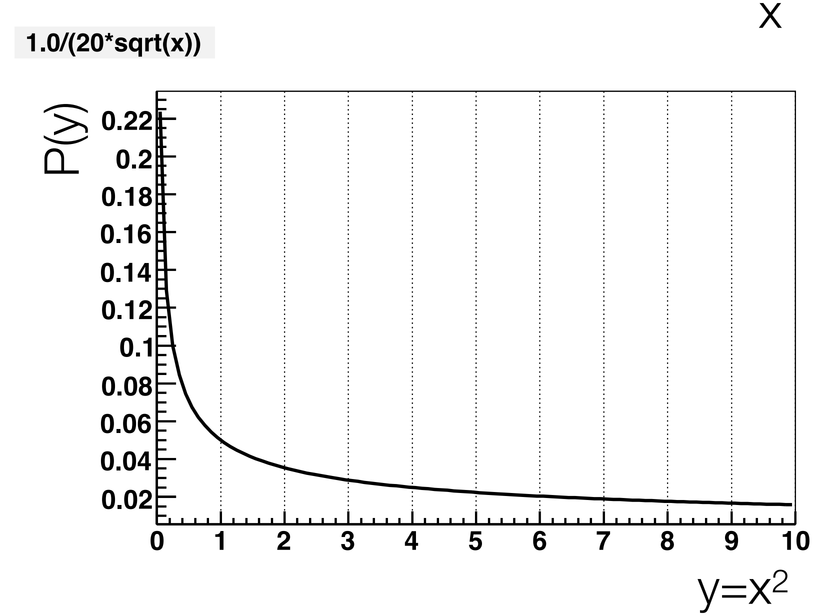
$$y = x^2 \Leftrightarrow x = \sqrt{y} \text{ for } x > 0$$

$$P(y)\, dy = P(x)\, dx$$

$$P(y) = P(x)\,\frac{dx}{dy}$$

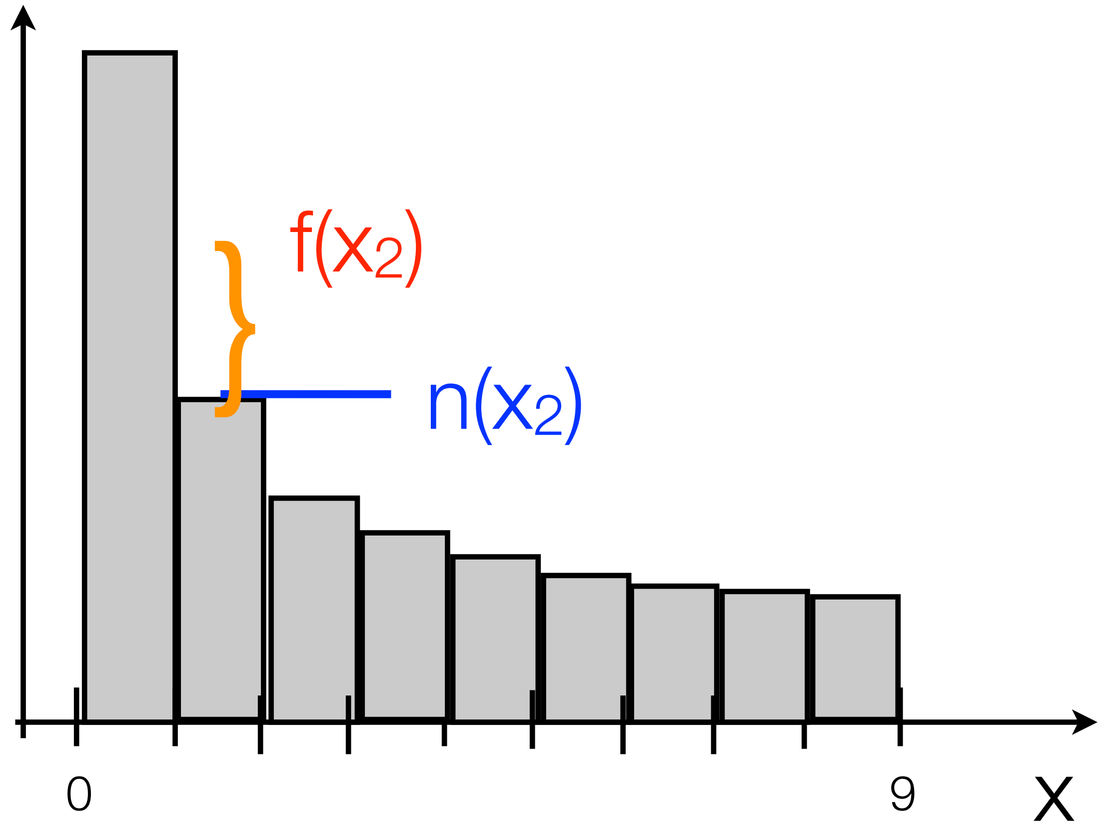
$$= P(x)\frac{1}{2\sqrt{y}}$$

$$= \frac{1}{20\sqrt{y}}$$



Last time: χ^2 Fitting

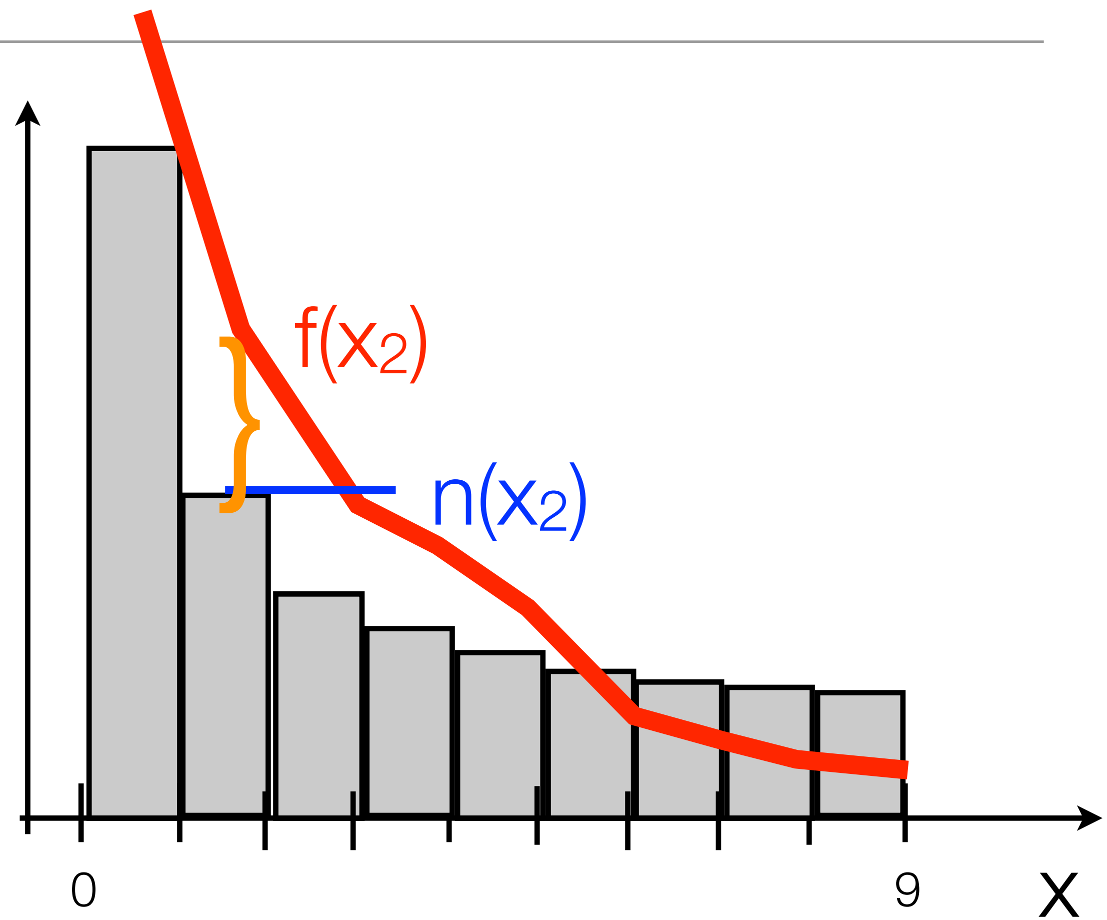
- **Use for binned data**



usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

Last time: χ^2 Fitting

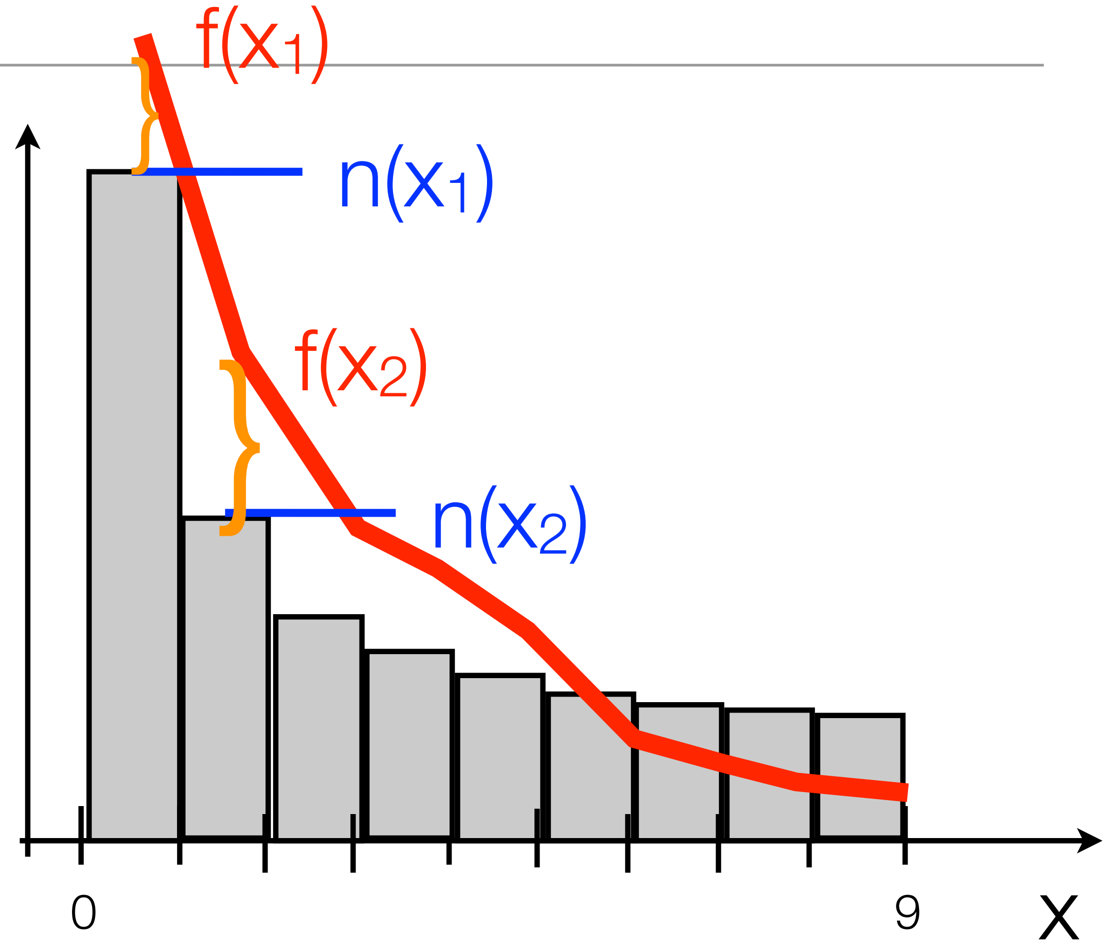
- **Use for binned data**



usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

Last time: χ^2 Fitting

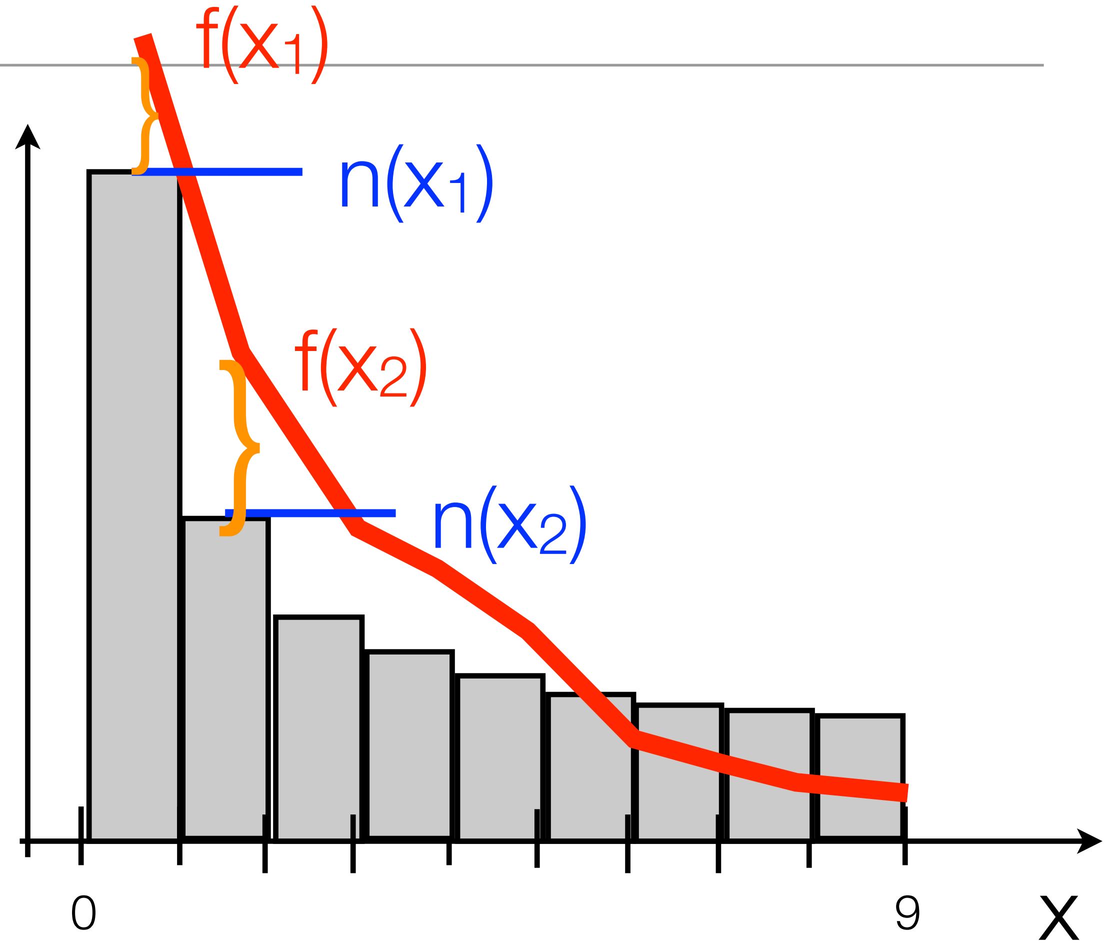
- **Use for binned data**



usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

Last time: χ^2 Fitting

- **Use for binned data**

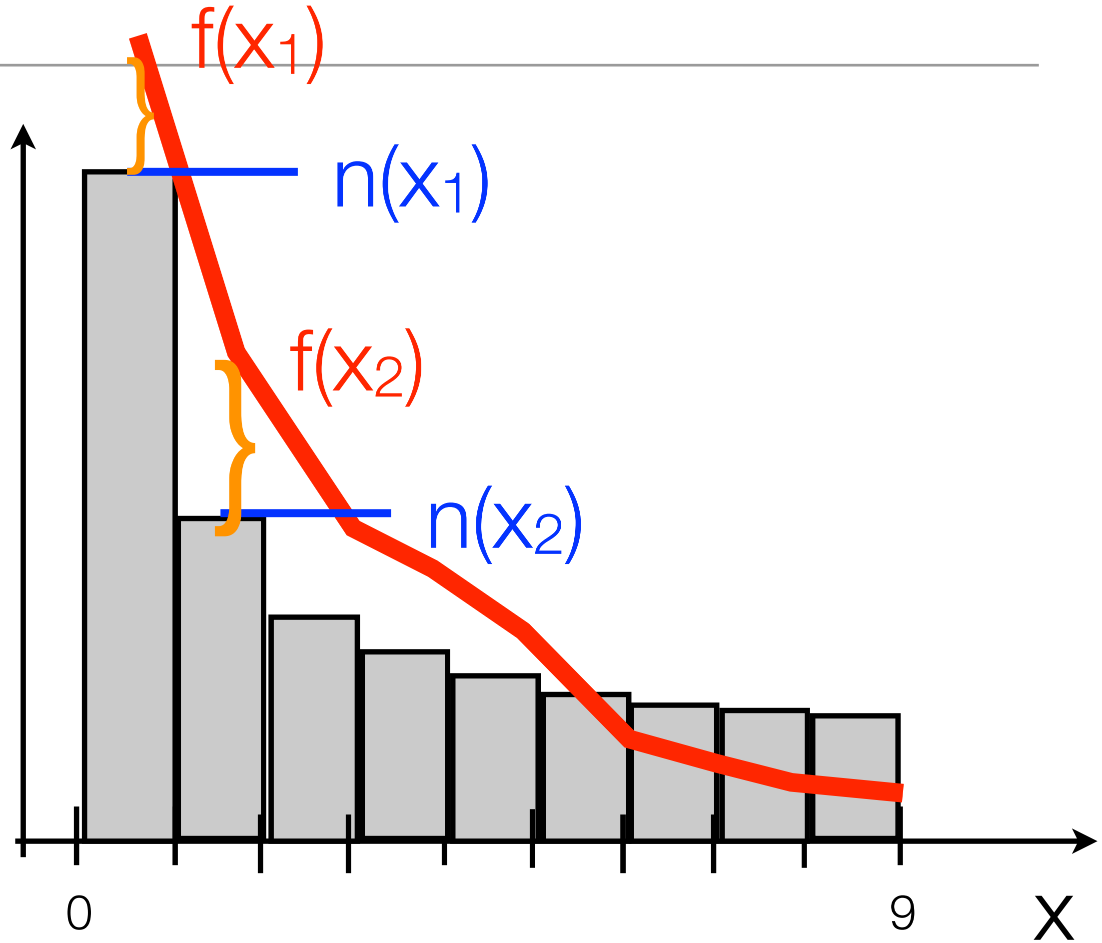


usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

Last time: χ^2 Fitting

- **Use for binned data**

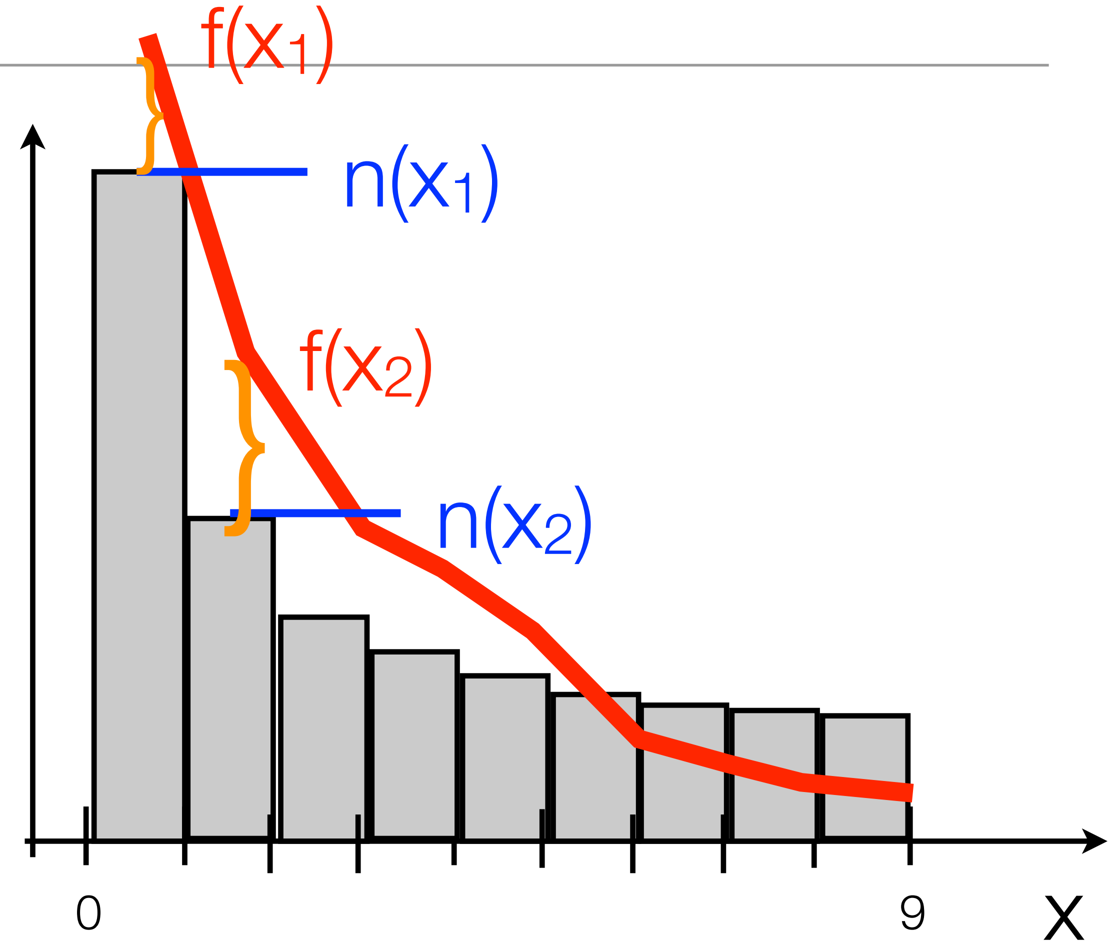
- **Minimise weighted distance between data and function that describes data.**



usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

Last time: χ^2 Fitting

- **Use for binned data**
- **Minimise weighted distance between data and function that describes data.**



$$\chi^2 \equiv \sum_{\text{all bins}} \frac{\left(n_{\text{meas}}(x_i) - f(x_i)\right)^2}{\sigma^2}$$

usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

Likelihood fits

- **Define the likelihood:**

$$\mathcal{L} \equiv \prod_{\text{all data points}} P(t_i)$$

- **View this as a function of the parameters of the PDF, here τ:**

$$\mathcal{L}(\tau) \equiv \prod_{\text{all data points}} P(t_i; \tau)$$

- **This gives us the probability that, given τ, we see the data we see. We adjust τ to maximise this.**

- **Note that this does not give us the probability that τ is the right value (although we would probably quite like to know that - too bad, it's not what it tells us).**

Likelihood fits

- **Rather than maximising this product:**

$$\mathcal{L}(\tau) \equiv \prod_{\text{all data points}} P(t_i;\tau)$$

- **it is usually easier (and equivalent), to maximise the logarithm of the likelihood, since this turns the product into a sum**

$$\ln \mathcal{L}(\tau) = \sum_{\text{all data points}} \ln P(t_i;\tau)$$

Normalising your PDF

- **This property:**

$$\int_{-\infty}^{+\infty} P(x)\, dx = 1$$

is crucial! Often you have a function f(x) you want to fit to the data that is not normalised. Before you can use it in your likelihood fit, you must always normalise it

$$P(x) = \frac{f(x)}{\int_{-\infty}^{+\infty} f(x')\, dx'}$$

$$\int_{-\infty}^{+\infty} P(x')\, dx = \frac{\int_{-\infty}^{+\infty} f(x')\, dx'}{\int_{-\infty}^{+\infty} f(x')\, dx'} = 1$$

Normalising your PDF

- **This property:**

$$\int_{-\infty}^{+\infty} P(x)\,dx = 1$$

is crucial! Often you have a function f(x) you want to fit to the data that is not normalised. Before you can use it in your likelihood fit, you must always normalise it

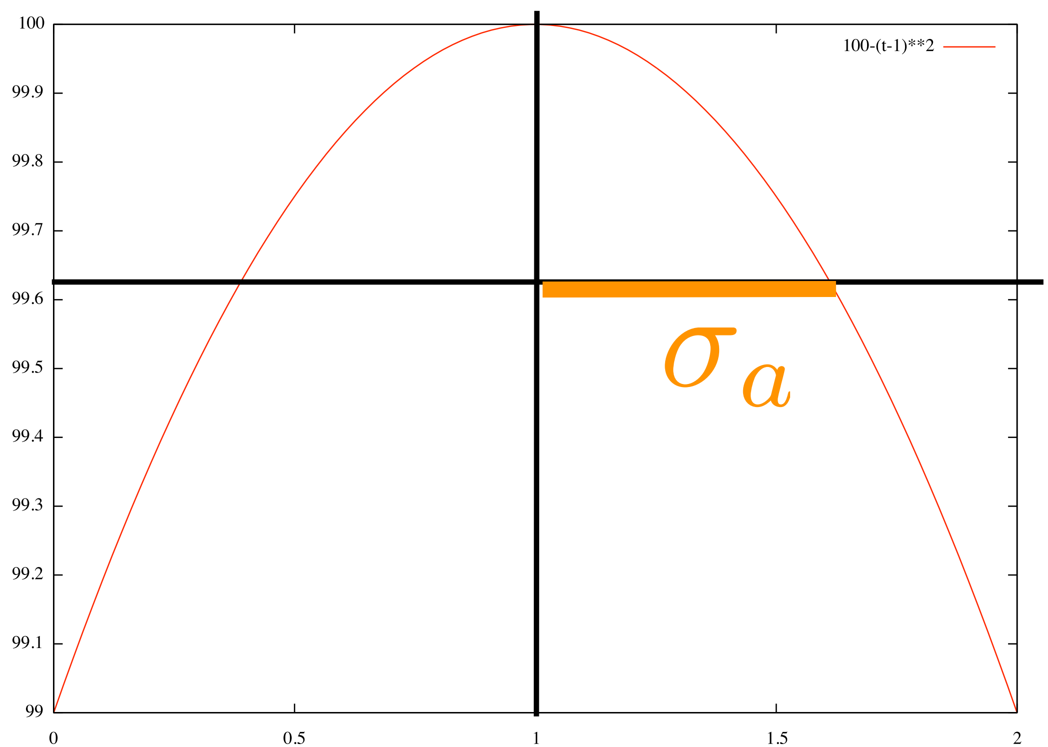
$$P(x) = \frac{f(x)}{\int_{-\infty}^{+\infty} f(x')\,dx'}$$

$$\int_{-\infty}^{+\infty} P(x')\,dx = \frac{\int_{-\infty}^{+\infty} f(x')\,dx'}{\int_{-\infty}^{+\infty} f(x')\,dx'} = 1$$

Likelihood Shape

- L should be Gaussian, and L should be a parabola (near the maximum) from which you can read off the uncertainty

$$\ln \mathcal{L} = -\frac{(a-\hat{a})^2}{2\sigma_a^2} + (\text{meaningless constant})$$



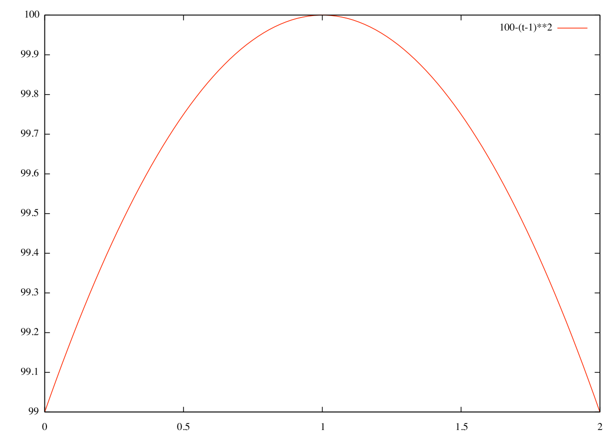
Uncertainty from likelihood "Parabolic Error"

- **You can also calculate the uncertainty directly from**

$$\ln \mathcal{L} = -\frac{(a-\hat{a})^2}{2\sigma_a^2} + (\text{meaningless constant})$$

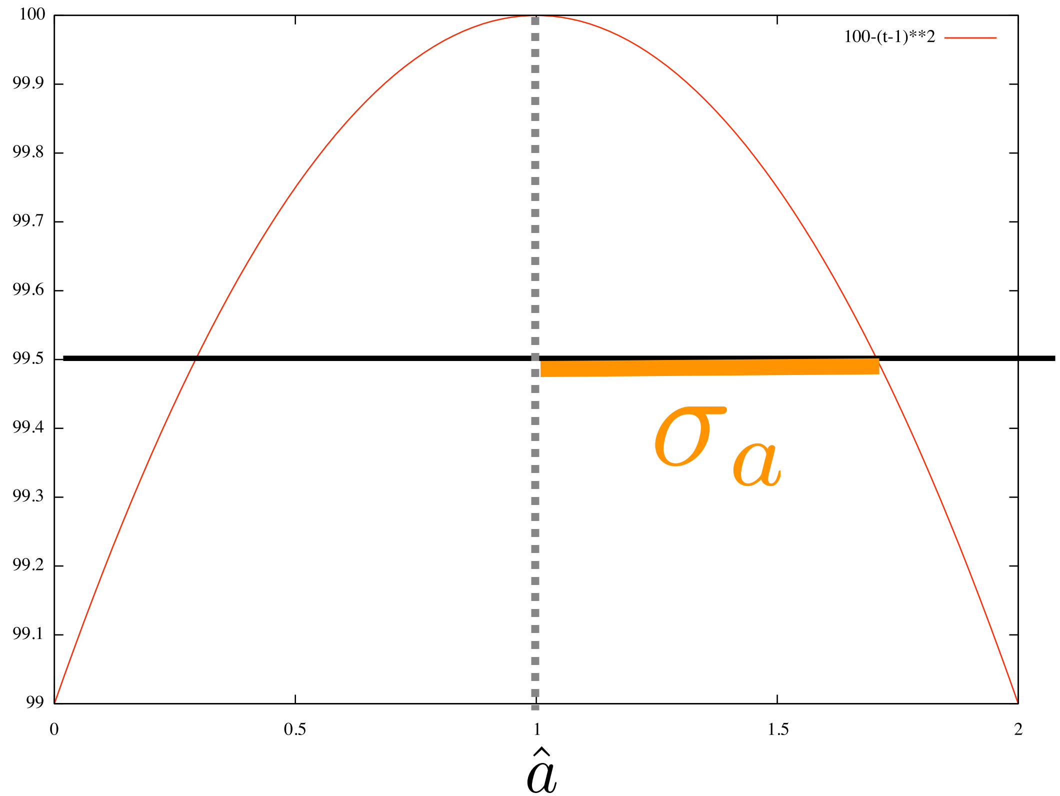
$$\left.\frac{d^2(\ln \mathcal{L})}{d\,a^2}\right|_{\text{at a=â}} = -\frac{1}{\sigma_a^2}$$

$$\sigma_a = \sqrt{\frac{1}{-\left.\frac{d^2(\ln \mathcal{L})}{d\,a^2}\right|_{\text{at a=â}}}}$$



Error Estimate

$$\ln \mathcal{L} = -\frac{(a-\hat{a})^2}{2\sigma_a^2} + (\text{meaningless constant})$$

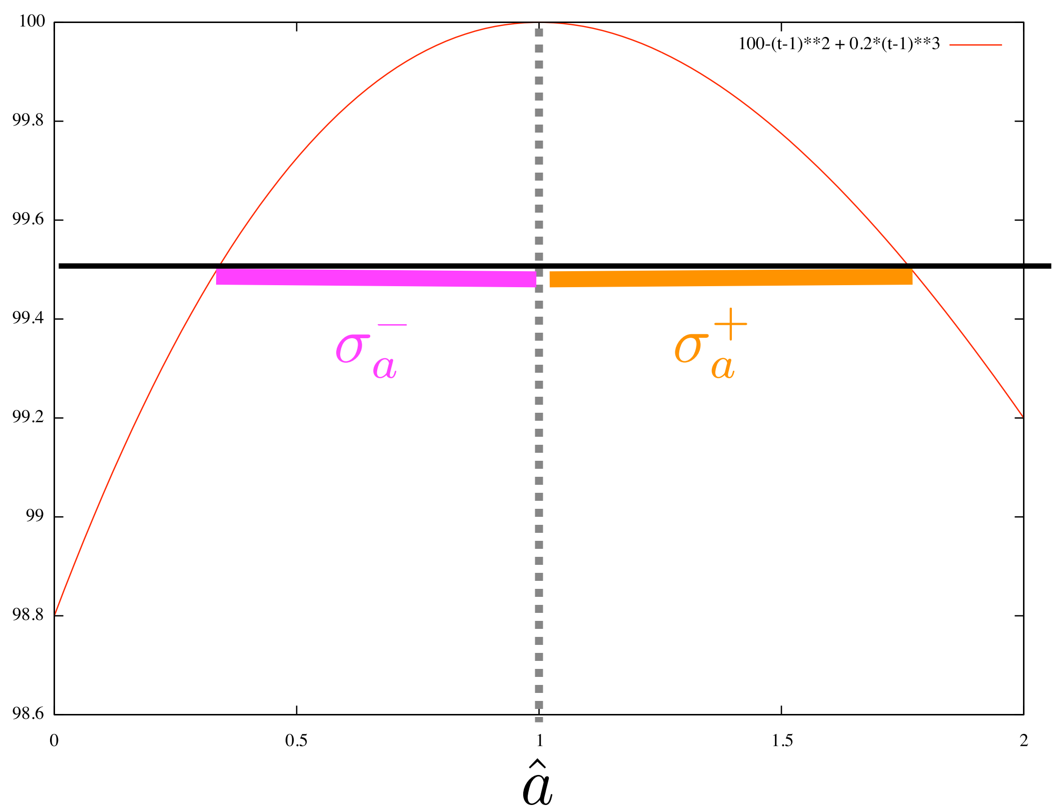


$$\Delta \ln \mathcal{L} = \frac{1}{2}$$

Error Estimate for low N

- **If it's not a Gaussian, you get asymmetric errors.**

$$a = \hat{a}^{+\sigma_a^+}_{-\sigma_a^-}$$

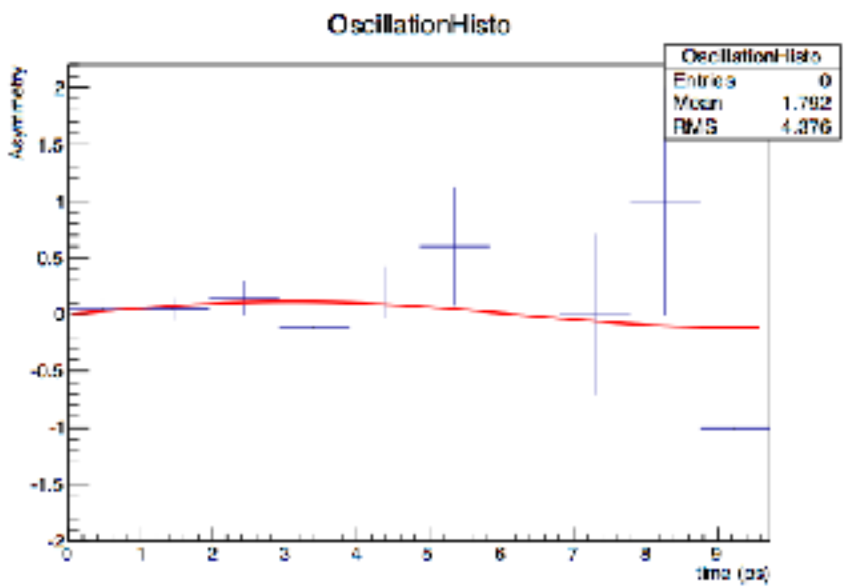


$$\Delta \ln \mathcal{L} = \frac{1}{2}$$

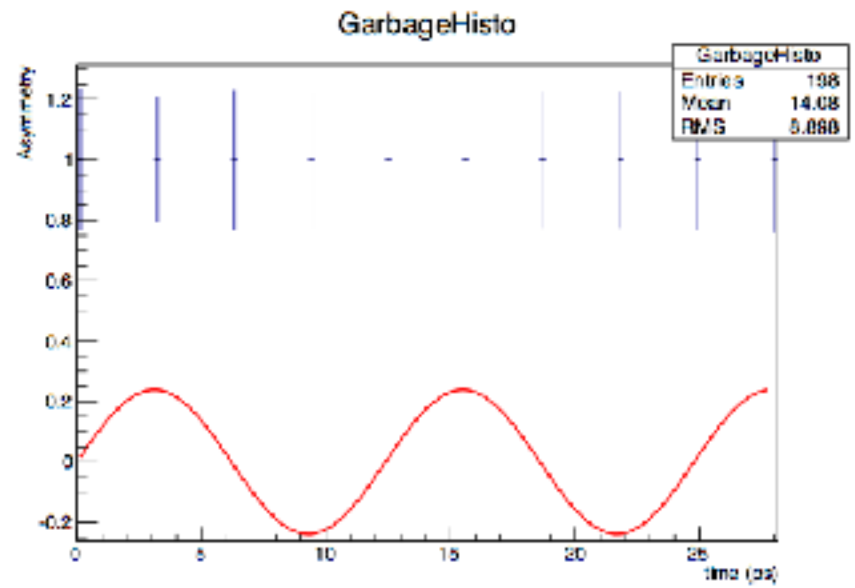
Quality of Fit

- **Very tricky for likelihood fits. The value of the likelihood function does not tell you anything at all about the quality of the fit.**

lnL = –276.3



lnL = –271.4

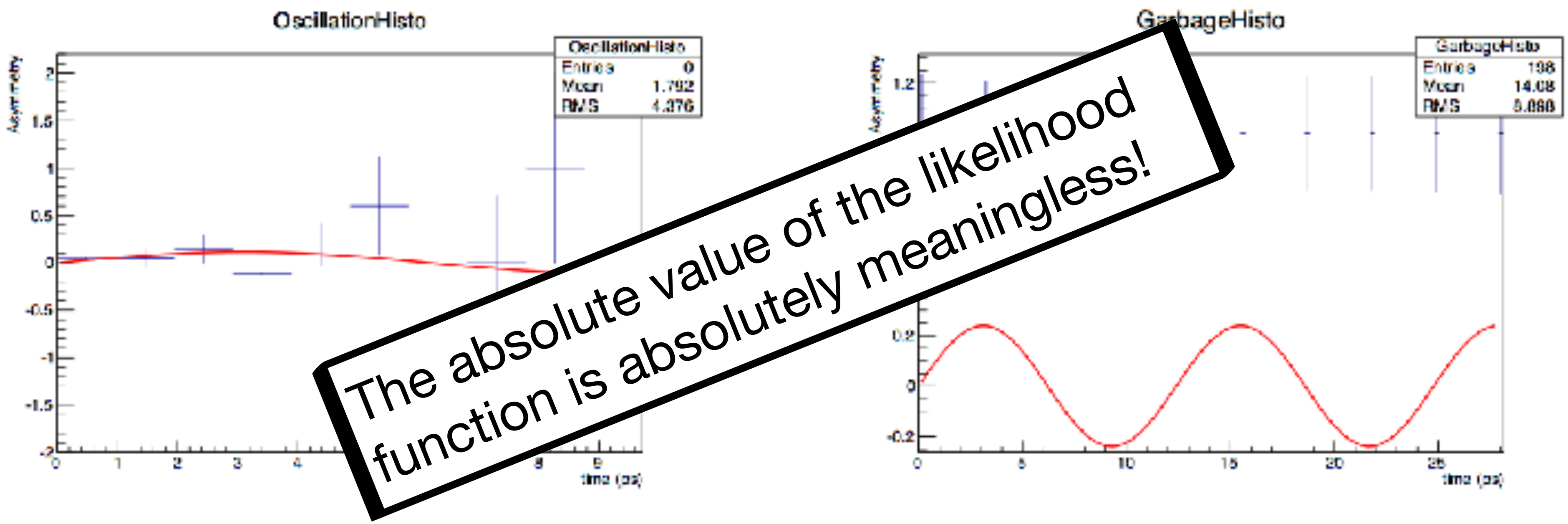


- **One solution: After doing an un-binned likelihood fit, bin the data and calculate the χ^2 between data and fit.**

Quality of Fit

- **Very tricky for likelihood fits. The value of the likelihood function does not tell you anything at all about the quality of the fit.**

lnL = –276.3 lnL = –271.4



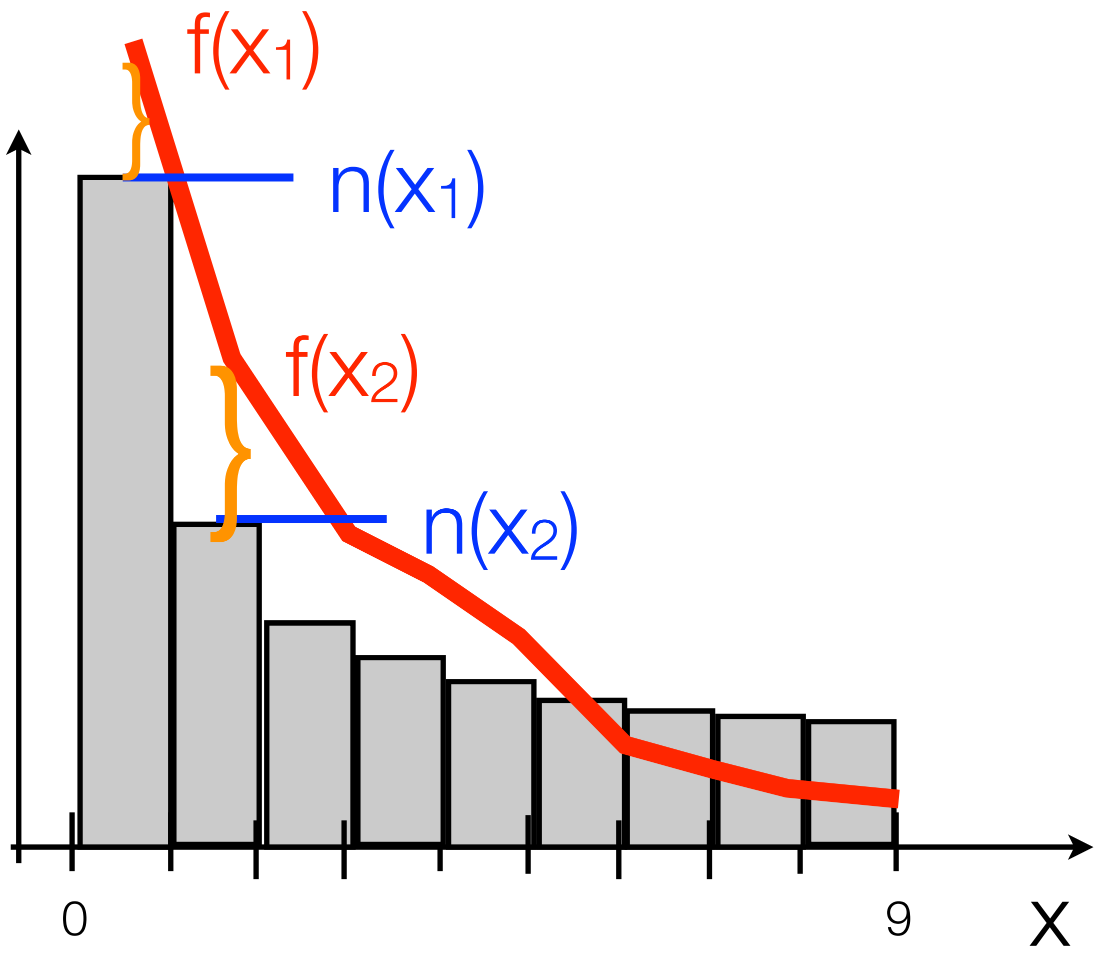
- **One solution: After doing an un-binned likelihood fit, bin the data and calculate the χ^2 between data and fit.**

χ^2 Fitting and likelihood.

