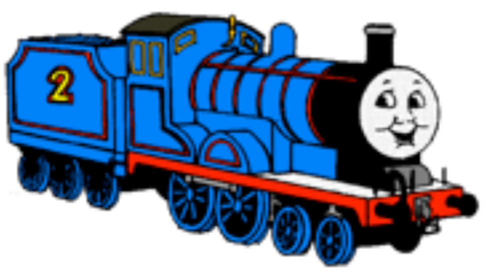
Statistics

Jonas Rademacker at TESHEP 2024

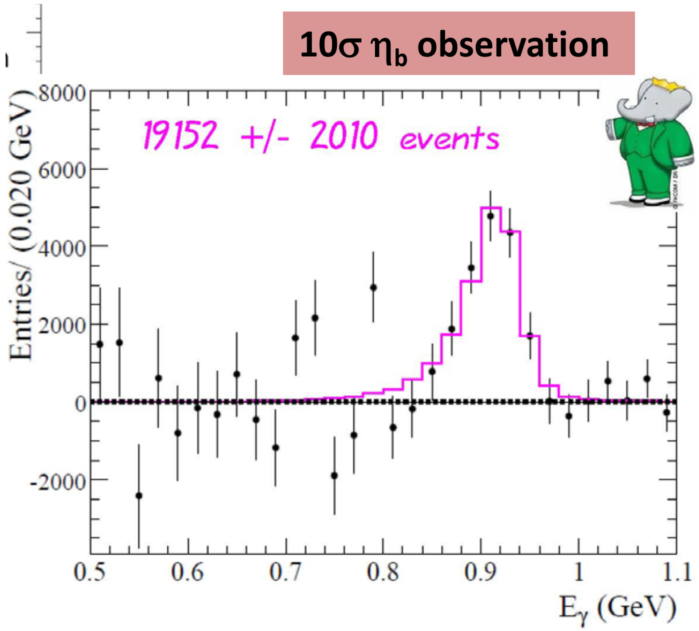
Statistics,Probability and Physics

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

Quantum Mechanics



Thermodynamics



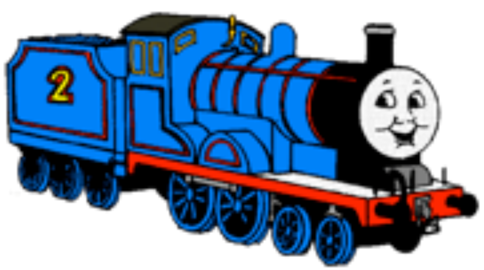
Interpretation of data

Statistics,Probability and Physics

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

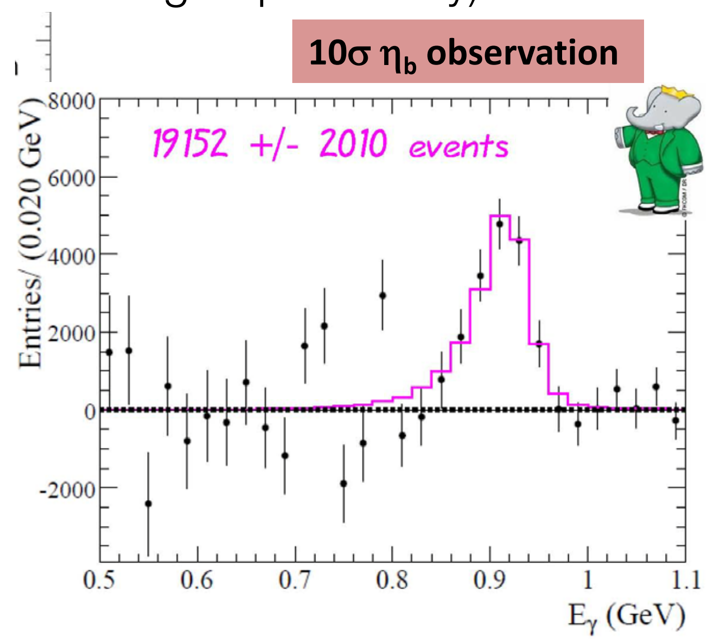
Quantum Mechanics

(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?

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

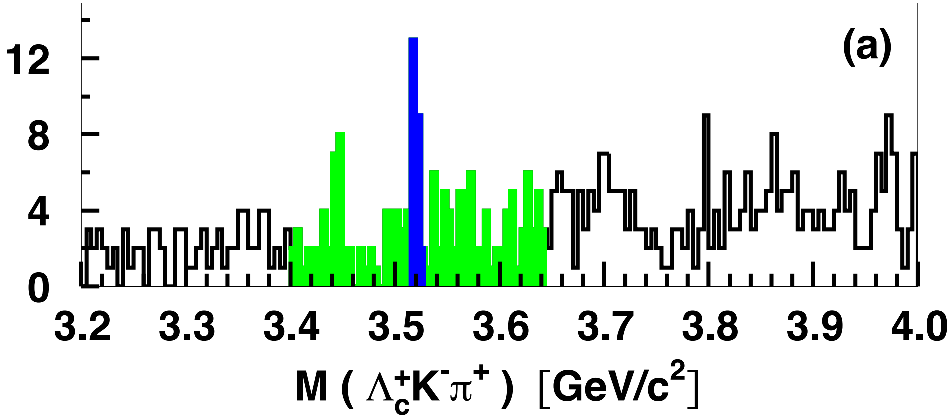
For a physics Masters/Ph.D….

- 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?

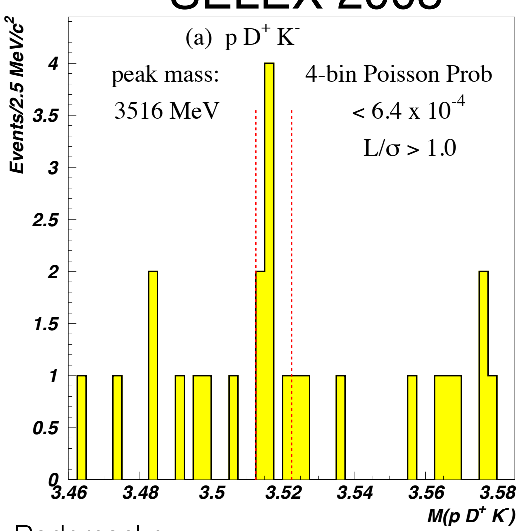
A Ξ_{cc} at 3.5 GeV?

SELEX see it twice

SELEX 2002



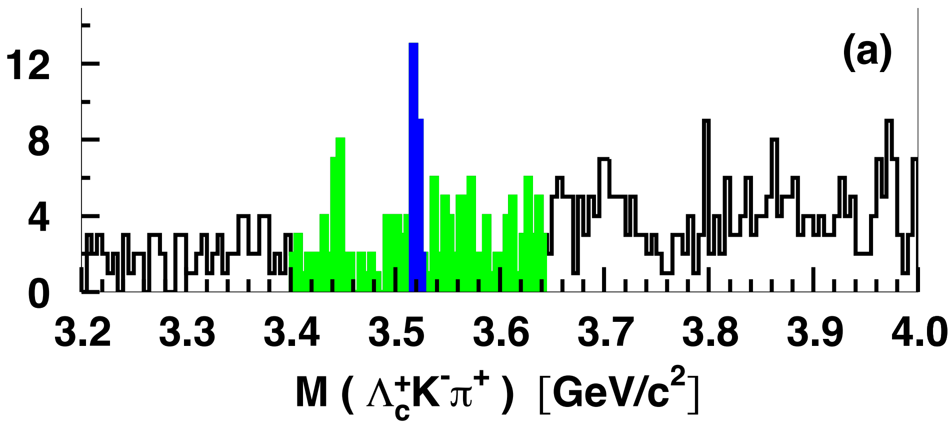
SELEX 2005



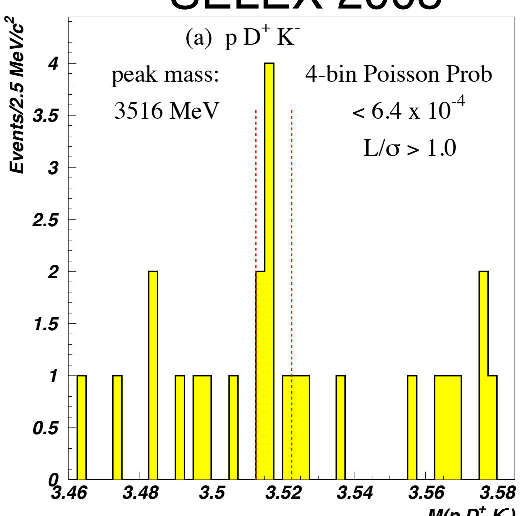
A Ξ_{cc} at 3.5 GeV?

SELEX see it twice

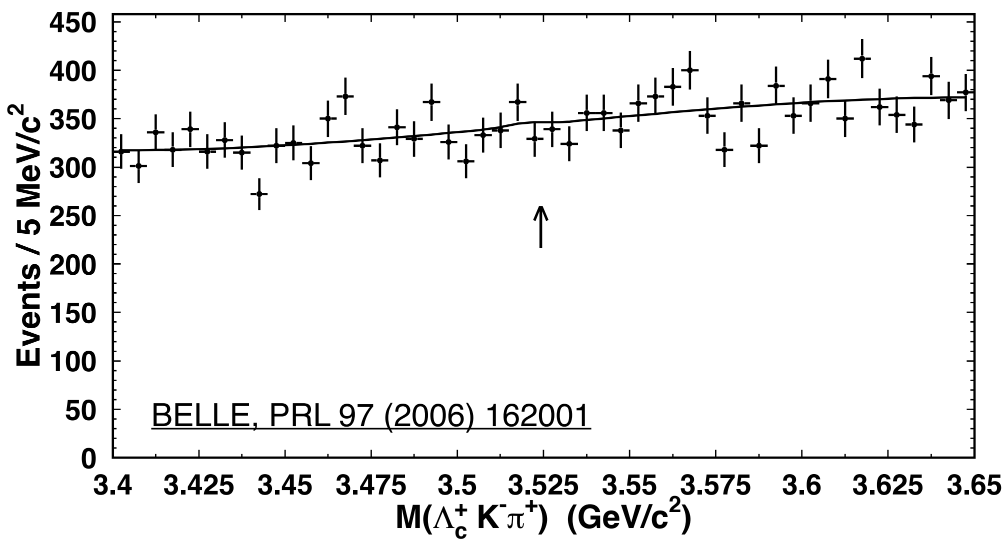
SELEX 2002



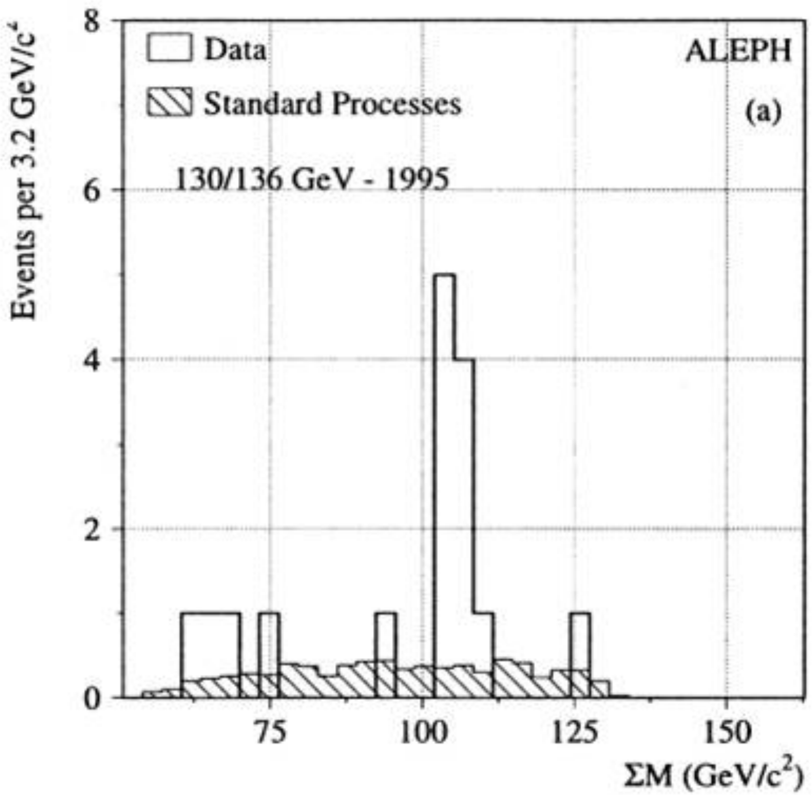
SELEX 2005



FOCUS, BaBar, BELLE, LHCb don't



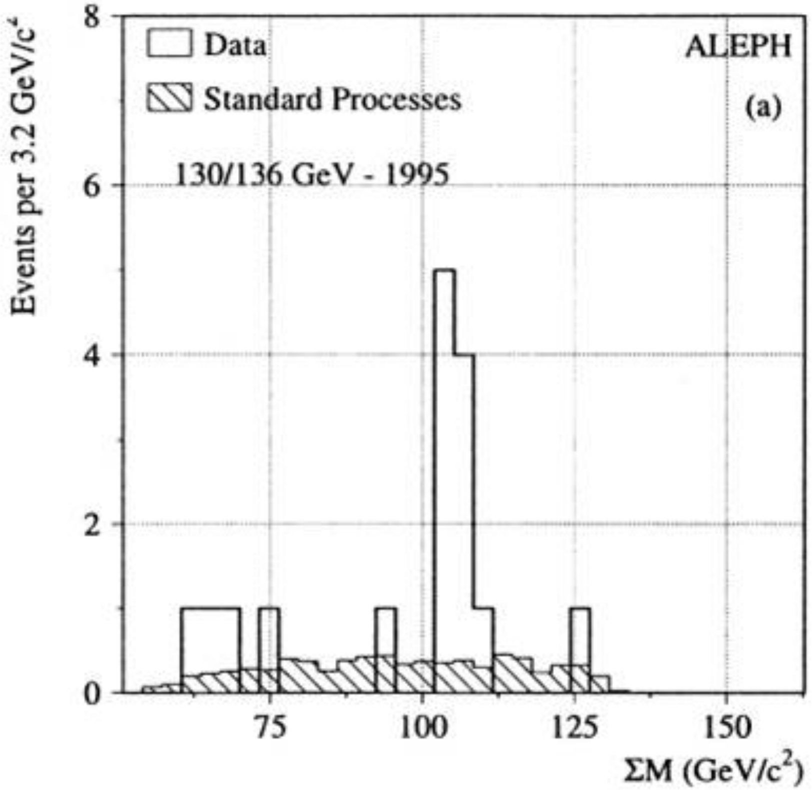
Higgs: true or false?



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)

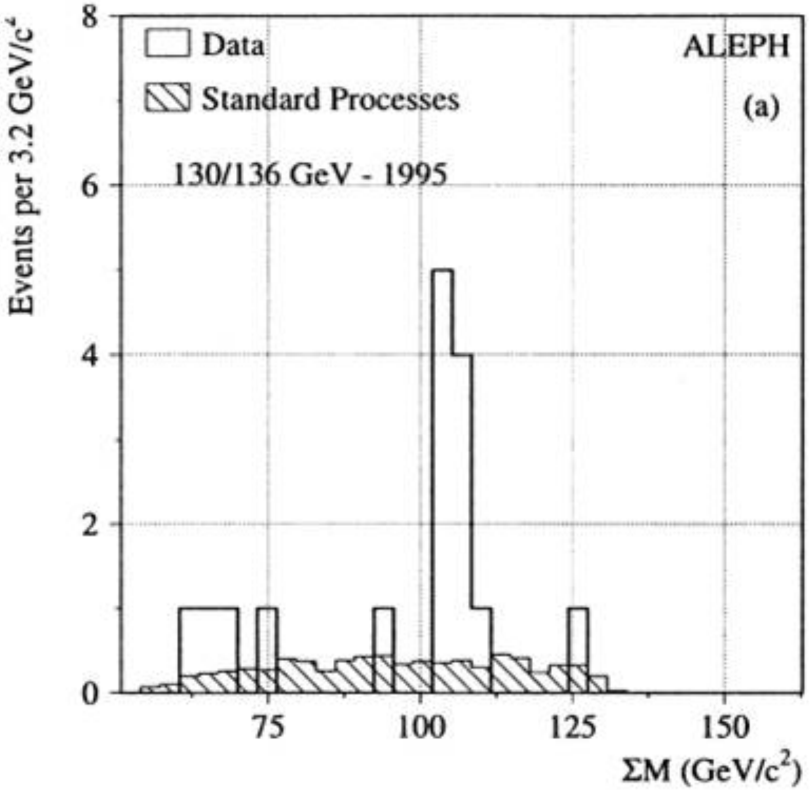


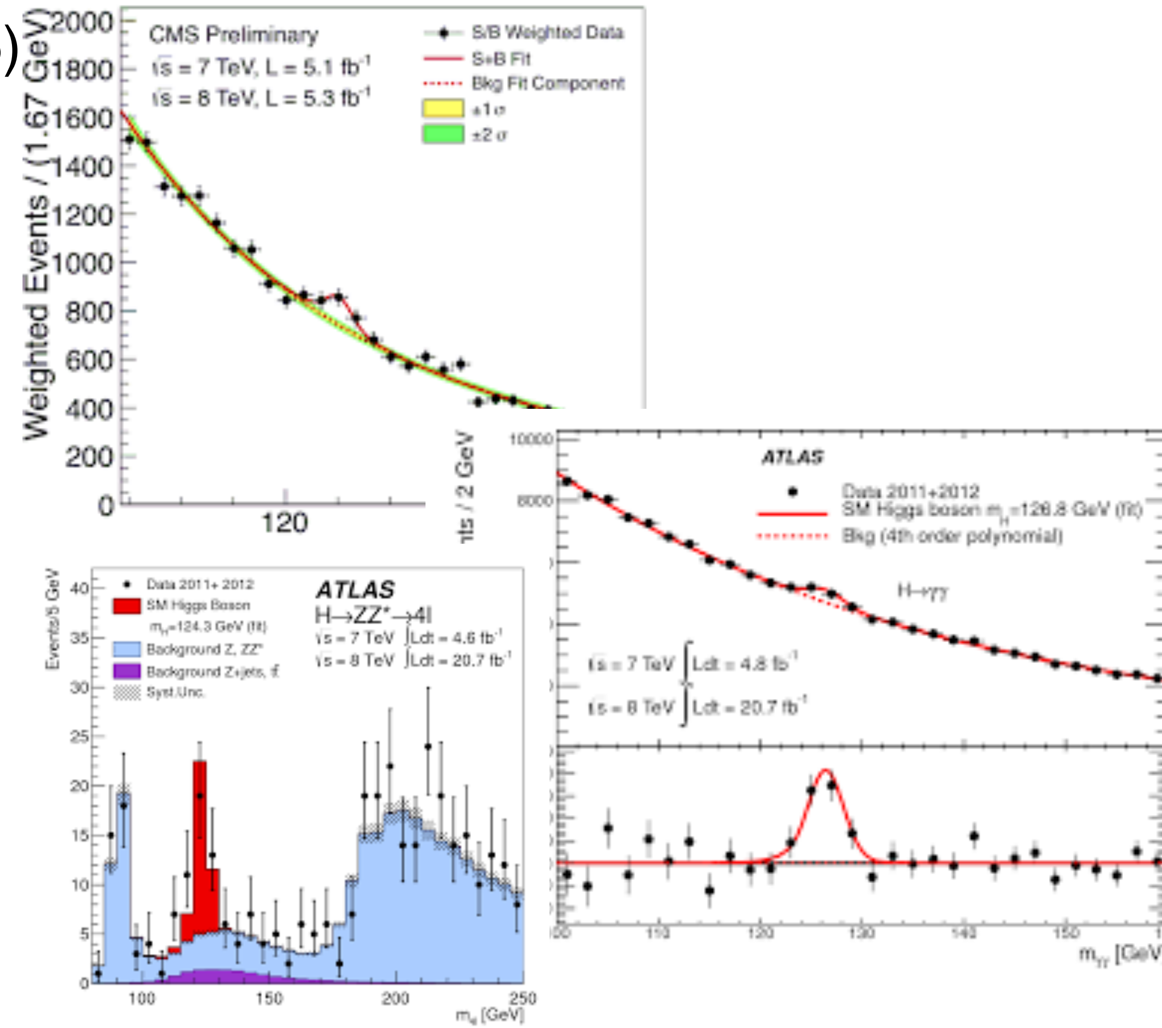
see: http://www.science20.com/a_quantum_diaries_survivor/true_and_false_discoveries_how_to_tell_them_apart-141024

Higgs: true or false?

real Higgs (LHC 1996)

false Higgs (ALEPH/LEP 1996)

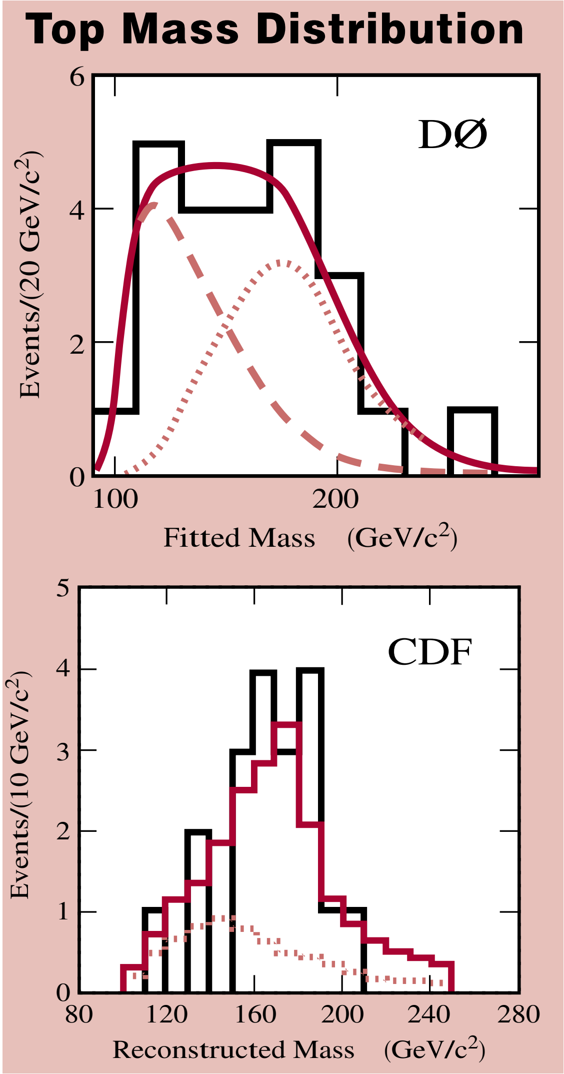




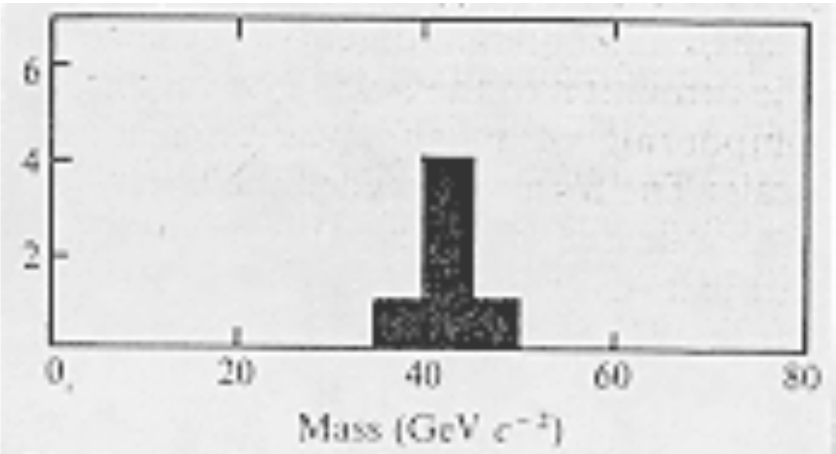
see: http://www.science20.com/a_quantum_diaries_survivor/true_and_false_discoveries_how_to_tell_them_apart-141024

True and False

True top (1996)



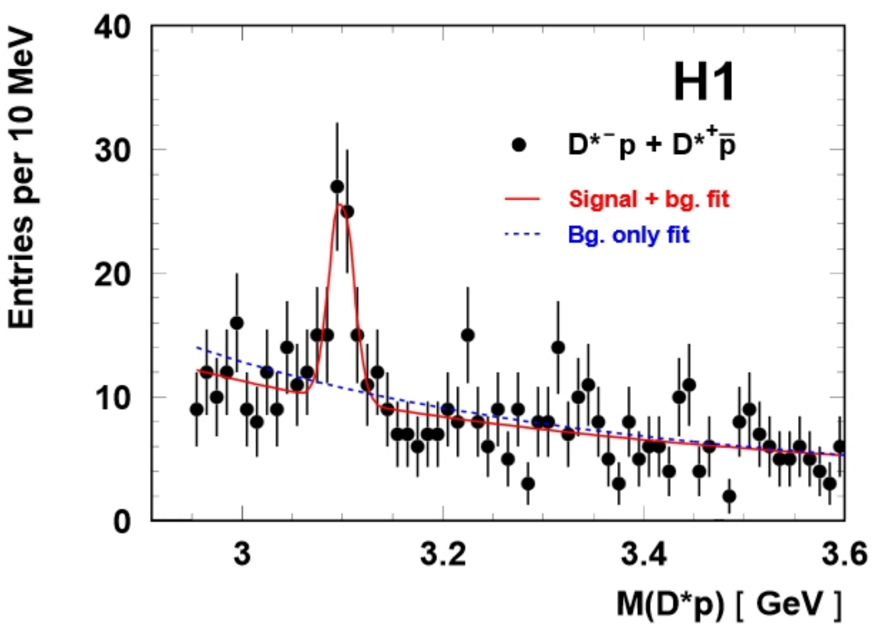
False top (1985)



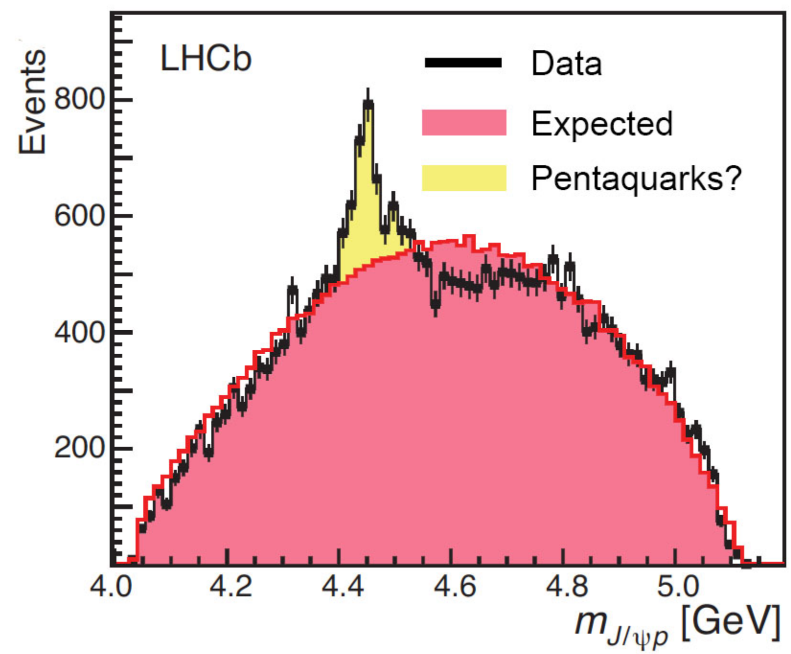
Distribution of measured top quark mass.

True & False: Pentaquark

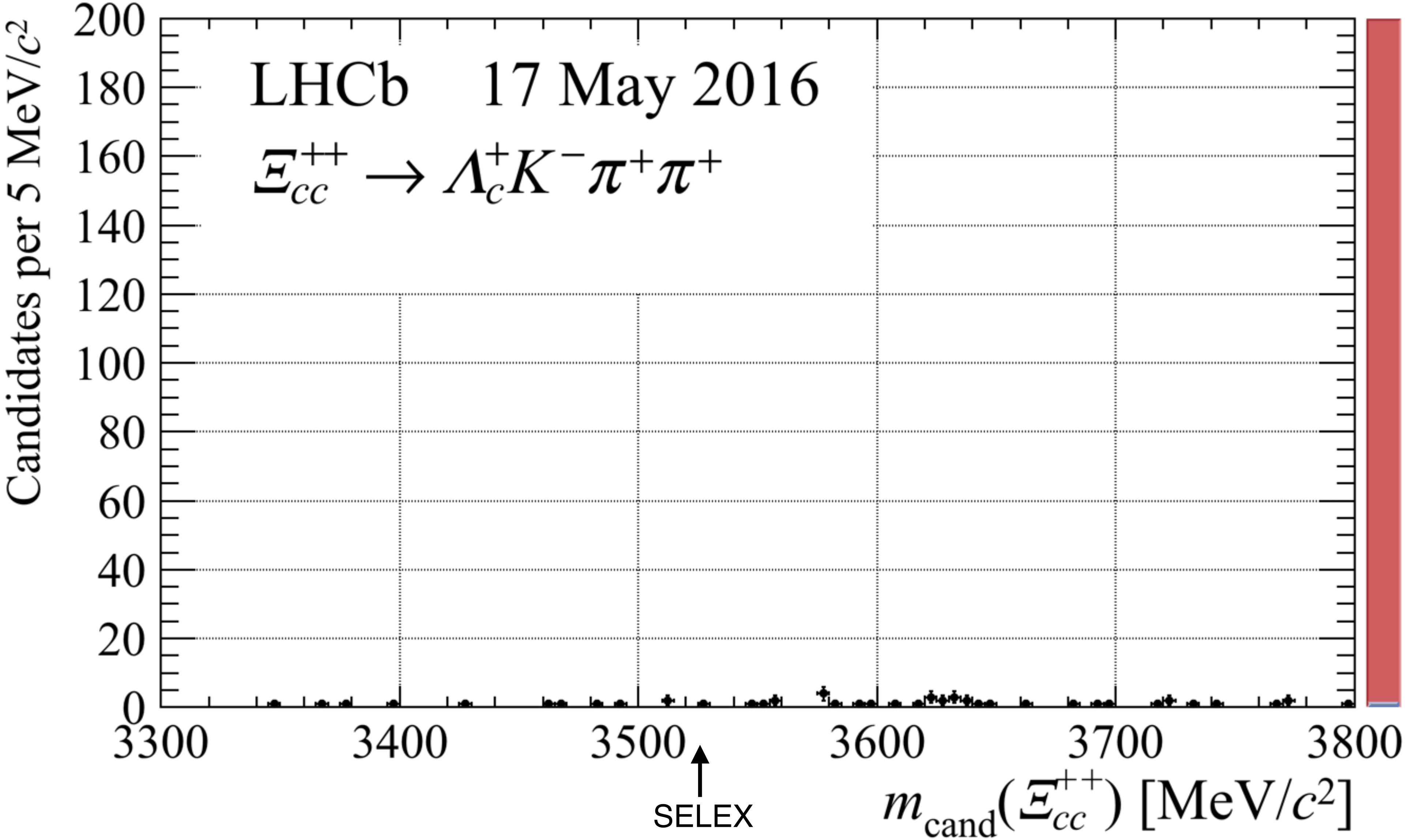
false, 2004, H1 (DESY)



true (LHCb, 2015)

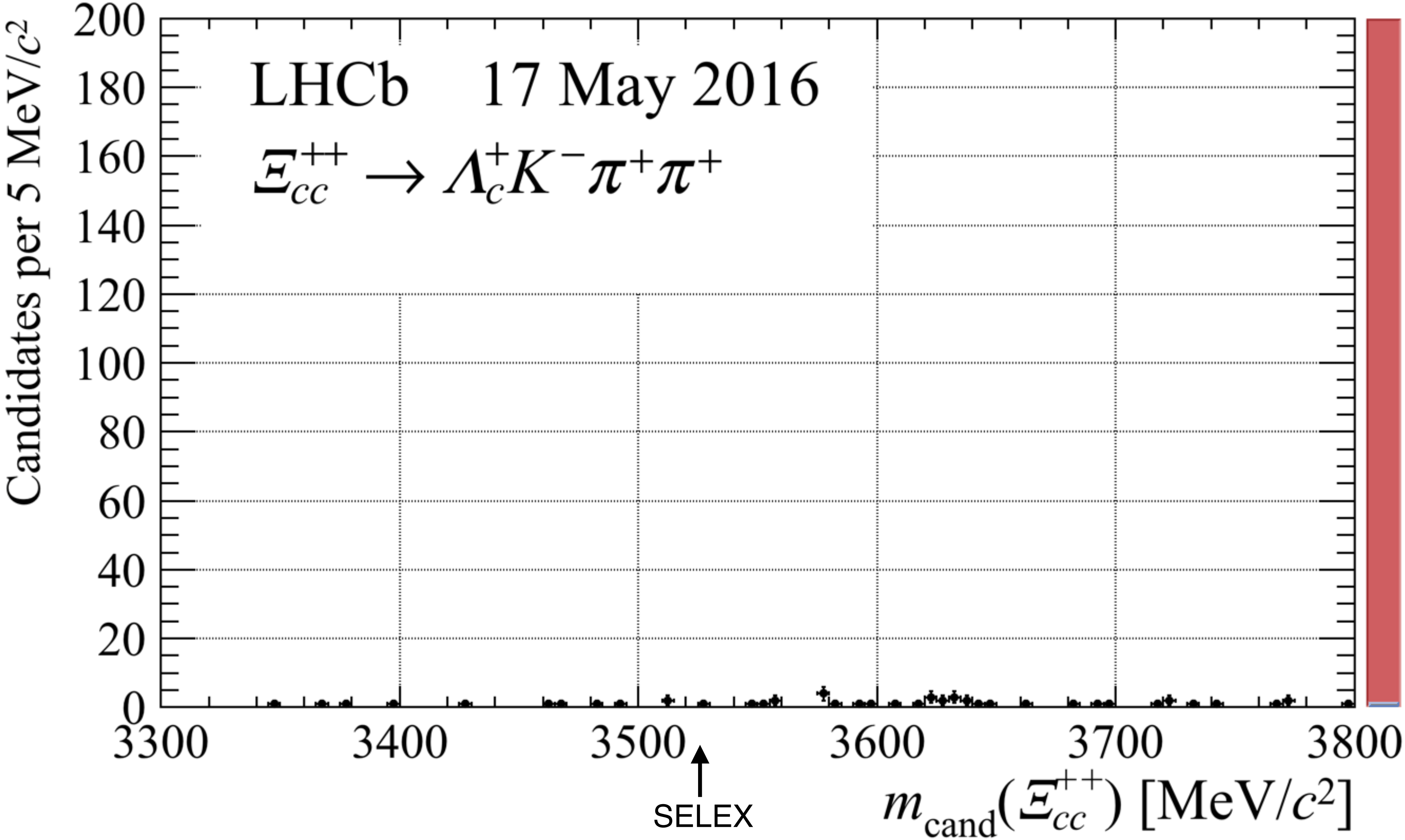


Ξ_{cc} at LHCb?



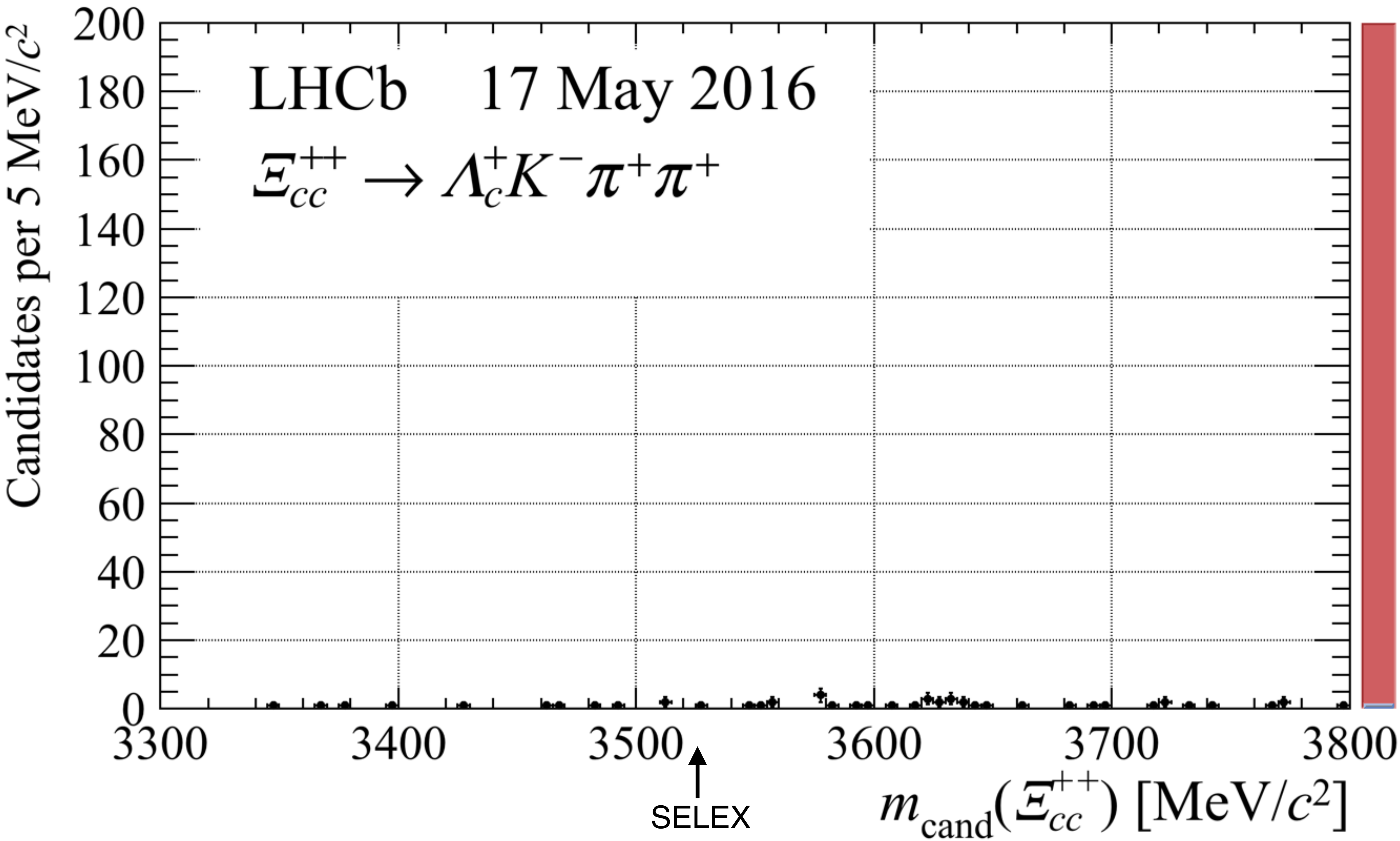
LHCb 17 May 2016
$\Xi_{cc}^{++} \to \Lambda_c^+ K^- \pi^+ \pi^+$
Candidates per 5 MeV/c^2
0
20
40
60
80
100
120
140
160
180
200
3300
3400
3500
3600
3700
3800
$m_{\text{cand}}(\Xi_{cc}^{++})$ [MeV/c^2]
SELEX

Ξ_{cc} at LHCb?