- **Let's do a binned likelihood fit. Our model predicts *f*(x1) events for bin centred at x1.**

- **The probability to see n_i events given that we expect *f(xi)* is given by a Poisson distribution**

$$P(n_i; f(x_i)) = e^{-f(x_i)} \frac{f(x_i)^{n_i}}{n_i!}$$



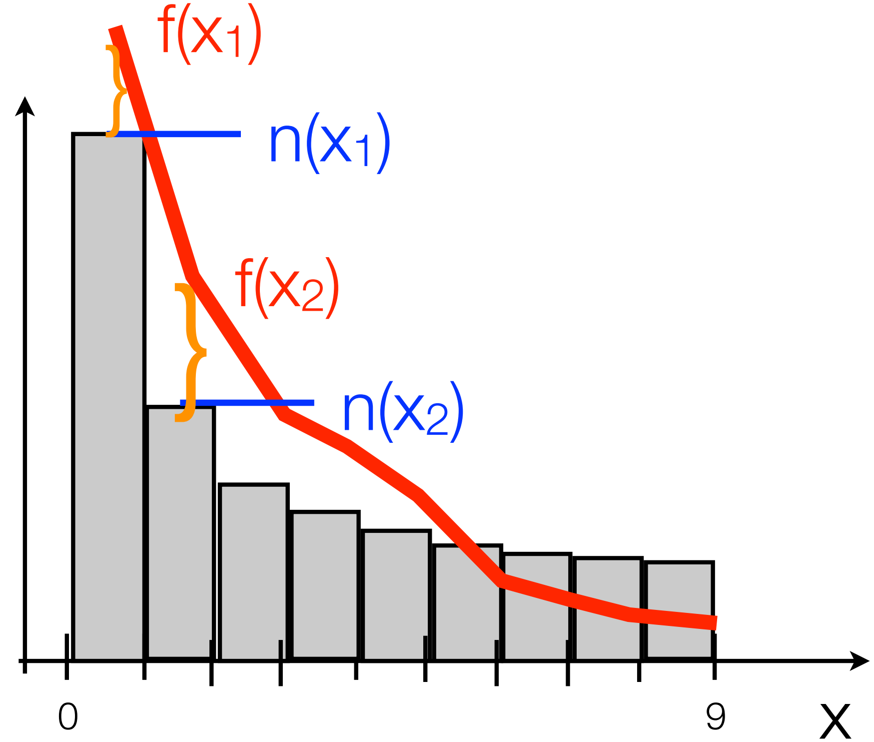
χ^2 Fitting and likelihood.

- **Binned likelihood:**

$$P(n_i; f(x_i)) = e^{-f(x_i)} \frac{f(x_i)^{n_i}}{n_i!}$$

- **if n_i is large, approximate**

$$P(n_i; f(x_i)) = \frac{1}{\sqrt{2\pi}\sqrt{n_i}} e^{-\frac{1}{2}\left(\frac{f(x_i)-n_i}{\sqrt{n_i}}\right)^2}$$



- **log-likelihood**

$$\log \mathcal{L} = \sum_i P(n_i; f(x_i)) = \sum_i -\frac{1}{2}\left(\frac{f(x_i)-n_i}{\sqrt{n_i}}\right)^2 + C$$

$$-2\ln \mathcal{L} = \sum_i \ln P(n_i; f(x_i)) = \sum_i \left(\frac{f(x_i)-n_i}{\sqrt{n_i}}\right)^2 + K$$

meaningless constants

χ^2 Fitting and likelihood.

- The χ^2 fit is equivalent to a binned likelihood fit for large numbers of events. The interpretation of the χ^2 in terms probabilities etc is based on that.
- Conversely, χ^2 fits only work properly if you have a large number of events in each bin. Say at least 10.
- What to do if you have fewer than 10 events in a bin:
  - Merge bins until you have at least 10 events per bin.
  - Do a binned likelihood fit (i.e. simply do not approximate the Poisson with the Gaussian).
  - Do an unbinned likelihood fit.

Testing your fit

Whatever you do, test your fit!

Pull study

- **Simulate a lot of datasets using Monte-Carlo simulation.**

- **Fit each dataset and calculate the**

pull = $$\frac{(\text{fit result}) - (\text{true value})}{(\text{error estimate})}$$

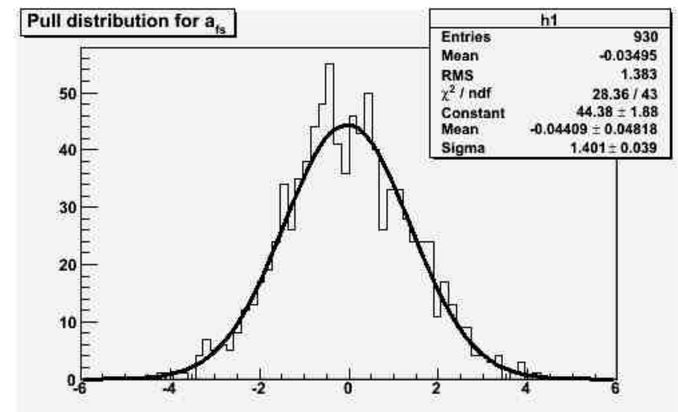
and put it in a histogram.

- **For a good, unbiased fitter, you get:**

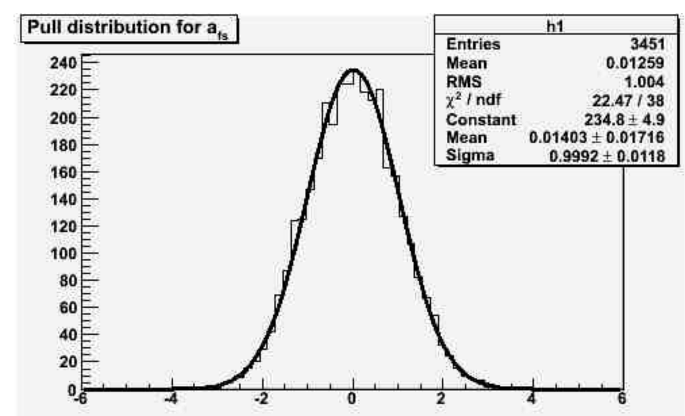
$$\text{Mean} = 0 \pm \frac{1}{\sqrt{N_{\text{exp}}}}$$

$$\sigma = 1 \pm \frac{1}{\sqrt{2N_{\text{exp}}}}$$

σ=1.4 for 1k events ⇒ wrong errors



σ=1.0 for 1k events ⇒ correct errors



Monte Carlo

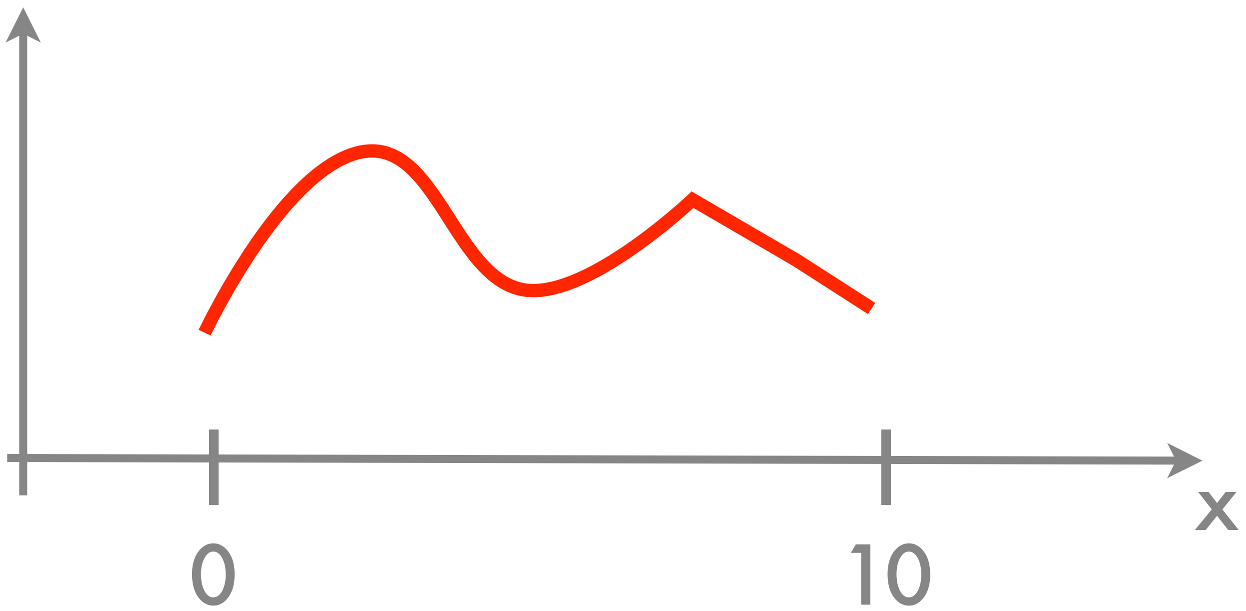
Gamblers in the casino at Monte-Carlo. c.1910

Monte Carlo Simulations

- To test your fit, you need to try it out on simulated data.
- To really test it properly, you cannot rely on the experiment's detailed simulation - you want to run thousands of simulated experiments and see if your fitter behaves as expected. You need a simplified, fast Monte Carlo for that.
- Today:
  - How do generate any distribution
  - How to do it a bit more efficiently

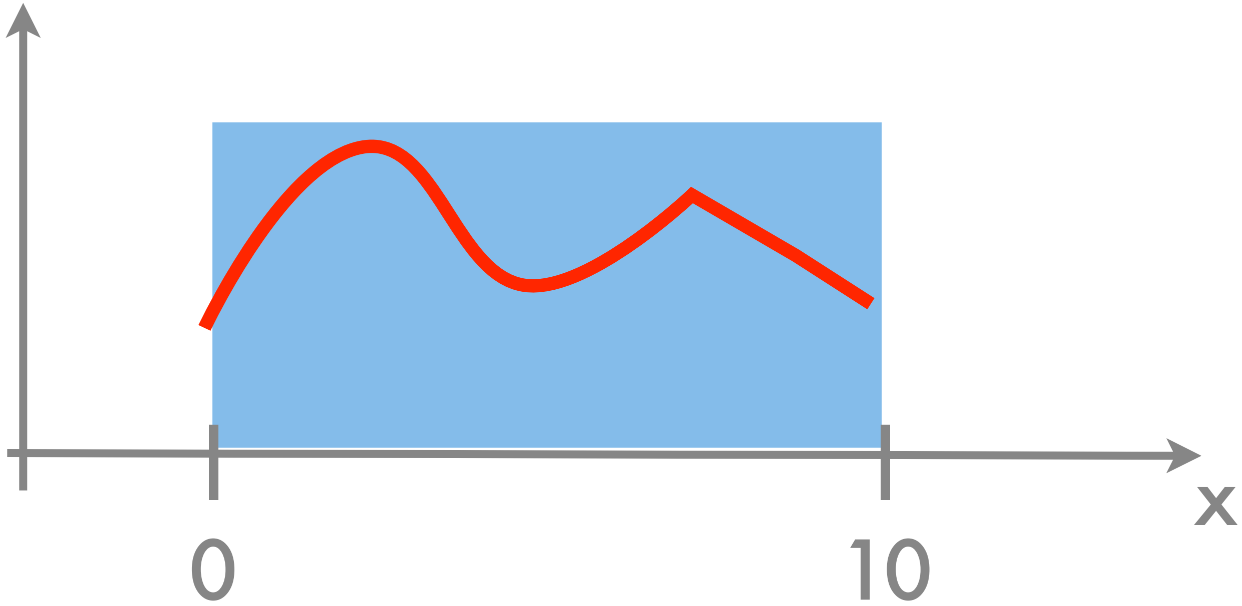
Von Neumann Accept-Reject

- Aim: Generate f(x) between 0 and 10



Von Neumann Accept-Reject

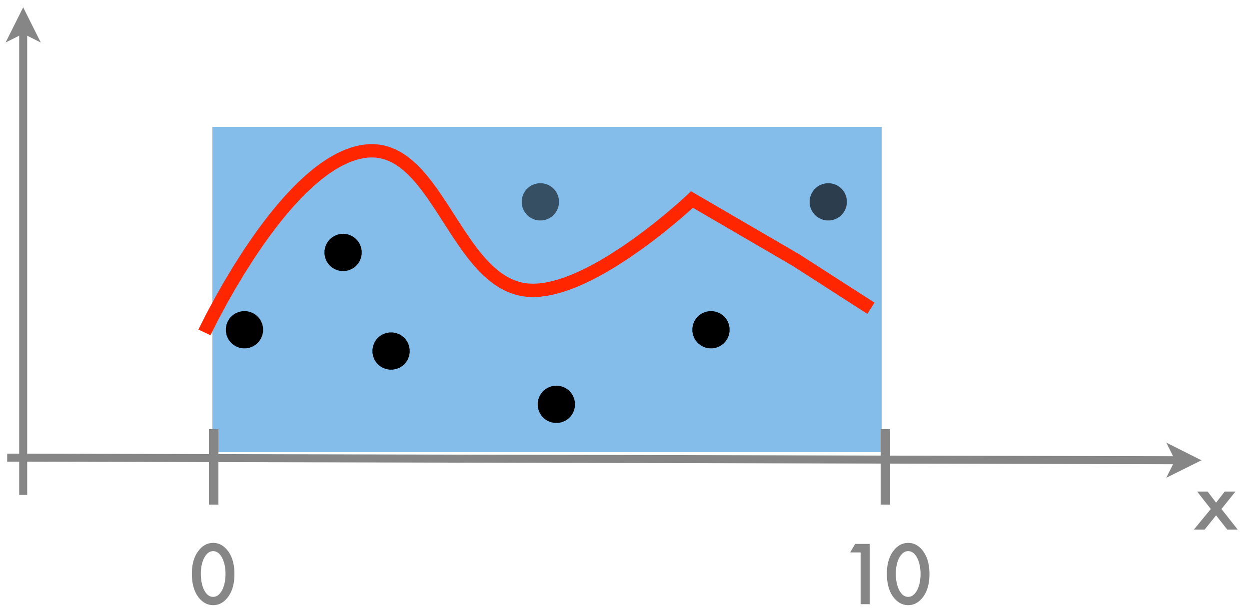
- Aim: Generate f(x) between 0 and 10



- Define a box from 0 and 10, such that f(x) is always below the box (i.e. you need to know f(x)'s maximum in the are of interest).

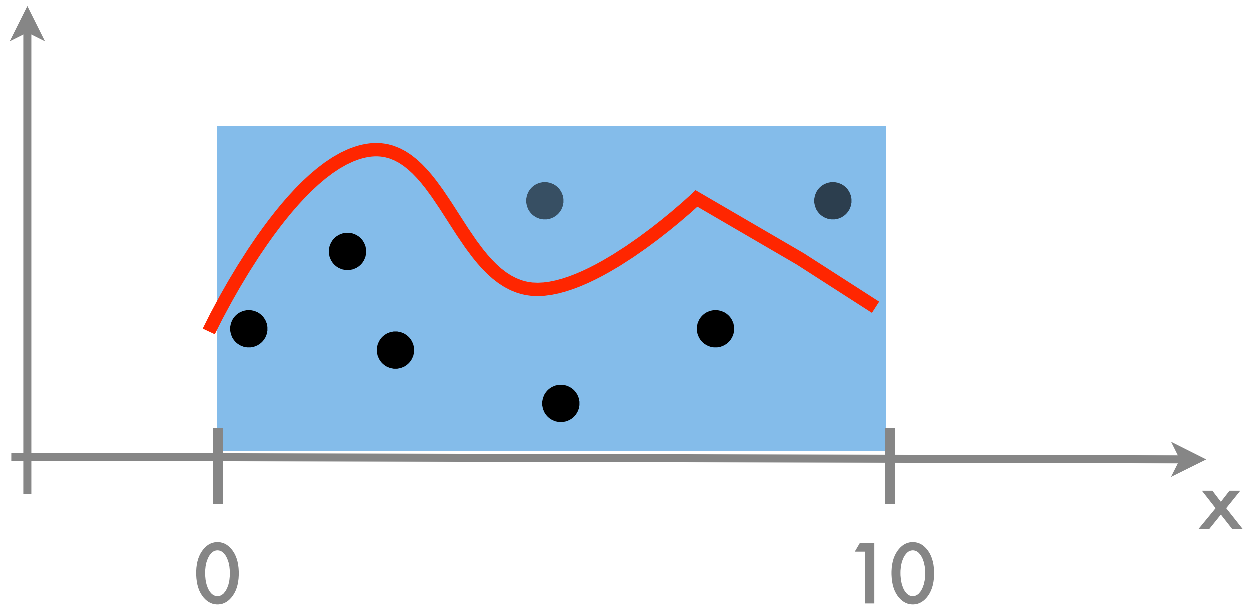
Von Neumann Accept-Reject

- Aim: Generate f(x) between 0 and 10



- Randomly shoot into the box. Accept those events that are below the red line.

Von Neumann Accept-Reject

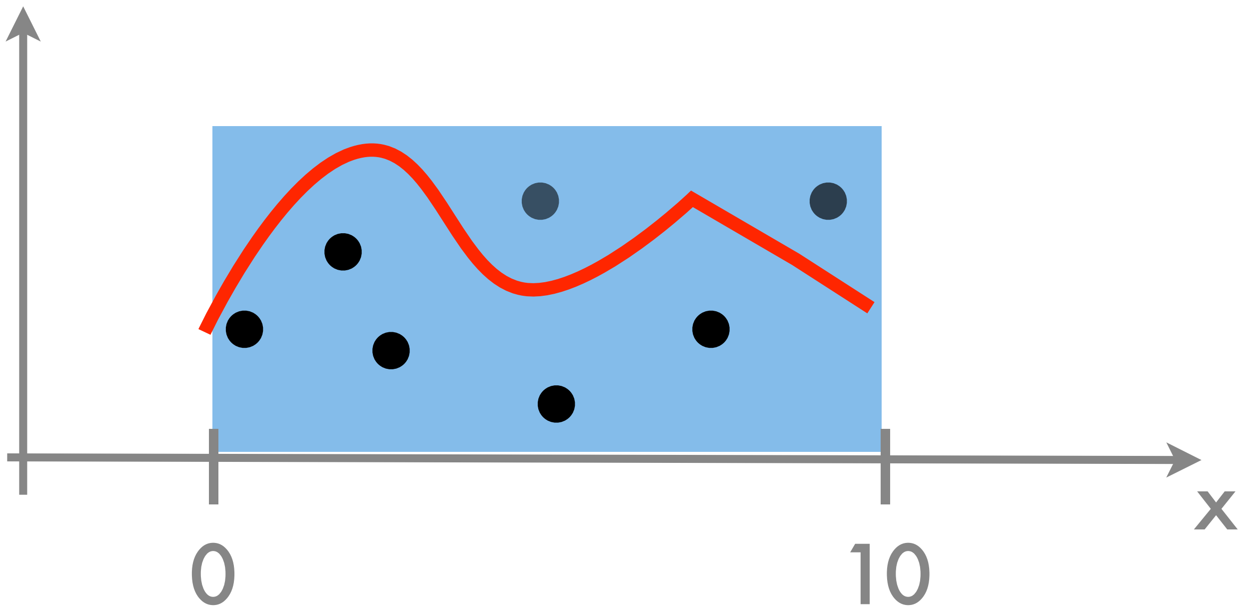


- $x = \text{rnd->Rndm()} \cdot 10;$

  $y = \text{rnd->Rndm()} \cdot \text{fmax};$

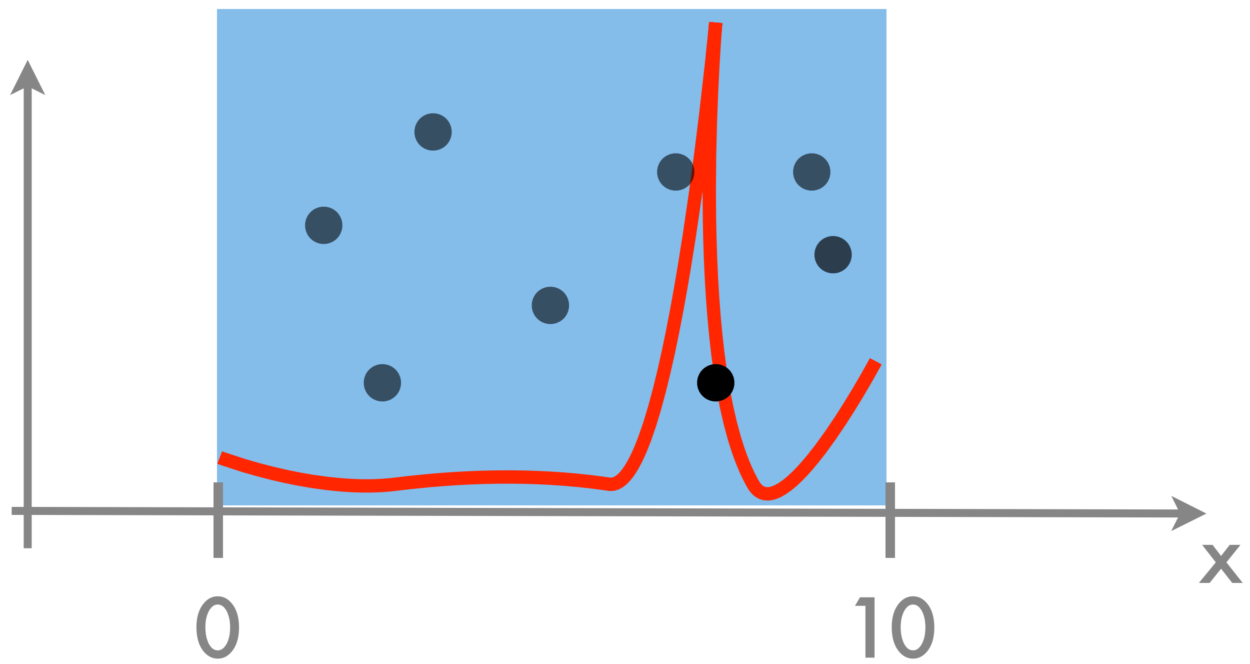
  $\text{if}(y < f(x))\ \text{acceptEvent}(x,y)$

MC-integration

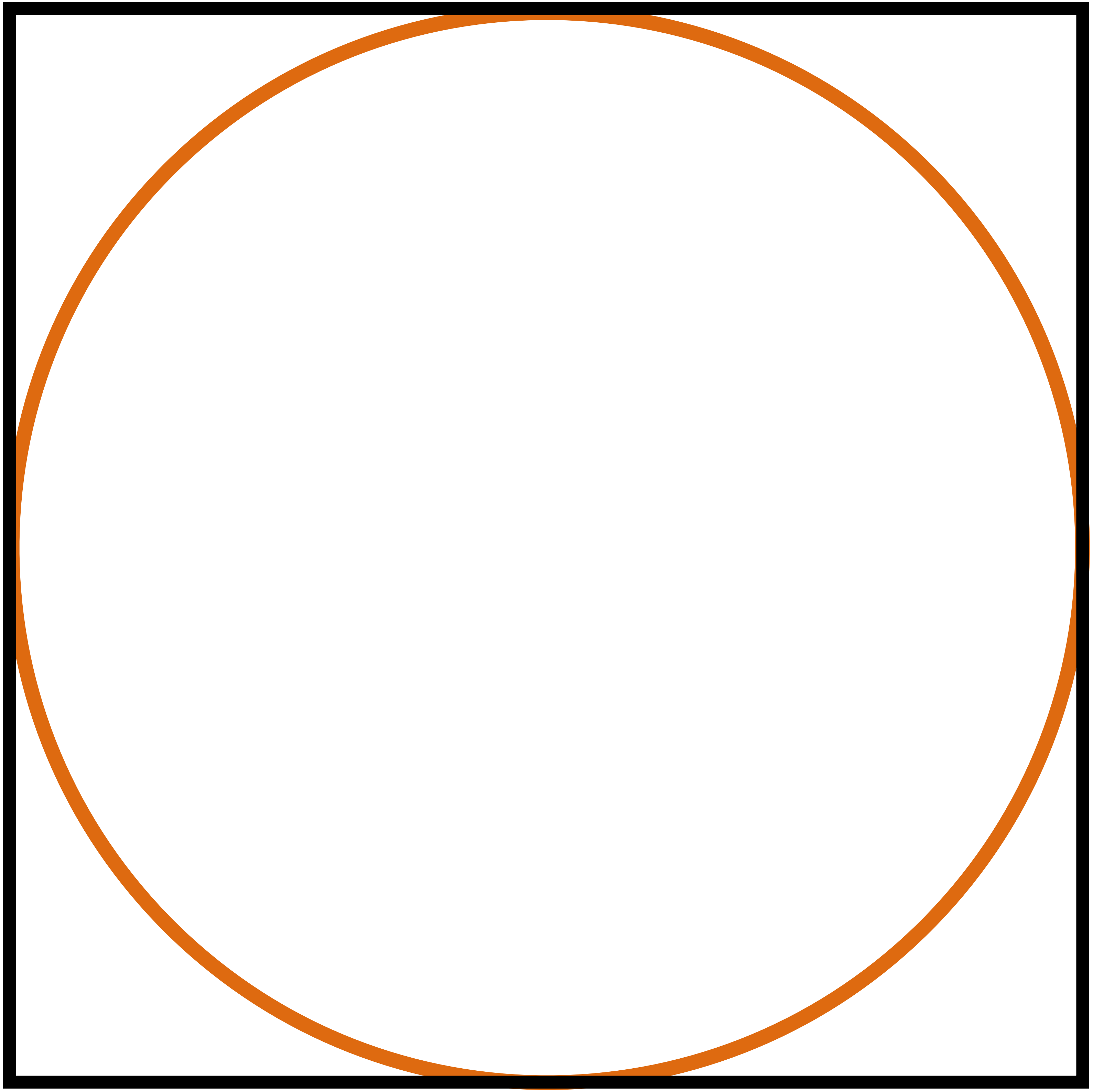


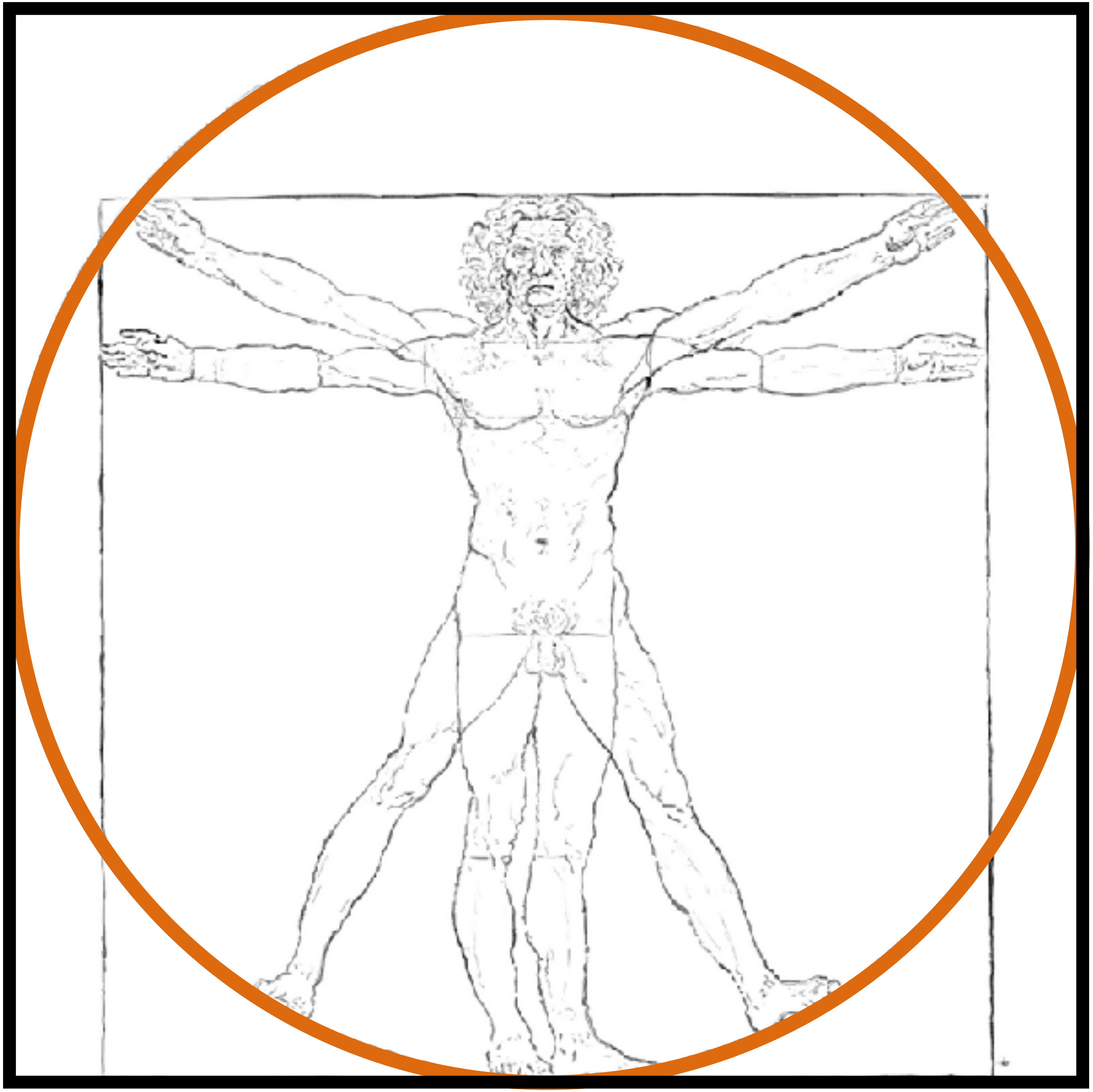
- This can be used for MC integration - the fraction of points accepted is $\propto$ to the area under the curve.
- This is the most efficient method of numerical integration in many dimensions (say more than 3).

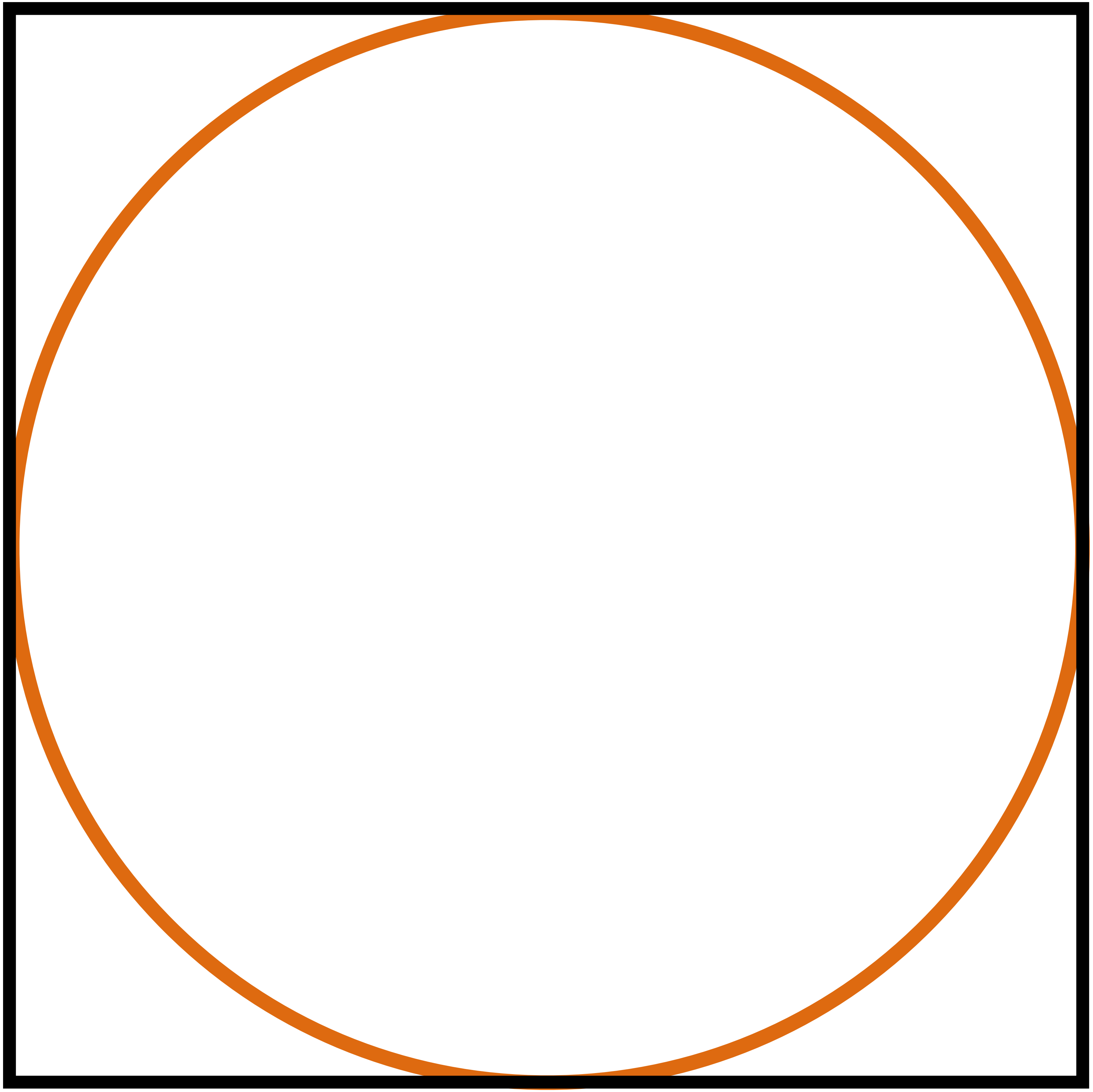
Von Neumann Accept-Reject

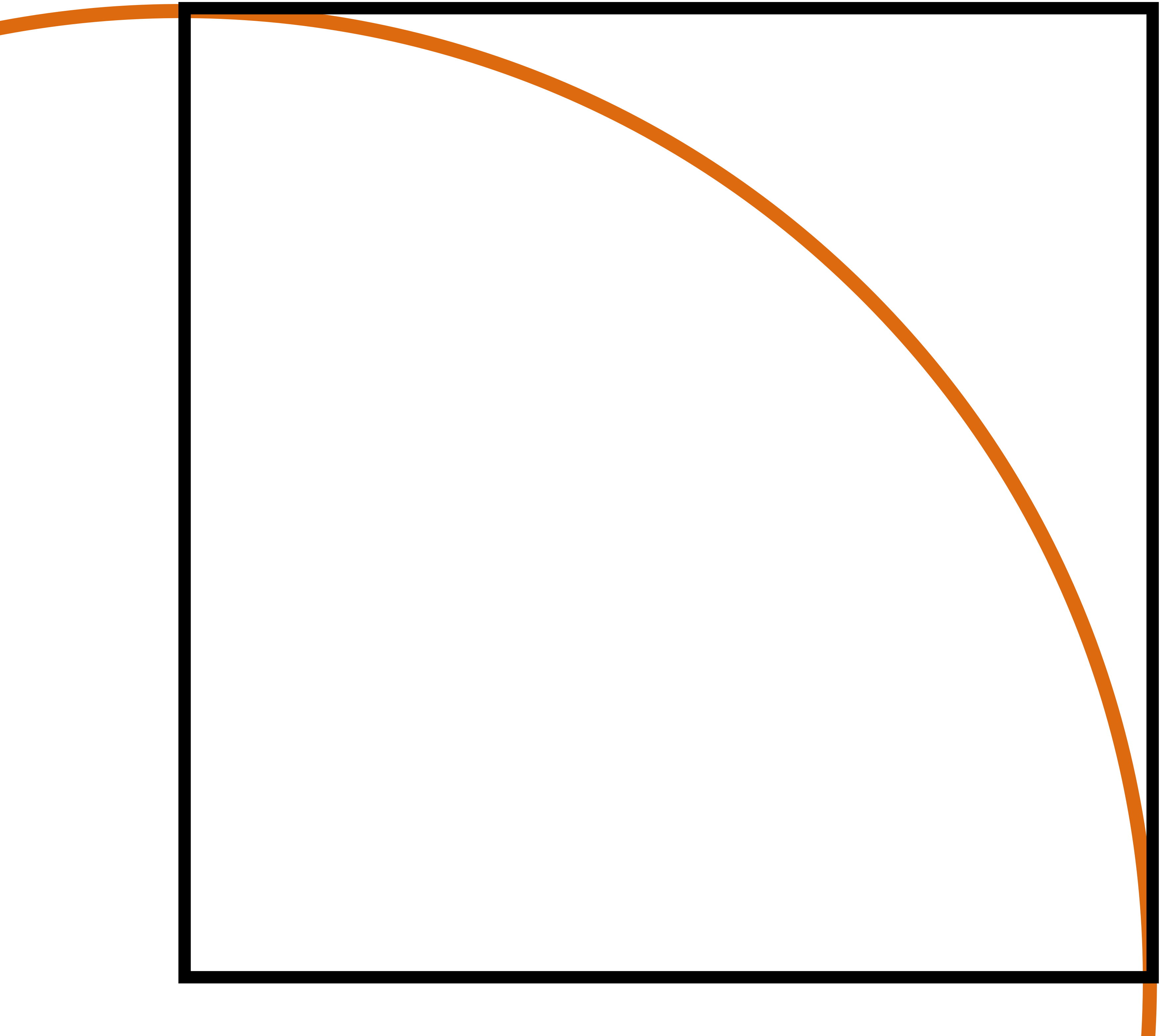


- Can be very inefficient for peaky distributions









Resources for today's problem classes:

This course:

https://tinyurl.com/TeshepProblems

https://tinyurl.com/TeshepStatCode

https://tinyurl.com/TeshepMC

External:

http://jupyter.readthedocs.io/en/latest/install.html

https://docs.anaconda.com/anaconda/

The End

Backup

Worked example

- In a large medical trial on 10,000 patients, 100 people would be expected to die without treatment. They find that with treatment, 80 die.