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

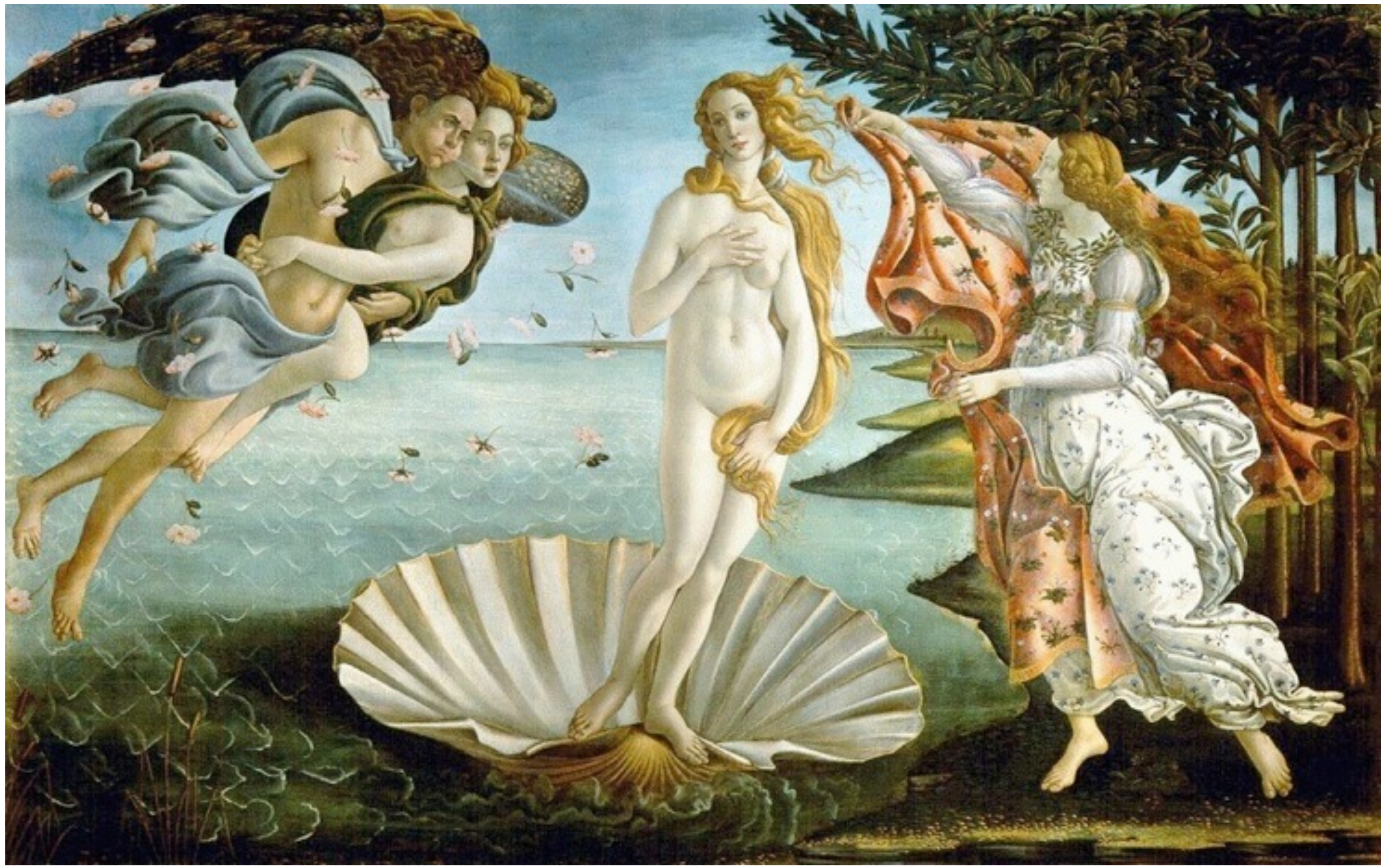
Discoveries…

- 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.

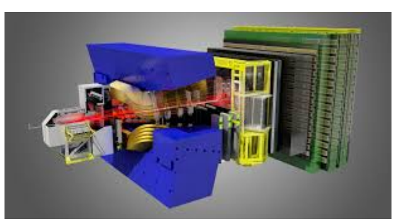
What we want

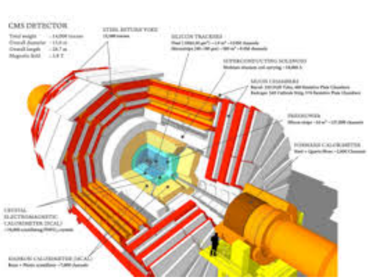
$$\mathscr{L} =$$



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

What we get





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.

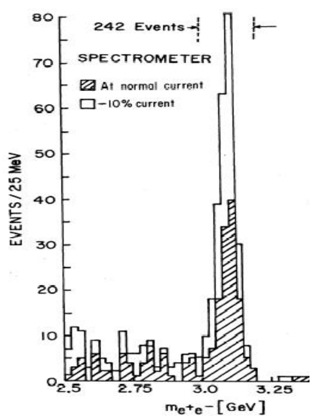
Statistics and Measurements

- 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!

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

Books

- R. J. Barlow: “Statistics”, John Wiley & Sons, ISBN 0-471-92295-1.
- Louis Lyons: “Statistics 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

Problems

Problem sheets:

https://tinyurl.com/TeshepProblems

Code (Jupyter Notebooks):

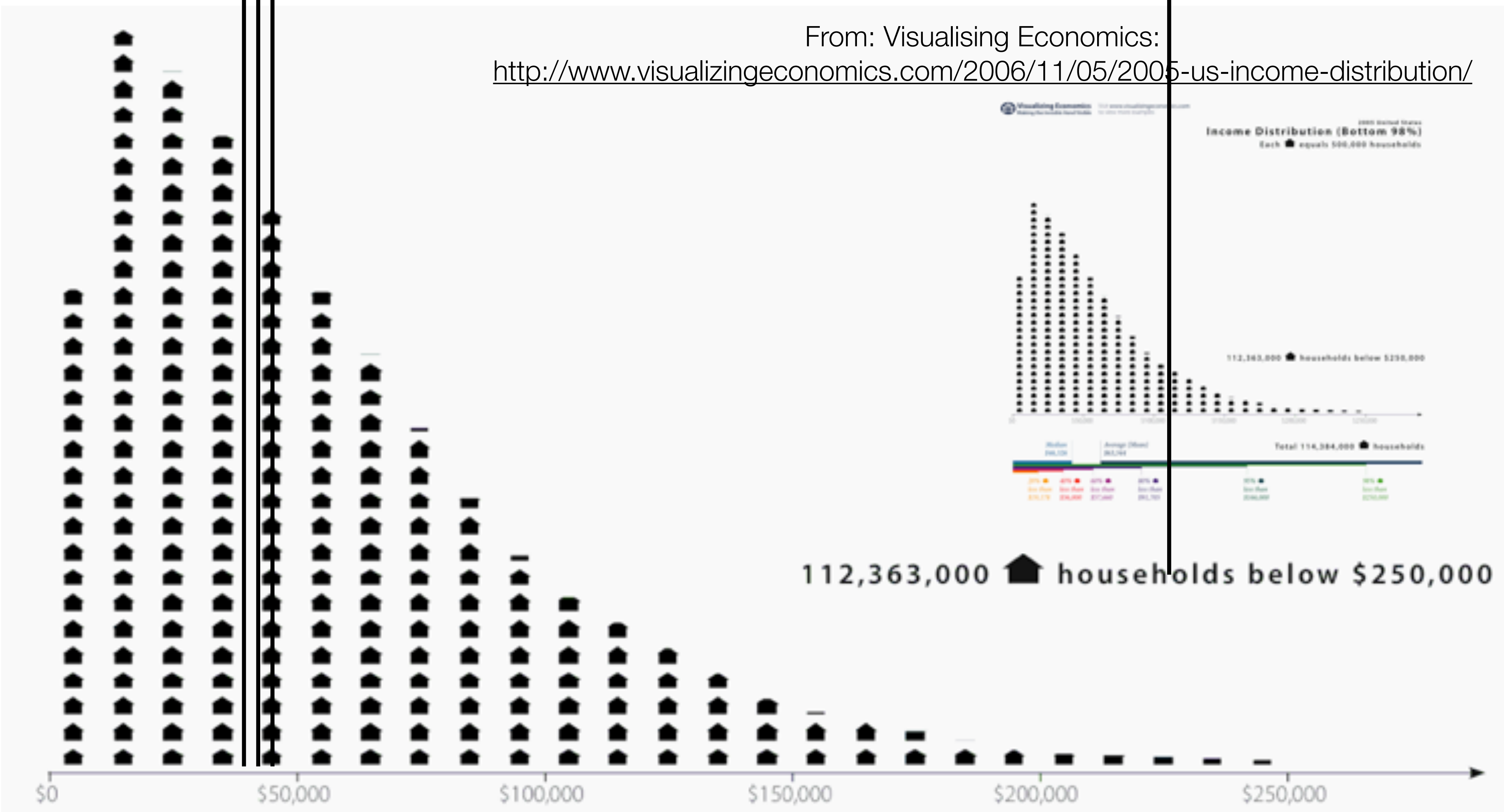
https://tinyurl.com/TeshepStatCode

Describing data with numbers

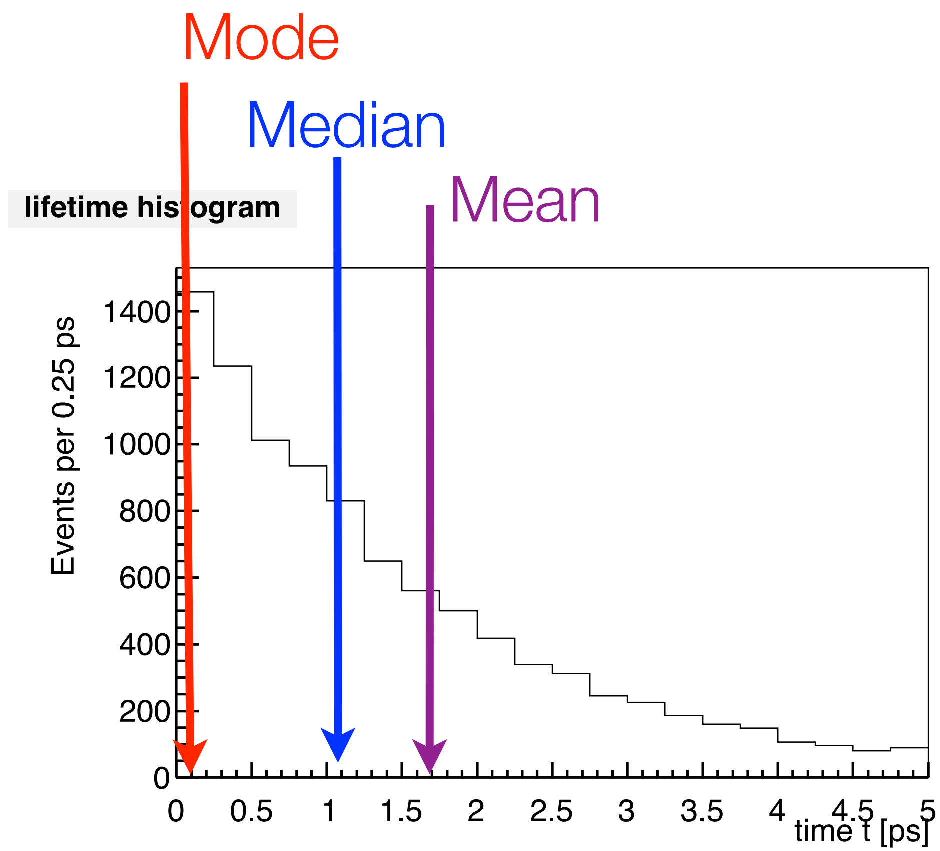
- **How do we describe a set of measurements with just a couple of characteristic, meaningful numbers?**