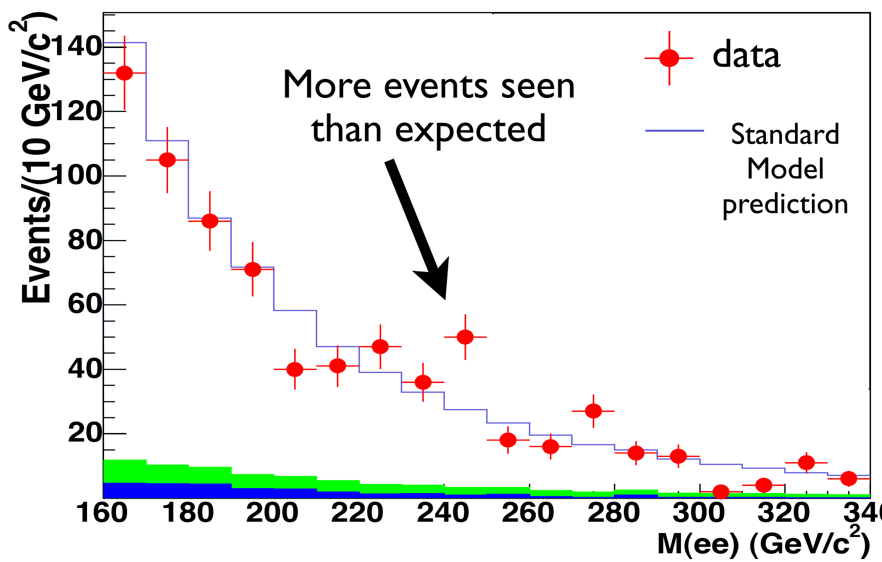
- Is this significant?

https://tinyurl.com/TeshepProblems

Calculate significances

- Estimate the significance of this observation:
  - Step 1: calculate the probability so see an upward fluctuation this big or bigger in the Standard Model, in this one bin
  - Step 2: take into account that they looked in 84 bins (tricky!)
- You should get a fairly small number. Why, do you think, have you not read in the news about the discovery of the Z' at CDF?

Z' search at CDF



- In the bin with the arrow, we expect 28 events without the Z'
- See 48 events.

Worked example

- **In a large medical trial on 10,000 patients, 100 people would be expected to die without treatment. They find that with treatment, 80 die.**
- **Is this significant?**
- **Now you learn that the researches have performed performed 50 such trials with different medicines, and only published the one that looked like a success. Why does this affect the significance of this result? Calculate how likely it is to have at least one result with 80 or fewer death**

Also, have a look at this: **http://tinyurl.com/y837ke92**

Problem 2

- 0.01% of the population is infected with a nasty, contagious virus

  A test for this virus is developed. This test identifies correctly 100% of those carrying the virus. Amongst those that do not carry the virus, it gives the correct result in 99.8% of the cases.

- If you test positive, how worried should you be? Are you likely to be infected?

Problem 2

- 0.01% of the population is infected with a nasty, contagious virus

  A test for this virus is developed. This test identifies correctly 100% of those carrying the virus. Amongst those that do not carry the virus, it gives the correct result in 99.8% of the cases.

- If you test positive, how worried should you be? Are you likely to be infected?

- Task: calculate how likely you are infected if the test is positive

Probability

Probability

- **As an average UK citizen, at the age of 20, the probability that you die within a year is 0.048%.**

Probability

- **As an average UK citizen, at the age of 20, the probability that you die within a year is 0.048%.**

- **But who is average?**

Probability

- As an average UK citizen, at the age of 20, the probability that you die within a year is 0.048%.
- But who is average?
- If you are female, it is only 0.026% (male: 0.069%)

Probability

- As an average UK citizen, at the age of 20, the probability that you die within a year is 0.048%.
- But who is average?
- If you are female, it is only 0.026% (male: 0.069%)
- If you are a male in Scotland, it is 0.1%

Probability

- As an average UK citizen, at the age of 20, the probability that you die within a year is 0.048%.
- But who is average?
- If you are female, it is only 0.026% (male: 0.069%)
- If you are a male in Scotland, it is 0.1%
- But what if you smoke? If you don't? If you are a heroin-addicted bomb-disposal expert?

What is Probability?

- Mathematically: Defines basic properties such as $0 \leq P \leq 1$ and calculation rules; all other definitions must satisfy also this one. But: No meaning.

- Frequentist: How many times n_E does something (event E) happen if I try N times? $P(E) = n_E/N$ for $N \to \infty$
Problem: What if I can try only once?

- Bayesian: Probability is a measure for the "degree of belief" that event E happens. One possible definition: I'd bet up to € n_E that E happens, if I get € N if I win: $P(E) = (£\ n_E)/(£\ N)$.
Problem: Subjective (not good for science, but occasionally unavoidable, e.g. for systematics.)

Probabilities nomenclatura

- P(A) = probability that A happens
- P(A or B) = probability that A happens, or B happens, or both.
- P(A & B) = P(A and B) probability that both A and B happen.
- P(A|B) = “P of A given B”, the probability that A happens given that B happens.
  - Note: while P(A & B) = P(B & A), P(A or B) = P(B or A), P(A|B) ≠ P(B|A), for example:
    P(pregnant | woman) ≈ a few %
    P(woman | pregnant) ≈ 100%

Bayes Theorem

- **$P(A \text{ and } B) = P(A)\, P(B|A) = P(B)\, P(A|B)$**

  - **From this follows Bayes' theorem:**

    $P(A|B) = P(B|A)\, P(A)/P(B)$

Bayes Theorem

Very important theorem. Also worth noting: This is not Bayesian statistics (every frequentist will happily use Bayes theorem)

- P(A and B) = P(A) P(B|A) = P

- From this follows Bayes' theorem:

  P(A|B) = P(B|A) P(A)/P(B)

Problem

- 0.01% of the population is infected with a nasty, contagious virus

  A test for this virus is developed. This test identifies correctly 100% of those carrying the virus. Amongst those that do not carry the virus, it gives the correct result in 99.8% of the cases.

- If you test positive, how worried should you be? Are you likely to be infected?

Problem

- 0.01% of the population is infected with a nasty, contagious virus

  A test for this virus is developed. This test identifies correctly 100% of those carrying the virus. Amongst those that do not carry the virus, it gives the correct result in 99.8% of the cases.

- If you test positive, how worried should you be? Are you likely to be infected?

- Task: calculate how likely you are infected if the test is positive

Bayes Theorem

- P(A|B) = P(B|A) P(A)/P(B)

- P(person carries virus GIVEN test says person carries virus) =

  P(test says person carries virus GIVEN person carries virus

  × P(person carries virus)
  / P(test says person carries virus)

Bayes Theorem

- $P(A|B) = P(B|A)\,P(A)/P(B)$

- P(person carries virus GIVEN test says person carries virus) =

  ~~P(test says person carries virus GIVEN person carries virus~~ 1

  × P(person carries virus)

  / P(test says person carries virus)

Bayes Theorem

- P(A|B) = P(B|A) P(A)/P(B)

- P(person carries virus GIVEN test says person carries virus) =

  P(test says person carries virus GIVEN person carries virus 1

  × P(person carries virus) 0.0001
  / P(test says person carries virus)

Bayes Theorem

- P(test says person carries virus)

  = P(test says person carries virus & person carries virus)
  + P(test says person carries virus & person does not virus)

  = P(person carries virus)
  × P(test says person carries virus GIVEN person carries virus)

  + P(person does not carry virus)
  × P(test says person carries virus GIVEN person does not virus)

  = 0.0001 × 1
  + 0.9999 × 0.002

  = 0.0021

Bayes Theorem

- P(A|B) = P(B|A) P(A)/P(B)

- P(person carries virus GIVEN test says person carries virus) =

  ~~P(test says person carries virus GIVEN person carries virus~~ 1

  × ~~P(person carries virus)~~ 0.0001

  / P(test says person carries virus)

Bayes Theorem

- P(A|B) = P(B|A) P(A)/P(B)

- P(person carries virus GIVEN test says person carries virus) =

  ~~P(test says person carries virus GIVEN person carries virus~~ 1

  × ~~P(person carries virus)~~ 0.0001

  / P(test says person carries virus)

=0.047 = 4.7%

Bayes Theorem

- P(A|B) = P(B|A) P(A)/P(B)

- P(person carries virus GIVEN test says person carries virus) =

  ~~P(test says person carries virus GIVEN person carries virus~~ 1

  × ~~P(person carries virus)~~ 0.0001

  / ~~P(test says person carries virus)~~ 0.0021

=0.047 = 4.7%

Bayes Theorem

- P(A|B) = P(B|A) P(A)/P(B)

- P(person carries virus GIVEN test says person carries virus) =

P(test says person carries virus GIVEN person carries virus 1

× P(person carries virus) 0.0001

/ P(test says person carries virus) 0.0021

=0.047 = 4.7%

This means >95% of the patients with a positive test are in fact healthy.

Bayes Theorem

- $P(A|B) = P(B|A)\, P(A)/P(B)$

- P(person carries virus GIVEN test says person carries virus) =

  ~~P(test says person carries virus GIVEN person carries virus~~ 1

  × ~~P(person carries virus)~~ 0.0001
  / ~~P(test says person carries virus)~~ 0.0021

=0.047 = 4.7%

This means >95% of the patients with a positive test are in fact healthy.

The test is complete rubbish!!

Pull study

- **Simulate a lot of datasets using Monte-Carlo simulation.**
- **Fit each dataset and calculate the**

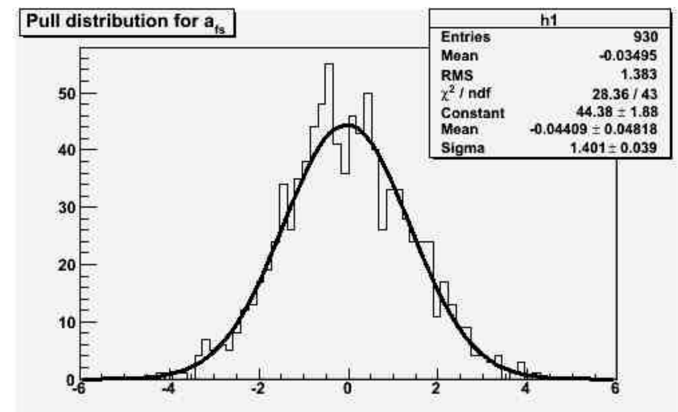
pull = $$\frac{(\text{fit result}) - (\text{true value})}{(\text{error estimate})}$$

and put it in a histogram.

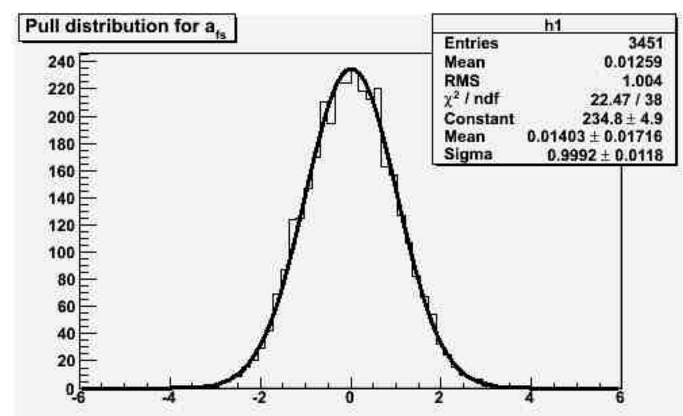
- **For a good, unbiased fitter, you get:**

$$\text{Mean} = 0 \pm \frac{1}{\sqrt{N_{\text{exp}}}}$$
$$\sigma = 1 \pm \frac{1}{\sqrt{2N_{\text{exp}}}}$$

σ=1.4 for 1k events ⇒ wrong errors



σ=1.0 for 1k events ⇒ correct errors



Monte Carlo

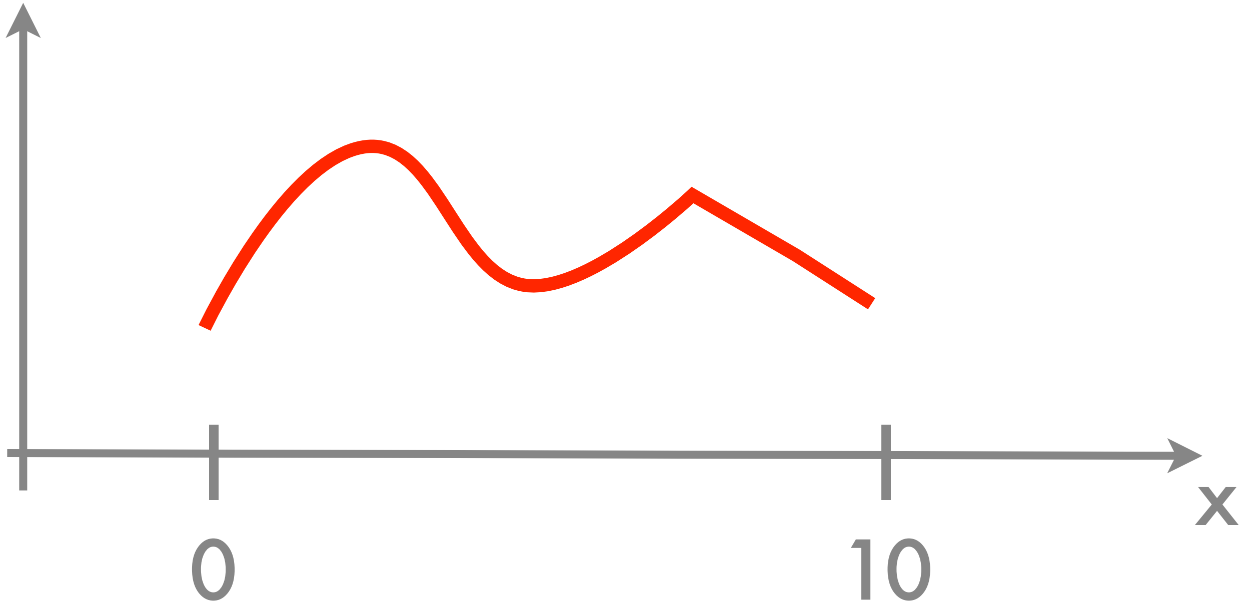
Gamblers in the casino at Monte-Carlo. c.1910

Monte Carlo Simulations

- To test your fit, you need to try it out on simulated data.
- To really test it properly, you cannot rely on the experiment's detailed simulation - you want to run thousands of simulated experiments and see if your fitter behaves as expected. You need a simplified, fast Monte Carlo for that.
- Today:
    - How do generate any distribution
    - How to do it a bit more efficiently

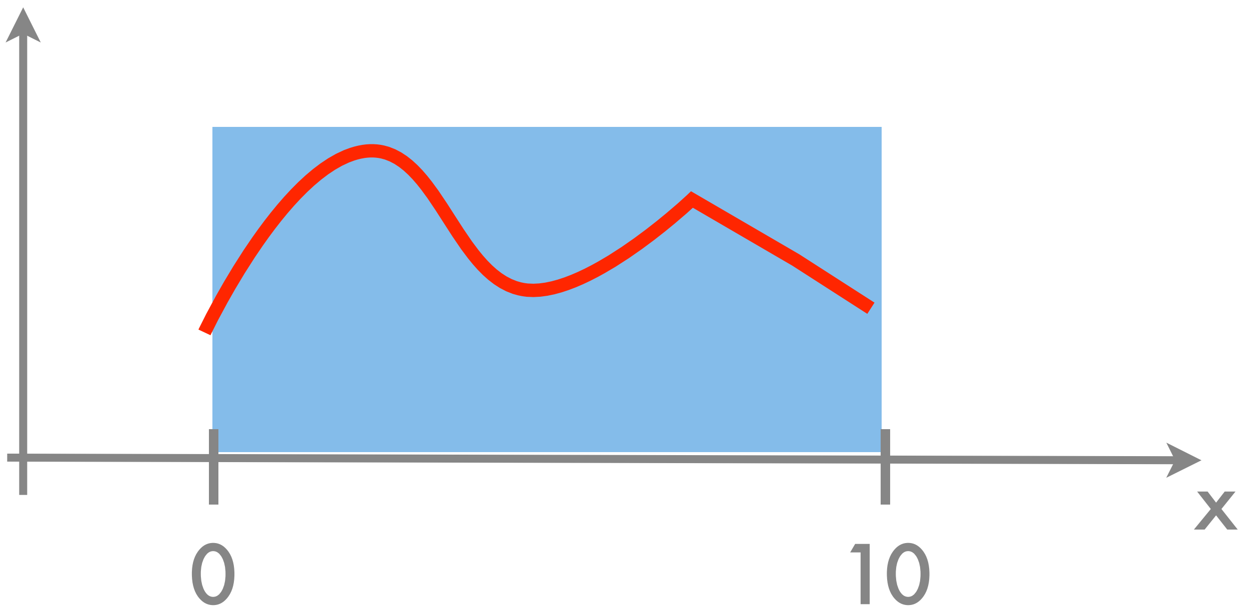
Von Neumann Accept-Reject

- Aim: Generate f(x) between 0 and 10



Von Neumann Accept-Reject

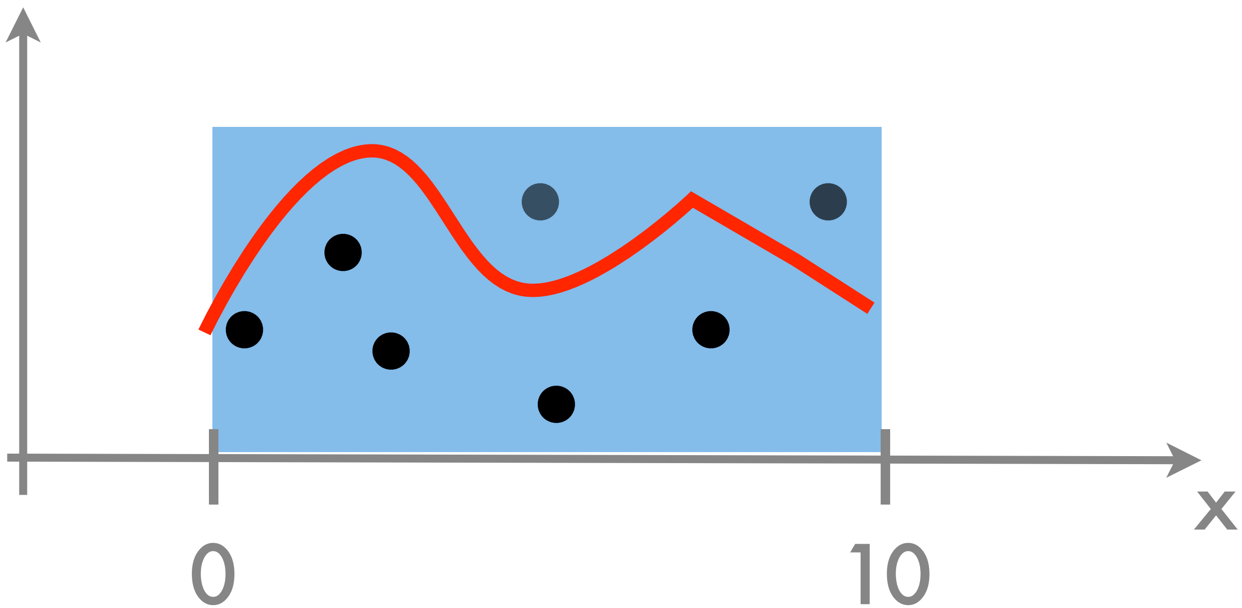
- Aim: Generate f(x) between 0 and 10



- Define a box from 0 and 10, such that f(x) is always below the box (i.e. you need to know f(x)'s maximum in the are of interest).

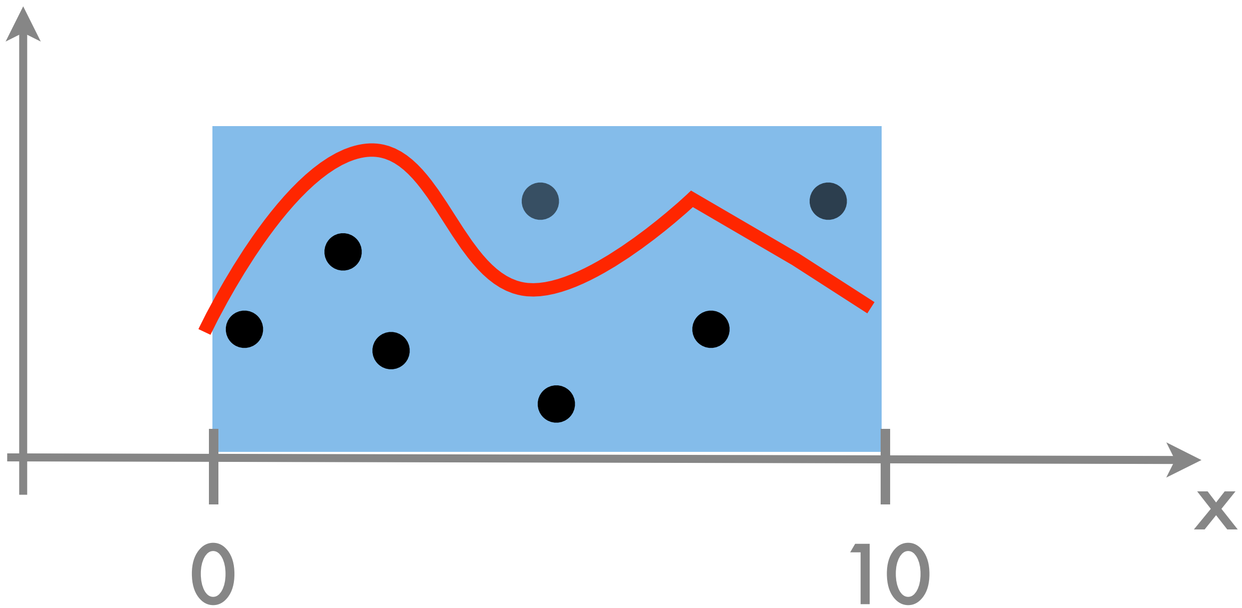
Von Neumann Accept-Reject

- Aim: Generate f(x) between 0 and 10



- Randomly shoot into the box. Accept those events that are below the red line.

Von Neumann Accept-Reject

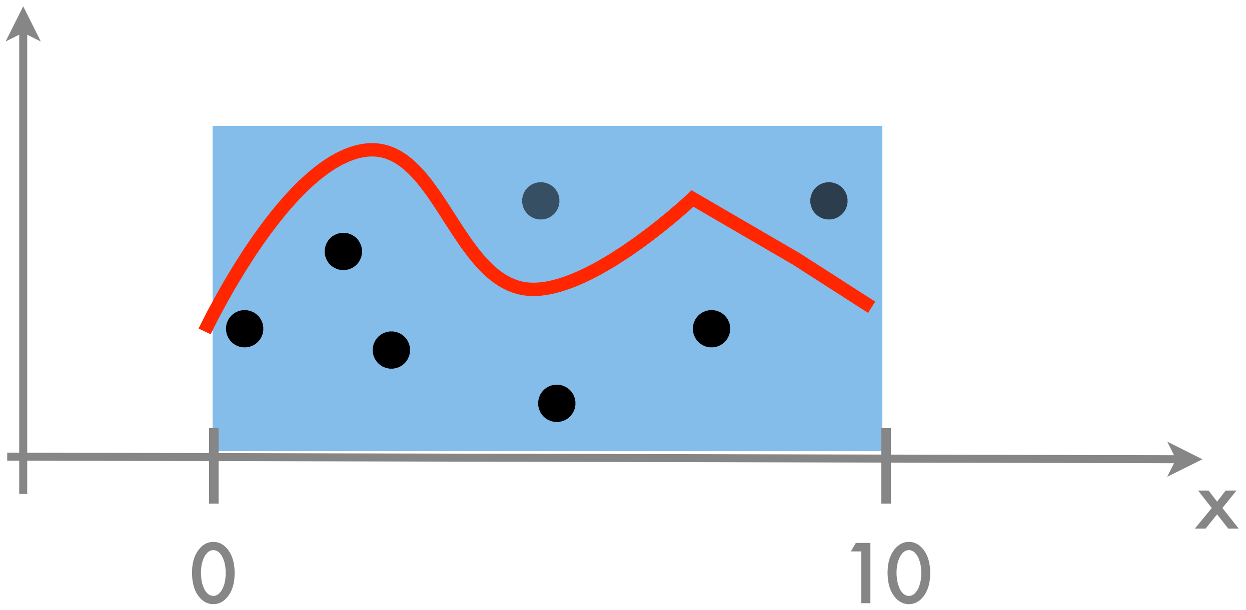


- x = rnd->Rndm() · 10;

  y = rnd->Rndm() · fmax;

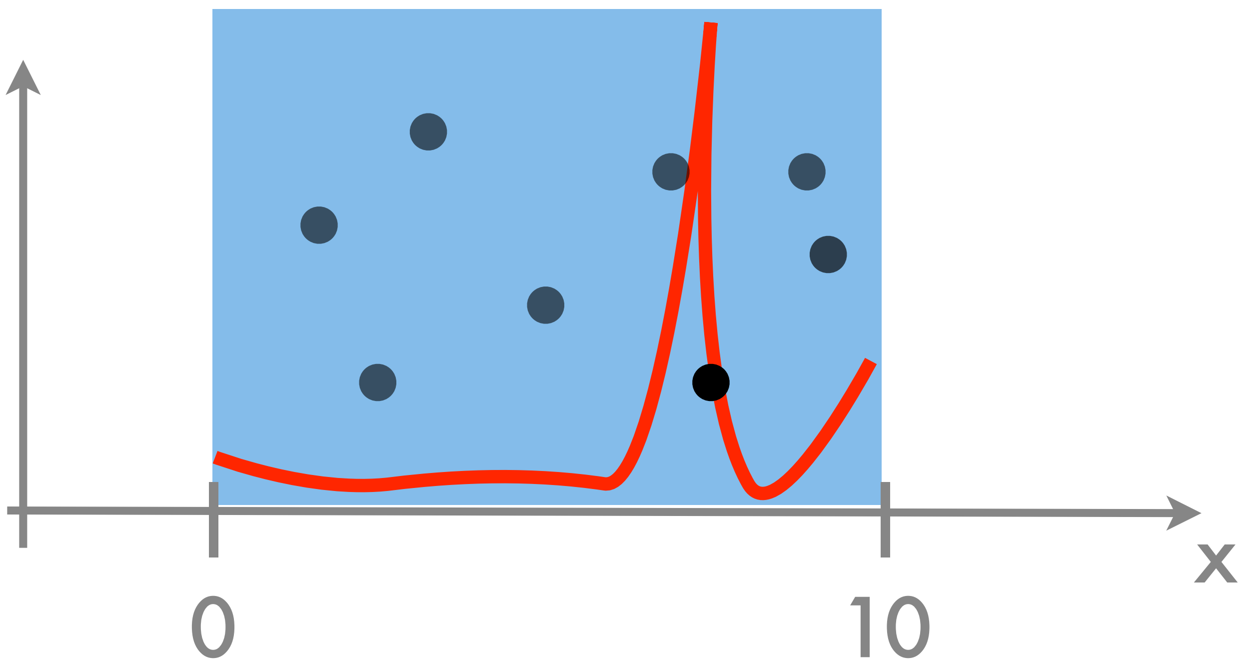
  if(y < f(x)) acceptEvent(x,y)

MC-integration

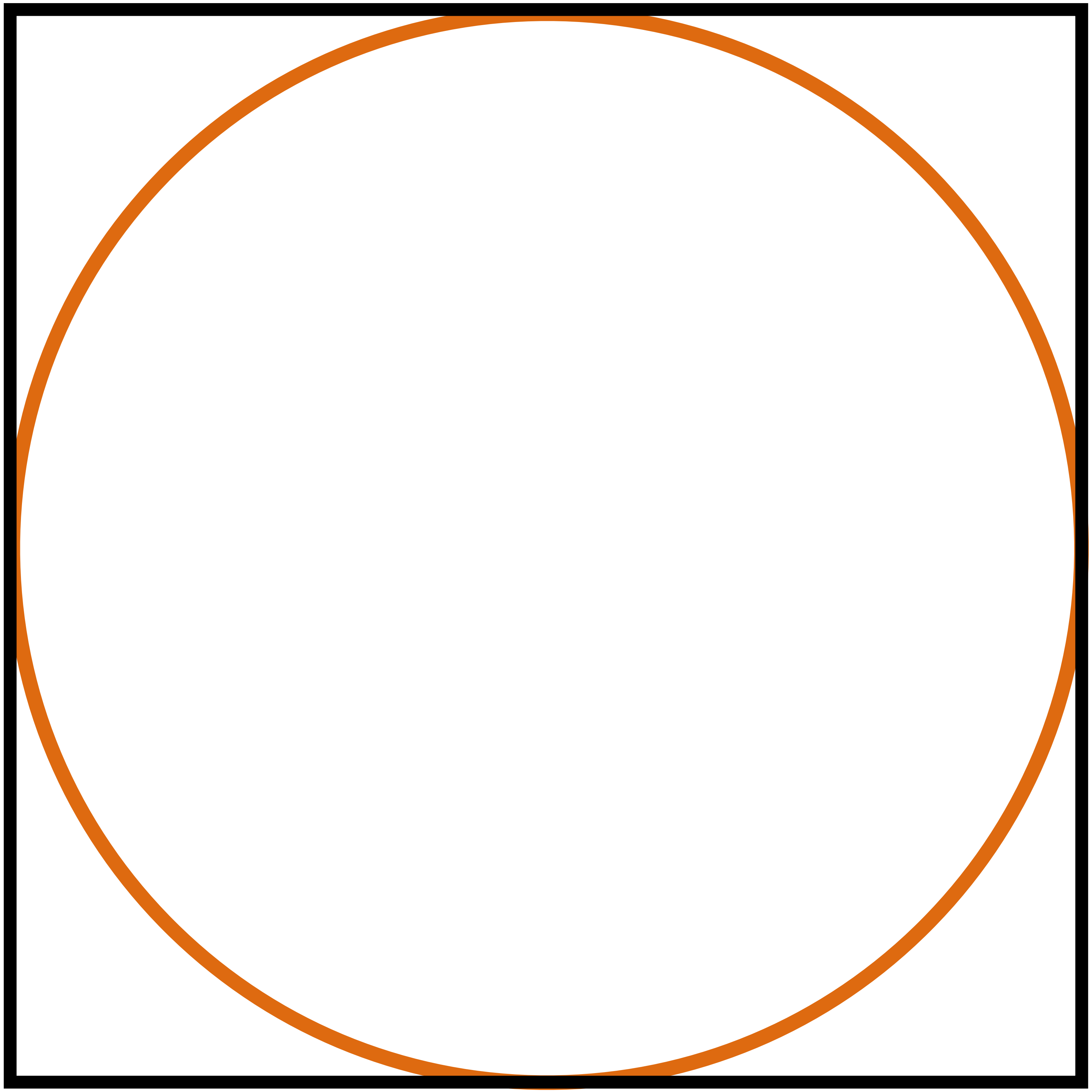


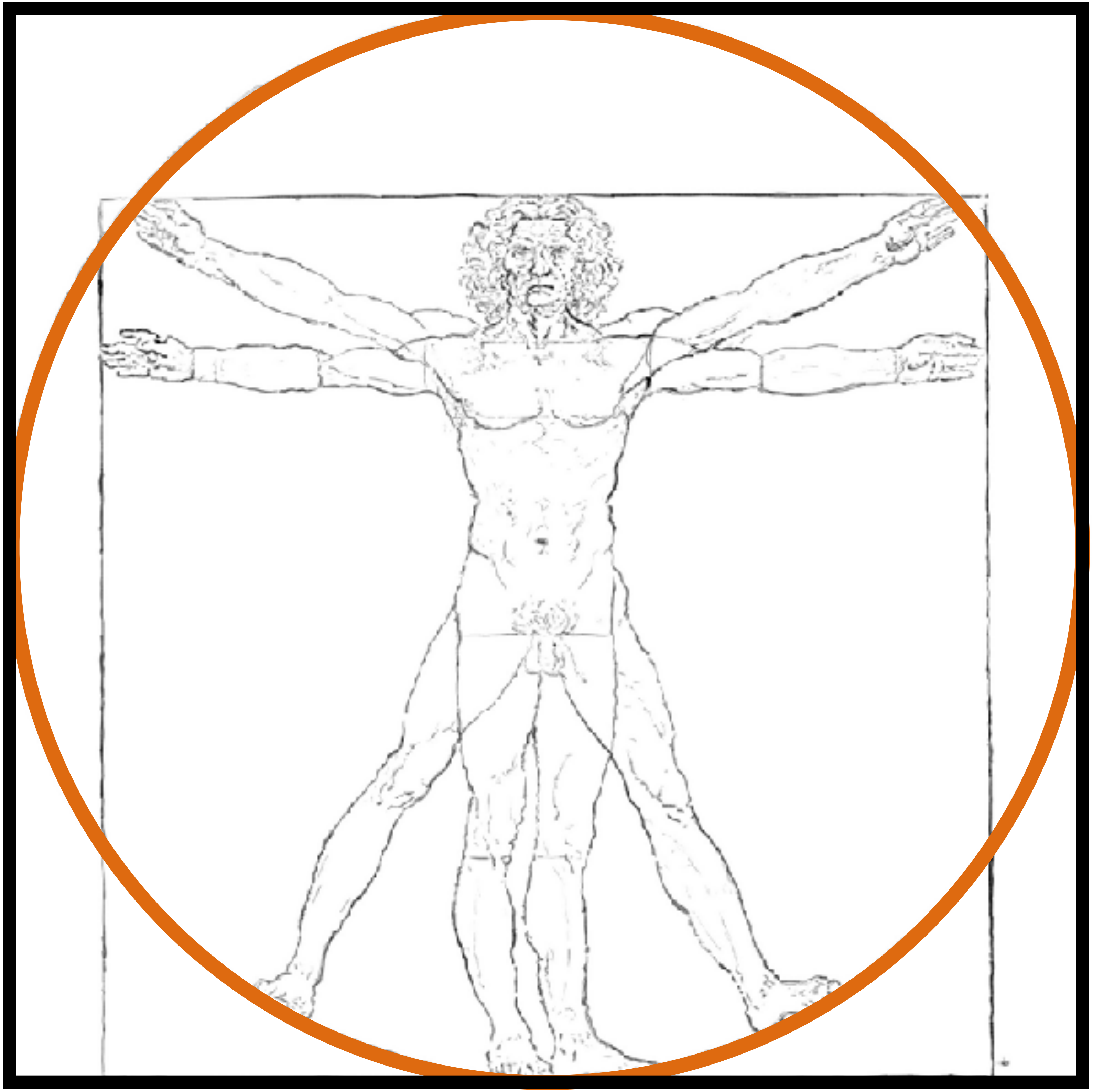
- This can be used for MC integration - the fraction of points accepted is ∝ to the area under the curve.
- This is the most efficient method of numerical integration in many dimensions (say more than 3).

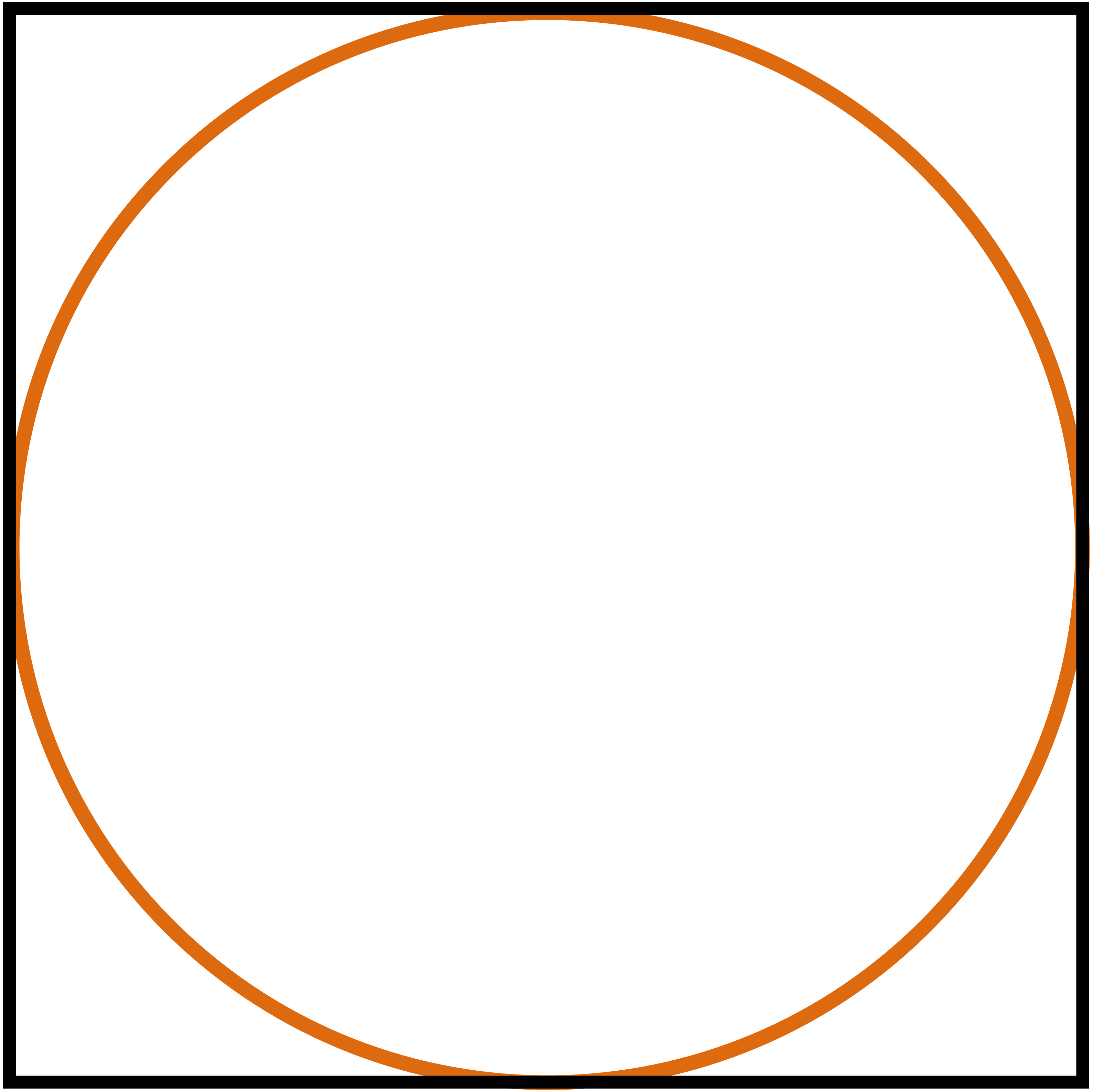
Von Neumann Accept-Reject

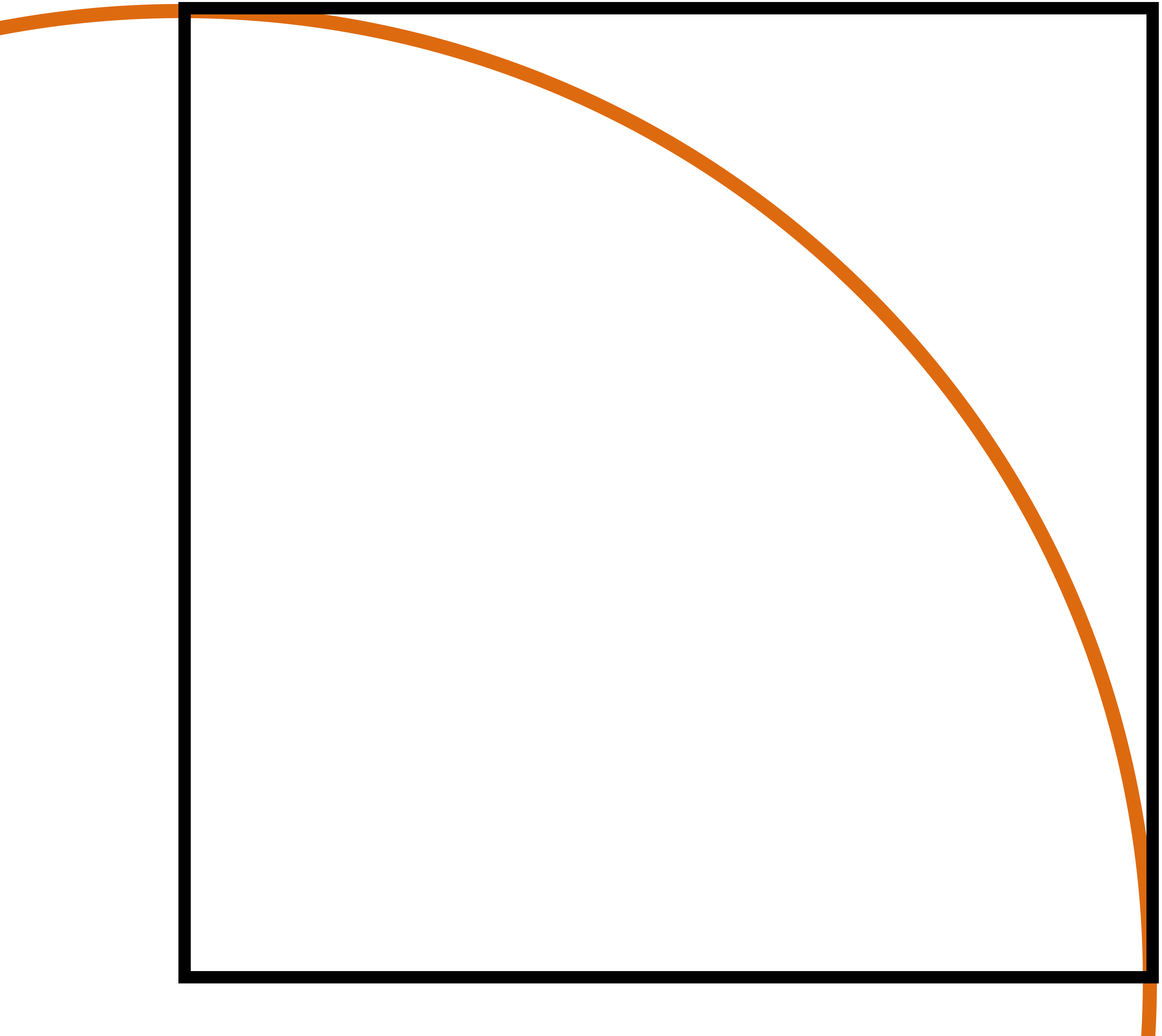


- Can be very inefficient for peaky distributions









Loss-free generation

- Assume you start from a random number generator that generates a flat distribution between 0 and 1.
- Task: Generate a an exponential without having to reject any events.
- Trick: Solve this (for x = flat distribution) for t:

$$P(t) \quad = \quad P(x)\frac{dx}{dt}$$

Using your MC to test your fit

Homework

https://tinyurl.com/TeshepProblems

The government gets tough on crime.

Because most violent crime takes place within the closest circle of friends and family, it is decided that anybody above the age of 18 who wants to engage in any kind of personal relationship must first obtain a permit to do so. The decision whether a permit is granted is based on a detailed background check.

When the method is tested on a sample of known violent offenders and another sample of innocent people, it seems to work surprisingly well: 80% of violent offenders are refused the permit. Only 0.1% of non-violent people are refused the permit.

Assume that in 2084, 1 in 10,000 of the adult population is (criminally) violent, and that violent and non-violent people are equally likely to ask for a permit.

What fraction of those who are refused a permit are in fact non-violent?

Further examples

- A Ring Imaging CHerenkov (RICH) detector differentiates pions from kaons. It identifies 85% of pions correctly, and 85% of kaons correctly. At a typical LHC collision, 95% of particles passing through the detector are pions.

  Given the RICH identifies a kaon, what is the probability that it is a kaon?

Further examples

- A Ring Imaging CHerenkov (RICH) detector differentiates pions from kaons. It identifies 85% of pions correctly, and 85% of kaons correctly. At a typical LHC collision, 95% of particles passing through the detector are pions.

  Given the RICH identifies a kaon, what is the probability that it is a kaon?

  Assume there are only kaons and pions, for simplicity, so we have 5% kaons.

  Then the answer is 22%   (it's even less if you take into account other particles)

  So most particles identified as kaons are pions. You need a better detector (or at least adjust your particle ID cuts for higher kaon purity)

Medical Study

a) In a large medical trial on 10,000 patients, 100 people would be expected to die without treatment. They find that with treatment, 80 die. Is this significant? To estimate how much evidence for the treatment this constitutes, calculate the probability to find 80 or fewer events when one expects 100.

b) Now you learn that the researches have performed performed 50 such trials with different medicines, and only published the one that looked like a success. Why does this affect the significance of this result? Calculate how likely it is to have at least one result with 80 or fewer death

Also, have a look at this:

http://tinyurl.com/y837ke92

Philosophical Musings

What are probabilities anyway?

Frequentists vs Bayesians

Frequentists vs Bayesians

- Frequentist probability: $P(x) = (\text{Number of times x happens})/N$ for $N \rightarrow \infty$

- Bayesian: "degree of belief that x will happen", I'd bet N_x € if I get N € if x happens. $P(x) = N_x/N$

Frequentists vs Bayesians

- **Applied to fitting: Frequentists maximise the likelihood**

$$\mathcal{L}(\text{theory}) \equiv \prod_{\text{all data points}} P(\text{datapoint}_i|\text{theory})$$

- **This is in fact the probability to see the data, given the theory**

$$P(\text{data}|\text{theory}) = \prod_{\text{all data points}} P(\text{datapoint}_i|\text{theory})$$

- **But wouldn't we rather want know the probability of the theory given the data, i.e. *P*(theory | data)?**
    - **Note once more that this is very different!**
      P(sits in this room | scientist) ≈ few%
      P(scientist | sits in this room) ≈ 100%

Frequentists vs Bayesians

- **Frequentists:** $P(\text{data}|\text{theory})$

- **Bayes' theorem to the rescue:**

$$P(\text{theory}|\text{data}) = \frac{P(\text{data}|\text{theory})P(\text{theory})}{P(\text{data})}$$

- **Using variables again (it gets too messy otherwise):**

  - **τ = theory = fit parameter (say the mean lifetime of the D meson)**

  - **t_i = measured data** $P(\tau|t_i) = \dfrac{P(t_i|\tau)P(\tau)}{P(t_i)}$

Bayesian statistics

$$P(\text{theory}|\text{data}) = \frac{P(\text{data}|\text{theory})P(\text{theory})}{P(\text{data})}$$

$$P(\tau|t_i) = \frac{P(t_i|\tau)P(\tau)}{P(t_i)}$$

- **What are the terms?**
  - **P(data | theory) = our well-known likelihood**
  - **P(data) =** $P(t_i) = \int P(t_i|\tau)P(\tau)d\tau$
  - **P(theory)?**

Bayesian statistics

$$P(\text{theory}|\text{data}) = \frac{P(\text{data}|\text{theory})P(\text{theory})}{P(\text{data})}$$

$$P(\tau|t_i) = \frac{P(t_i|\tau)P(\tau)}{P(t_i)}$$

← Bayesians maximise this

- **P(theory) is your prior belief of what you expect, i.e. how likely you think given values of the true mean lifetime are before looking at the data. P(theory) is called “the prior”, while P(theory|data) is the posterior probability.**

Bayesian statistics

arbitrary prior

$$P(\text{theory}|\text{data}) = \frac{P(\text{data}|\text{theory})P(\text{theory})}{P(\text{data})}$$

$$P(\tau|t_i) = \frac{P(t_i|\tau)P(\tau)}{P(t_i)}$$

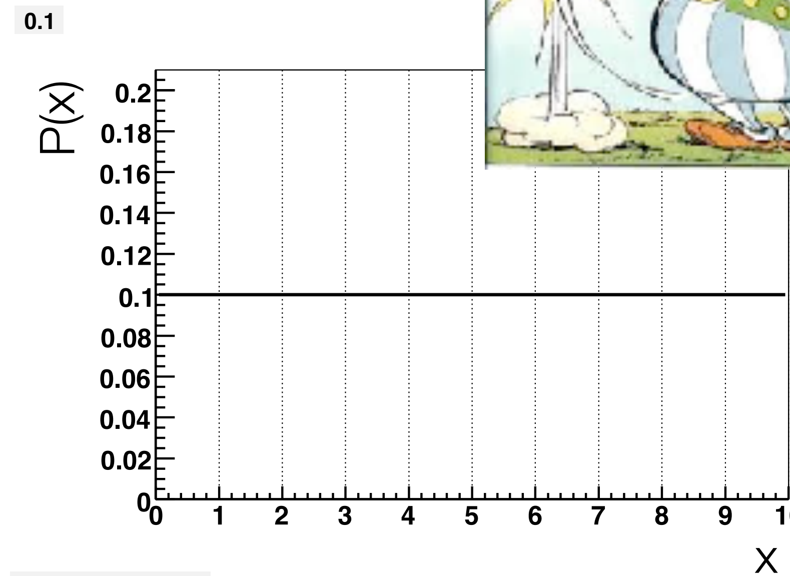
Bayesians maximise this

- **What prior you choose affects the result you get. There is no right choice of prior. Flat, $1/\sqrt{\tau}$, $\log(\tau)$ are all equally sensible. It also depends on your variable. <u>Flat in θ is not flat in cosθ</u>. What prior you choose is a matter of opinion.**

- **The good news is, for large statistics, results get less dependent on the prior, and tend towards the Frequentist result.**

Flat? Who's flat? Flat in x is not flat in y.

$$P(x) = \left\{ \begin{array}{cr} \frac{1}{10} & \text{between 0 and 10} \\ 0 & \text{otherwise} \end{array} \right\}$$



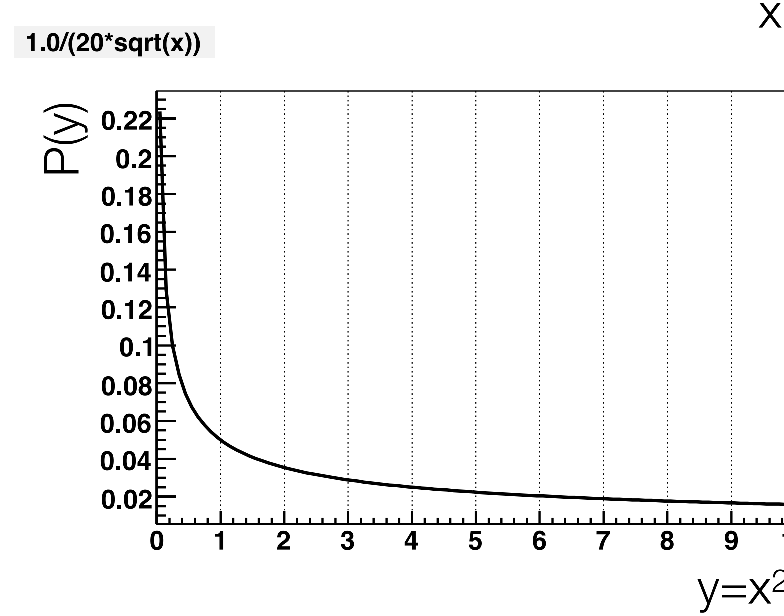
$$y = x^2 \Leftrightarrow x = \sqrt{y} \text{ for } x > 0$$

$$P(y)\, dy = P(x)\, dx$$

$$P(y) = P(x)\, \frac{dx}{dy}$$

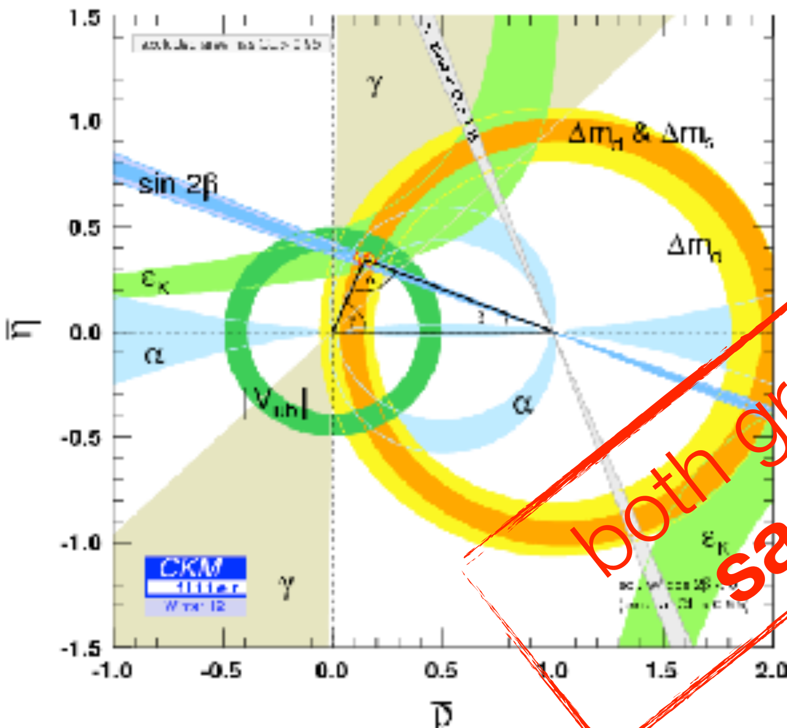
$$= P(x) \frac{1}{2\sqrt{y}}$$

$$= \frac{1}{20\sqrt{y}}$$

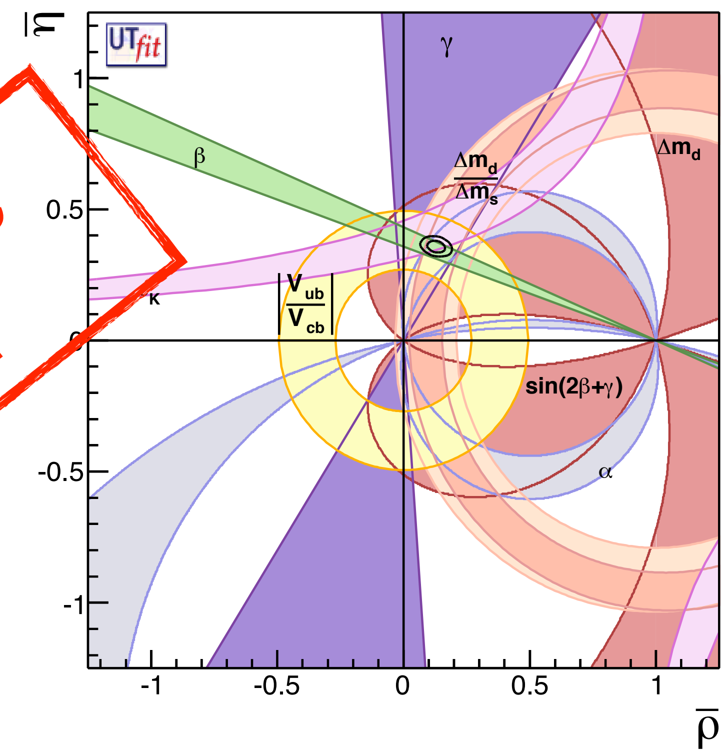


Frequentist vs Bayesian: It matters (for large errors)

CKM fitter (Frequentist)



UTFit (Bayesian)



both groups use **same** data!

$\gamma = 66^\circ \pm 12^\circ$

$\beta = 21^\circ \pm 1^\circ$

$\gamma = 76^\circ \pm 11^\circ$

$\beta = 21^\circ \pm 1^\circ$

Confidence Levels for Frequentists

- Measure the Higgs mass
- There is only one true value - that we can attempt to measure many times, with different results due to measurement errors and the intrinsic width of the Higgs.
- Frequentists might one day say: “The Higgs mass is within 120GeV and 130GeV at 90%CL”.
- Frequentists mean: If I keep repeating the experiment and follow the same prescription of defining 90% CL limits, the true Higgs mass will be inside these limits 90% of the time.

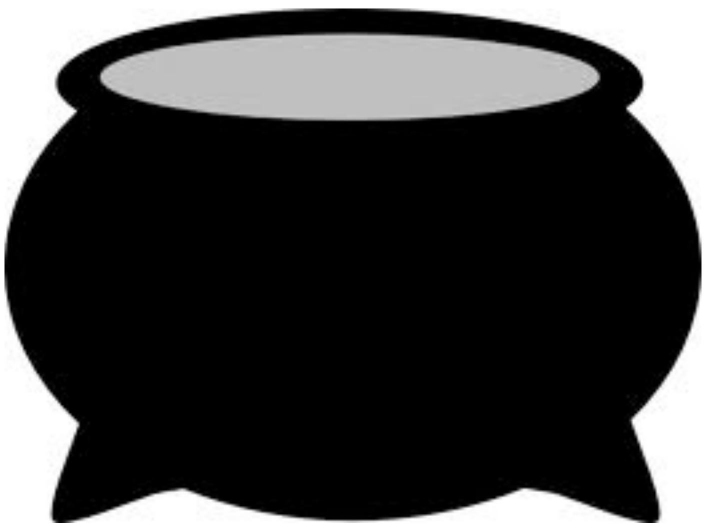
Frequentist vs Bayesian

- Frequentist probability: P(x) = (Number of times x happens)/N for $N \to \infty$
    - For frequentists, talking about P(theory), as in P(true mass of Higgs) does not make sense, because there is only one true mass of the Higgs - no ensemble, hence no frequentist probability. Instead, talk about the probability to find the data (repeatable) given the theory.
- Bayesian: "degree of belief that x will happen", I'd bet N_x quid if I get N quid if x happens. $P(x) = N_x/N$
    - Bayesians can quite happily talk about P(true mass of Higgs) and give a result for the most likely mass of the Higgs, but unfortunately this result is just an opinion.

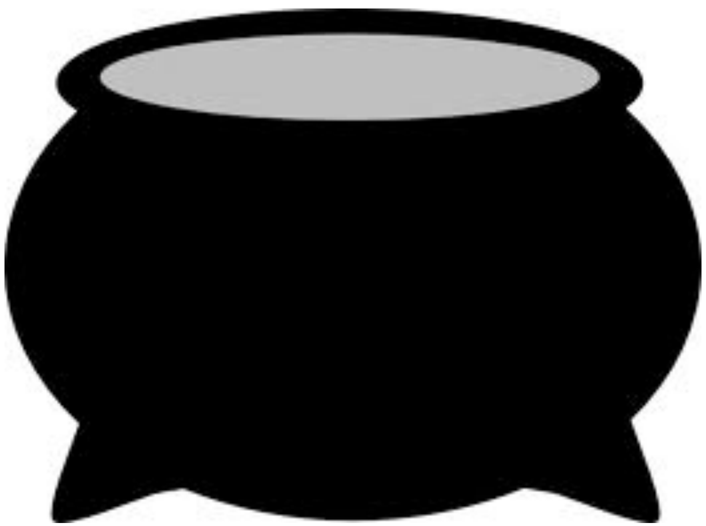
Confidence Levels for Frequentists: Coverage

- Coverage: If I say the true value is within the A and B at a given confidence value of p (say 90%) I must be right in p (say 90%) of the time,
- If I repeat the experiment N times, as $N \to \infty$, the true value is inside the 90% confidence limit 0.9*N times, and is outside 0.1*N times.
- Getting it right is called exact coverage
- If it is outside more often, this is called under-coverage
- If it is inside more often, this is overcoverage.
- Ideally: Achieve exact coverage. Overcoverage better than under-coverage.

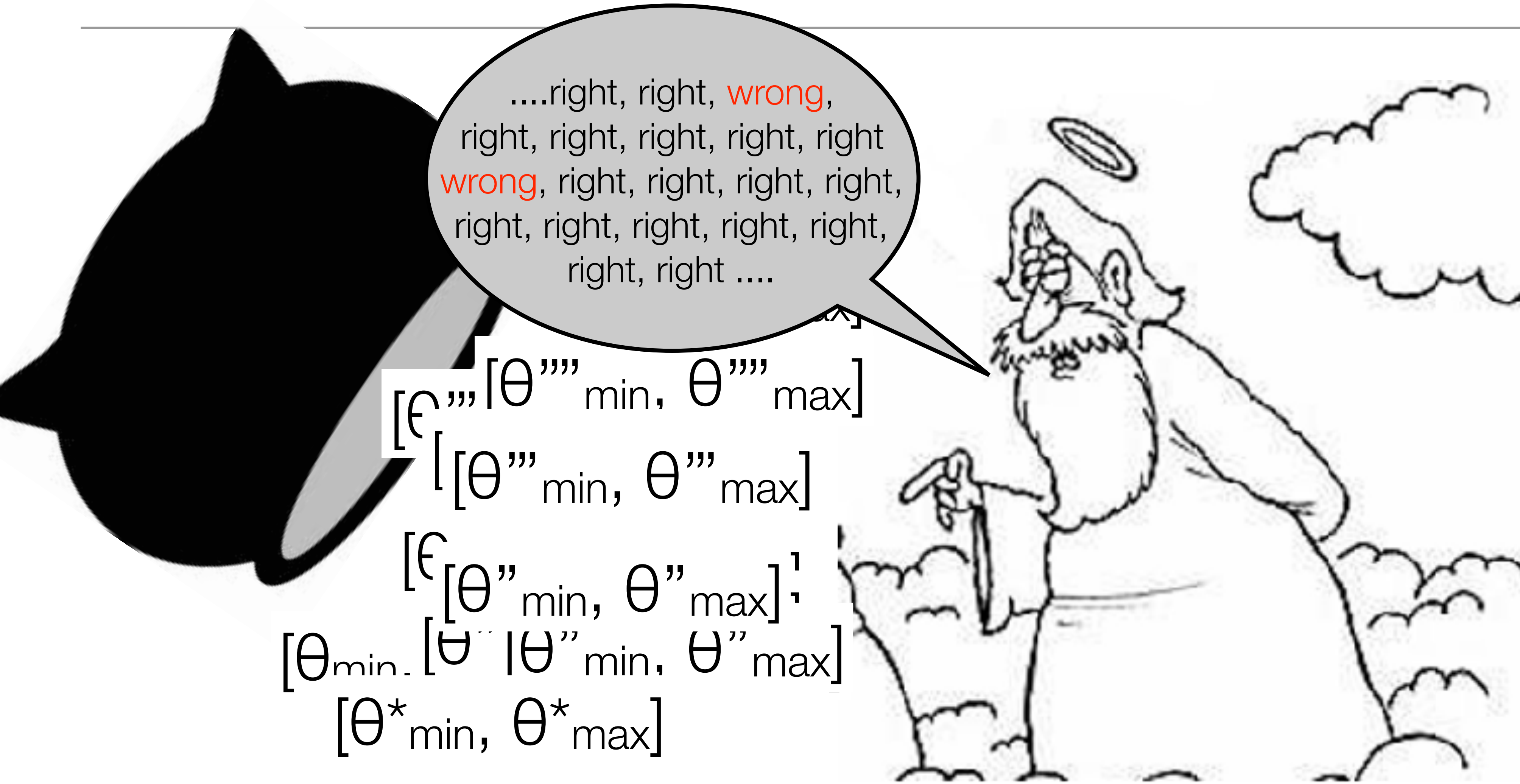
Confidence Levels for Frequentists: Coverage



Confidence Levels for Frequentists: Coverage



Confidence Levels for Frequentists: 90% Coverage



$[\theta''''_{min}, \theta''''_{max}]$

$[\theta'''_{min}, \theta'''_{max}]$

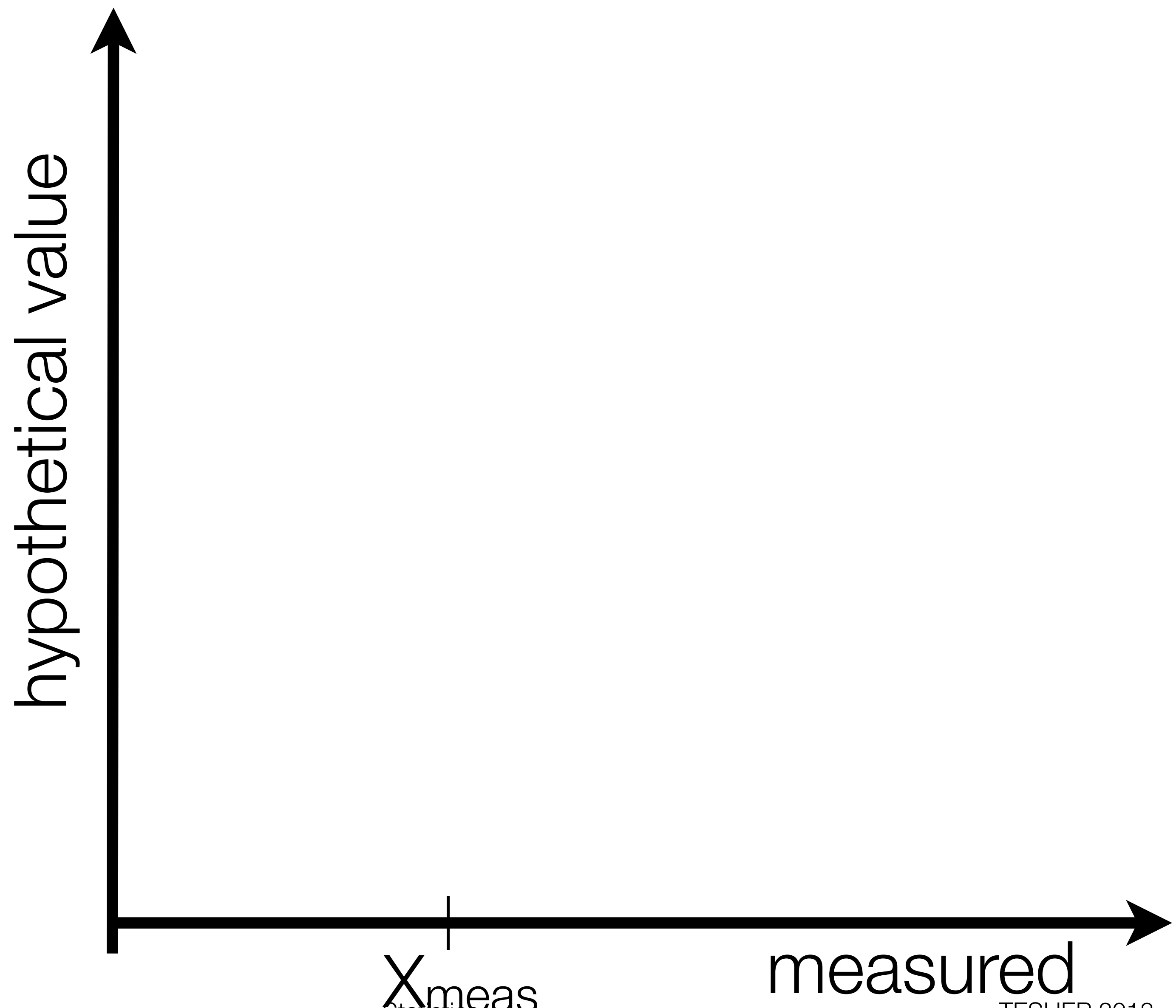
$[\theta''_{min}, \theta''_{max}]$

$[\theta''_{min}, \theta''_{max}]$

$[\theta^*_{min}, \theta^*_{max}]$

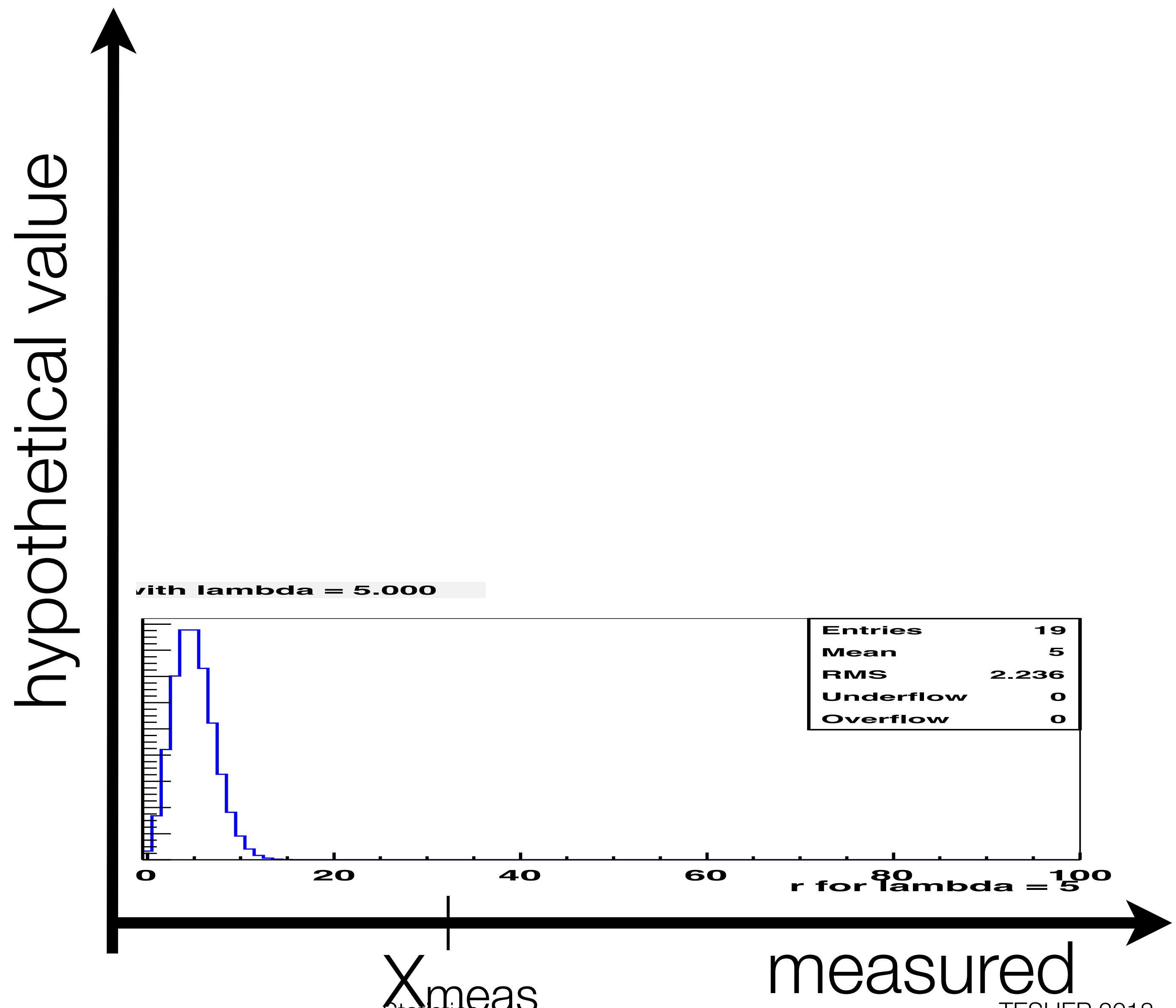
Constructing Confidence Limits (Neyman)

- **Construct horizontally, Read Vertically**
- **Achieves exact coverage.**



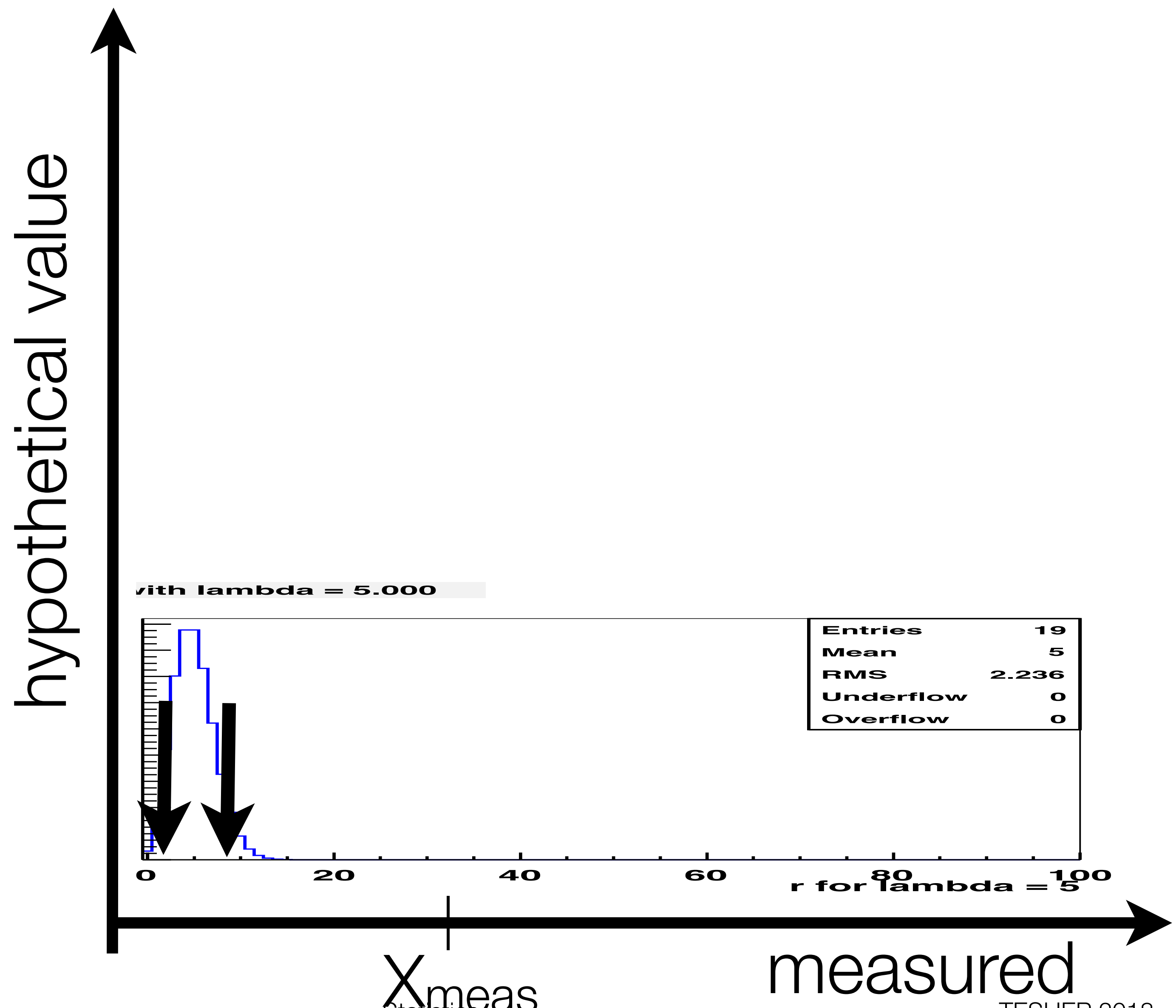
Constructing Confidence Limits (Neyman)

- **Construct horizontally, Read Vertically**
- **Achieves exact coverage.**



Constructing Confidence Limits (Neyman)

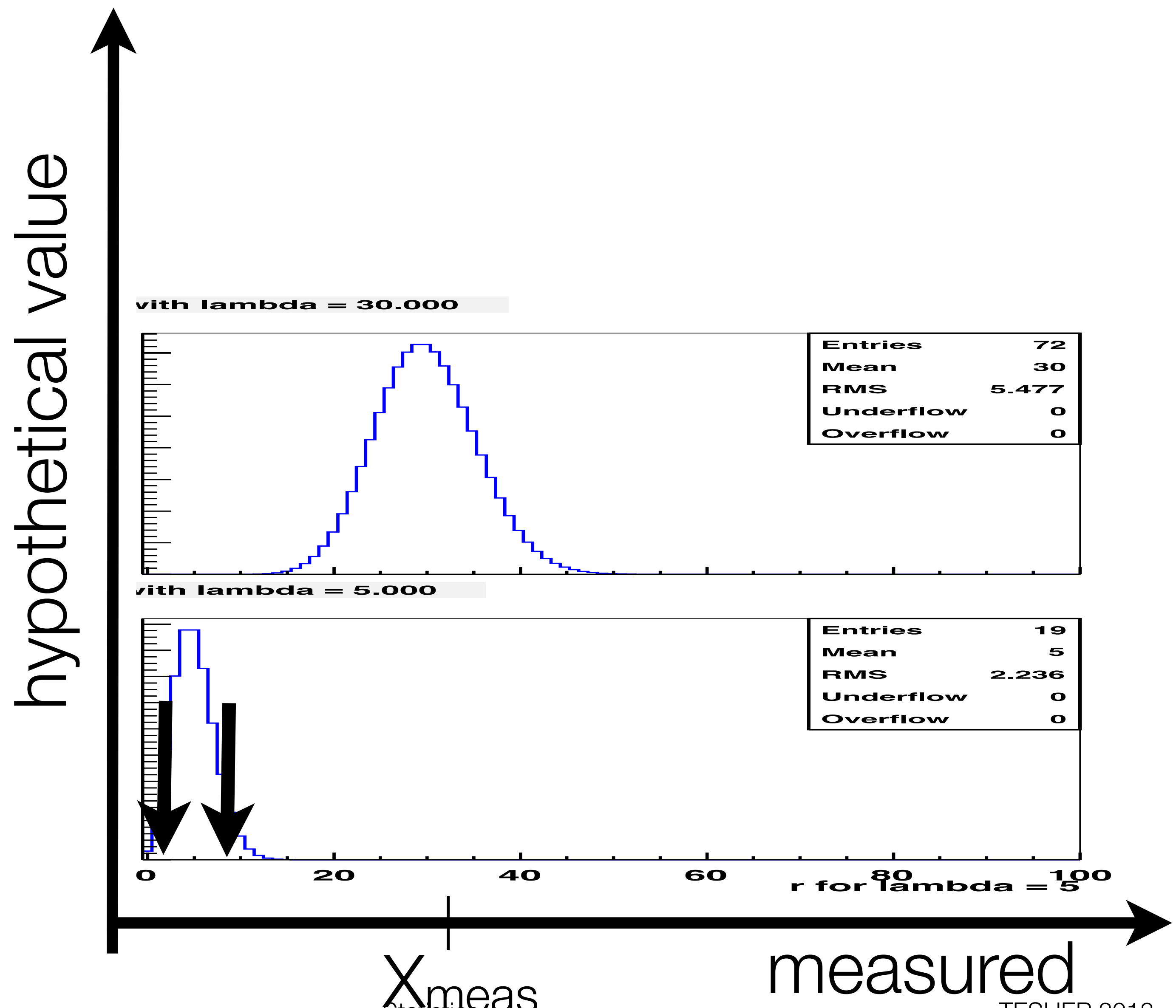
- Construct horizontally, Read Vertically
- Achieves exact coverage.