Annual Income

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



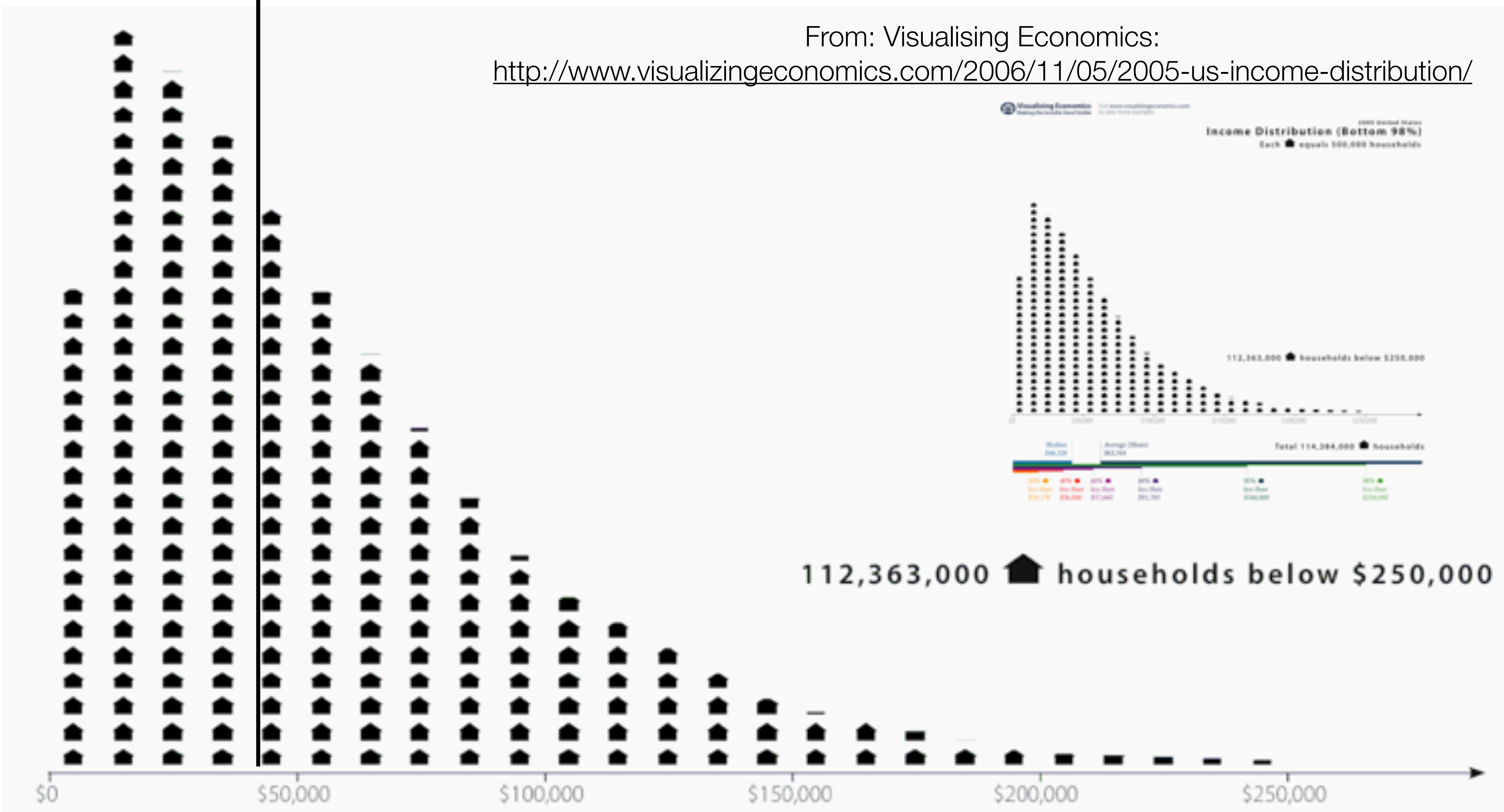
Central Values



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

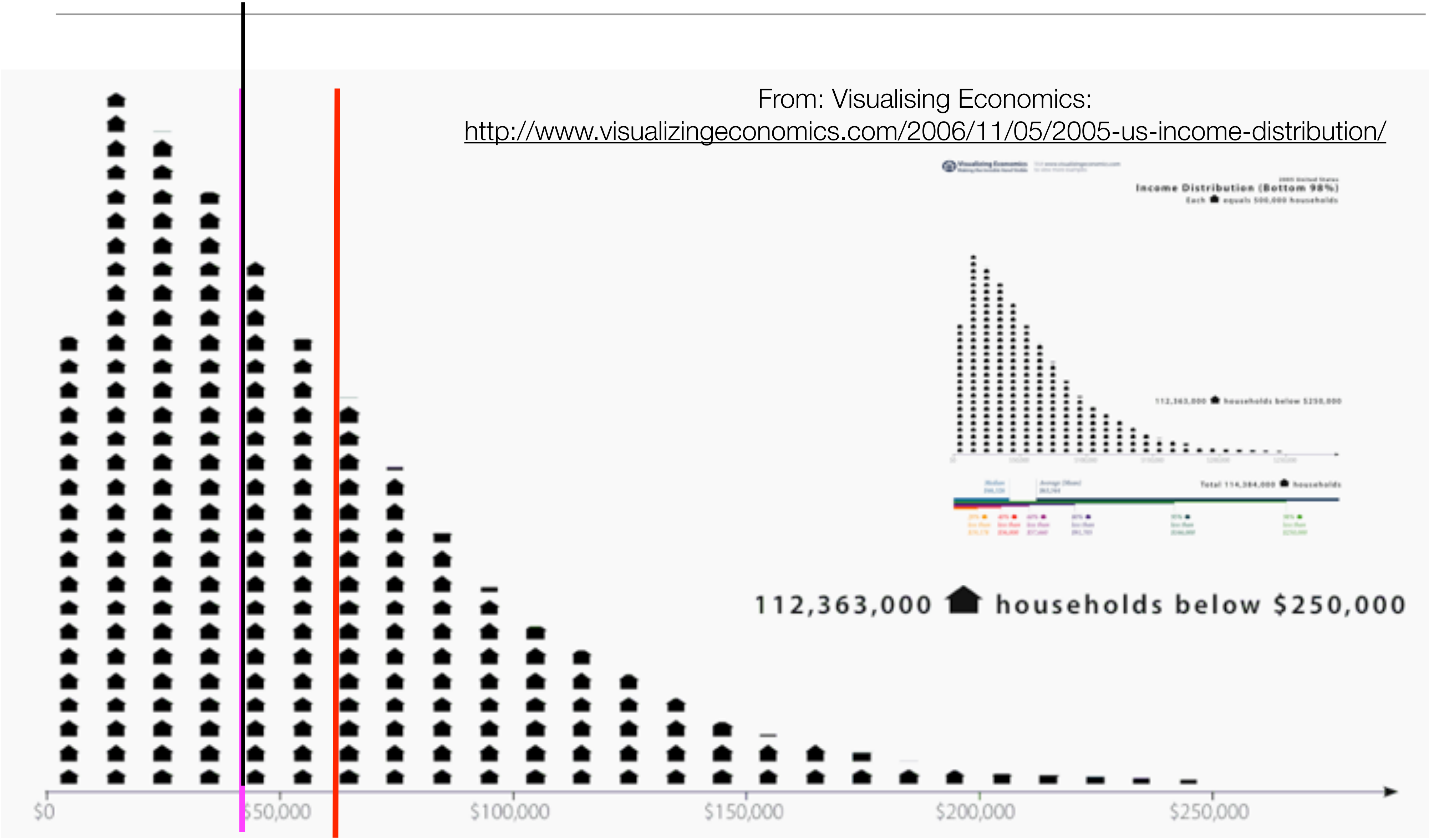
Annual Income

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



Annual Income

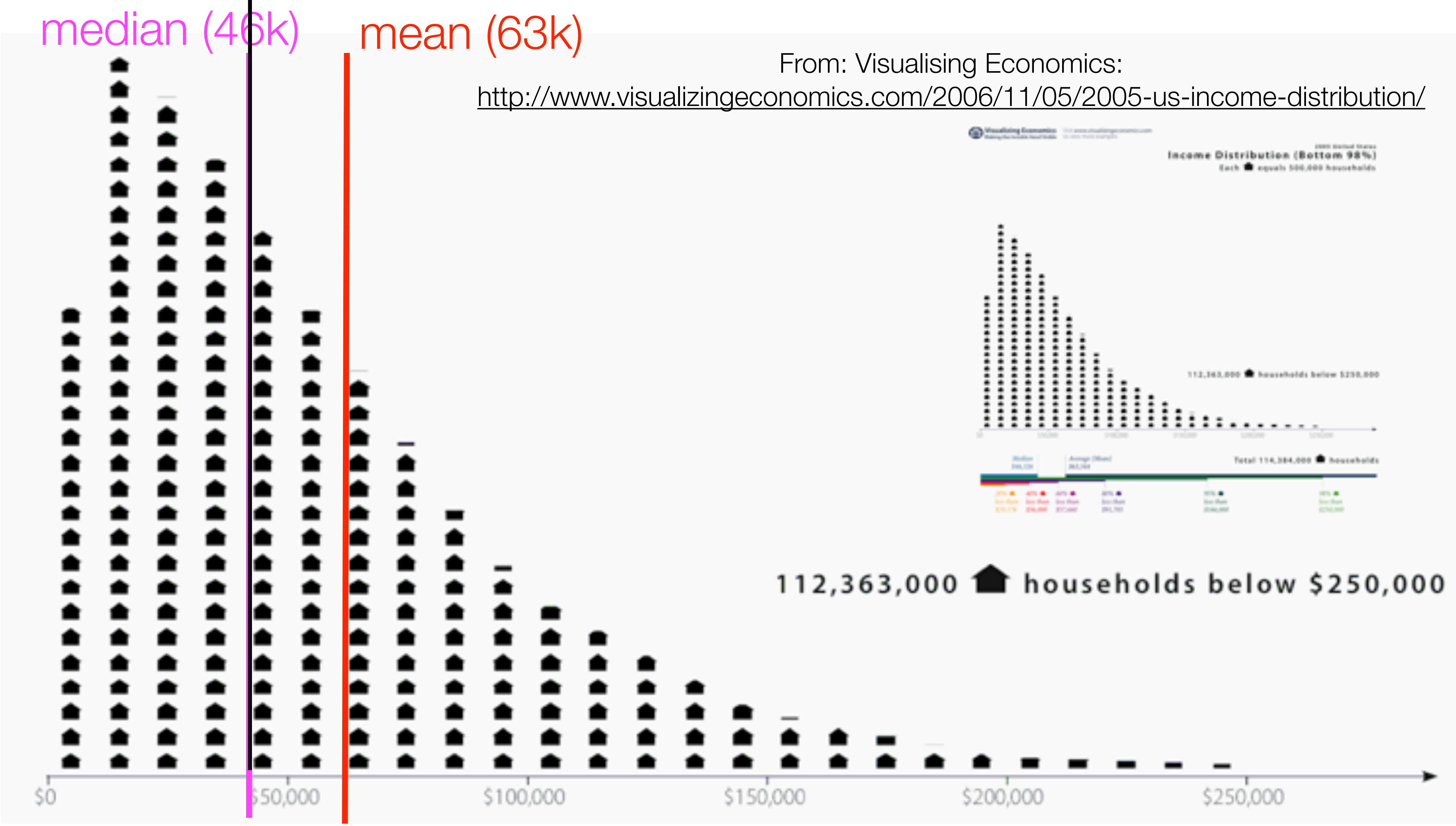
From: Visualising Economics:
http://www.visualizingeconomics.com/2006/11/05/2005-us-income-distribution/



Annual Income

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

Annual Income



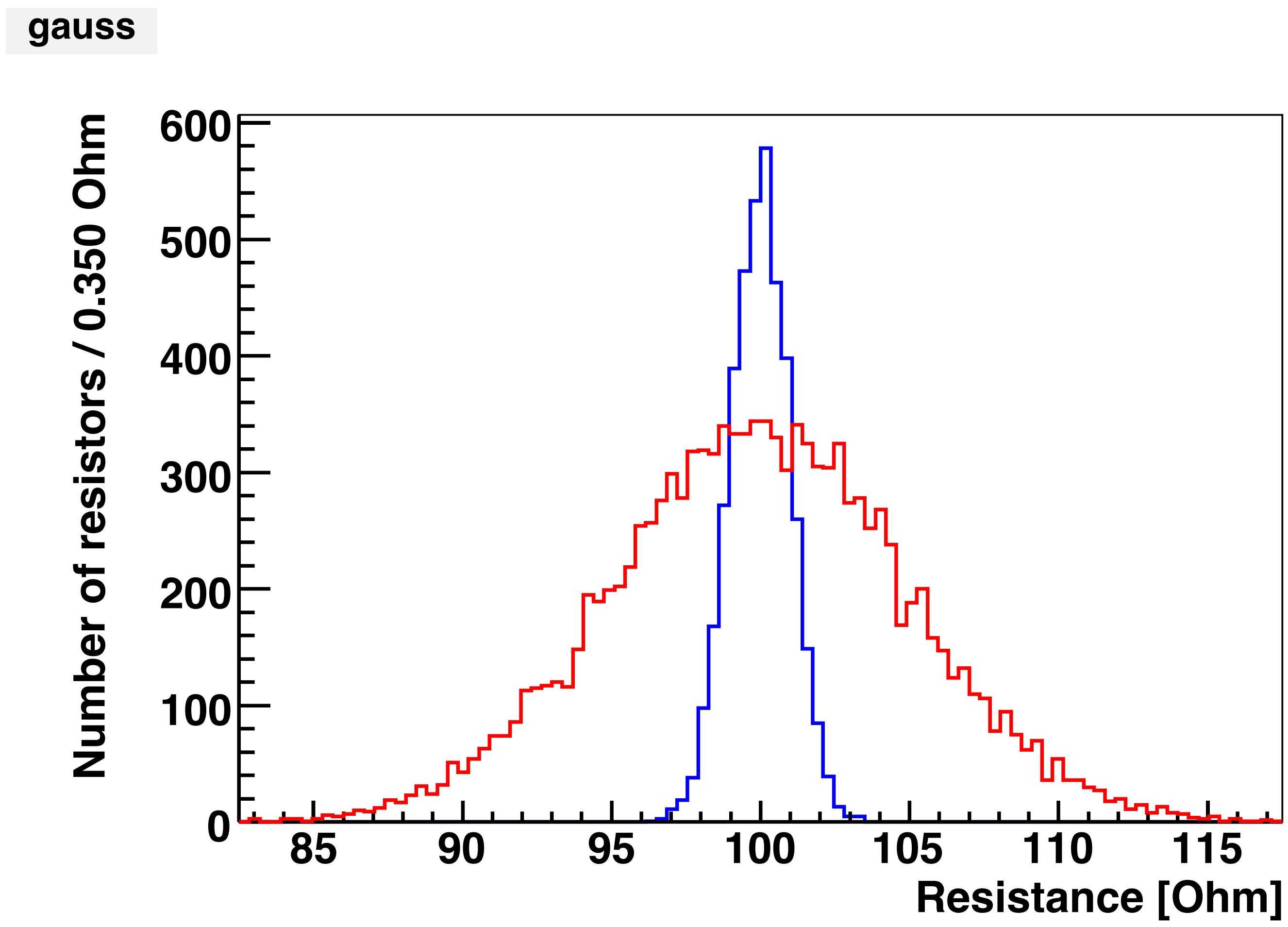
Mean

- 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:**

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

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

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

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:

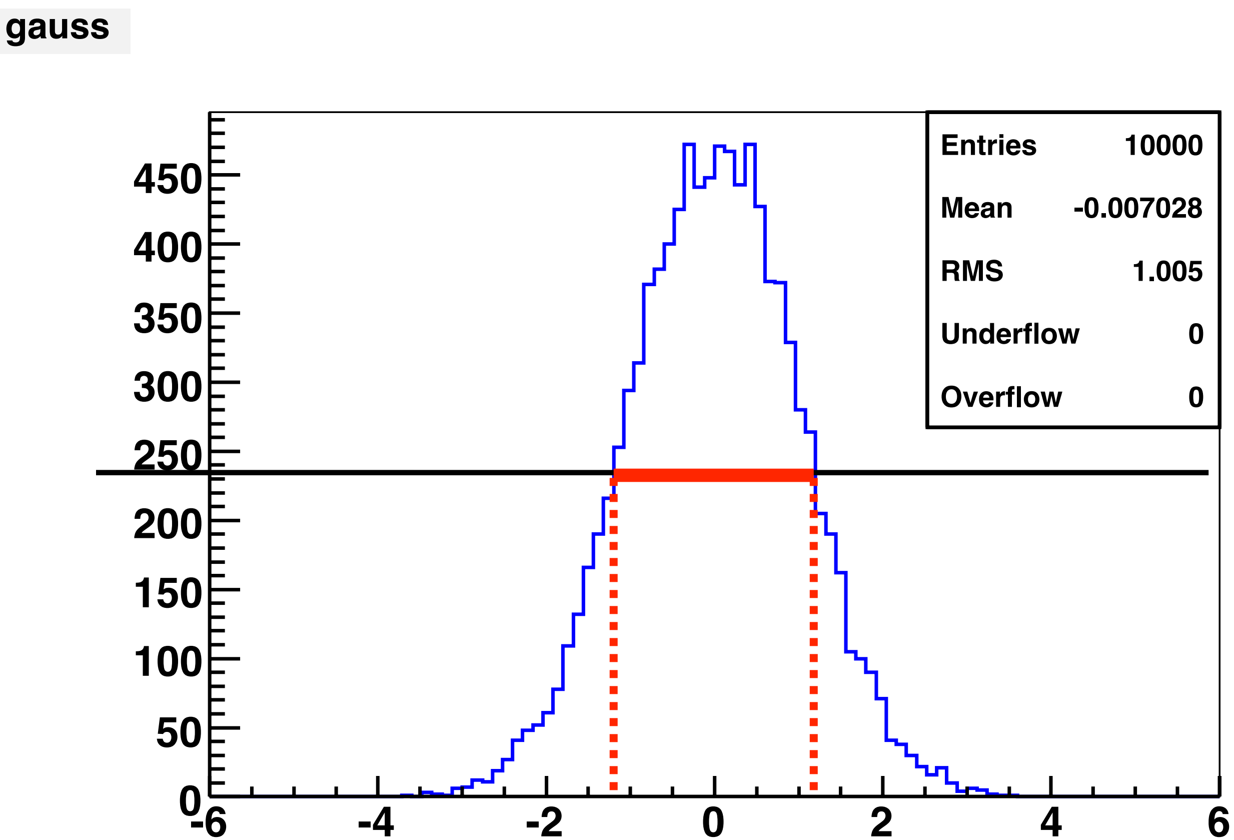
$$\sigma = \sqrt{V}$$

- 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:

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

- 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:

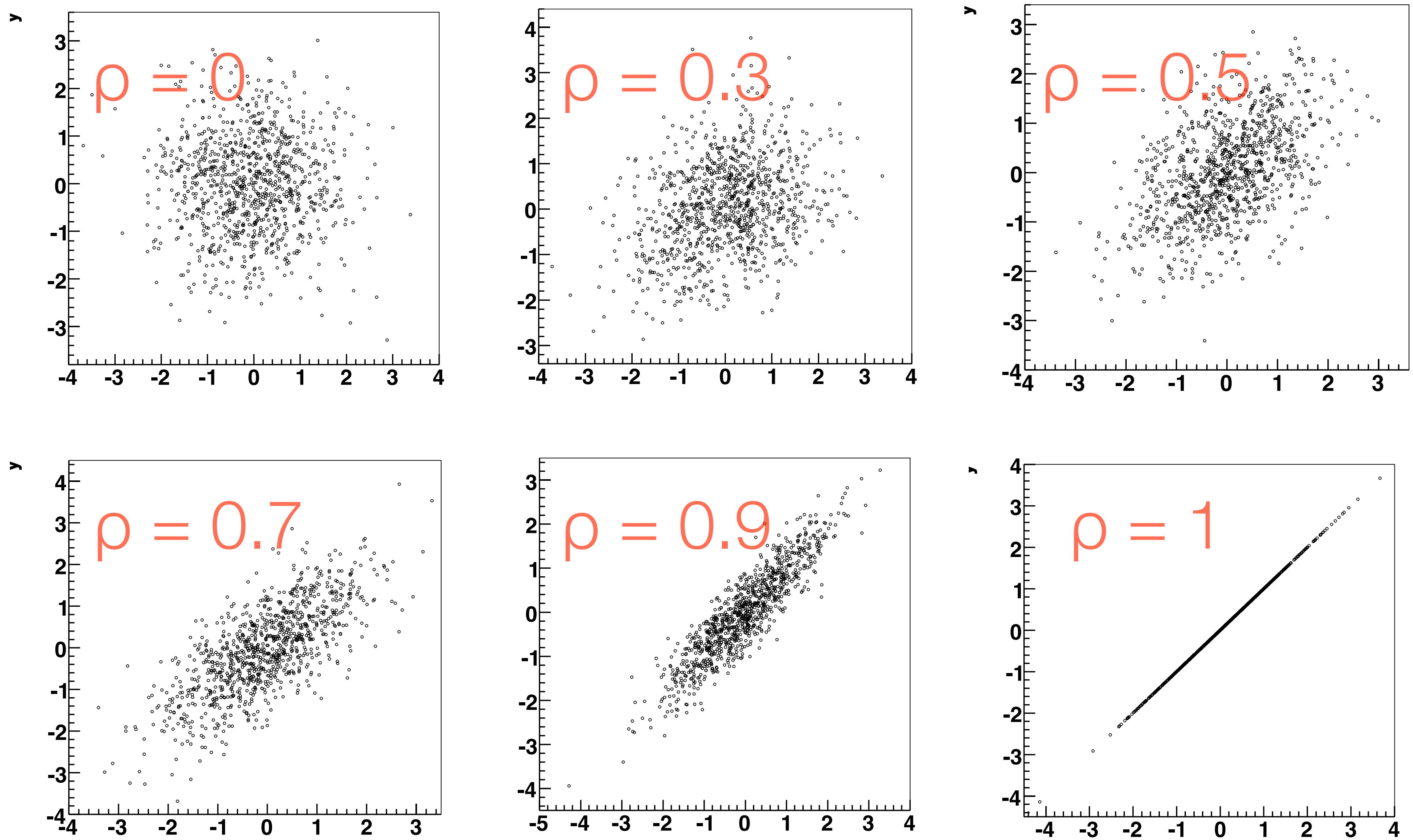
$$\rho_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

- 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:

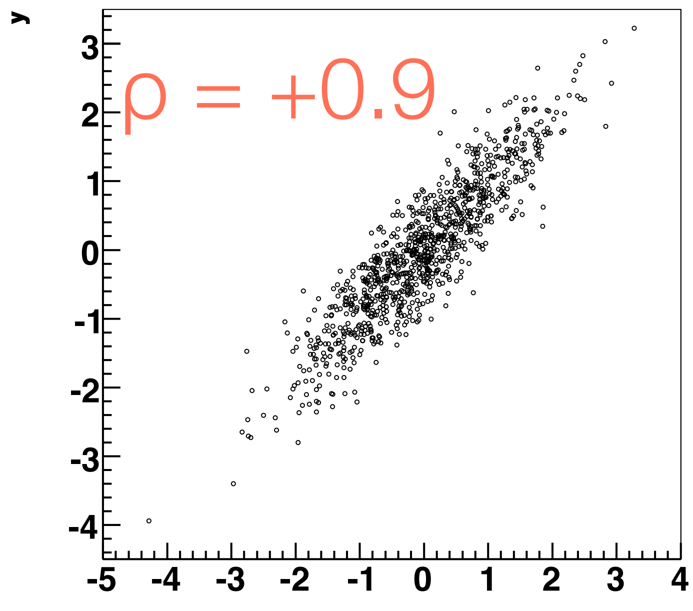
$$\begin{aligned} \rho_{xx} &= \frac{\text{cov}(x, x)}{\sigma_x \cdot \sigma_x} \\ &= \frac{\text{Var}(x)}{\sigma_x^2} = \frac{\sigma_x^2}{\sigma_x^2} = 1 \end{aligned}$$