Constructing Confidence Limits (Neyman)

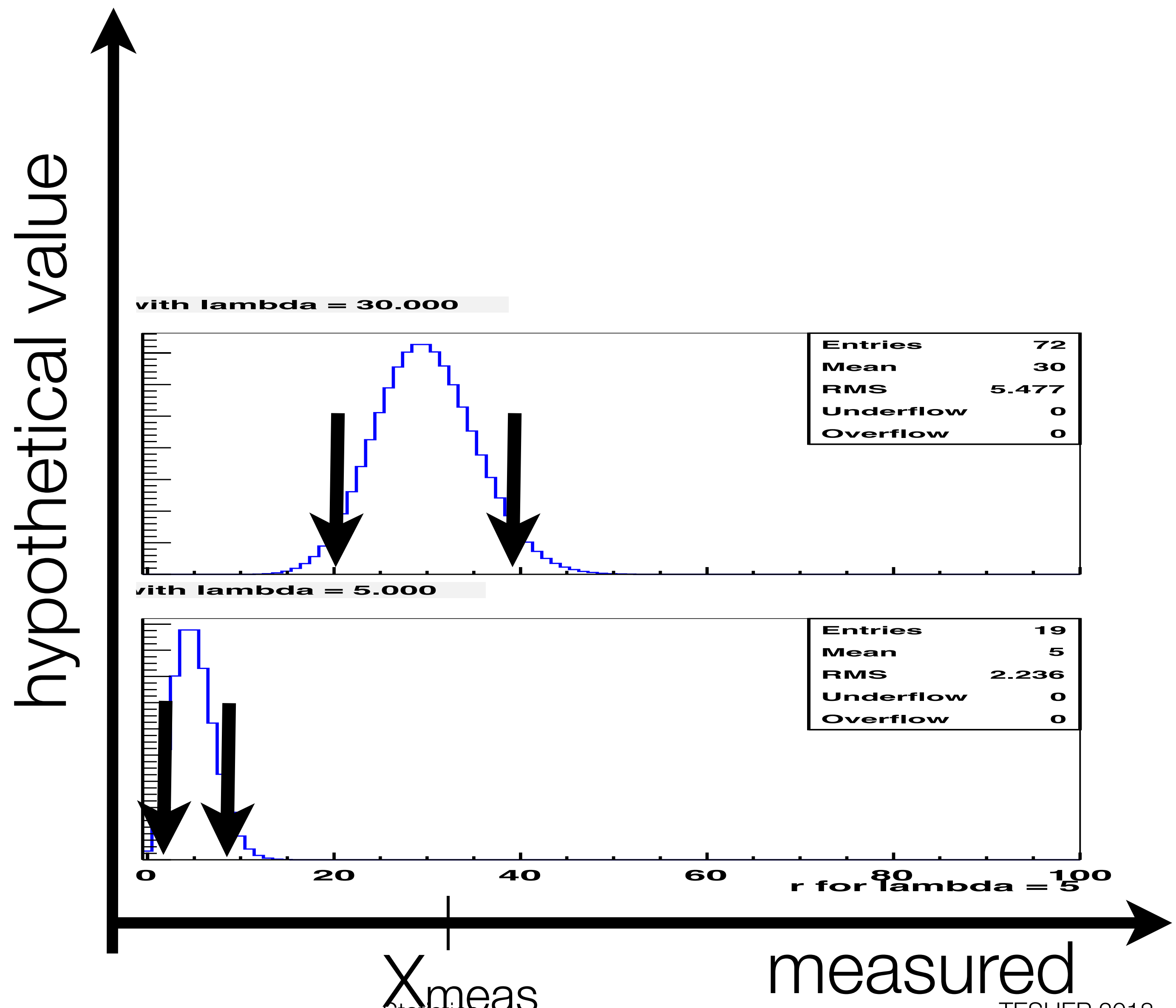
- **Construct horizontally, Read Vertically**

- **Achieves exact coverage.**



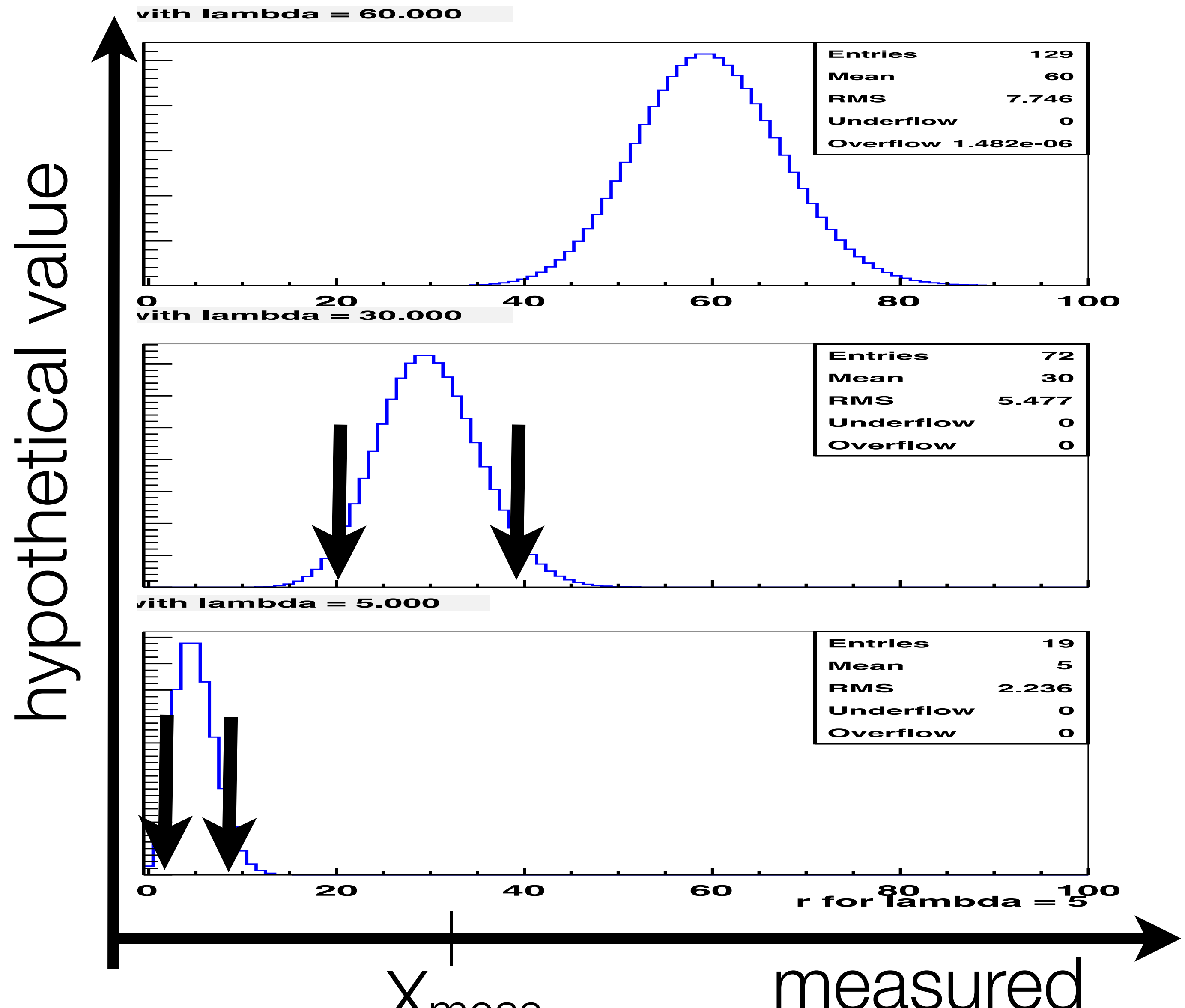
Constructing Confidence Limits (Neyman)

- **Construct horizontally, Read Vertically**
- **Achieves exact coverage.**



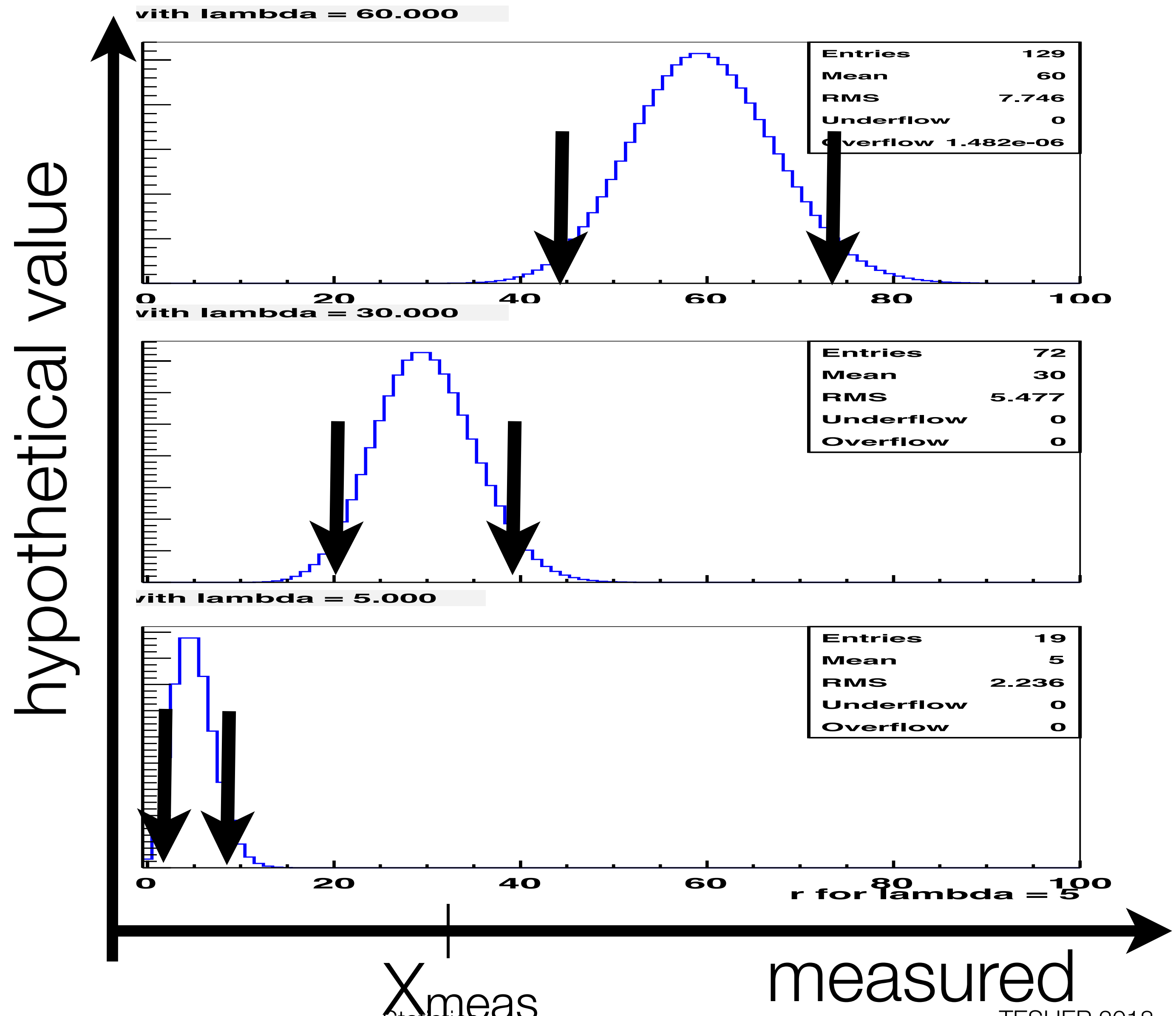
Constructing Confidence Limits (Neyman)

- **Construct horizontally, Read Vertically**
- **Achieves exact coverage.**



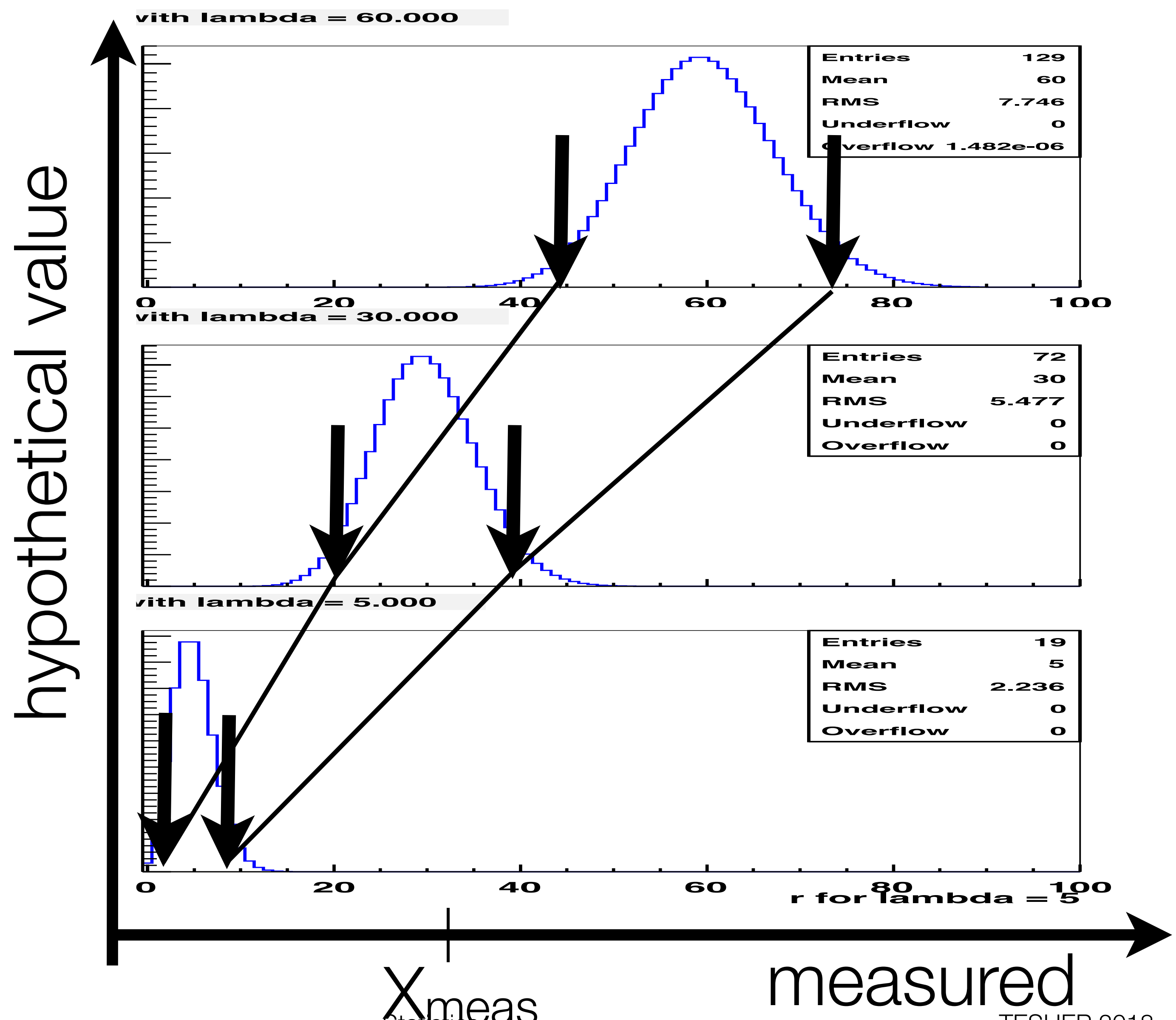
Constructing Confidence Limits (Neyman)

- **Construct horizontally, Read Vertically**
- **Achieves exact coverage.**



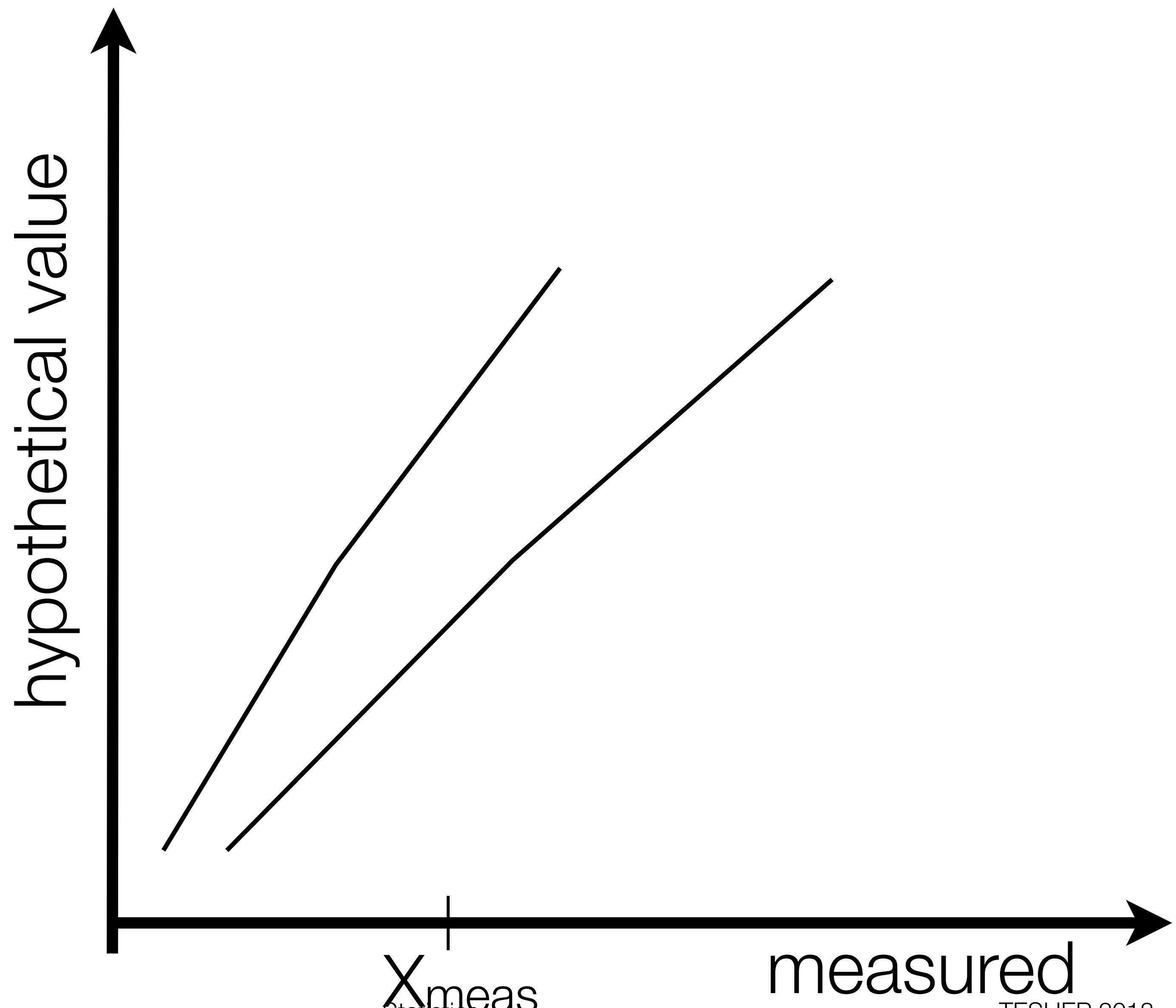
Constructing Confidence Limits (Neyman)

- **Construct horizontally, Read Vertically**
- **Achieves exact coverage.**



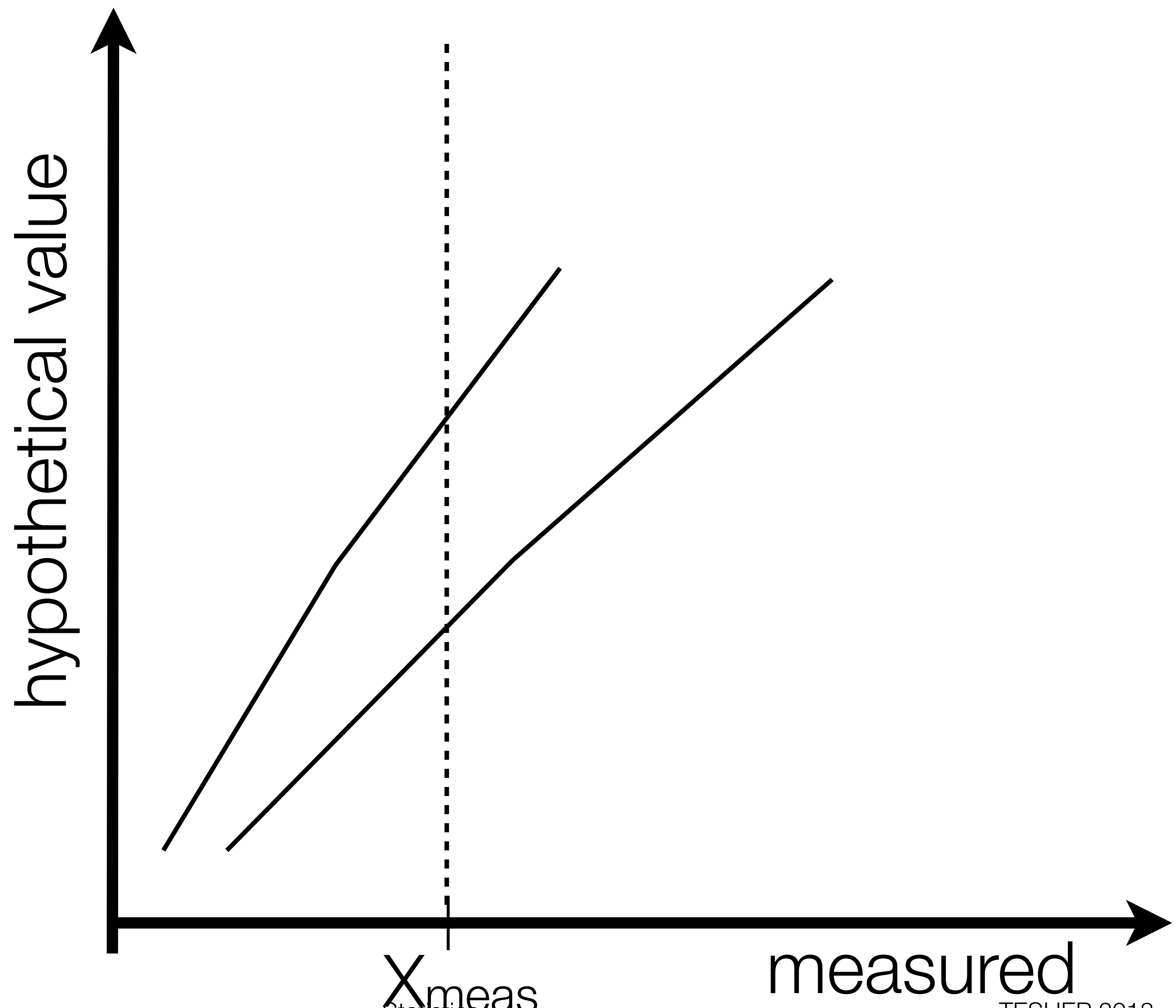
Constructing Confidence Limits (Neyman)

- **Construct horizontally, Read Vertically**
- **Achieves exact coverage.**



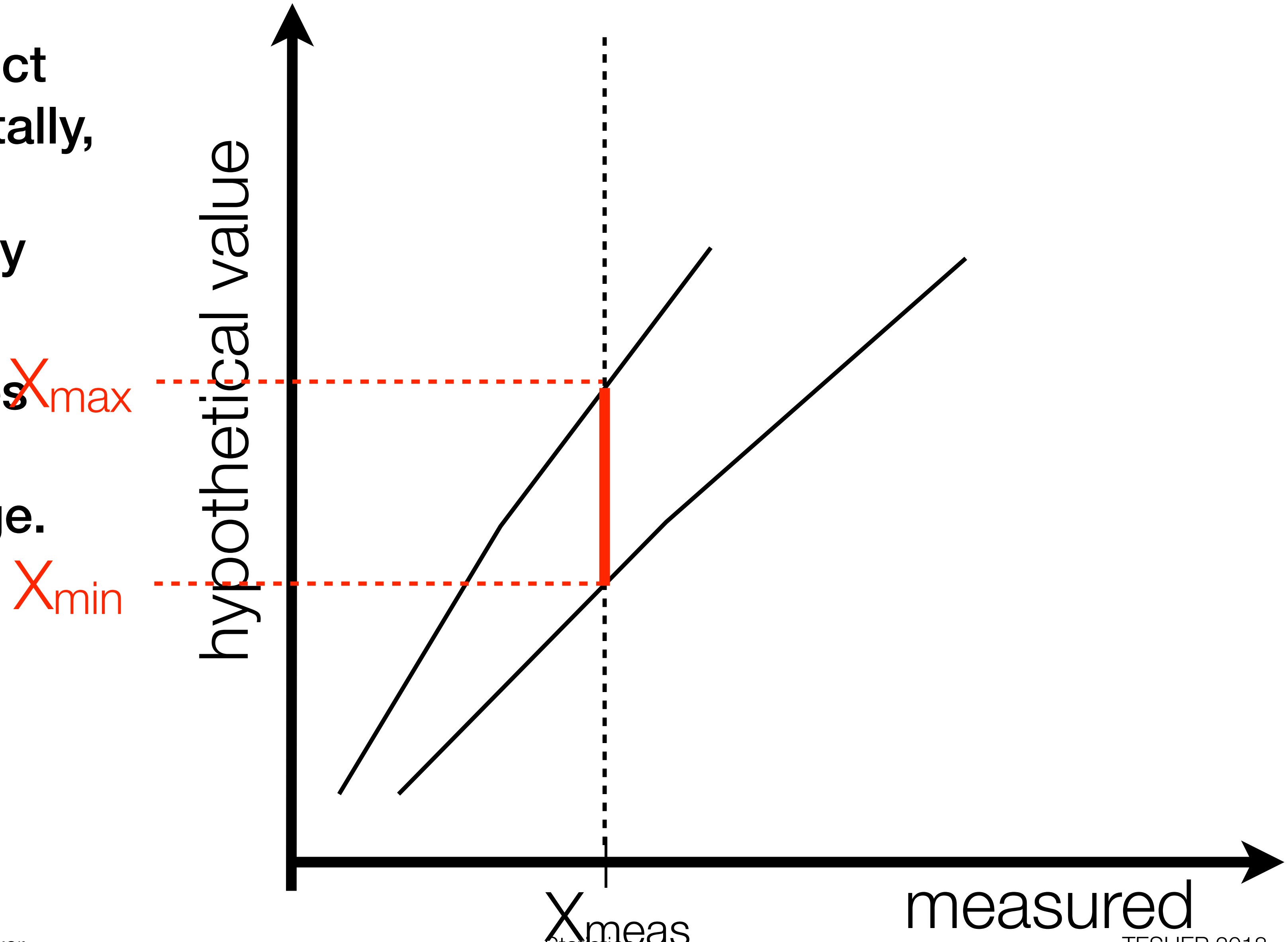
Constructing Confidence Limits (Neyman)

- **Construct horizontally, Read Vertically**
- **Achieves exact coverage.**



Constructing Confidence Limits (Neyman)

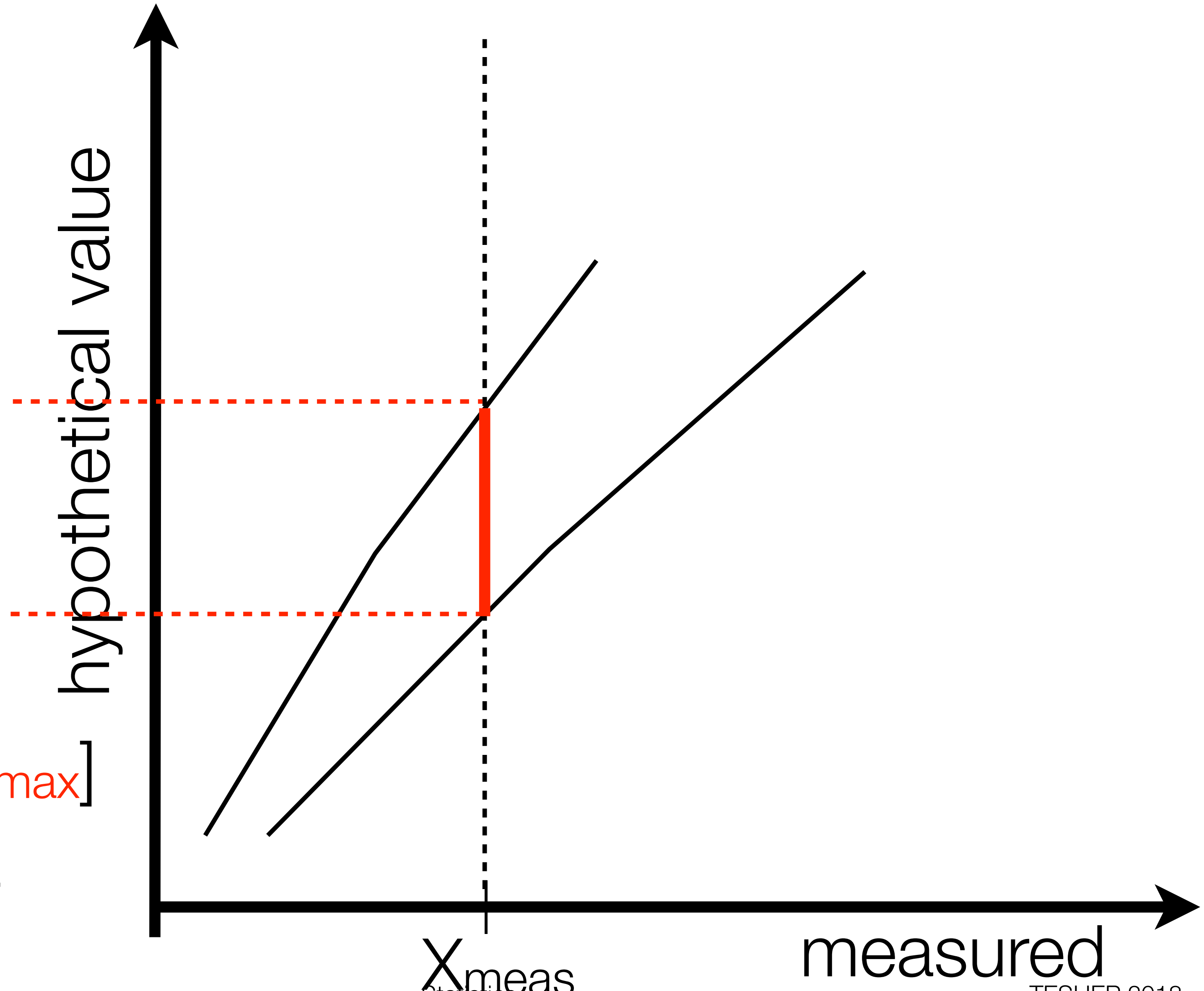
- **Construct horizontally, Read Vertically**
- **Achieves exact coverage.**



Constructing Confidence Limits (Neyman)

- **Construct horizontally, Read Vertically**

- **Achieves exact coverage.**

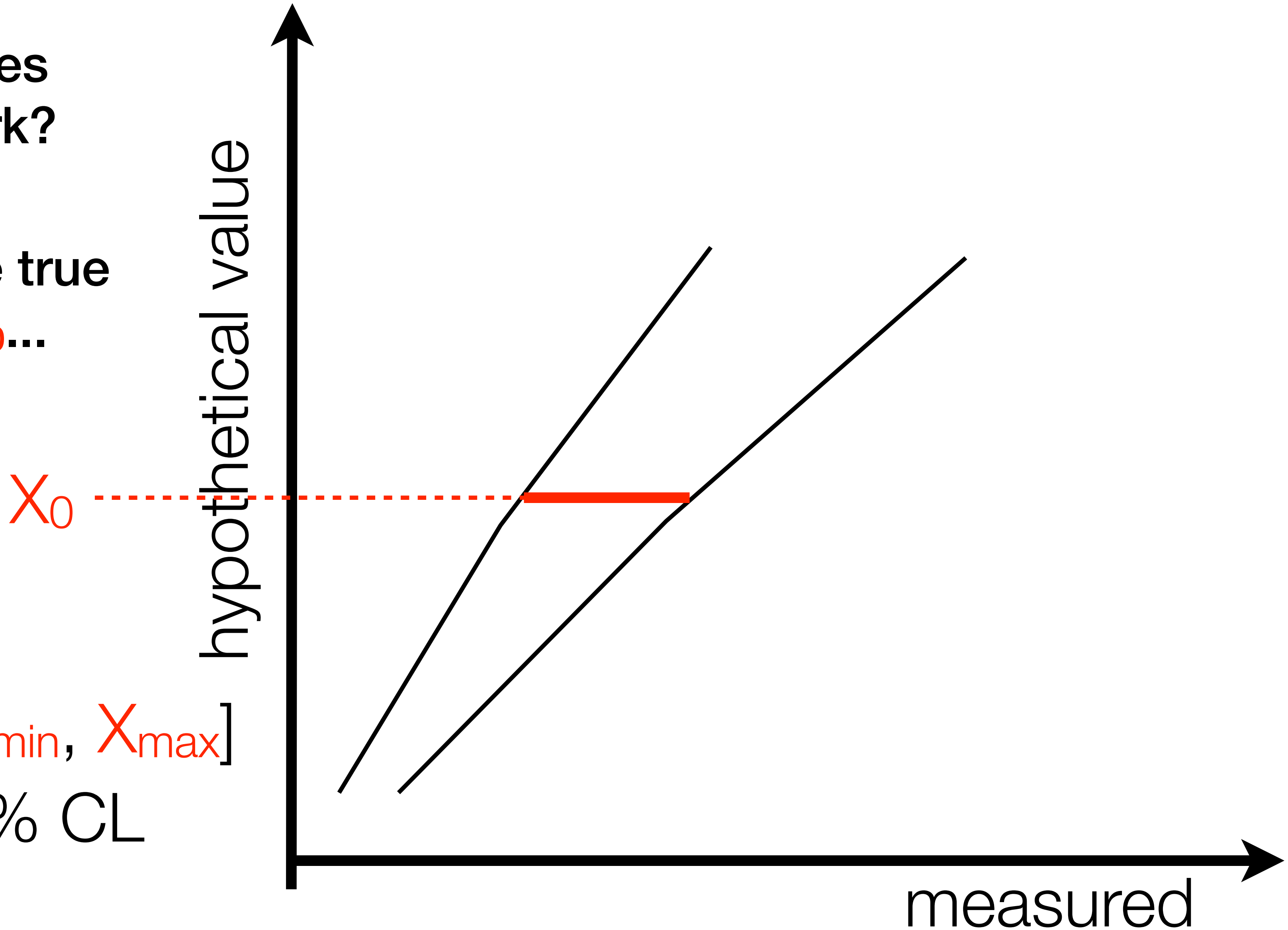


$X_{true} \in [X_{min}, X_{max}]$

at n% CL

Constructing Confidence Limits (Neyman)

- **Why does this work?**
- **Assume true value X_0...**

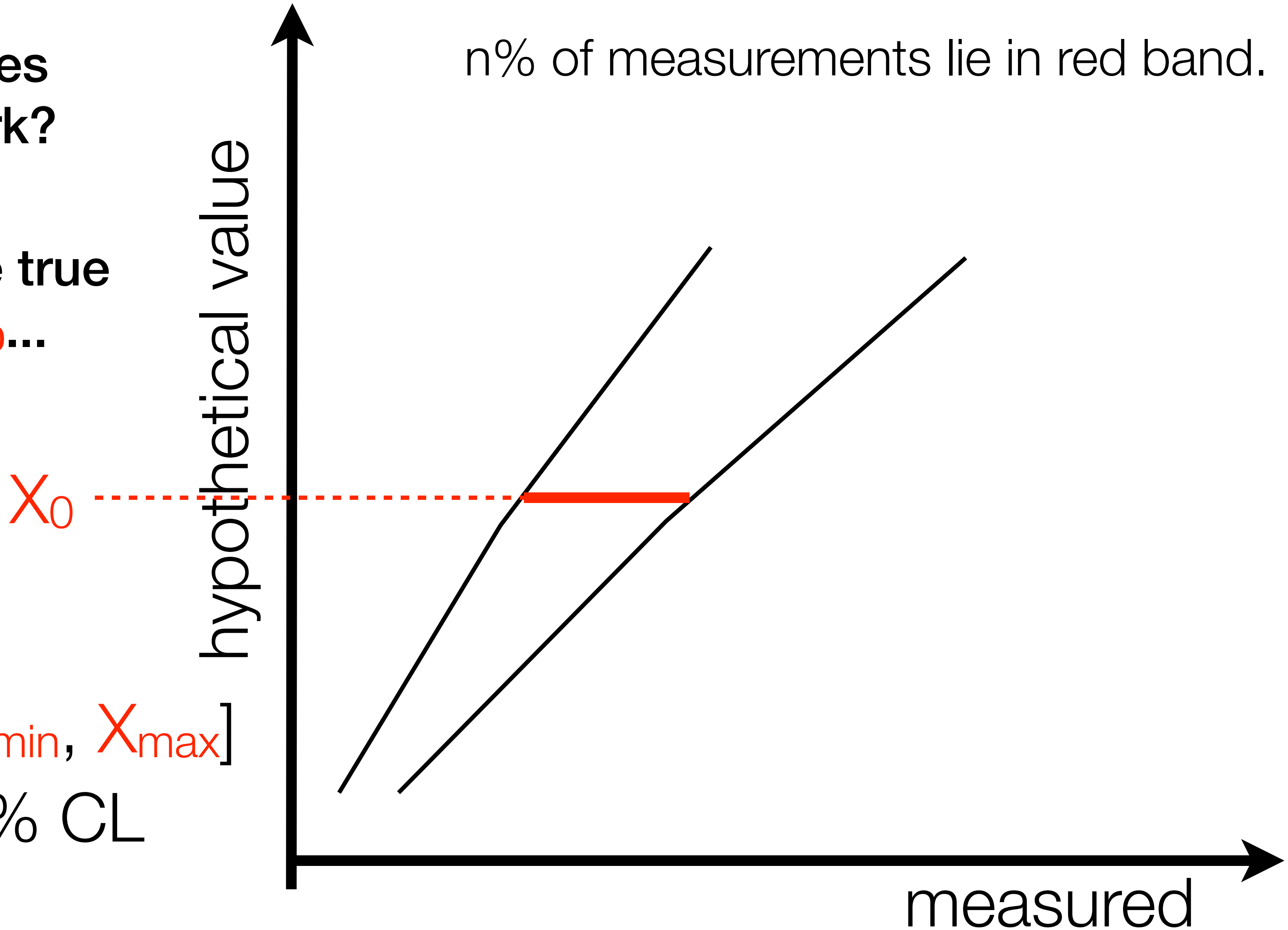


$X_{true} \in [X_{min}, X_{max}]$

at n% CL

Constructing Confidence Limits (Neyman)

- **Why does this work?**
- **Assume true value X_0...**

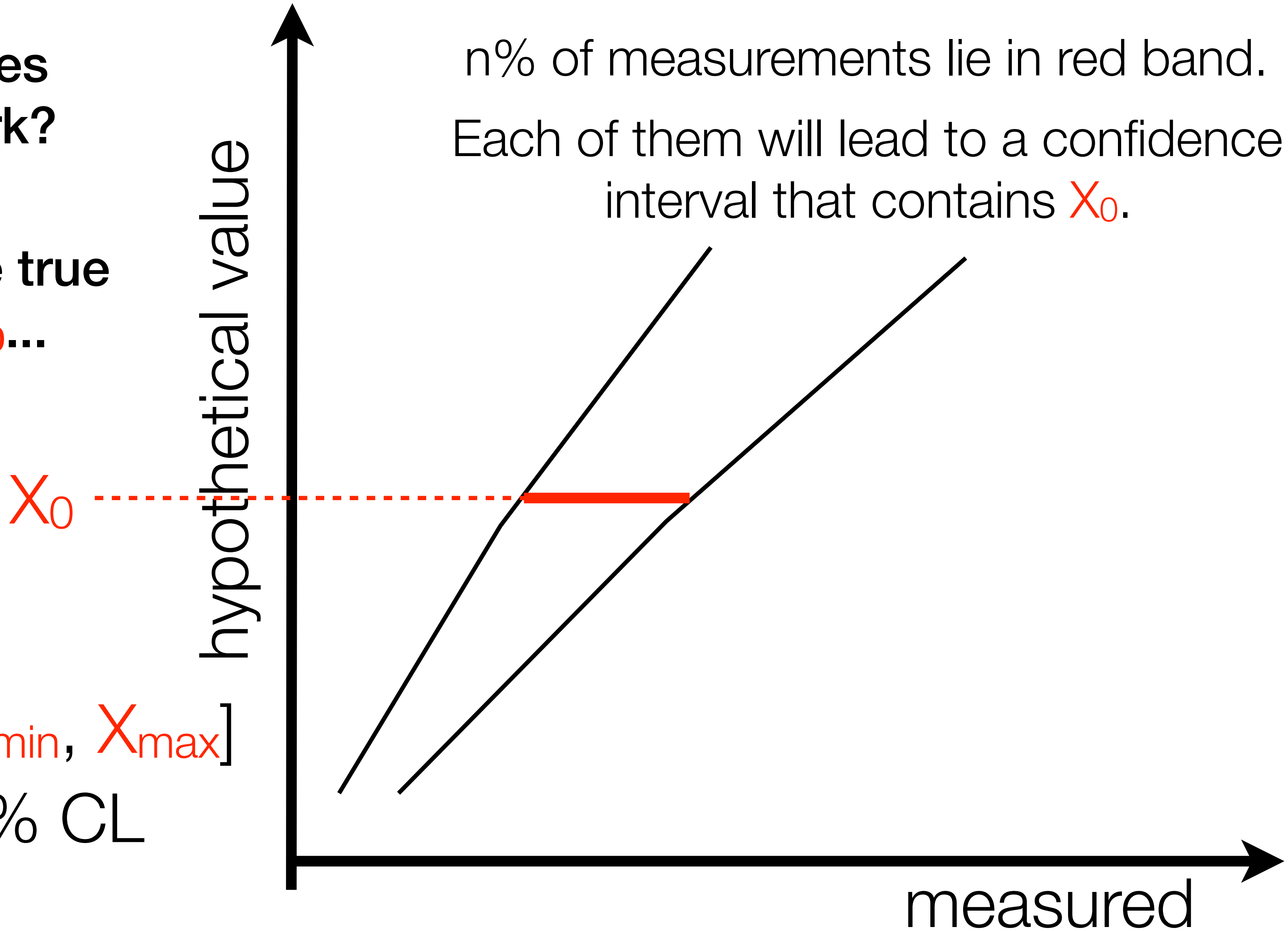


$X_{true} \in [X_{min}, X_{max}]$ at n% CL

Constructing Confidence Limits (Neyman)

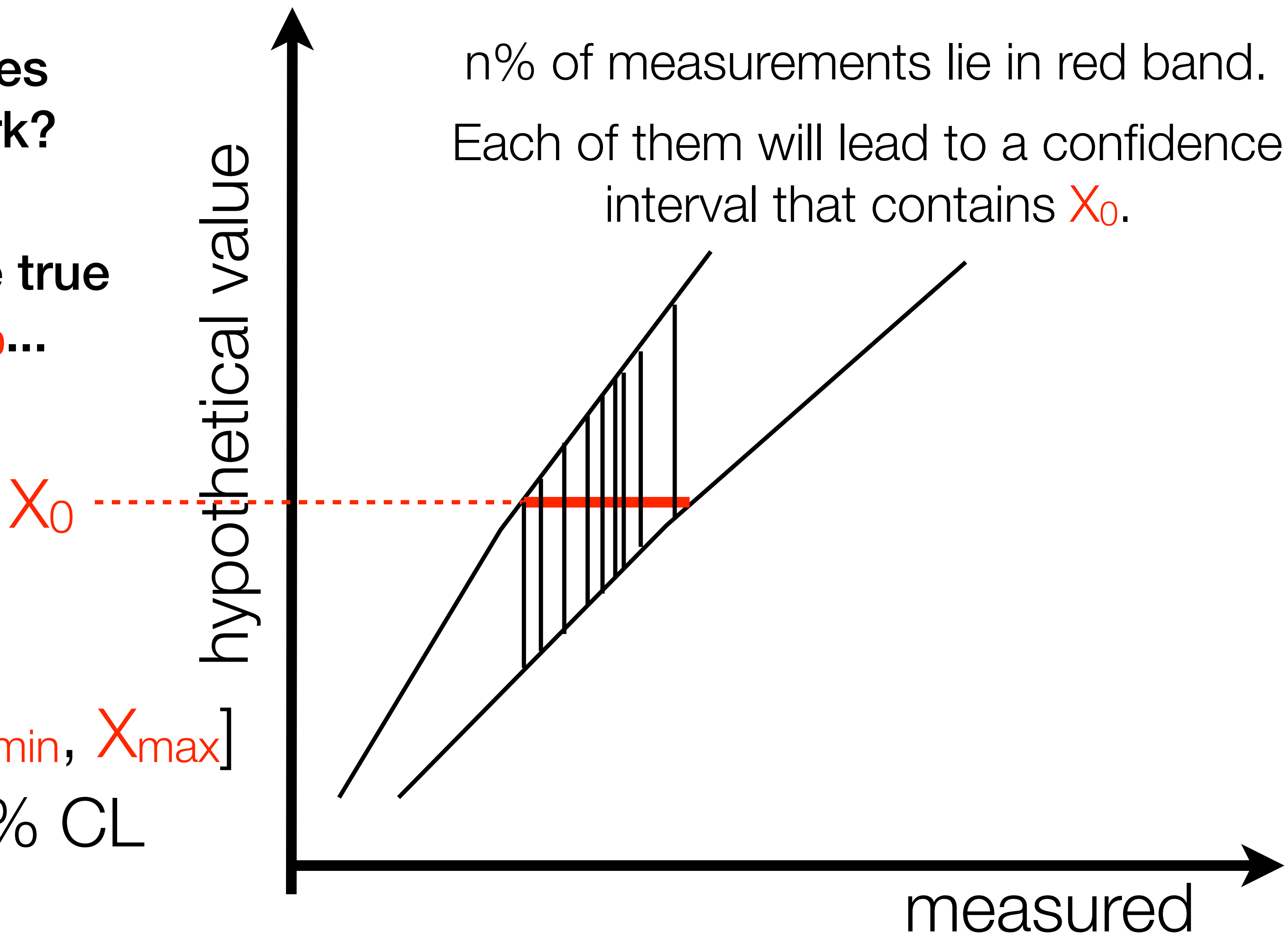
- **Why does this work?**
- **Assume true value X_0...**

$X_{true} \in [X_{min}, X_{max}]$ at n% CL



Constructing Confidence Limits (Neyman)

- **Why does this work?**
- **Assume true value X_0...**

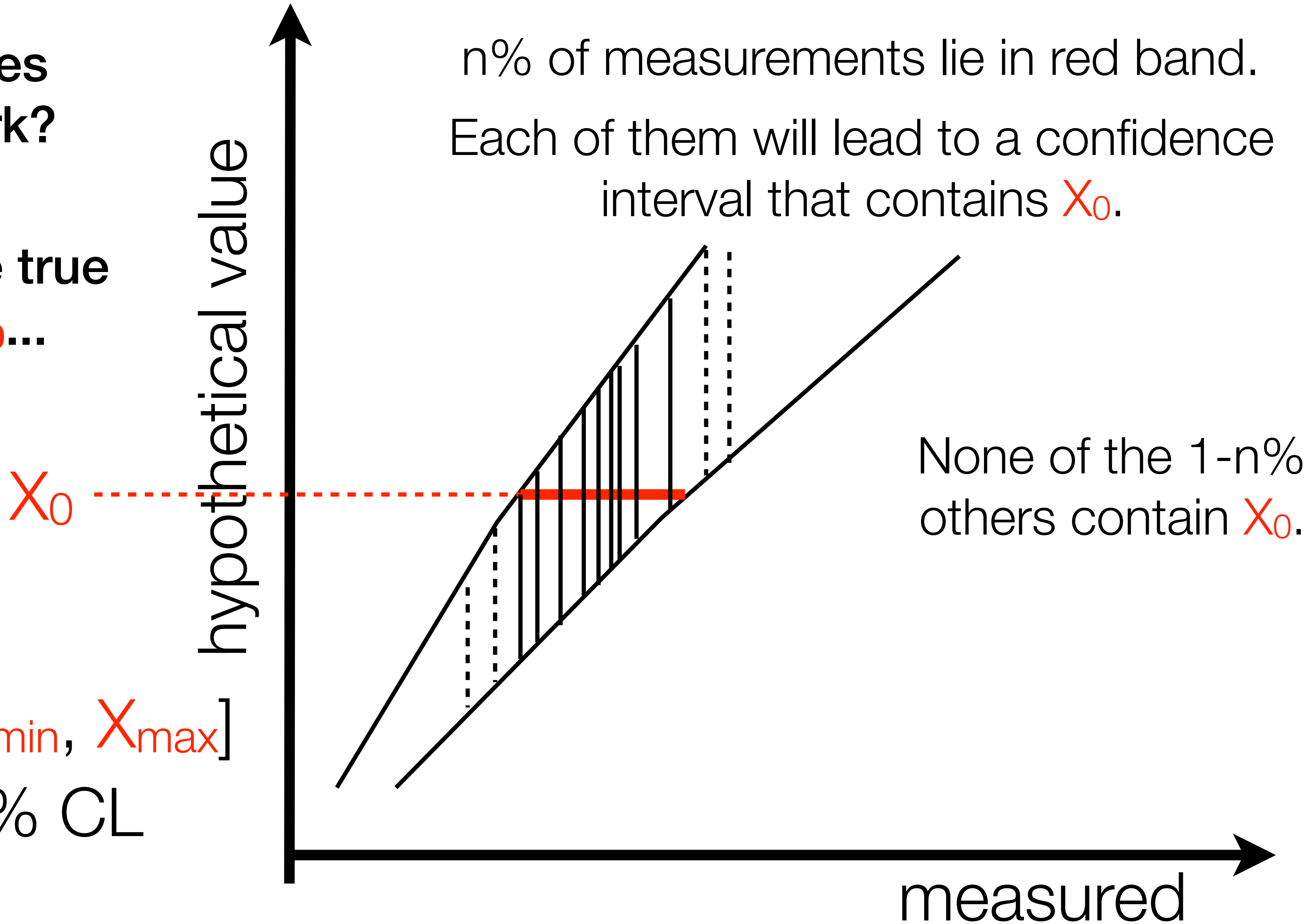


$X_{true} \in [X_{min}, X_{max}]$ at n% CL

Constructing Confidence Limits (Neyman)

- **Why does this work?**
- **Assume true value X_0...**

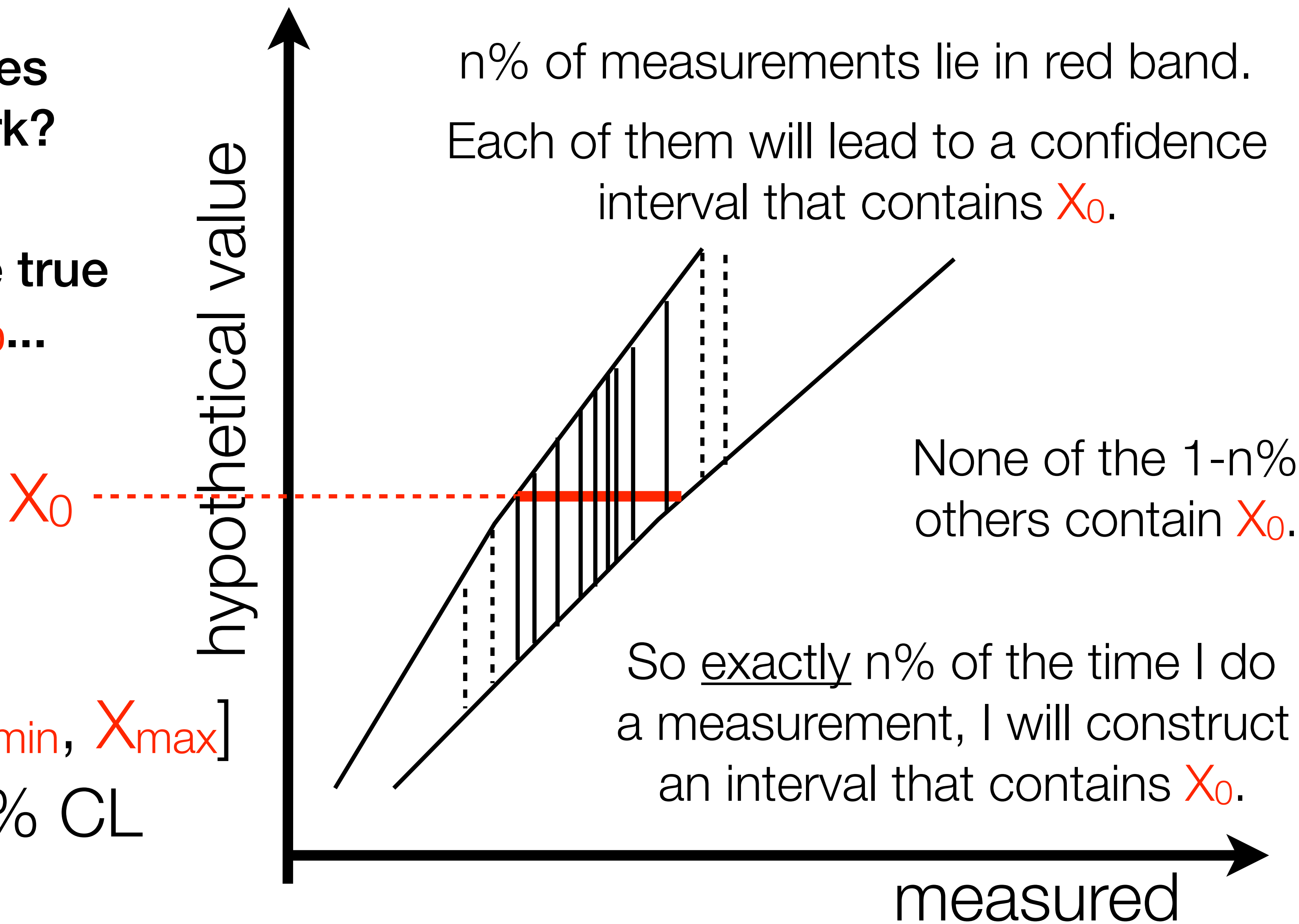
$X_{true} \in [X_{min}, X_{max}]$ at n% CL



Constructing Confidence Limits (Neyman)

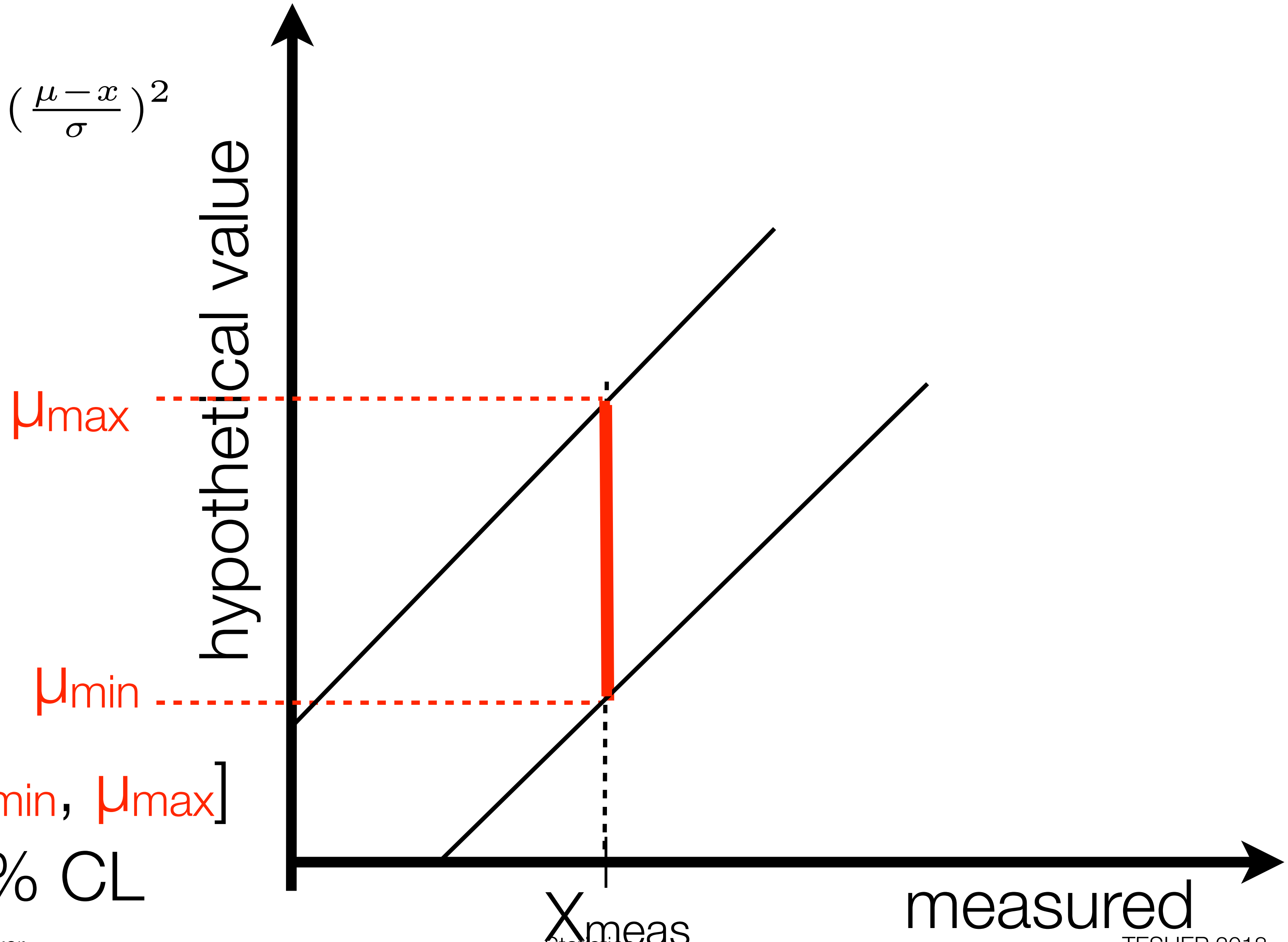
- **Why does this work?**
- **Assume true value X_0...**

$X_{true} \in [X_{min}, X_{max}]$ at n% CL



Very easy for Gaussians

$$\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{1}{2}(\frac{\mu-x}{\sigma})^2}$$

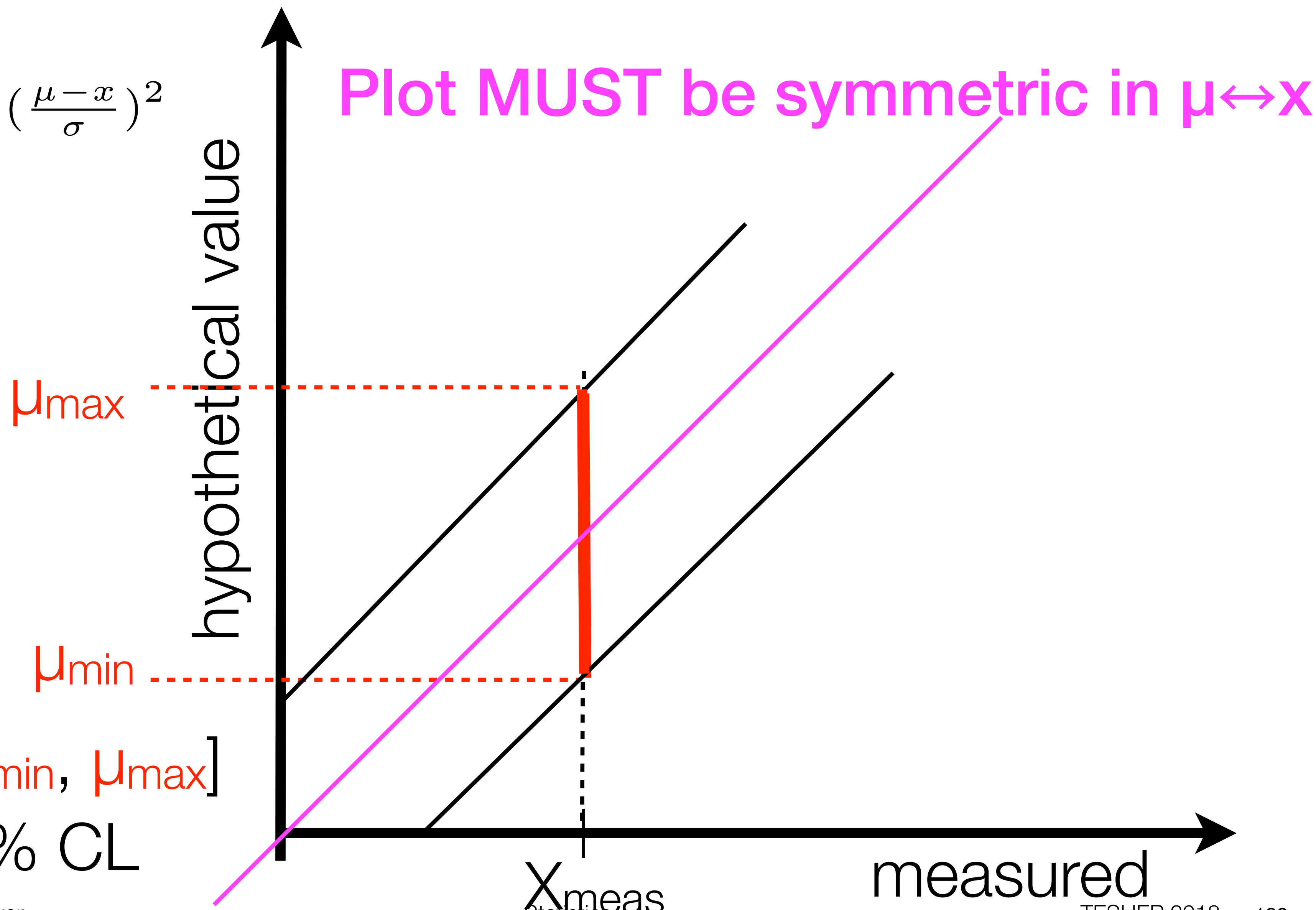


$\mu_{true} \in [\mu_{min}, \mu_{max}]$

at n% CL

Very easy for Gaussians

$$\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{\mu - x}{\sigma}\right)^2}$$

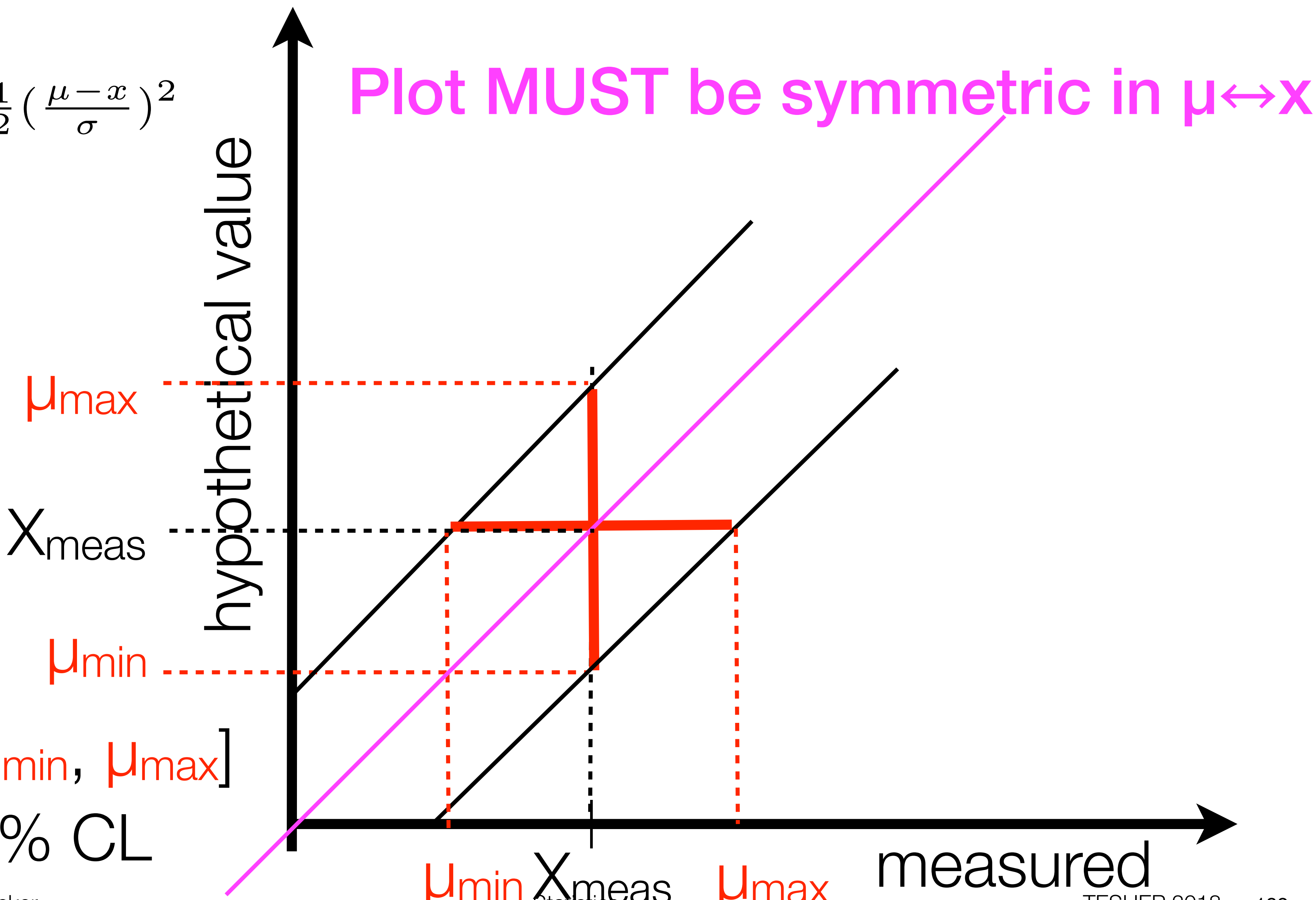


$\mu_{true} \in [\mu_{min}, \mu_{max}]$

at n% CL

Very easy for Gaussians

$$\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{\mu-x}{\sigma}\right)^2}$$

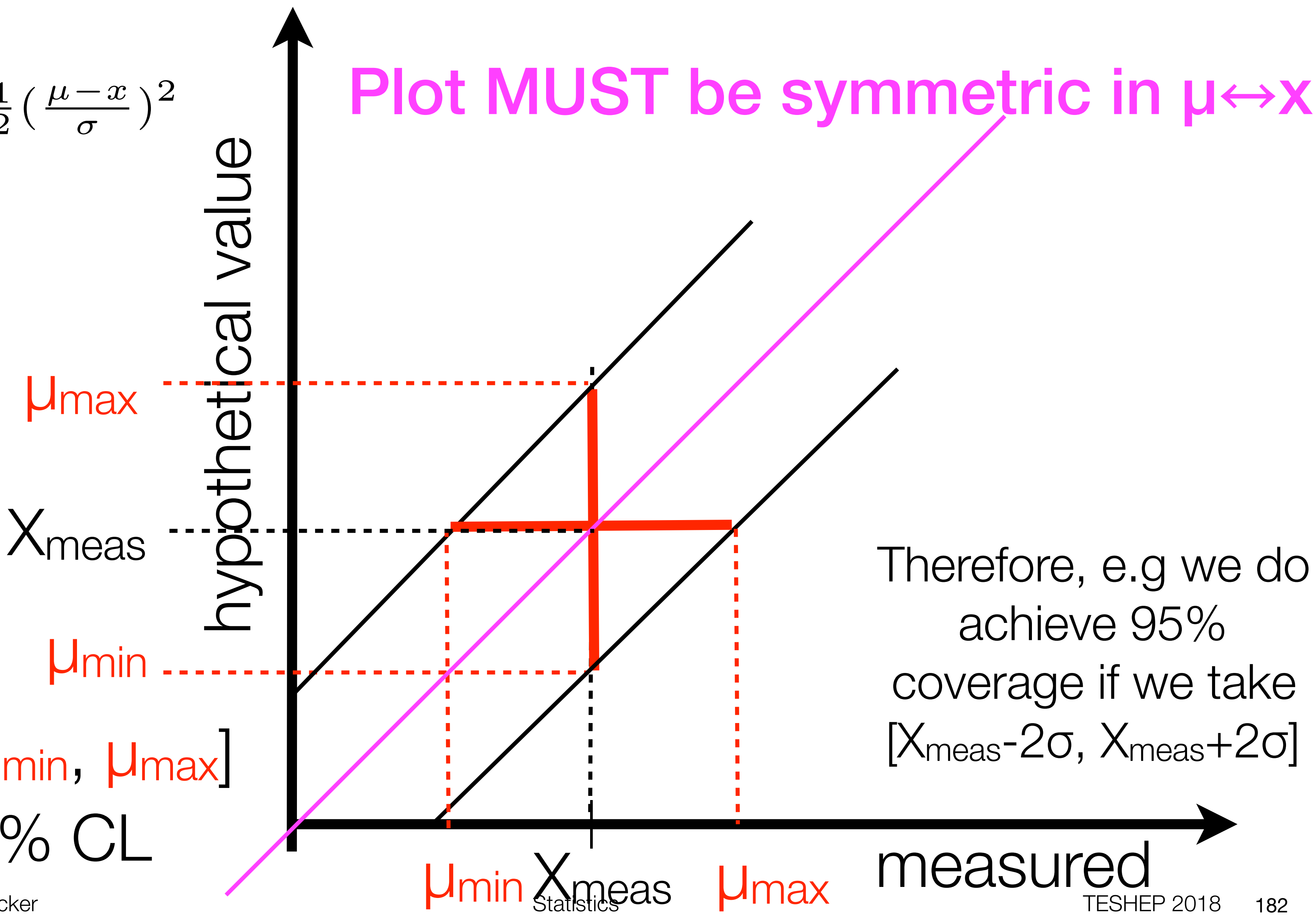


$\mu_{true} \in [\mu_{min}, \mu_{max}]$

at n% CL

Very easy for Gaussians

$$\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{\mu - x}{\sigma}\right)^2}$$



$\mu_{true} \in [\mu_{min}, \mu_{max}]$ at n% CL

Therefore, e.g we do achieve 95% coverage if we take $[X_{meas}-2\sigma, X_{meas}+2\sigma]$

Problems

Problems

- **Estimated background (say from sideband): 2.5 events**

Problems

- **Estimated background (say from sideband): 2.5 events**
- **Detect 3 events:**

Problems

- Estimated background (say from sideband): 2.5 events
- Detect 3 events:
    - Total < 6.68 at 95% CL

Problems

- Estimated background (say from sideband): 2.5 events
- Detect 3 events:
    - Total < 6.68 at 95% CL
    - Signal < 4.18 at 95% CL

Problems

- Estimated background (say from sideband): 2.5 events
- Detect 3 events:
  - Total < 6.68 at 95% CL
  - Signal < 4.18 at 95% CL
- Detect 0 events

Problems

- Estimated background (say from sideband): 2.5 events
- Detect 3 events:
  - Total < 6.68 at 95% CL
  - Signal < 4.18 at 95% CL
- Detect 0 events
  - Total < 2.3 at 95% CL

Problems

- Estimated background (say from sideband): 2.5 events
- Detect 3 events:
  - Total < 6.68 at 95% CL
  - Signal < 4.18 at 95% CL
- Detect 0 events
  - Total < 2.3 at 95% CL
  - Signal < -0.2 at 95% CL

Problems

- **Estimated background (say from sideband): 2.5 events**
- **Detect 3 events:**
  - **Total < 6.68 at 95% CL**
  - **Signal < 4.18 at 95% CL**
- **Detect 0 events**
  - **Total < 2.3 at 95% CL**
  - **Signal < -0.2 at 95% CL**

This is allowed since we are allowed to get it wrong 5% of the time. However, it's silly to make such a statement, because in this case, we already know it's wrong.

What is Bayesian statistics good for?

What is Bayesian statistics good for?

- **Convenient when estimating parameters near a physical boundary.**

What is Bayesian statistics good for?

- **Convenient when estimating parameters near a physical boundary.**
- **Detect 3 events:**

What is Bayesian statistics good for?

- **Convenient when estimating parameters near a physical boundary.**

- **Detect 3 events:**

  - **Total < 6.68 at 95% CL**

What is Bayesian statistics good for?

- **Convenient when estimating parameters near a physical boundary.**
- **Detect 3 events:**
    - **Total < 6.68 at 95% CL**
    - **Signal < 4.18 at 95% CL**

What is Bayesian statistics good for?

- Convenient when estimating parameters near a physical boundary.
- Detect 3 events:
  - Total < 6.68 at 95% CL
  - Signal < 4.18 at 95% CL
- Detect 0 events

What is Bayesian statistics good for?

- **Convenient when estimating parameters near a physical boundary.**
- **Detect 3 events:**
  - **Total < 6.68 at 95% CL**
  - **Signal < 4.18 at 95% CL**
- **Detect 0 events**
  - **Total < 2.3 at 95% CL**

What is Bayesian statistics good for?

- Convenient when estimating parameters near a physical boundary.
- Detect 3 events:
  - Total < 6.68 at 95% CL
  - Signal < 4.18 at 95% CL
- Detect 0 events
  - Total < 2.3 at 95% CL
  - Signal < -0.2 at 95% CL

What is Bayesian statistics good for?

- **Convenient when estimating parameters near a physical boundary.**
- **Detect 3 events:**
  - **Total < 6.68 at 95% CL**
  - **Signal < 4.18 at 95% CL**
- **Detect 0 events**
  - **Total < 2.3 at 95% CL**
  - **Signal < -0.2 at 95% CL**

There are frequentist (and thus objective) solutions to this (e.g. Feldman Cousins), and I highly recommend them. But let's see what Bayesians do.