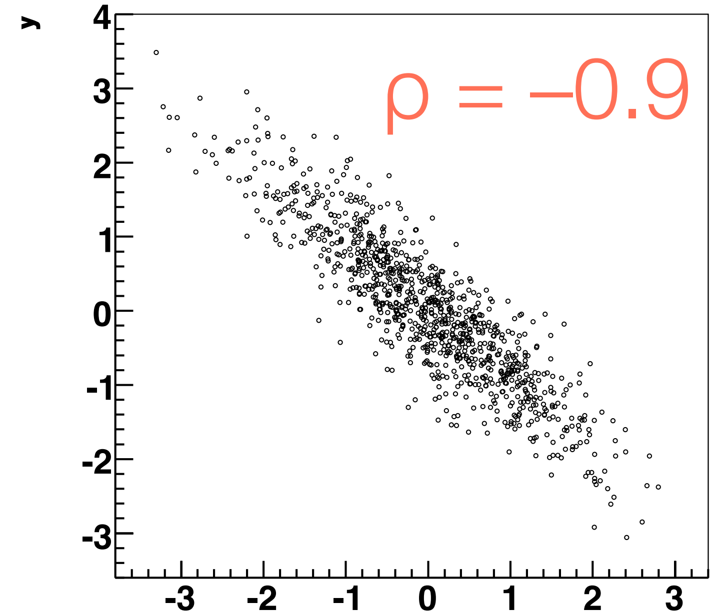
Correlation Coefficient Examples



Correlation Coefficients Examples

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





Make these plots yourself:

https://tinyurl.com/TeshepStatCode

https://github.com/JonasRademacker/JupyterNotebooksForTeachingMath/blob/master/CovarianceAndCorrelation.ipynb

The Covariance/Error Matrix

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

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

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

- **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 $x^{(1)}, \ldots, x^{(N)}$**

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

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

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

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

Correlation and Causality

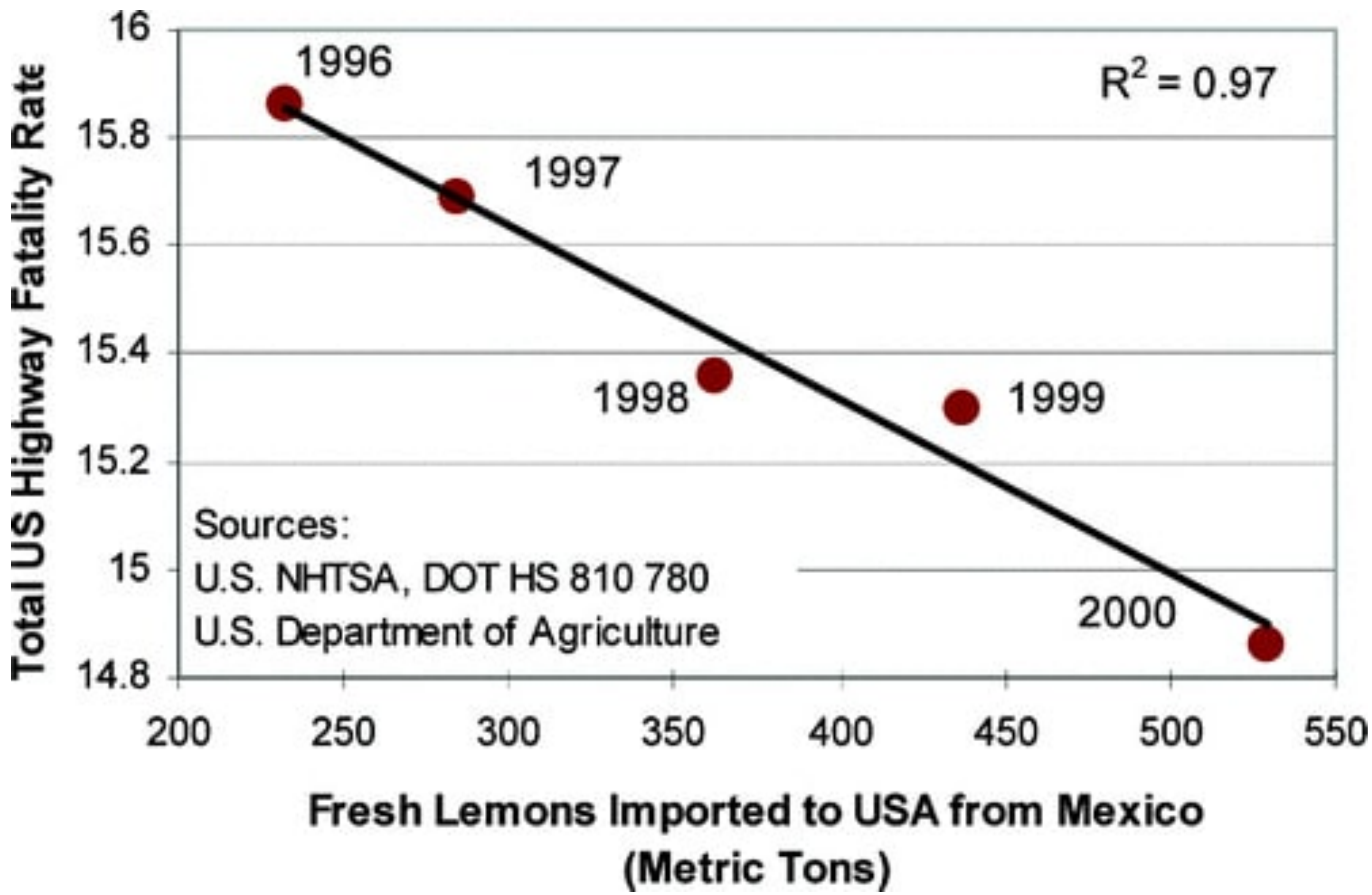
- Among my favourite correlations is this one:

- 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

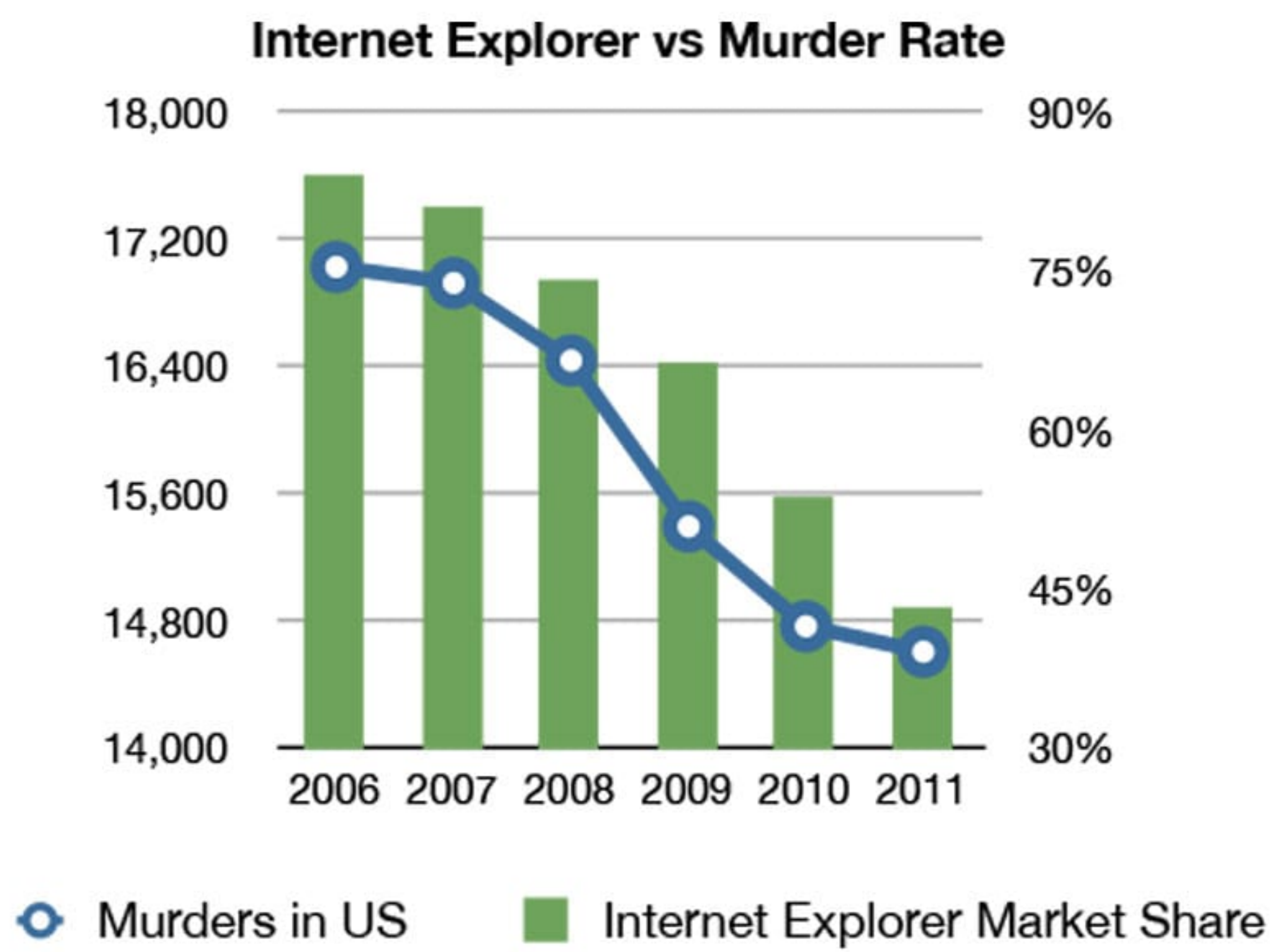
Lemons prevent traffic deaths



http://pubs.acs.org/doi/abs/10.1021/ci700332k

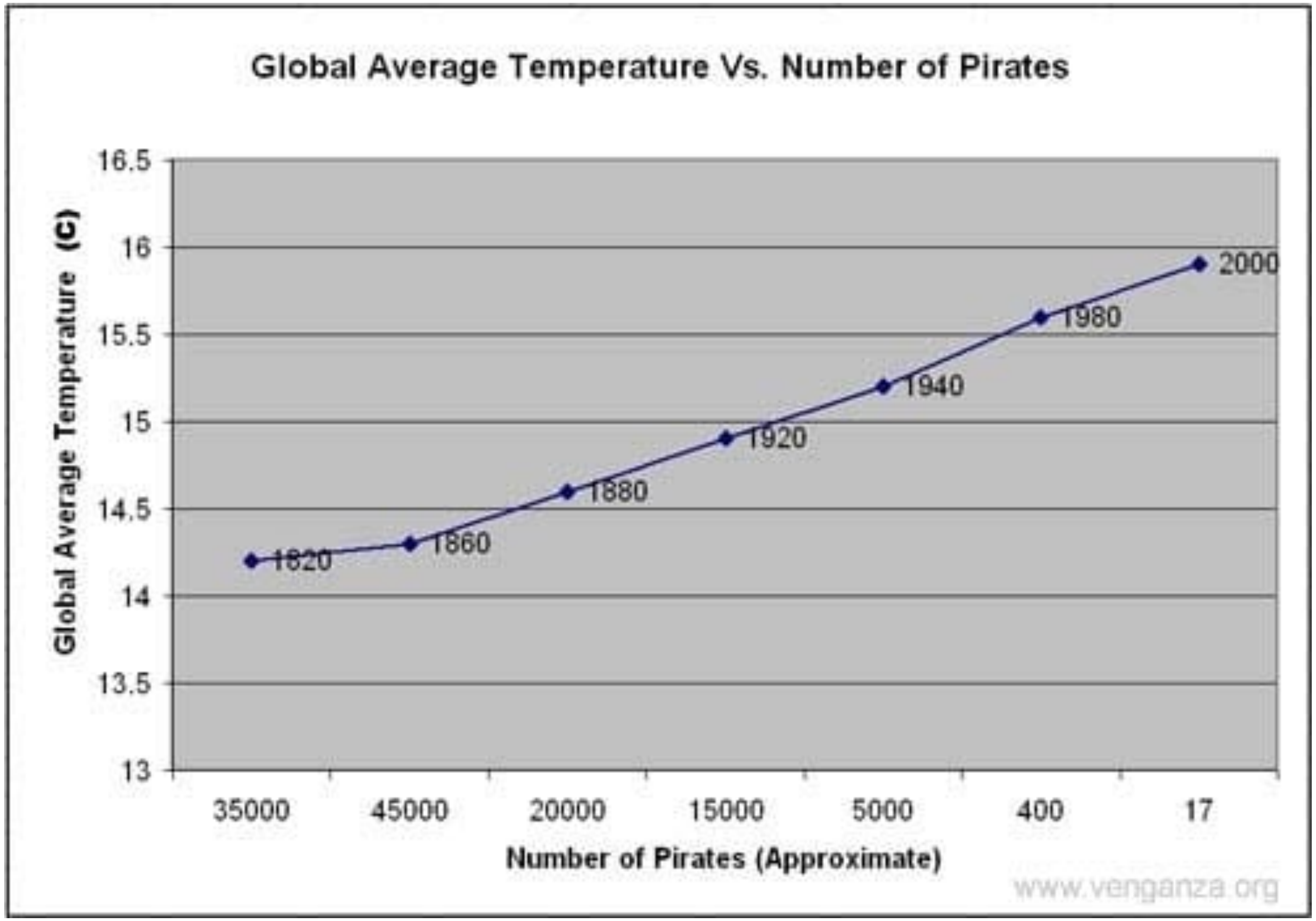
find this and other weird correlations at: https://www.buzzfeednews.com/article/kjh2110/the-10-most-bizarre-correlati

Internet Explorer causes murder



http://gizmodo.com/5977989/internet-explorer-vs-murder-rate-will-be-your-favorite-chart-today

Lack of (Caribbean) pirates causes global warming



http://www.venganza.org/about/open-letter/

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.**

Homework

- **Write down 100 times:**

  “Correlation is not causation”

Summary: Representing Data

- **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 variables x and y are uncorrelated).**

- **Multiparameter distributions: Covariance, Correlation.**

Correlation is not causation.
Correlation is not causation.

Summary: Representing Data

- **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 variables x and y are uncorrelated).**

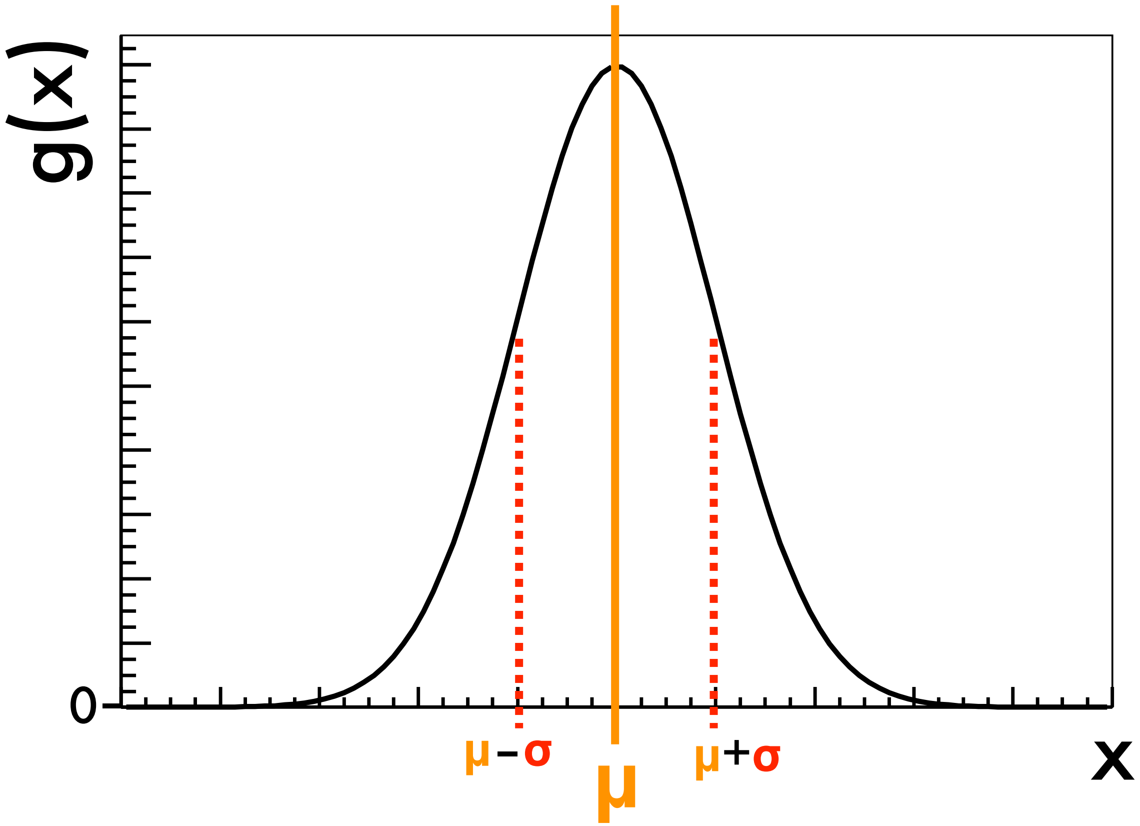
- **Multiparameter distributions: Covariance, Correlation.**

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



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

Gauss & me hanging out in Göttingen

Gauss on old money

The Central Limit Theorem

- Consider random variable $Y = \sum_i x_i$, where each x_i is taken from a distribution with mean $\langle x_i \rangle$ and variance $V_i = \sigma_i^2$
- Then
  - Y has an expectation value $\langle Y \rangle = \sum_i \langle x_i \rangle$
  - Y has a variance $V_Y = \sum_i V_i$. Equivalently: $\sigma_Y^2 = \sum_i \sigma_i^2$
  - The distribution of *Y* becomes Gaussian as *N*→∞.

Variances add up! (Standard deviations don't)

Roll some Dice, submit results, here

https://tinyurl.com/DiceTESHEP

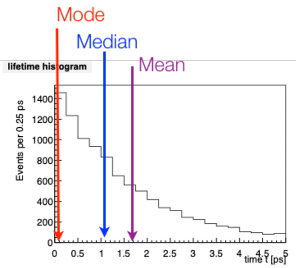
Largest number of entries wins!

Rolling Dice, *predict* results, here

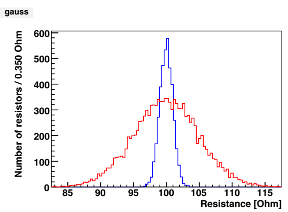
https://tinyurl.com/PredictDiceTESHEP

First (few) correct answers win

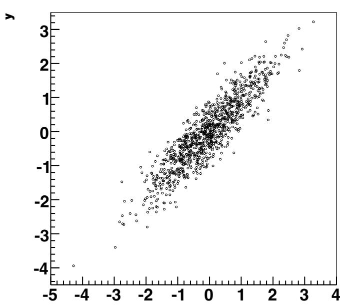
Summary



- Averages: Mean, Median, Mode - usually we chose arithmetic mean, but there are use cases for alternatives.

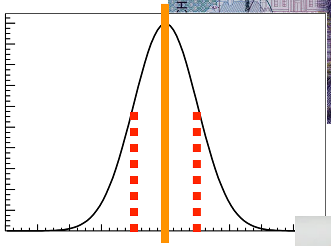


- Width: Standard deviation, Variance, FWHM

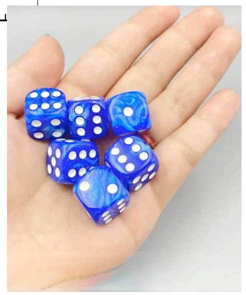


- Covariance, correlation (is not causation, but still informative)

- CLT, transforms ignorance to well-defined uncertainty.

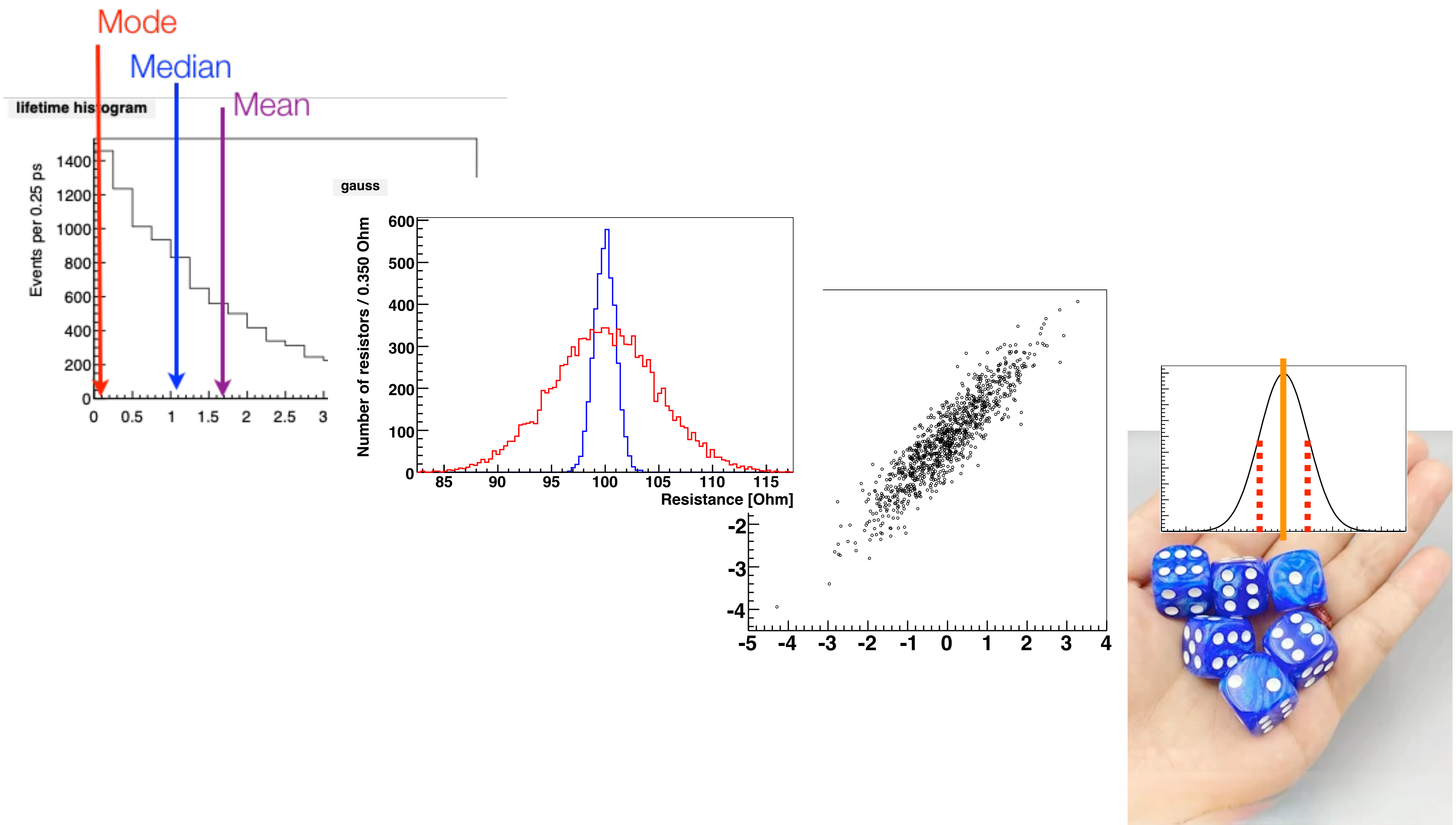


- Do your bit for the CLT and win a prize!
  - Roll dice: https://tinyurl.com/DiceTESHEP
  - Predict results: https://tinyurl.com/PredictDiceTESHEP

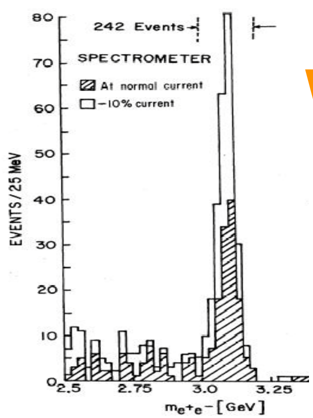


Lecture 2

Recap


Mode
Median
Mean
lifetime histogram
Events per 0.25 ps
0 200 400 600 800 1000 1200 1400
0 0.5 1 1.5 2 2.5 3
gauss
Number of resistors / 0.350 Ohm
0 100 200 300 400 500 600
85 90 95 100 105 110 115
Resistance [Ohm]
-2 -3 -4
-5 -4 -3 -2 -1 0 1 2 3 4

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

- Analyse yesterday's data, and discuss their implications
- Fitting
- Monte Carlo

The Central Limit Theorem

- Consider random variable $Y = \sum_i x_i$, where each x_i is taken from a distribution with mean $\langle x_i \rangle$ and variance $V_i = \sigma_i^2$, and all x_i are INDEPENDENT

- Then

  - Y has an expectation value $\langle Y \rangle = \sum_i \langle x_i \rangle$

  - Y has a variance $V_Y = \sum_i V_i$. Equivalently: $\sigma_Y^2 = \sum_i \sigma_i^2$

  - The distribution of *Y* becomes Gaussian as *N*→∞.

Variances add up! (Standard deviations don't)

Rolling Dice

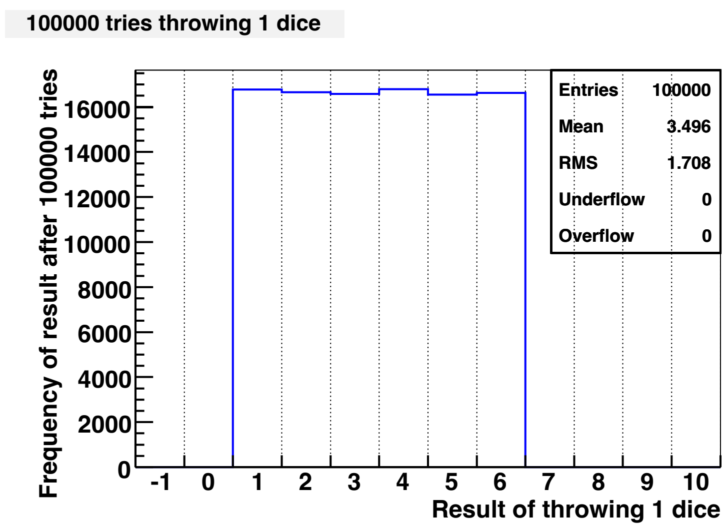
Your data: https://tinyurl.com/TESHEP24DiceResults

Code to analyse data: https://tinyurl.com/RealDiceTESHEP

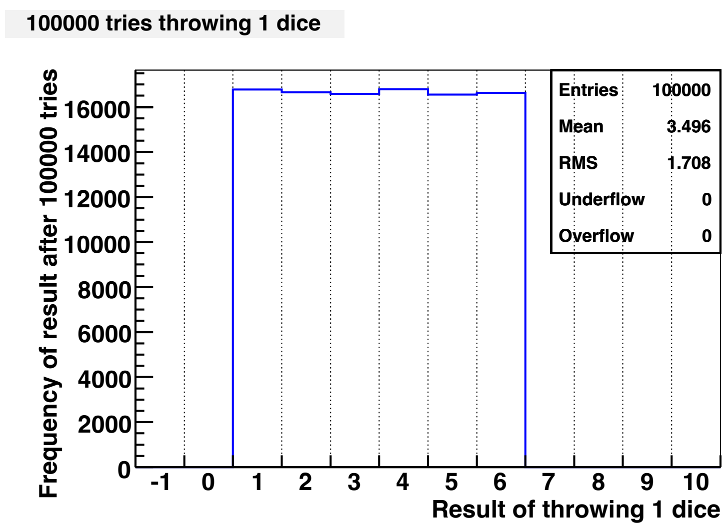
Code to generate more data: https://tinyurl.com/SimDiceTESHEP