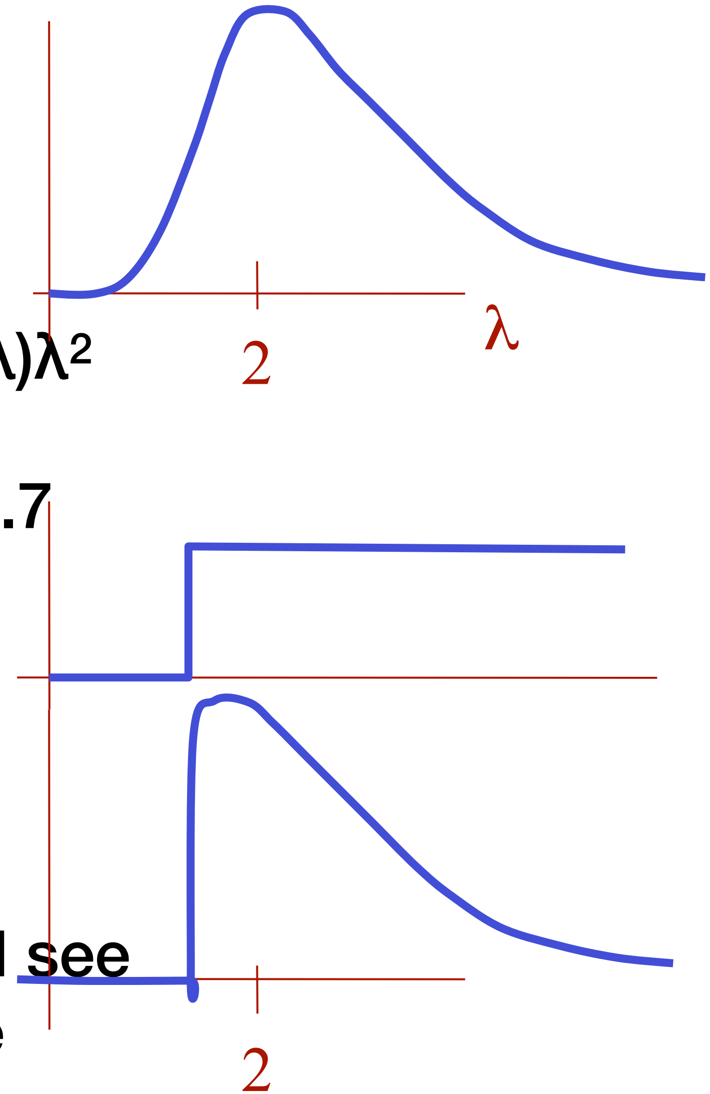
Bayesian statistics offers a neat solution

$$P(\text{theory}|\text{data}) = \frac{P(\text{data}|\text{theory})P(\text{theory})}{P(\text{data})}$$

- **Pick a prior P(theory) that excludes values < 0.**
- **P(theory|data) can be used directly to construct confidence intervals.**
- **But beware: there is no unique prior.**

Bayesian solution

- Background: 1.7 events
- Total mean: $\lambda = 1.7 + \lambda_{signal}$
- Observe 2 events: $P(2,\lambda)=0.5*\exp(-\lambda)\lambda^2$
- Put in prior taking into account $\lambda>1.7$
- Multiply, and normalise
- Interpret this as probability for λ.
- Many priors possible. Try a few and see if you get consistent results. Or use frequentist methods.



Summary Frequentist vs Bayesian

- Frequentists need ensembles/repeatable experiments. There is only one true theory (even if we don't know it). But data can be taken many times. Work with P(data | theory).
- Bayesians can make sense of P(theory | data). Can be very convenient, especially if you have a fit result near a physical boundary, as it can be easily accommodated in the prior. But it is not objective as the prior is not unique.
- For most problems, there are frequentist (i.e. objective) solutions, e.g. Feldman-Cousins' approach.
- If you use Bayesian statistics, try different priors and, as always, describe exactly what you did.

Summary

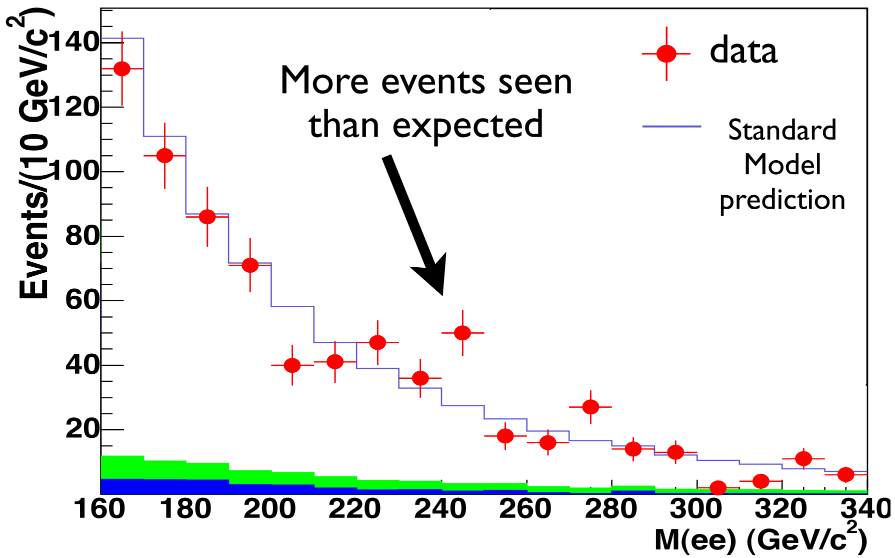
- Looked at a few statistical issues that will become part of your daily analysis life
  - Central Limit Theorem
  - Basic fitting methods - whatever you use, test your fit!
  - Monte Carlo event generation and MC integration
  - Maybe we even looked into evaluating confidence intervals.
- We also had a look at some philosophical aspects of statistics. What is probability? What are the differences between the Bayesian and Frequentist approach?

Homework:

http://goo.gl/COvCmK

- Estimate the significance of this observation
    - Step 1: Calculate the probability so see an upward fluctuation this big or bigger in the Standard Model, in this one bin
    - Step 2 take into account that they looked in 84 bins (tricky!)
- You should get a fairly small number. Why, do you think, have you not read in the news about the discovery of the Z' at CDF?

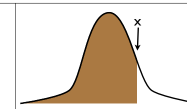
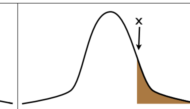
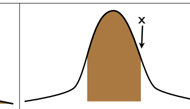
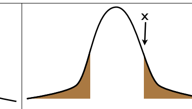
Z' search at CDF



- In the bin with the arrow, we expect 28 events without the Z'
- See 48 events.

First steps...

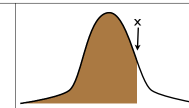
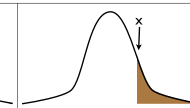
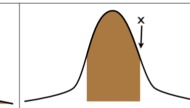
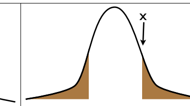
Integals over Gaussians

				
$s = \frac{x-\mu}{\sigma}$	$\int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$1 - \int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$\int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$1 - \int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$
0	0.500	5.0E-01	0.000	1.0E+00
0.1	0.540	4.6E-01	0.080	9.2E-01
0.2	0.579	4.2E-01	0.159	8.4E-01
0.3	0.618	3.8E-01	0.236	7.6E-01
0.4	0.655	3.4E-01	0.311	6.9E-01
0.5	0.691	3.1E-01	0.383	6.2E-01
0.6	0.726	2.7E-01	0.451	5.5E-01
0.7	0.758	2.4E-01	0.516	4.8E-01
0.8	0.788	2.1E-01	0.576	4.2E-01
0.9	0.816	1.8E-01	0.632	3.7E-01
1	0.841	1.6E-01	0.683	3.2E-01
1.1	0.864	1.4E-01	0.729	2.7E-01
1.2	0.885	1.2E-01	0.770	2.3E-01
1.3	0.903	9.7E-02	0.806	1.9E-01
1.4	0.919	8.1E-02	0.838	1.6E-01
1.5	0.933	6.7E-02	0.866	1.3E-01
1.6	0.945	5.5E-02	0.890	1.1E-01
1.7	0.955	4.5E-02	0.911	8.9E-02
1.8	0.964	3.6E-02	0.928	7.2E-02
1.9	0.971	2.9E-02	0.943	5.7E-02
2	0.977	2.3E-02	0.954	4.6E-02
2.1	0.982	1.8E-02	0.964	3.6E-02
2.2	0.986	1.4E-02	0.972	2.8E-02
2.3	0.9893	1.1E-02	0.9786	2.1E-02
2.4	0.9918	8.2E-03	0.9836	1.6E-02
2.5	0.9938	6.2E-03	0.9876	1.2E-02
2.6	0.9953	4.7E-03	0.9907	9.3E-03
2.7	0.9965	3.5E-03	0.9931	6.9E-03
2.8	0.9974	2.6E-03	0.9949	5.1E-03
2.9	0.9981	1.9E-03	0.9963	3.7E-03
3	0.99865	1.3E-03	0.99730	2.7E-03
3.1	0.99903	9.7E-04	0.99806	1.9E-03
3.2	0.99931	6.9E-04	0.99863	1.4E-03
3.3	0.99952	4.8E-04	0.99903	9.7E-04
3.4	0.99966	3.4E-04	0.99933	6.7E-04
3.5	0.99977	2.3E-04	0.99953	4.7E-04
3.6	0.999841	1.6E-04	0.999682	3.2E-04
3.7	0.999892	1.1E-04	0.999784	2.2E-04
3.8	0.999928	7.2E-05	0.999855	1.4E-04
3.9	0.999952	4.8E-05	0.999904	9.6E-05
4	0.999968	3.2E-05	0.999937	6.3E-05

Integals over Gaussians

First steps...

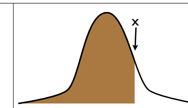
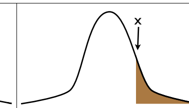
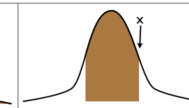
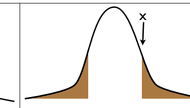
- **20/sqrt(28) = 3.8**

				
$s = \frac{x-\mu}{\sigma}$	$\int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$1 - \int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$\int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$1 - \int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$
0	0.500	5.0E-01	0.000	1.0E+00
0.1	0.540	4.6E-01	0.080	9.2E-01
0.2	0.579	4.2E-01	0.159	8.4E-01
0.3	0.618	3.8E-01	0.236	7.6E-01
0.4	0.655	3.4E-01	0.311	6.9E-01
0.5	0.691	3.1E-01	0.383	6.2E-01
0.6	0.726	2.7E-01	0.451	5.5E-01
0.7	0.758	2.4E-01	0.516	4.8E-01
0.8	0.788	2.1E-01	0.576	4.2E-01
0.9	0.816	1.8E-01	0.632	3.7E-01
1	0.841	1.6E-01	0.683	3.2E-01
1.1	0.864	1.4E-01	0.729	2.7E-01
1.2	0.885	1.2E-01	0.770	2.3E-01
1.3	0.903	9.7E-02	0.806	1.9E-01
1.4	0.919	8.1E-02	0.838	1.6E-01
1.5	0.933	6.7E-02	0.866	1.3E-01
1.6	0.945	5.5E-02	0.890	1.1E-01
1.7	0.955	4.5E-02	0.911	8.9E-02
1.8	0.964	3.6E-02	0.928	7.2E-02
1.9	0.971	2.9E-02	0.943	5.7E-02
2	0.977	2.3E-02	0.954	4.6E-02
2.1	0.982	1.8E-02	0.964	3.6E-02
2.2	0.986	1.4E-02	0.972	2.8E-02
2.3	0.9893	1.1E-02	0.9786	2.1E-02
2.4	0.9918	8.2E-03	0.9836	1.6E-02
2.5	0.9938	6.2E-03	0.9876	1.2E-02
2.6	0.9953	4.7E-03	0.9907	9.3E-03
2.7	0.9965	3.5E-03	0.9931	6.9E-03
2.8	0.9974	2.6E-03	0.9949	5.1E-03
2.9	0.9981	1.9E-03	0.9963	3.7E-03
3	0.99865	1.3E-03	0.99730	2.7E-03
3.1	0.99903	9.7E-04	0.99806	1.9E-03
3.2	0.99931	6.9E-04	0.99863	1.4E-03
3.3	0.99952	4.8E-04	0.99903	9.7E-04
3.4	0.99966	3.4E-04	0.99933	6.7E-04
3.5	0.99977	2.3E-04	0.99953	4.7E-04
3.6	0.999841	1.6E-04	0.999682	3.2E-04
3.7	0.999892	1.1E-04	0.999784	2.2E-04
3.8	0.999928	7.2E-05	0.999855	1.4E-04
3.9	0.999952	4.8E-05	0.999904	9.6E-05
4	0.999968	3.2E-05	0.999937	6.3E-05

First steps...

Integals over Gaussians

- **20/sqrt(28) = 3.8**

				
$s = \frac{x-\mu}{\sigma}$	$\int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$1 - \int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$\int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$1 - \int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$
0	0.500	5.0E-01	0.000	1.0E+00
0.1	0.540	4.6E-01	0.080	9.2E-01
0.2	0.579	4.2E-01	0.159	8.4E-01
0.3	0.618	3.8E-01	0.236	7.6E-01
0.4	0.655	3.4E-01	0.311	6.9E-01
0.5	0.691	3.1E-01	0.383	6.2E-01
0.6	0.726	2.7E-01	0.451	5.5E-01
0.7	0.758	2.4E-01	0.516	4.8E-01
0.8	0.788	2.1E-01	0.576	4.2E-01
0.9	0.816	1.8E-01	0.632	3.7E-01
1	0.841	1.6E-01	0.683	3.2E-01
1.1	0.864	1.4E-01	0.729	2.7E-01
1.2	0.885	1.2E-01	0.770	2.3E-01
1.3	0.903	9.7E-02	0.806	1.9E-01
1.4	0.919	8.1E-02	0.838	1.6E-01
1.5	0.933	6.7E-02	0.866	1.3E-01
2.4	0.9918	8.2E-03	0.9836	1.6E-02
2.5	0.9938	6.2E-03	0.9876	1.2E-02
2.6	0.9953	4.7E-03	0.9907	9.3E-03
2.7	0.9965	3.5E-03	0.9931	6.9E-03
2.8	0.9974	2.6E-03	0.9949	5.1E-03
2.9	0.9981	1.9E-03	0.9963	3.7E-03
3	0.99865	1.3E-03	0.99730	2.7E-03
3.1	0.99903	9.7E-04	0.99806	1.9E-03
3.2	0.99931	6.9E-04	0.99863	1.4E-03
3.3	0.99952	4.8E-04	0.99903	9.7E-04
3.4	0.99966	3.4E-04	0.99933	6.7E-04
3.5	0.99977	2.3E-04	0.99953	4.7E-04
3.6	0.999841	1.6E-04	0.999682	3.2E-04
3.7	0.999892	1.1E-04	0.999784	2.2E-04
3.8	0.999928	7.2E-05	0.999855	1.4E-04
3.9	0.999952	4.8E-05	0.999904	9.6E-05
4	0.999968	3.2E-05	0.999937	6.3E-05

7.2E-05

First steps...

- 20/sqrt(28) = 3.8
- p = $7.2 \cdot 10^{-5}$

Integals over Gaussians

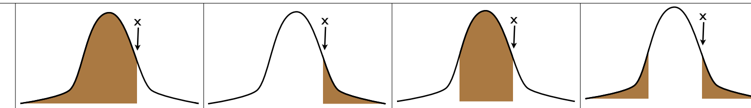
$s = \frac{x-\mu}{\sigma}$	$\int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$1 - \int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$\int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$1 - \int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$
0	0.500	5.0E-01	0.000	1.0E+00
0.1	0.540	4.6E-01	0.080	9.2E-01
0.2	0.579	4.2E-01	0.159	8.4E-01
0.3	0.618	3.8E-01	0.236	7.6E-01
0.4	0.655	3.4E-01	0.311	6.9E-01
0.5	0.691	3.1E-01	0.383	6.2E-01
0.6	0.726	2.7E-01	0.451	5.5E-01
0.7	0.758	2.4E-01	0.516	4.8E-01
0.8	0.788	2.1E-01	0.576	4.2E-01
0.9	0.816	1.8E-01	0.632	3.7E-01
1	0.841	1.6E-01	0.683	3.2E-01
1.1	0.864	1.4E-01	0.729	2.7E-01
1.2	0.885	1.2E-01	0.770	2.3E-01
1.3	0.903	9.7E-02	0.806	1.9E-01
1.4	0.919	8.1E-02	0.838	1.6E-01
1.5	0.933	6.7E-02	0.866	1.3E-01
2.4	0.9918	8.2E-03	0.9836	1.6E-02
2.5	0.9938	6.2E-03	0.9876	1.2E-02
2.6	0.9953	4.7E-03	0.9907	9.3E-03
2.7	0.9965	3.5E-03	0.9931	6.9E-03
2.8	0.9974	2.6E-03	0.9949	5.1E-03
2.9	0.9981	1.9E-03	0.9963	3.7E-03
3	0.99865	1.3E-03	0.99730	2.7E-03
3.1	0.99903	9.7E-04	0.99806	1.9E-03
3.2	0.99931	6.9E-04	0.99863	1.4E-03
3.3	0.99952	4.8E-04	0.99903	9.7E-04
3.4	0.99966	3.4E-04	0.99933	6.7E-04
3.5	0.99977	2.3E-04	0.99953	4.7E-04
3.6	0.999841	1.6E-04	0.999682	3.2E-04
3.7	0.999892	1.1E-04	0.999784	2.2E-04
3.8	0.999928	7.2E-05	0.999855	1.4E-04
3.9	0.999952	4.8E-05	0.999904	9.6E-05
4	0.999968	3.2E-05	0.999937	6.3E-05

7.2E-05

Integals over Gaussians

First steps...

- 20/sqrt(28) = 3.8
- p = $7.2 \cdot 10^{-5}$
- this is called the p-value, it the probability to see such a fluctuation in the SM (no Z'). It is very small (a "3.8 sigma effect").

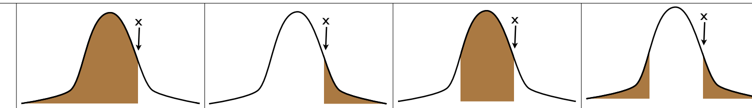


$s = \frac{x-\mu}{\sigma}$	$\int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$1 - \int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$\int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$1 - \int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$
0	0.500	5.0E-01	0.000	1.0E+00
0.1	0.540	4.6E-01	0.080	9.2E-01
0.2	0.579	4.2E-01	0.159	8.4E-01
0.3	0.618	3.8E-01	0.236	7.6E-01
0.4	0.655	3.4E-01	0.311	6.9E-01
0.5	0.691	3.1E-01	0.383	6.2E-01
0.6	0.726	2.7E-01	0.451	5.5E-01
0.7	0.758	2.4E-01	0.516	4.8E-01
0.8	0.788	2.1E-01	0.576	4.2E-01
0.9	0.816	1.8E-01	0.632	3.7E-01
1	0.841	1.6E-01	0.683	3.2E-01
1.1	0.864	1.4E-01	0.729	2.7E-01
1.2	0.885	1.2E-01	0.770	2.3E-01
1.3	0.903	9.7E-02	0.806	1.9E-01
1.4	0.919	8.1E-02	0.838	1.6E-01
1.5	0.933	6.7E-02	0.866	1.3E-01
2.4	0.9918	8.2E-03	0.9836	1.6E-02
2.5	0.9938	6.2E-03	0.9876	1.2E-02
2.6	0.9953	4.7E-03	0.9907	9.3E-03
2.7	0.9965	3.5E-03	0.9931	6.9E-03
2.8	0.9974	2.6E-03	0.9949	5.1E-03
2.9	0.9981	1.9E-03	0.9963	3.7E-03
3	0.99865	1.3E-03	0.99730	2.7E-03
3.1	0.99903	9.7E-04	0.99806	1.9E-03
3.2	0.99931	6.9E-04	0.99863	1.4E-03
3.3	0.99952	4.8E-04	0.99903	9.7E-04
3.4	0.99966	3.4E-04	0.99933	6.7E-04
3.5	0.99977	2.3E-04	0.99953	4.7E-04
3.6	0.999841	1.6E-04	0.999682	3.2E-04
3.7	0.999892	1.1E-04	0.999784	2.2E-04
3.8	0.999928	7.2E-05	0.999855	1.4E-04
3.9	0.999952	4.8E-05	0.999904	9.6E-05
4	0.999968	3.2E-05	0.999937	6.3E-05

7.2E-05

First steps...

- 20/sqrt(28) = 3.8
- p = $7.2\cdot 10^{-5}$
- this is called the p-value, it the probability to see such a fluctuation in the SM (no Z'). It is very small (a "3.8 sigma effect").
- But, we looked at many bins. If want to calculate how likely I am to make a wrong discovery, I need to know how likely am to get such an unlikely event (i.e. event with this p value) at in at least one bin.



$s = \frac{x-\mu}{\sigma}$	$\int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$1 - \int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$\int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$	$1 - \int_{-s}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$
0	0.500	5.0E-01	0.000	1.0E+00
0.1	0.540	4.6E-01	0.080	9.2E-01
0.2	0.579	4.2E-01	0.159	8.4E-01
0.3	0.618	3.8E-01	0.236	7.6E-01
0.4	0.655	3.4E-01	0.311	6.9E-01
0.5	0.691	3.1E-01	0.383	6.2E-01
0.6	0.726	2.7E-01	0.451	5.5E-01
0.7	0.758	2.4E-01	0.516	4.8E-01
0.8	0.788	2.1E-01	0.576	4.2E-01
0.9	0.816	1.8E-01	0.632	3.7E-01
1	0.841	1.6E-01	0.683	3.2E-01
1.1	0.864	1.4E-01	0.729	2.7E-01
1.2	0.885	1.2E-01	0.770	2.3E-01
1.3	0.903	9.7E-02	0.806	1.9E-01
1.4	0.919	8.1E-02	0.838	1.6E-01
1.5	0.933	6.7E-02	0.866	1.3E-01
2.4	0.9918	8.2E-03	0.9836	1.6E-02
2.5	0.9938	6.2E-03	0.9876	1.2E-02
2.6	0.9953	4.7E-03	0.9907	9.3E-03
2.7	0.9965	3.5E-03	0.9931	6.9E-03
2.8	0.9974	2.6E-03	0.9949	5.1E-03
2.9	0.9981	1.9E-03	0.9963	3.7E-03
3	0.99865	1.3E-03	0.99730	2.7E-03
3.1	0.99903	9.7E-04	0.99806	1.9E-03
3.2	0.99931	6.9E-04	0.9986[illegible]	1.4E-03
3.3	0.99952	4.8E-04	0.99903	9.7E-04
3.4	0.99966	3.4E-04	0.99933	6.7E-04
3.5	0.99977	2.3E-04	0.99953	4.7E-04
3.6	0.999841	1.6E-04	0.999682	3.2E-04
3.7	0.999892	1.1E-04	0.999784	2.2E-04
3.8	0.999928	7.2E-05	0.999855	1.4E-04
3.9	0.999952	4.8E-05	0.999904	9.6E-05
4	0.999968	3.2E-05	0.999937	6.3E-05

7.2E-05

PDFs: important properties

- **Probabilities by integrating PDFs** $P(x \in [a,b]) = \int_a^b P(x')\,dx'$

- **Normalisation - P(something happens) = 1:** $\int_{-\infty}^{+\infty} P(x')\,dx' = 1$

- **Expectation value of x, or any function of x, gives the average expected outcome for x (function of x)**

$$\langle x \rangle = \int x' P(x')\,dx' \qquad \langle f(x) \rangle = \int f(x') P(x')\,dx'$$

- **Variance** $V = \langle x^2 \rangle - \langle x \rangle^2$

- **Change of variables:** $P(y) = P(x) \left| \frac{dx}{dy} \right|$

Alternative Definitions of V and σ

- There are other definitions of V and σ on the market. One frequently encountered is this:

$$V_{N-1} = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{x} - x_i) = V_N \frac{N}{N-1}$$

- And correspondingly $\sigma_{N-1} = \sqrt{V_{N-1}} = \sigma_N \sqrt{\frac{N}{N-1}}$

- This has a slightly different value, and it has a subtly different meaning. Importantly, though, both definitions tend to the same value as N increases.

- In this course, the symbols V, σ stand for V_N, σ_N.

Subtleties of σ_N, σ_{N-1}

$$\lim_{N' \to \infty} \sigma_{N'}$$

Subtleties of σ_N, σ_{N-1}

- **As with many things of little importance, people can get very passionate about them.**

$$\lim_{N' \to \infty} \sigma_{N'}$$

Subtleties of σ_N, σ_{N-1}

- As with many things of little importance, people can get very passionate about them.

- σ_N represents the spread of the distribution as we measure it.

$$\lim_{N' \to \infty} \sigma_{N'}$$

Subtleties of σ_N, σ_{N-1}

- As with many things of little importance, people can get very passionate about them.

- σ_N represents the spread of the distribution as we measure it.

- What's behind σ_{N-1} is the concept that our measurements are drawn from a much larger (infinitely large), theoretical parent distribution. σ_{N-1} provides an unbiased estimate of the standard deviation of $\sigma_{\text{theor-parent}}$. So the point of σ_{N-1} is to estimate $\sigma_{\text{theor-parent}} = \lim_{N' \to \infty} \sigma_{N'}$ i.e. to estimate what σ_N would be had we got much more data.

Subtleties of σ_N, σ_{N-1}

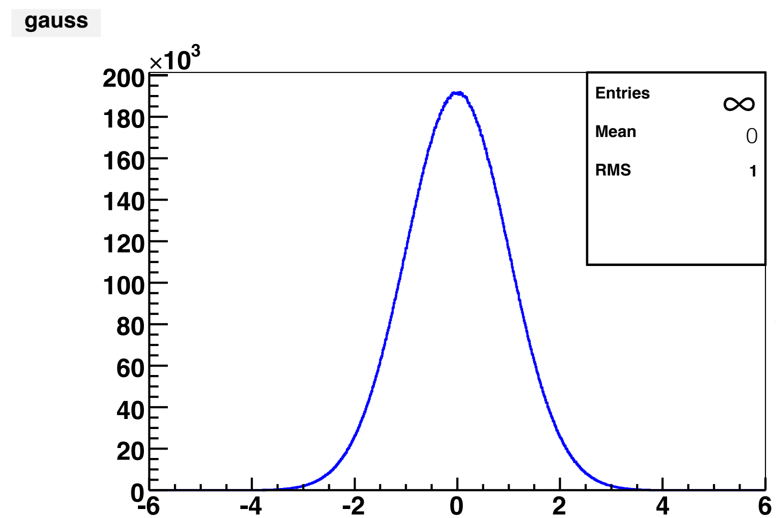
- As with many things of little importance, people can get very passionate about them.

- σ_N represents the spread of the distribution as we measure it.

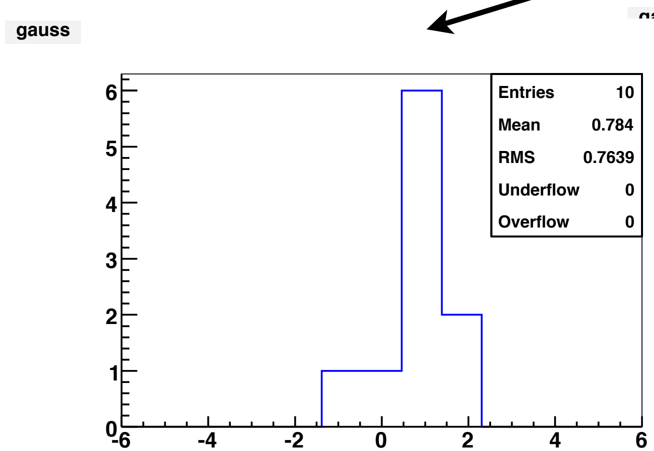
- What's behind σ_{N-1} is the concept that our measurements are drawn from a much larger (infinitely large), theoretical parent distribution. σ_{N-1} provides an unbiased estimate of the standard deviation of $\sigma_{\text{theor-parent}}$. So the point of σ_{N-1} is to estimate $\sigma_{\text{theor-parent}} = \lim_{N' \to \infty} \sigma_{N'}$ i.e. to estimate what σ_N would be had we got much more data.

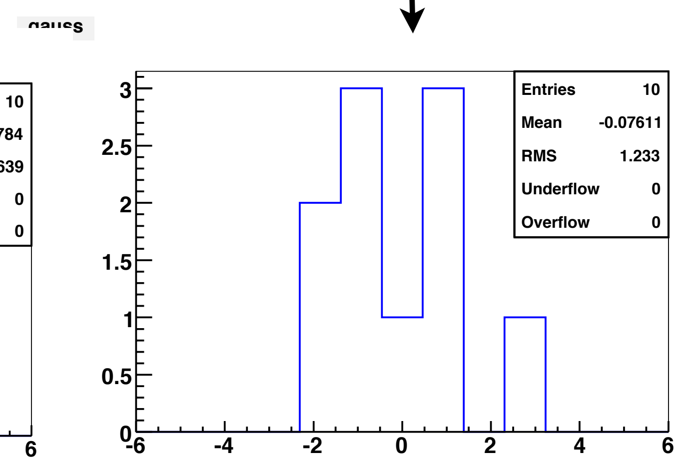
- So now we have 3 sigmas: σ_N, σ_{N-1}, $\sigma_{\text{theor-parent}}$.

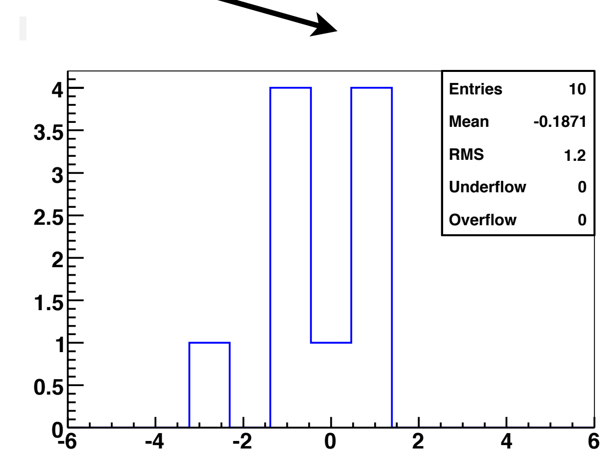
Idea of “ideal” parent sample



Ideal parent sample, in limit of infinite statistics (practically inaccessible)







Possible measured samples randomly drawn from parent

Subtleties of σ_N, σ_{N-1}

- There are a few problems with $\sigma_{theor\text{-}parent}$:
    - As we'll see later, not for all theoretical distributions is $\sigma_{theor\text{-}parent}$ defined (can be infinite), in which case there's not much point in estimating it - but σ_N is defined for all measured distributions with a finite number of data points.
    - An unbiased estimate is nice, however, being unbiased is only one of the many qualities of an estimate - efficiency (i.e. gets close to the truth fast) is another. σ_N is the most efficient estimator of $\sigma_{theor\text{-}parent}$, while σ_{N-1} is unbiased.
- σ_{N-1} gives larger values, i.e. is more conservative, and therefore liked by many cautious scientists.
- Here we'll take the view that data are the data and they have a well-defined sigma (σ_N), and that's that. Estimating the parameters of theoretical parent distributions from this is something different, that we will also look into, but separately and later.

Subtleties of σ_N, σ_{N-1}

- Don't loose any sleep over it. But it will come up again and again - now you know what it means.

- You got a first glimpse on important topics such as parameter estimation and parent sample.

- In this course, the variance and standard deviation of a sample are, respectively, $V=V_N$, $\sigma=\sigma_N$

- The important thing is that people know what convention you use and you stick to it.

- Even more importantly, remember $\sigma_N \approx \sigma_{N-1}$ as N gets large.

Moments

- The kth moment of a sample is just the average of the kth power of each data value:

$$m_k \equiv \frac{1}{N}\sum_{i=1}^{N} x_i^k \equiv \overline{x^k}$$

What is the 0th, 1st, moment?

- The kth central moment is

$$c_k \equiv \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^k \equiv \overline{(x - \bar{x})^k}$$

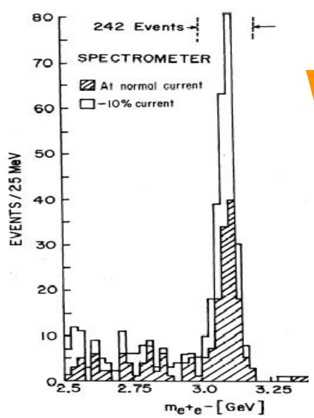
What is the 1st, 2nd central moment?

- So the mean is the 1st moment of the sample, the variance is the 2nd central moment.

- Higher moments play a marginal role in data analysis (with few exceptions), we won't consider them, here.

Lecture 2

Roadmap



What do I see?

Describing Data

What do I expect?

Probability and probability distributions, Probability density functions

Central Limit Theorem

Is what I see compatible with what I expect?

Discoveries
Confidence Levels
Hypothesis testing
Fitting

Monte Carlo simulation

Loss-free generation

- Trick: Solve this (for x = flat distribution) for t:

$$\begin{aligned} P(t) &= P(x)\frac{dx}{dt}, \quad \text{with } P(x) = 1, \quad P(t) = \frac{1}{\tau}e^{-t/\tau} \\ \frac{dx}{dt} &= \frac{1}{\tau}e^{-t/\tau} \\ x &= -e^{-t/\tau} + C \\ -\tau\ln(C-x) &= t \end{aligned}$$

- Integration constant C=1 determined by requirement to map $x \in [0,1]$ to $t \in [0,\infty]$.

$$t = -\tau\ln(1-x)$$

- This simple parameter transformation will change your flat distribution in x to an exponential. Neat.

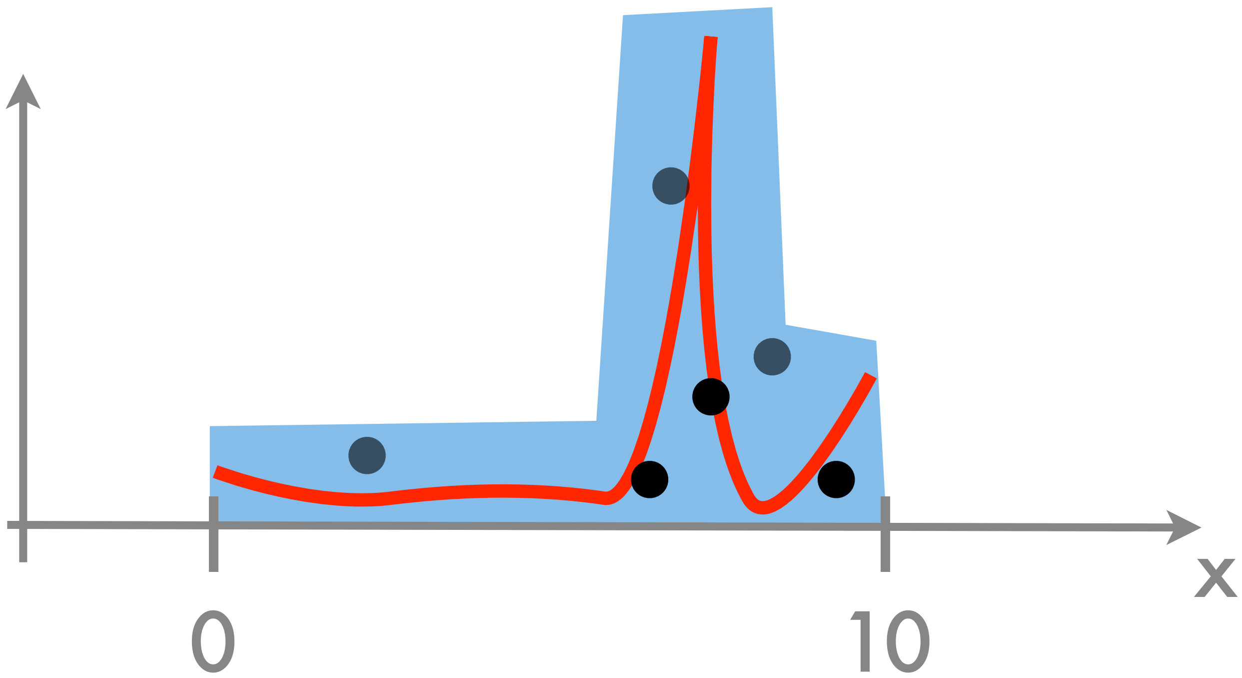
Efficient generation

- In reality, it is often not possible to find the correct parameter transformation - first of all you need to find an integral (not always easy), and then you have to invert the result (not always easy).
- What often works, though, is to generate something similar to the real distribution and then apply accept-reject on the ratio:

r = full_pdf(x)/pdf_used_for_efficient_generation(x);

if(y < r)acceptEvent(x,y)

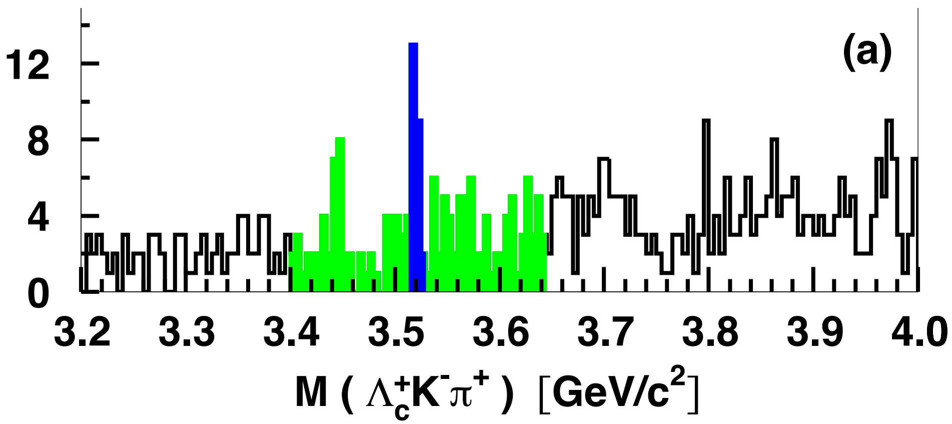
Mixed approach

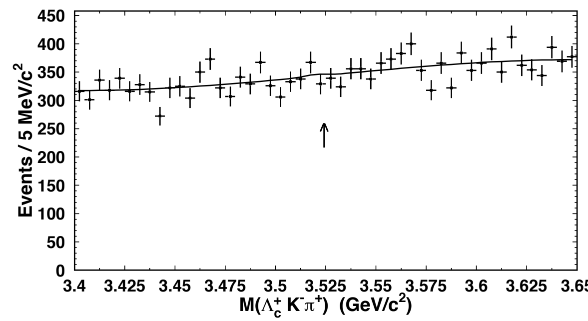


- First generate something similar to the true distribution and then use accept reject on that...

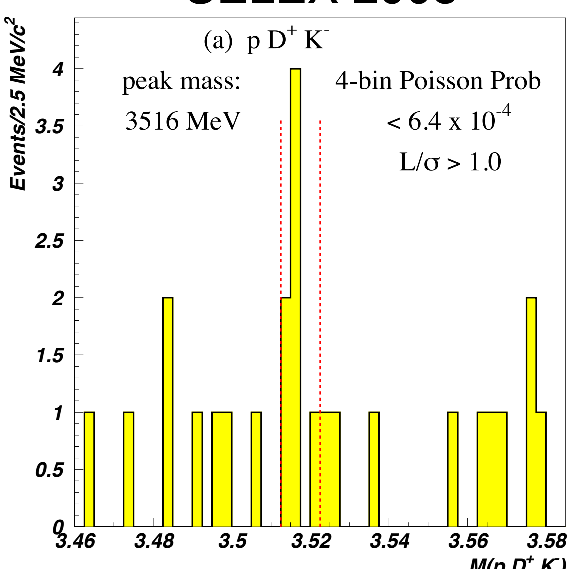
A Ξ_{cc} at 3.5 GeV?