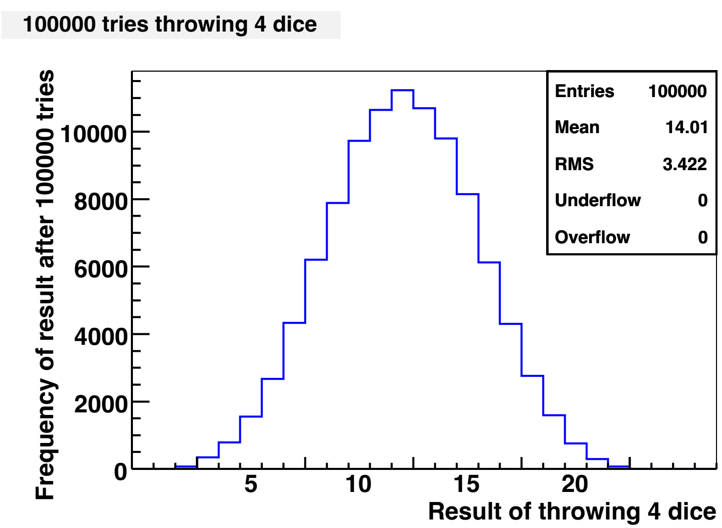
Rolling more and more dice

Rolling more and more dice

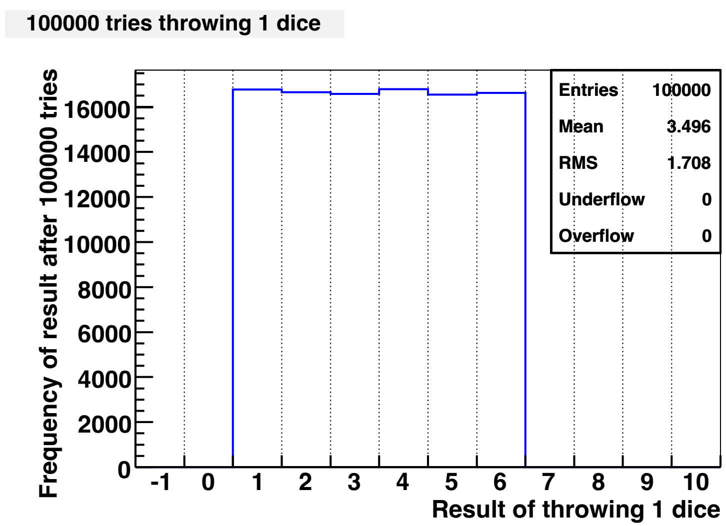


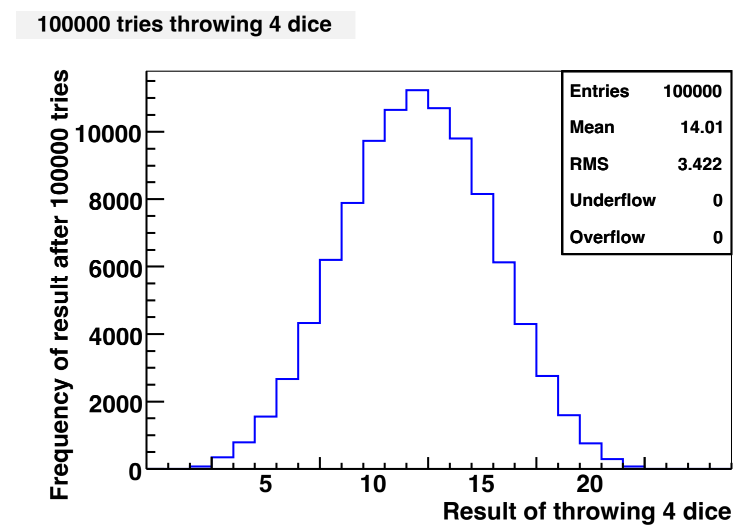
Rolling more and more dice

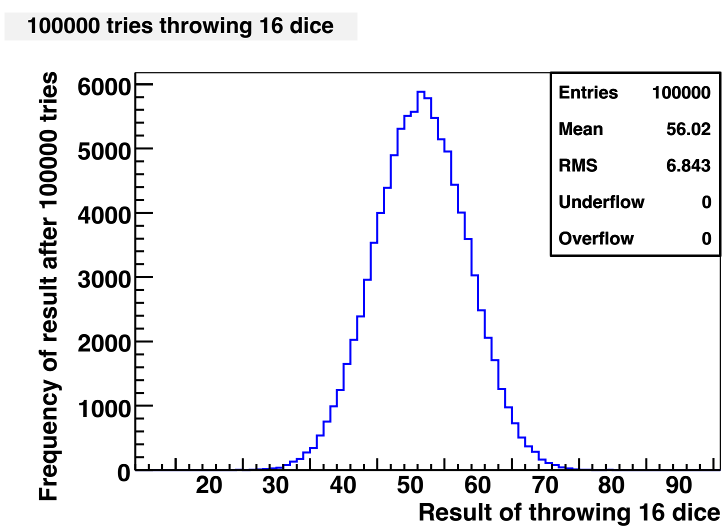




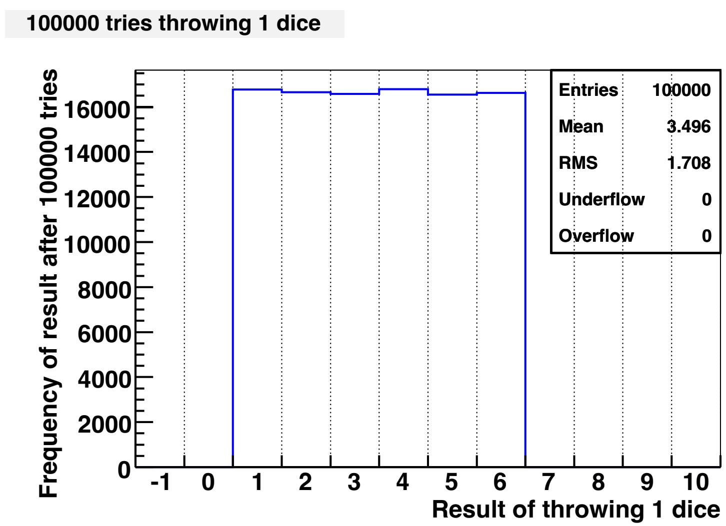
Rolling more and more dice

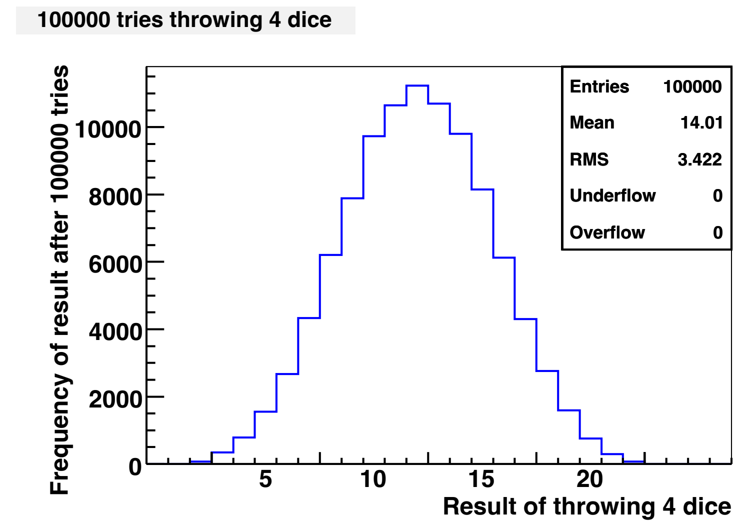


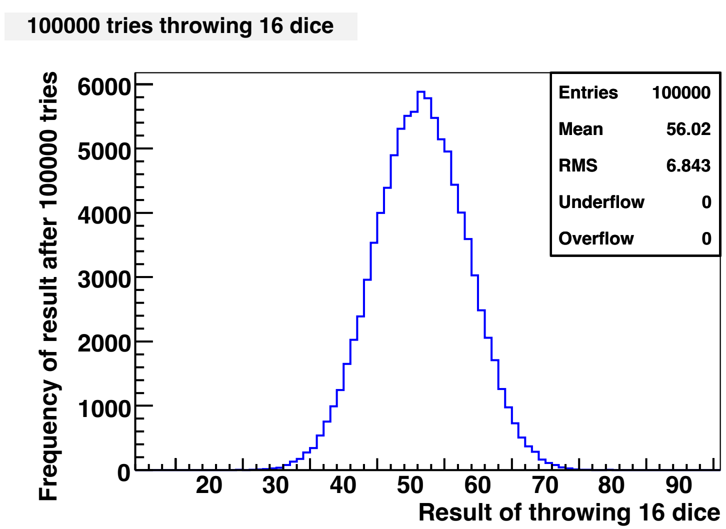


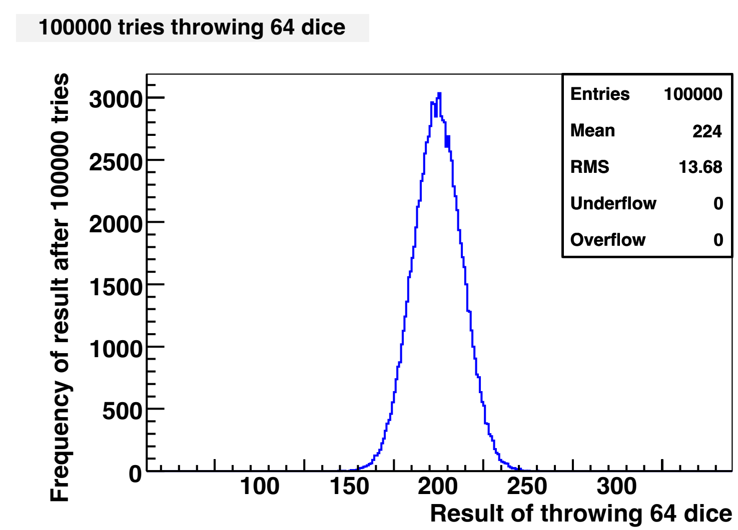


Rolling more and more dice

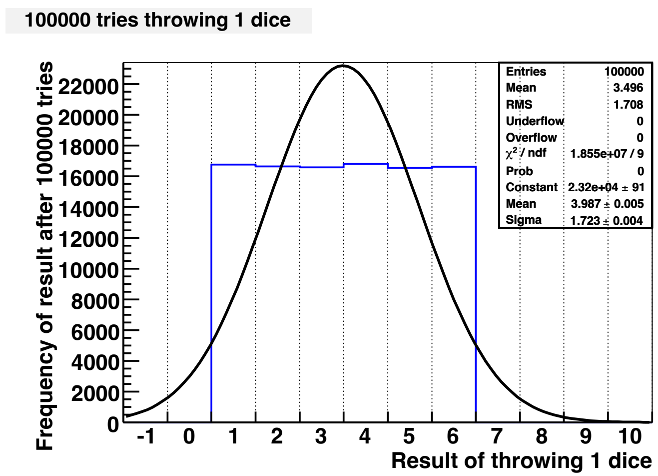


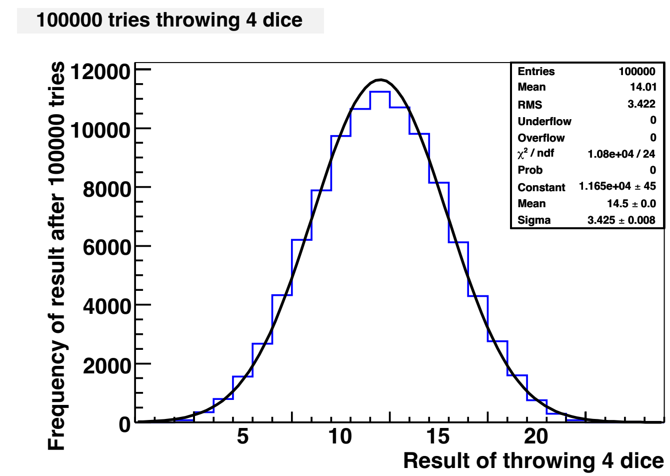


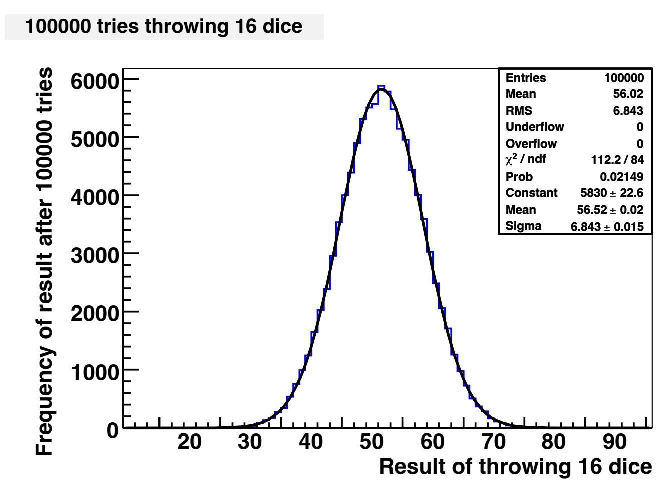


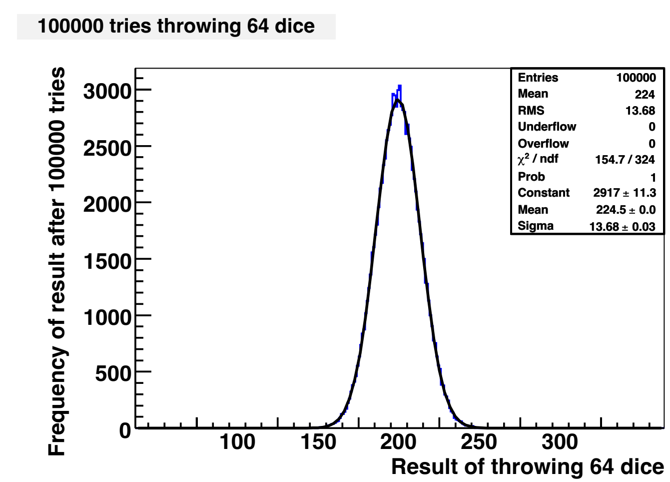


Comparing Gaussians to 1, 4, 16, 64-dice distributions

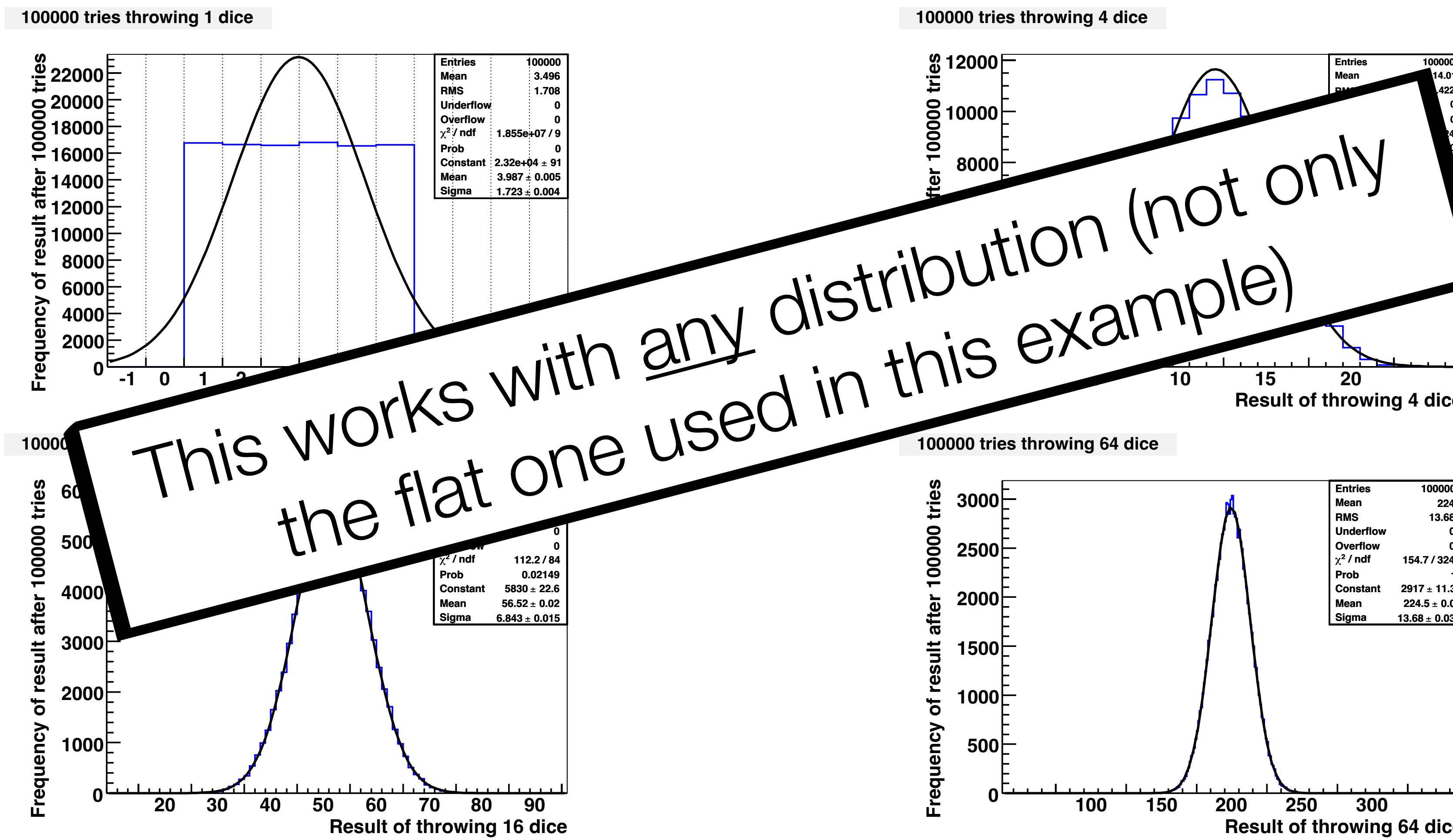








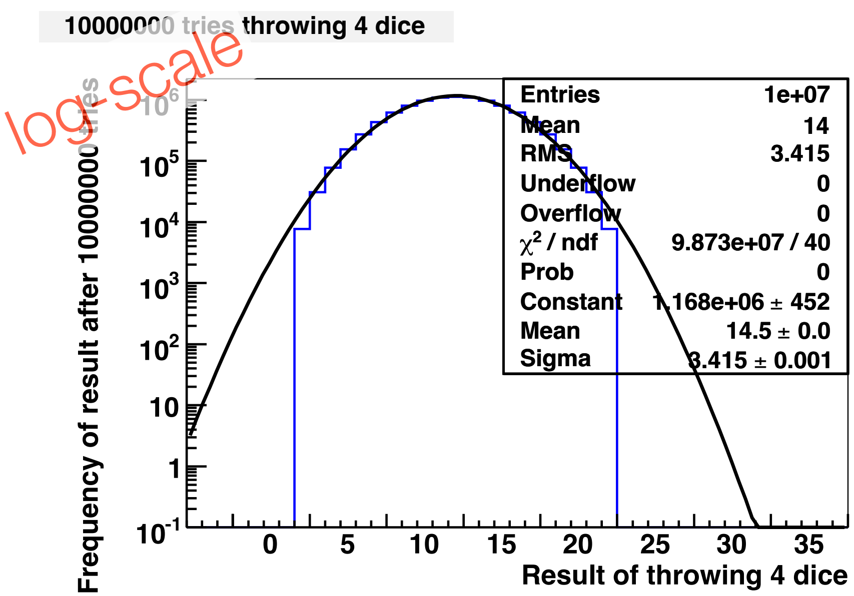
Comparing Gaussians to 1, 4, 16, 64-dice distributions

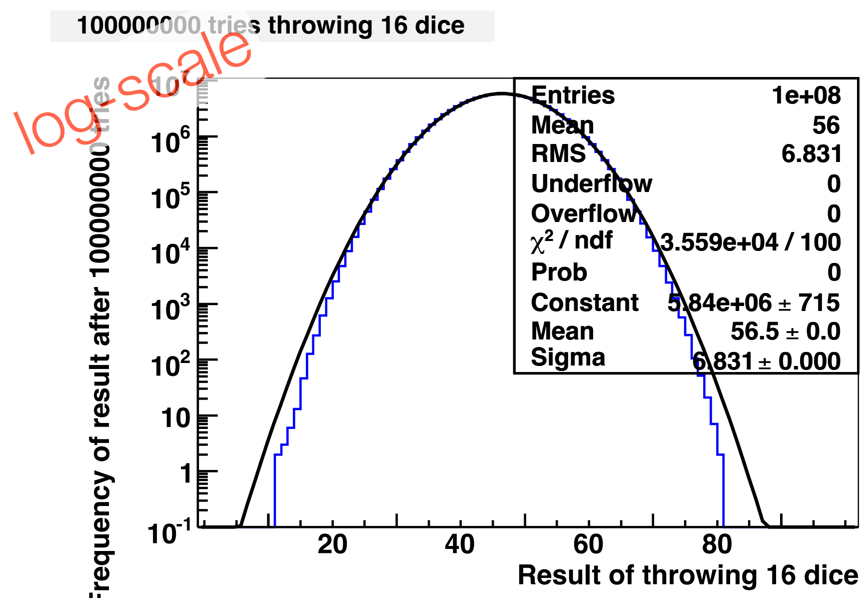


This works with <u>any</u> distribution (not only the flat one used in this example) bokeh serve jonas_singletoy.py

localhost:5006/jonas_singletoy

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"**

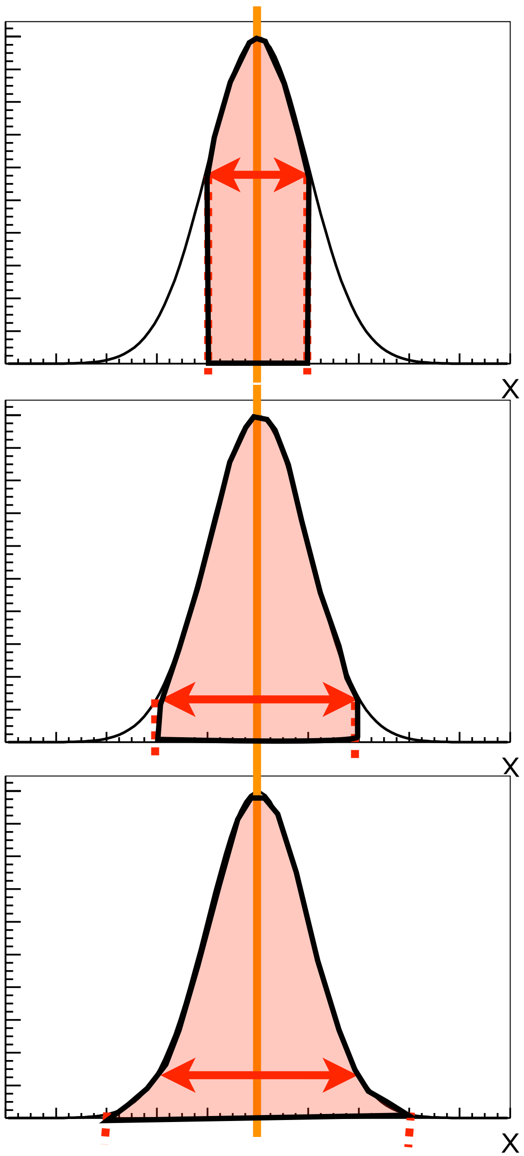
$$\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 68.27\%$$

- **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)

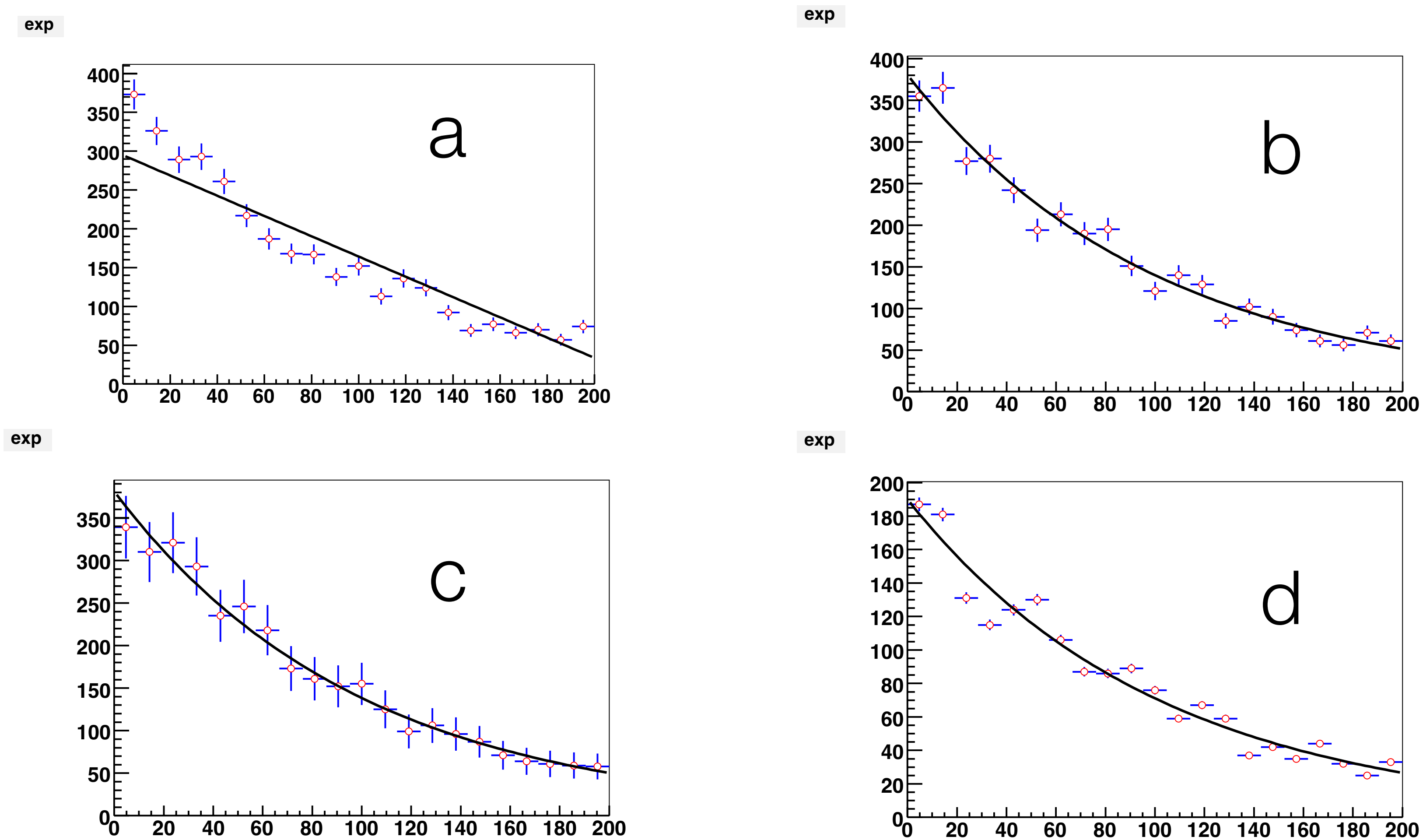
"What we've got here is...failure to communicate.

Some men you just can't reach."