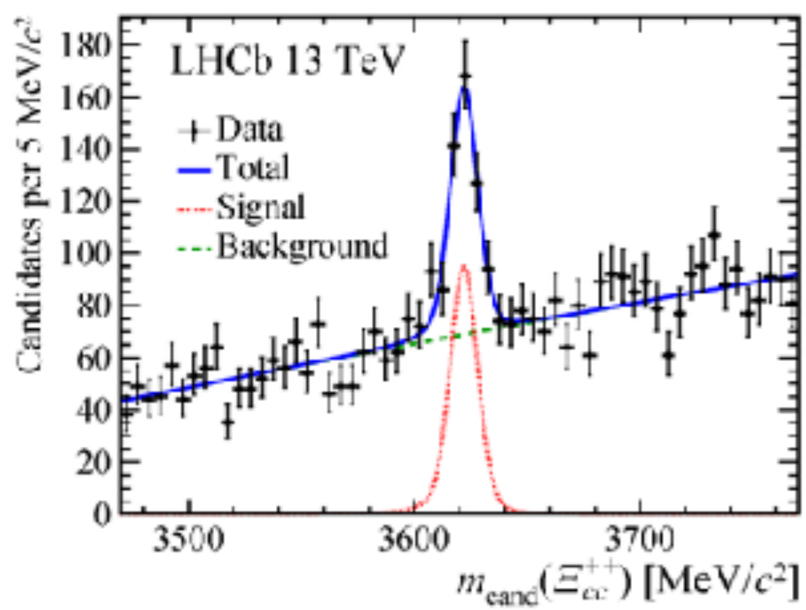
SELEX 2002





SELEX 2005

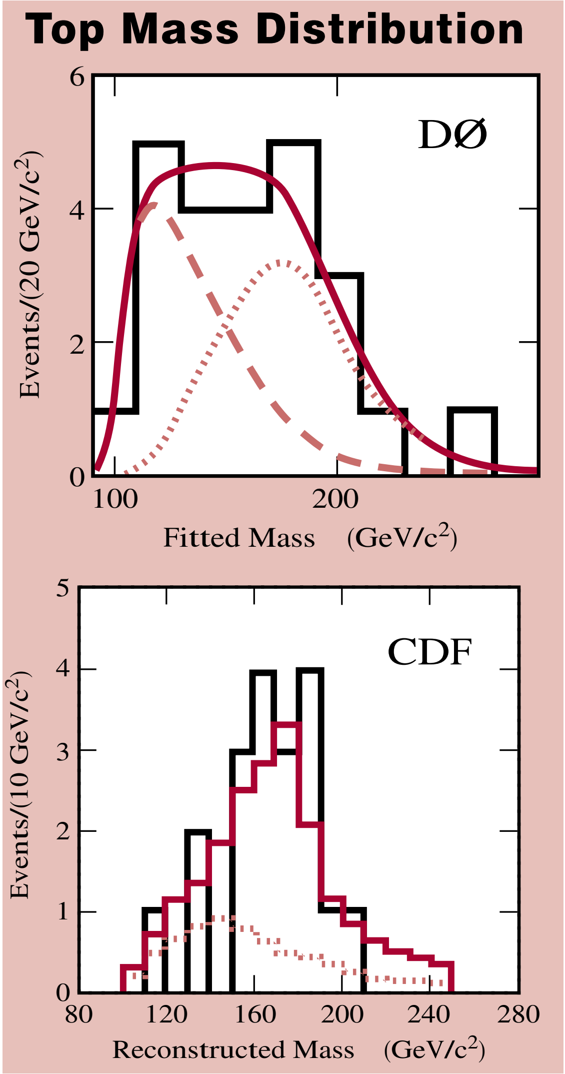




Discovery of Top at the Tevatron, Fermilab, in 1995

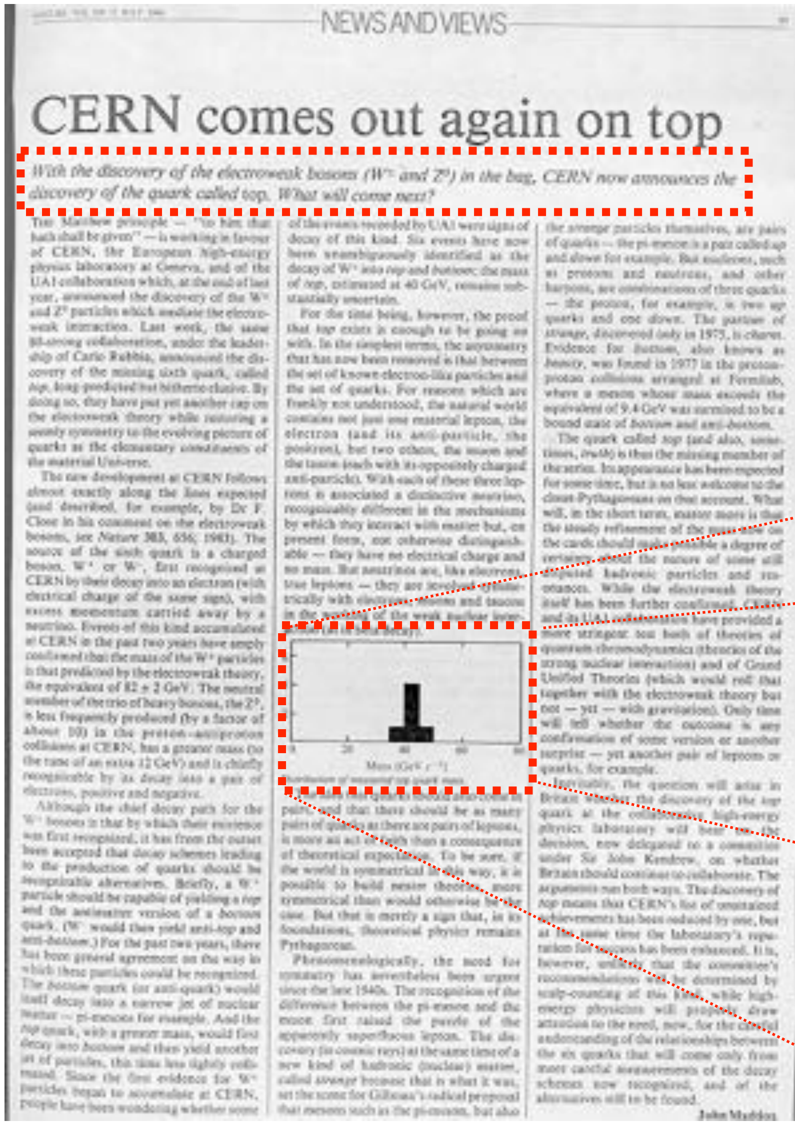
MANKIND has sought the elementary building blocks of matter ever since the days of the Greek philosophers. Over time, the quest has been successively refined from the original notion of indivisible "atoms" as the fundamental elements to the present idea that objects called quarks lie at the heart of all matter. So the recent news from Fermilab that the sixth—and possibly the last—of these quarks has finally been found may signal the end of one of our longest searches.

Fermilab Visual Media Services



Discovery of the Top at CERN

Discovery of the Top at CERN



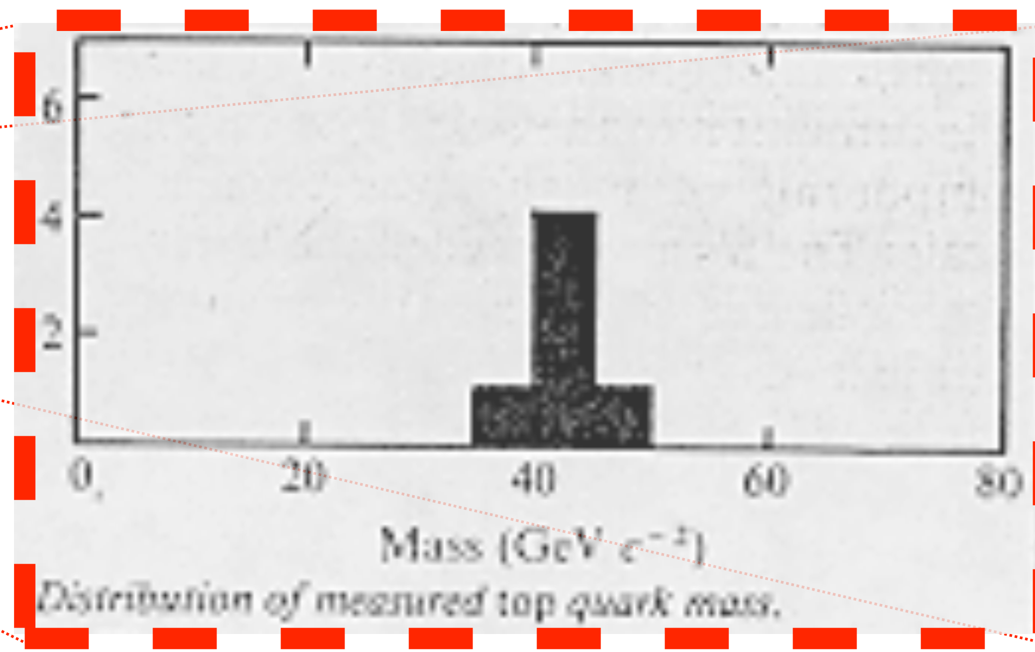
NEWS AND VIEWS

CERN comes out again on top

With the discovery of the electroweak bosons (W± and Z⁰) in the bag, CERN now announces the discovery of the quark called top. What will come next?

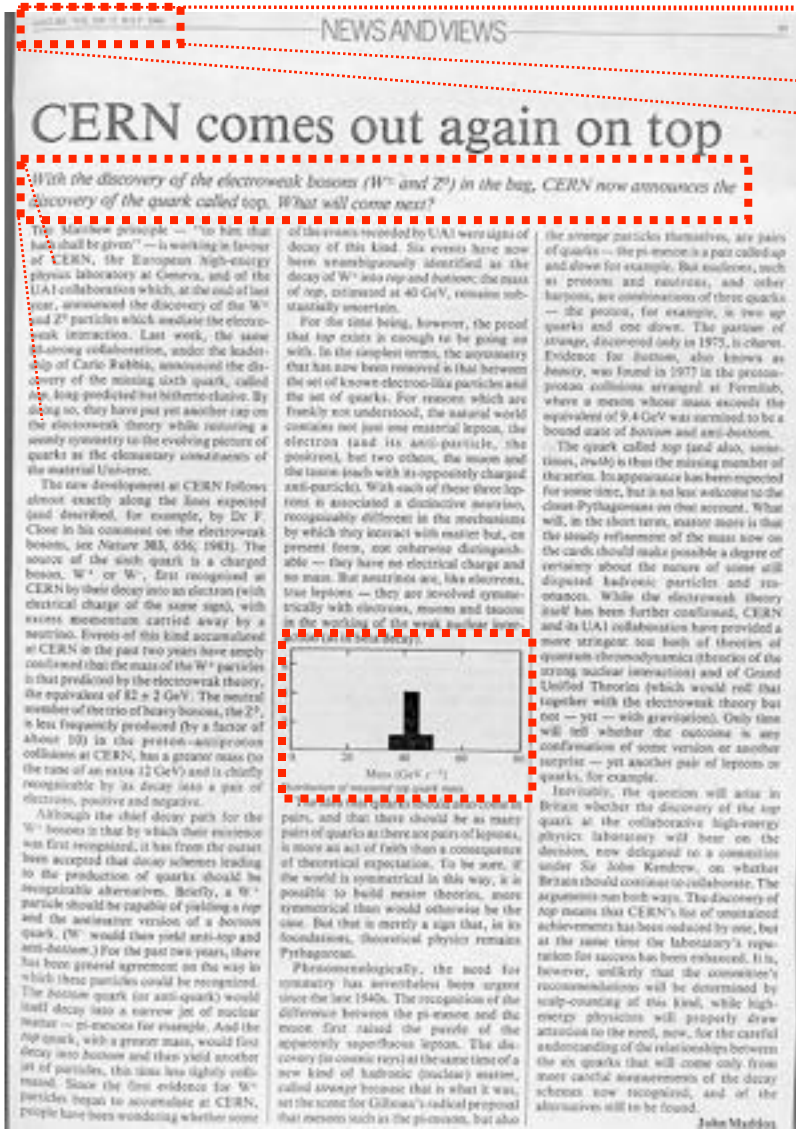
Distribution of measured top quark mass.

John Maddox



Distribution of measured top quark mass.

“Discovery” of the Top at CERN


NEWS AND VIEWS

CERN comes out again on top

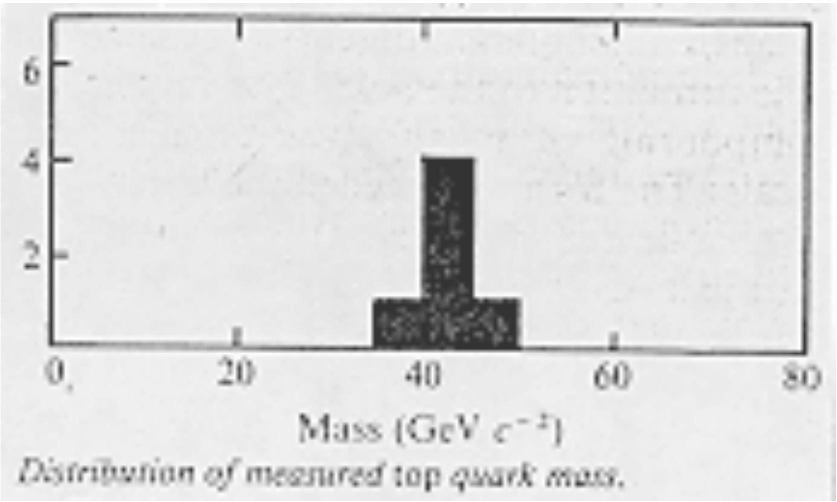
With the discovery of the electroweak bosons ($W^{\pm}$ and Z^0) in the bag, CERN now announces the discovery of the quark called top. What will come next?

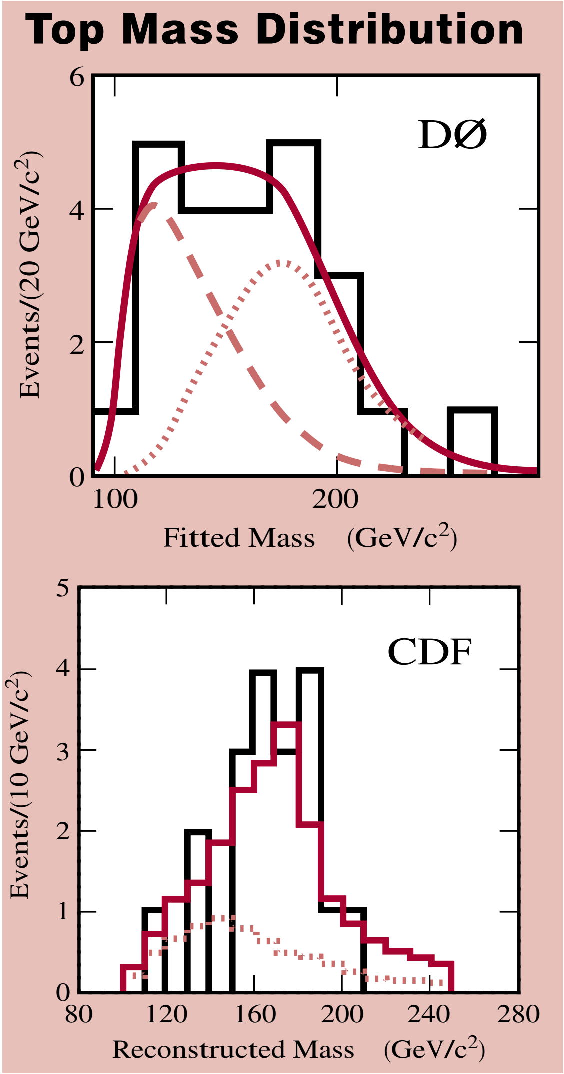
John Maddox

NATURE VOL 310 12 JULY 1984

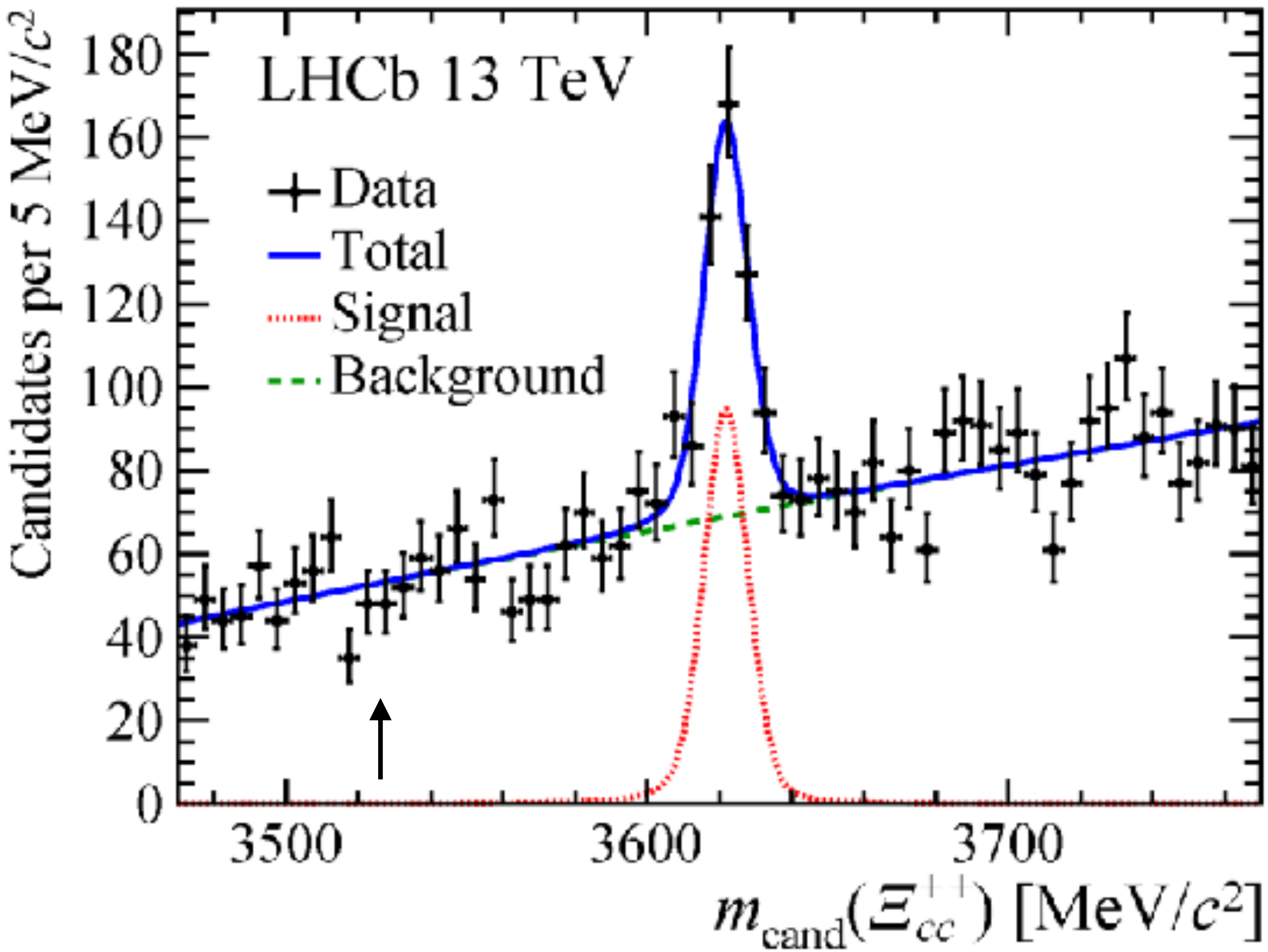
True and False

Can you see the difference?

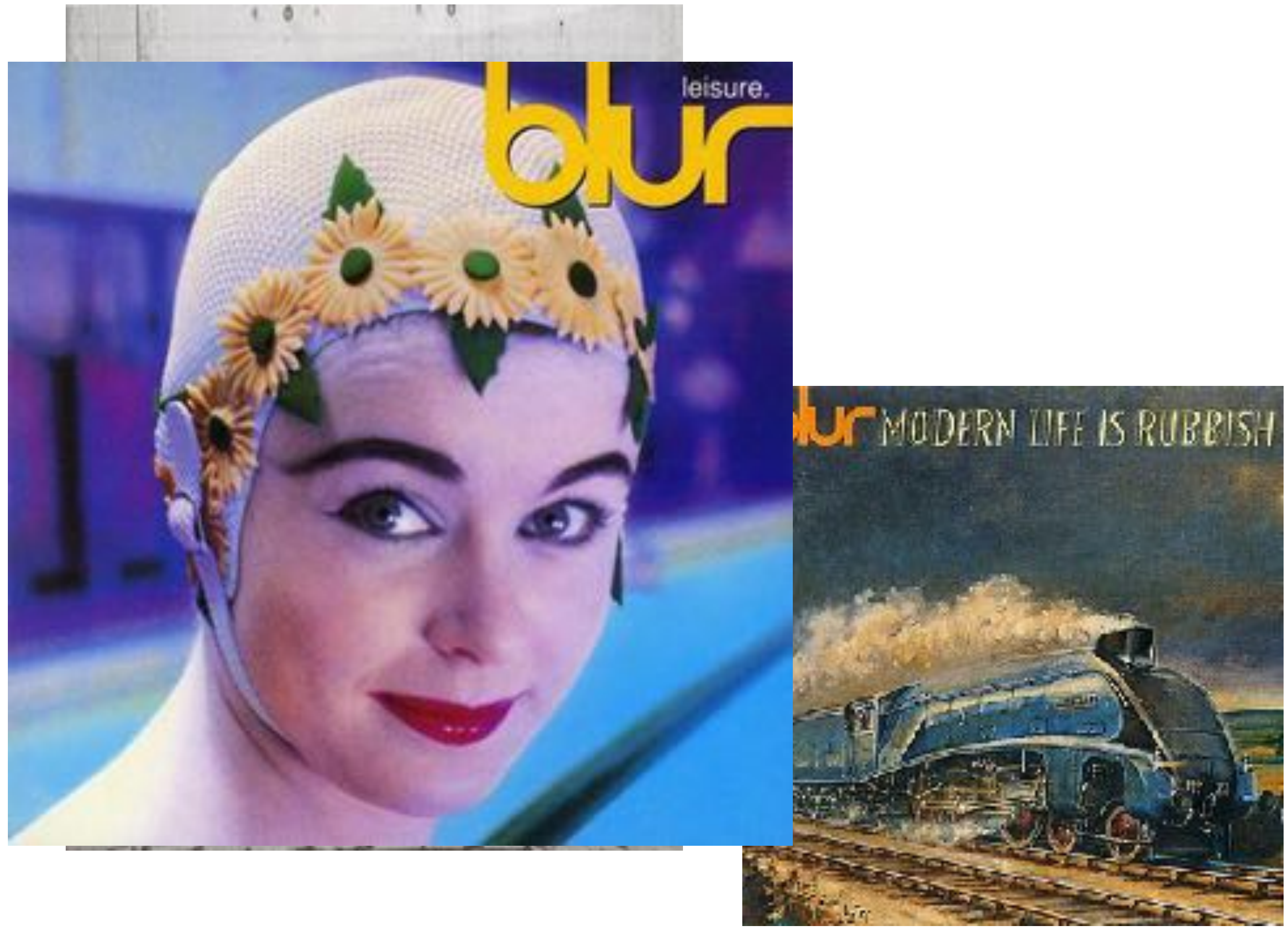


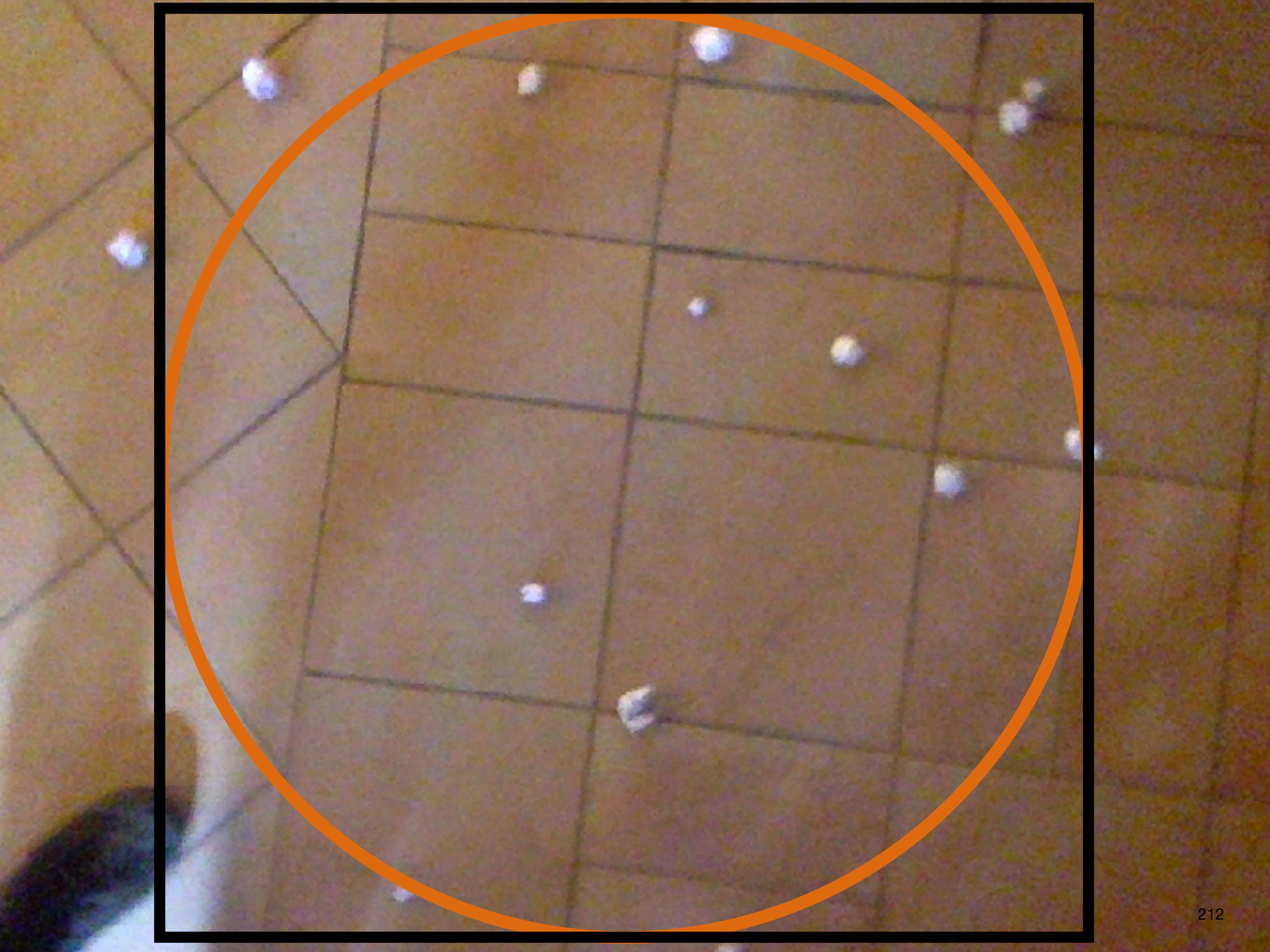


A Ξ_{cc} a few days ago



Blur





Discovery of the Top at CERN

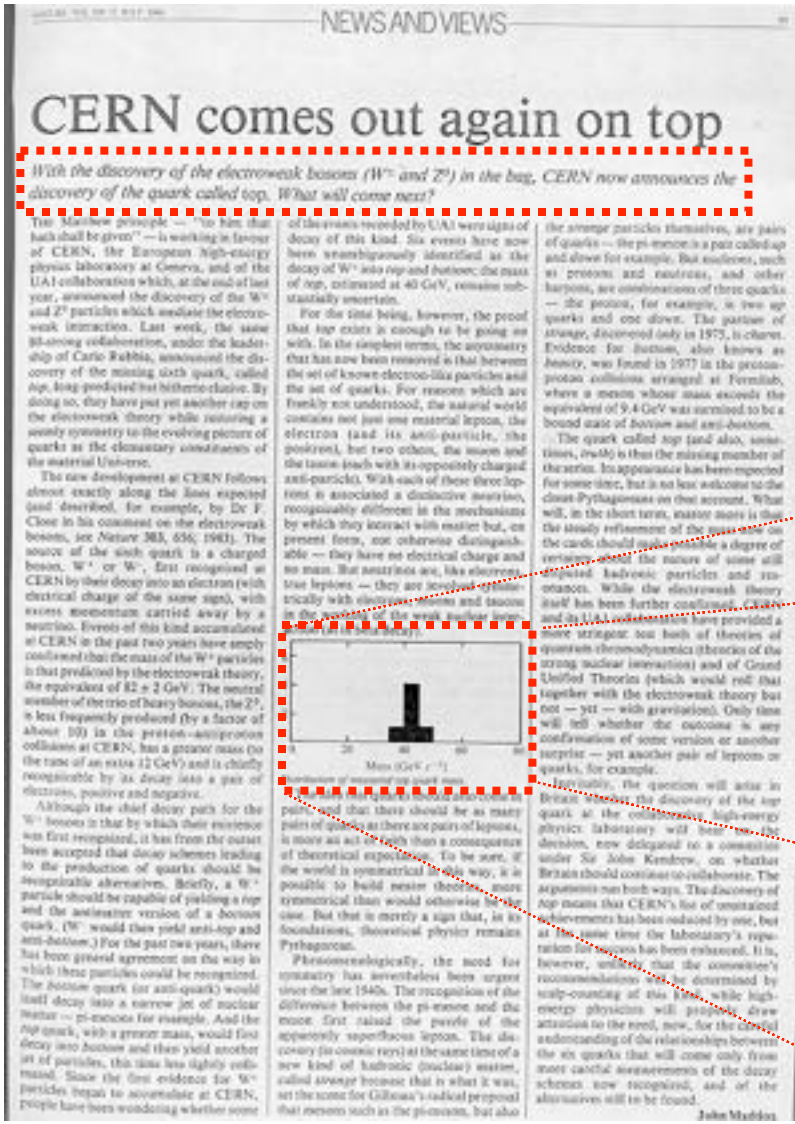
NEWS AND VIEWS

CERN comes out again on top

With the discovery of the electroweak bosons ($W^{\pm}$ and Z^0) in the bag, CERN now announces the discovery of the quark called top. What will come next?

With the discovery of the electroweak bosons ($W^{\pm}$ and Z^0) in the bag, CERN now announces the discovery of the quark called top. What will come next?

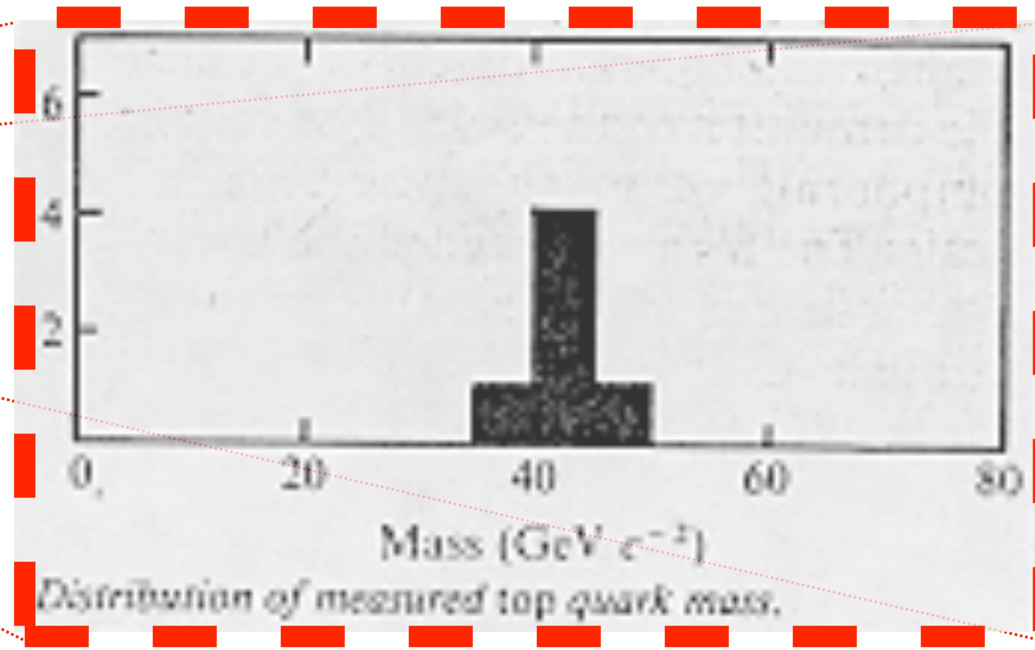
Discovery of the Top at CERN



NEWS AND VIEWS

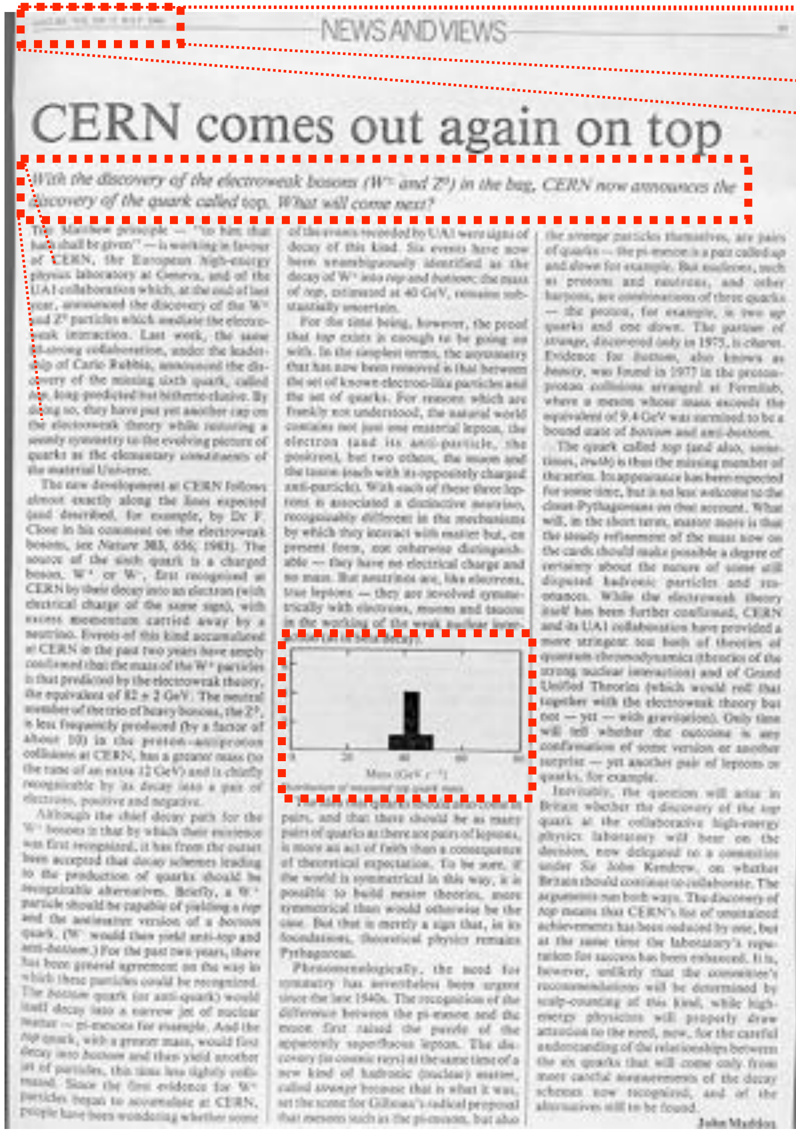
CERN comes out again on top

With the discovery of the electroweak bosons ($W^{\pm}$ and Z^0) in the bag, CERN now announces the discovery of the quark called top. What will come next?



Distribution of measured top quark mass.

“Discovery” of the Top at CERN

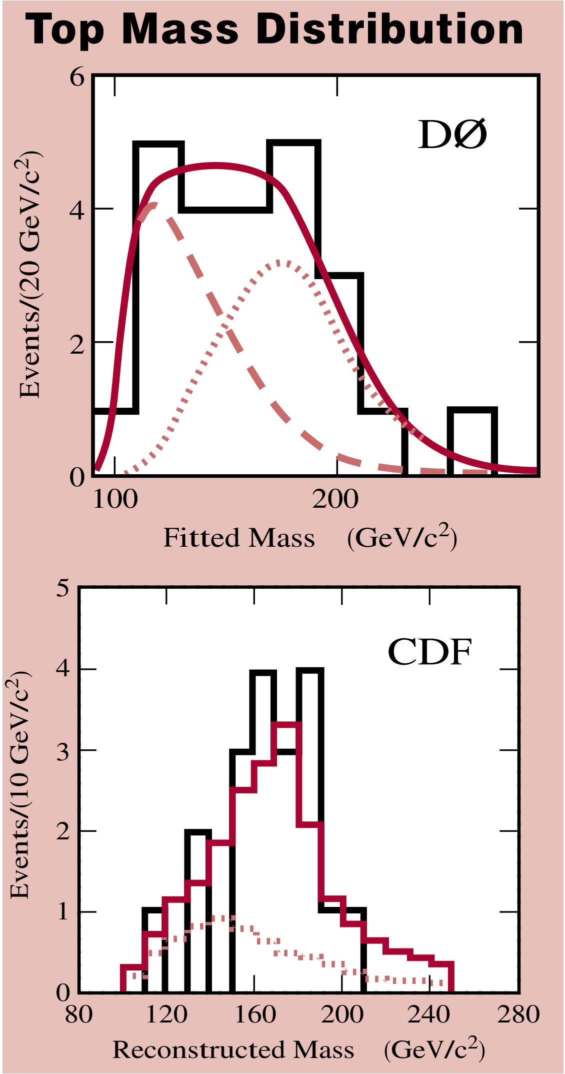


NATURE VOL 310 12 JULY 1984

Discovery of Top at the Tevatron, Fermilab, in 1995

MANKIND has sought the elementary building blocks of matter ever since the days of the Greek philosophers. Over time, the quest has been successively refined from the original notion of indivisible "atoms" as the fundamental elements to the present idea that objects called quarks lie at the heart of all matter. So the recent news from Fermilab that the sixth—and possibly the last—of these quarks has finally been found may signal the end of one of our longest searches.

Fermilab Visual Media Services



Problems & examples

To run the jupyter notebooks, you need to install some software. Follow the instructions here:

http://jupyter.readthedocs.io/en/latest/install.html

(note: there is an option to install the notebooks with python 2 or 3 - my notebooks use python 3.

I recommend the “anaconda route”. To install anaconda:

https://docs.anaconda.com/anaconda/