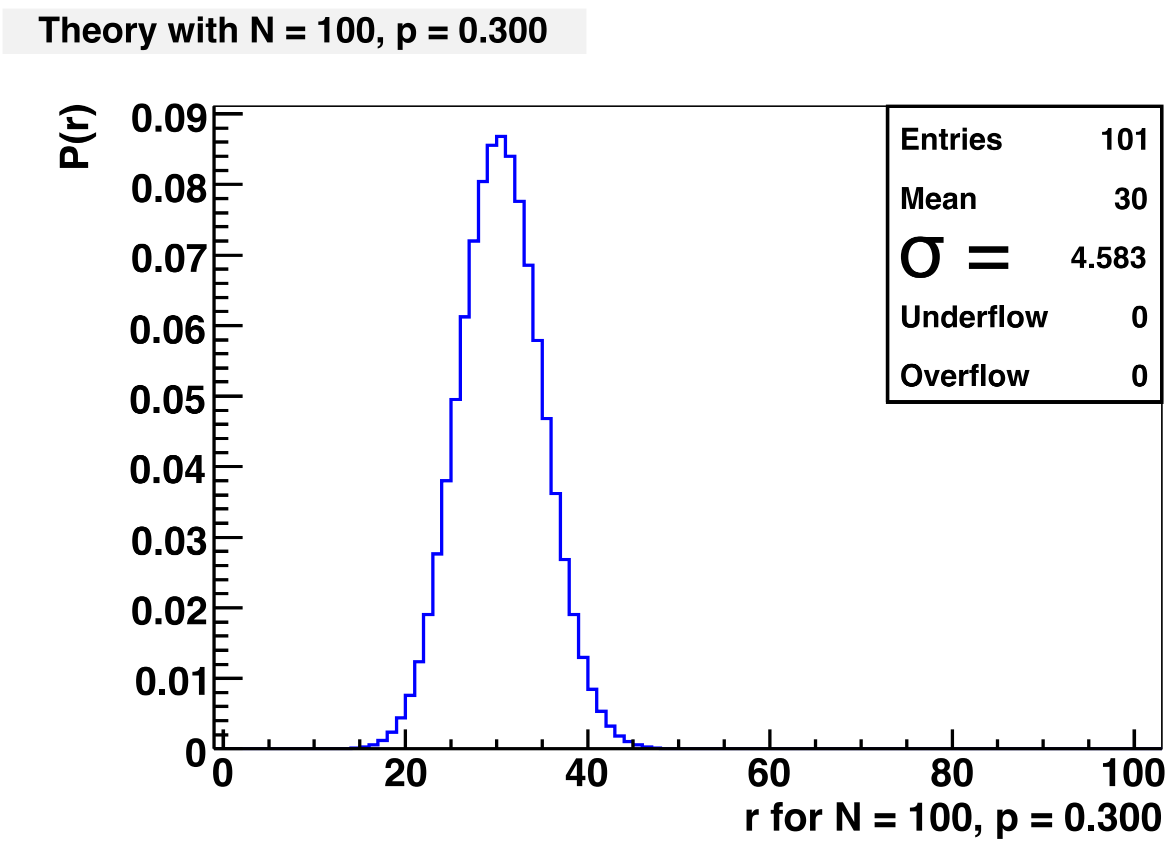
https://tinyurl.com/TeshepProblems

Which plot makes most sense?

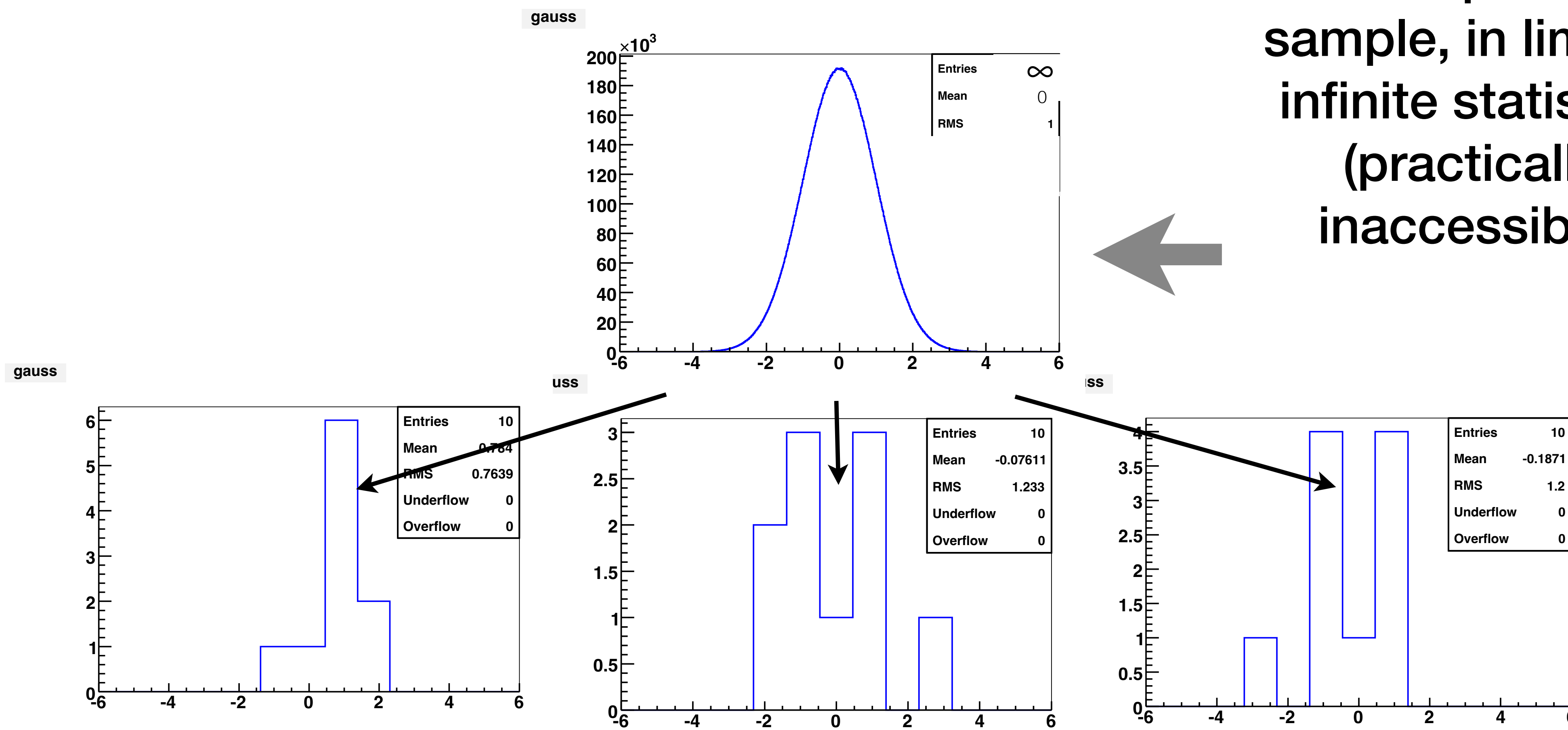
What is the most plausible plot if the line represents theory, dots data distributed according to that theory, and the vertical lines are 1σ error bars.



What's the uncertainty on the mean?



Uncertainty on the mean???



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 Y of N independent variables x_i $Y_{sum} \equiv \sum_{i=1}^{N} x_i$.

- $\langle Y_{sum} \rangle = \sum \langle x_i \rangle$

- Std dev. $\sigma_{Y_{sum}} = \sqrt{\sum \sigma_i^2}$

- Gaussian as $N \to \infty$.

Central Limit theorem

- Take the **sum** *Y* of *N* independent variables x_i $Y_{sum} \equiv \sum_{i=1}^{N} x_i$.
- $\langle Y_{sum} \rangle = \sum \langle x_i \rangle$
- **Std dev.** $\sigma_{Y_{sum}} = \sqrt{\sum \sigma_i^2}$
- **Gaussian as $N \rightarrow \infty$.**

- Take the average Y of N independent variables x_i: $Y_{av} \equiv \frac{1}{N} \sum_{i=1}^{N} x_i$.
- $\langle Y_{av} \rangle = \frac{1}{N} \sum \langle x_i \rangle$
- Std dev.: $\sigma_{Y_{av}} = \frac{1}{N} \sqrt{\sum \sigma_i^2}$

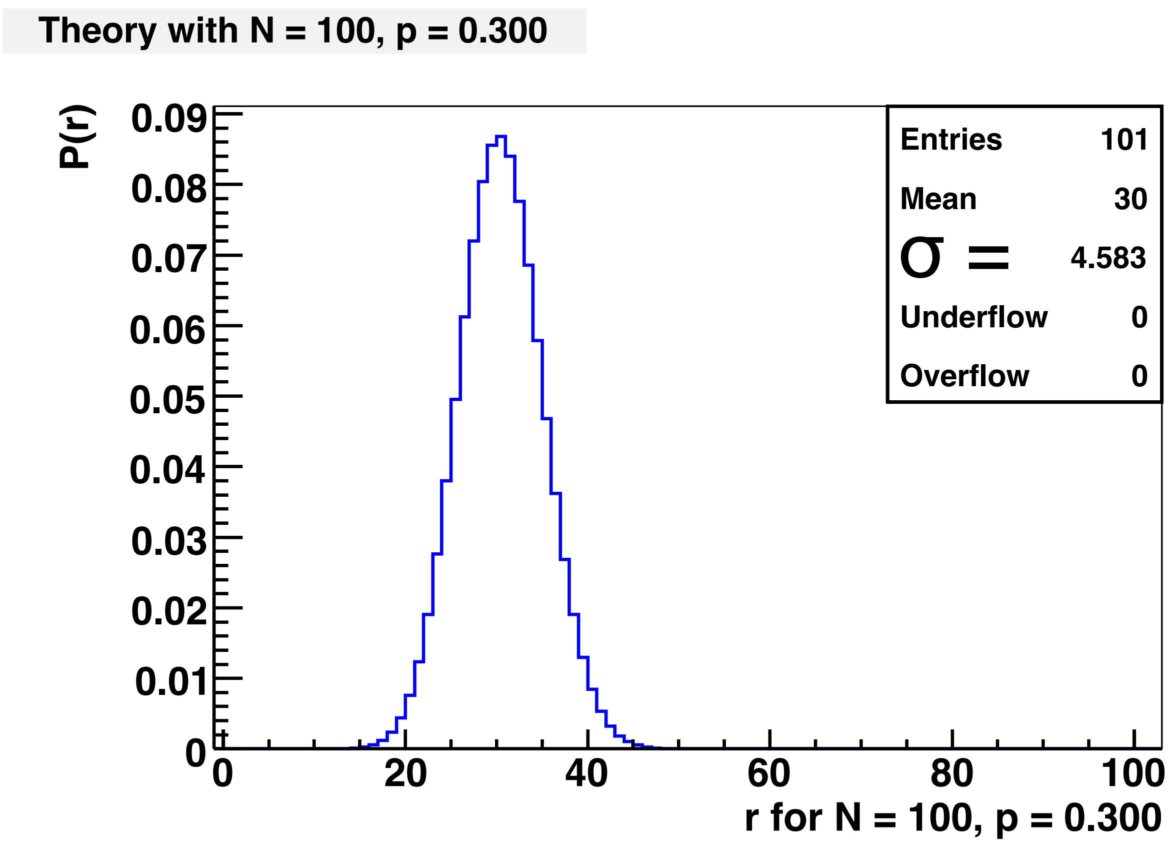
  if all σ_i the same: $= \frac{\sigma_i}{\sqrt{N}}$
- Gaussian as $N \rightarrow \infty$.

Central Limit theorem

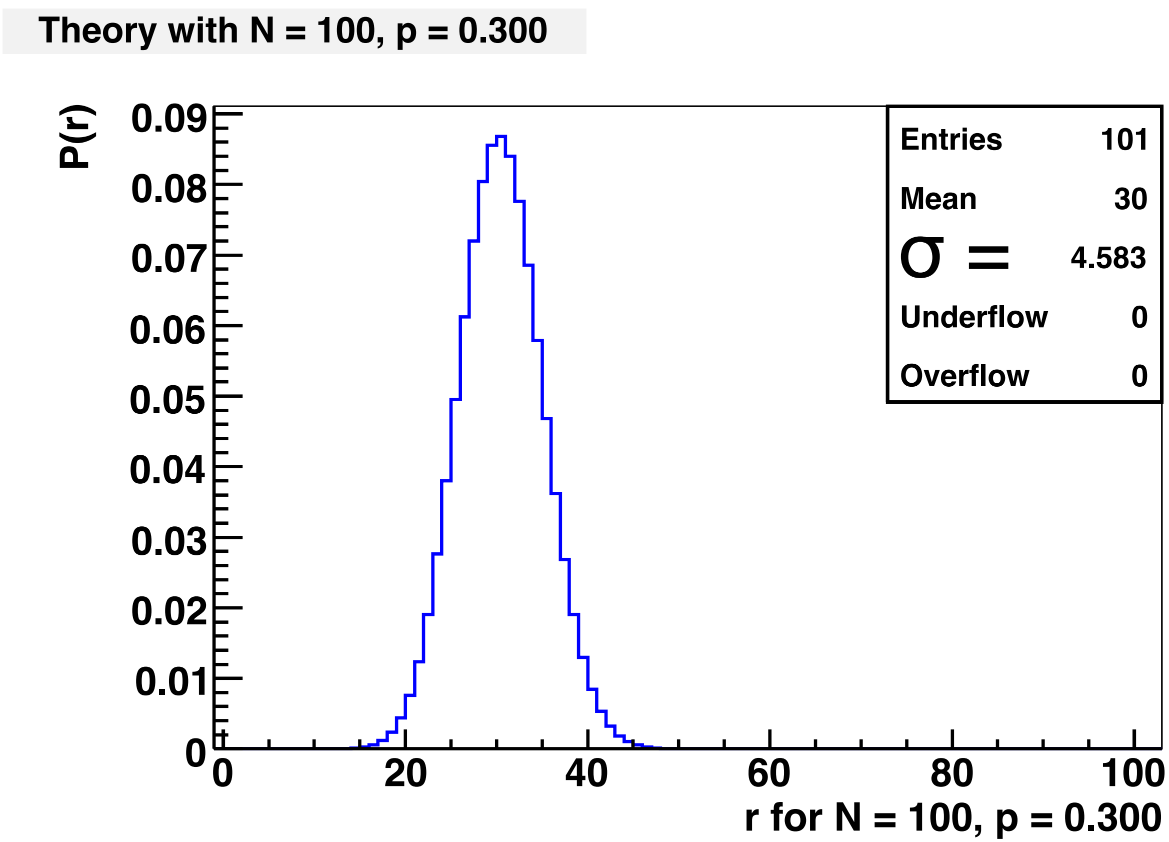
- Take the **sum** Y of N independent variables x_i $Y_{sum} \equiv \sum_{i=1}^{N} x_i$.
  - $\langle Y_{sum} \rangle = \sum \langle x_i \rangle$
  - Std dev. $\sigma_{Y_{sum}} = \sqrt{\sum \sigma_i^2}$
  - Gaussian as $N \to \infty$.
- Take the average Y of N independent variables x_i: $Y_{av} \equiv \frac{1}{N} \sum_{i=1}^{N} x_i$.
  - $\langle Y_{av} \rangle = \frac{1}{N} \sum \langle x_i \rangle$
  - Std dev.: $\sigma_{Y_{av}} = \frac{1}{N} \sqrt{\sum \sigma_i^2}$
    if all σ_i the same: $= \frac{\sigma_i}{\sqrt{N}}$
  - Gaussian as $N \to \infty$.

the 1st miracle of √N

What’s the uncertainty on the mean?

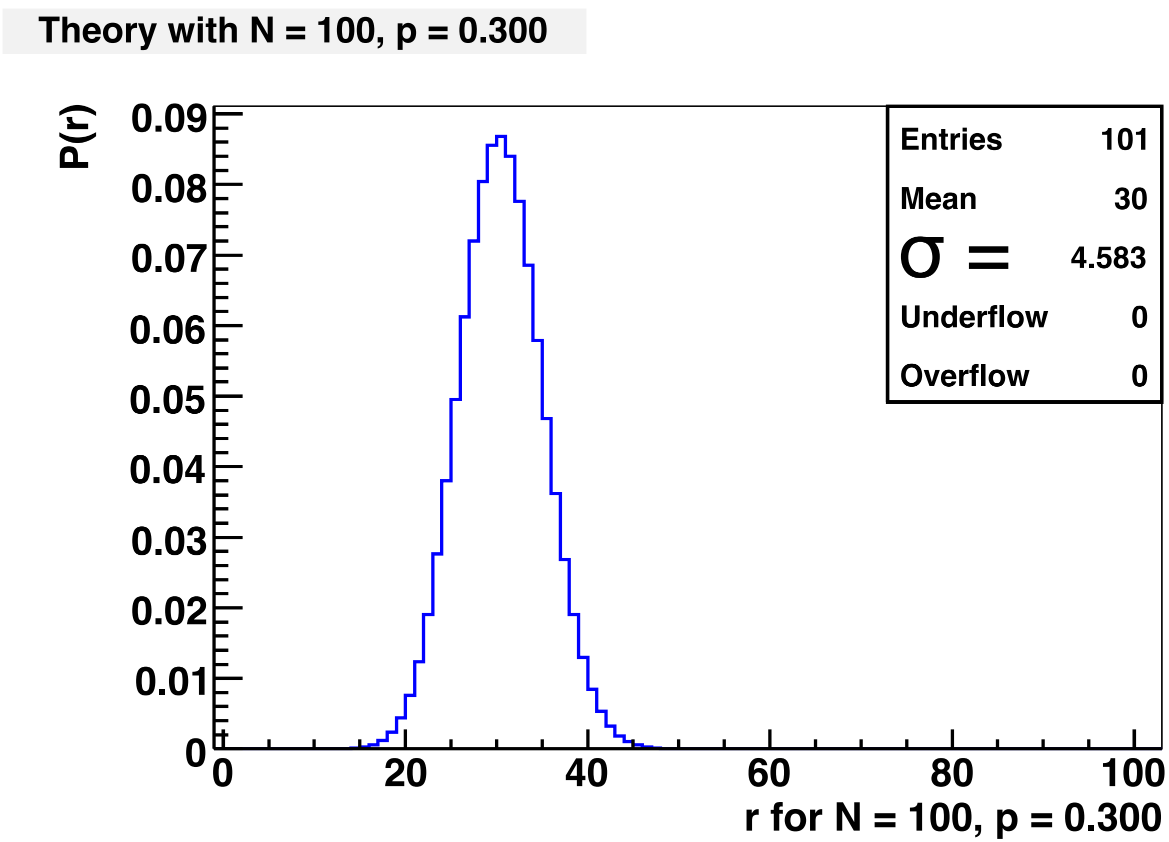


What's the uncertainty on the mean?



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

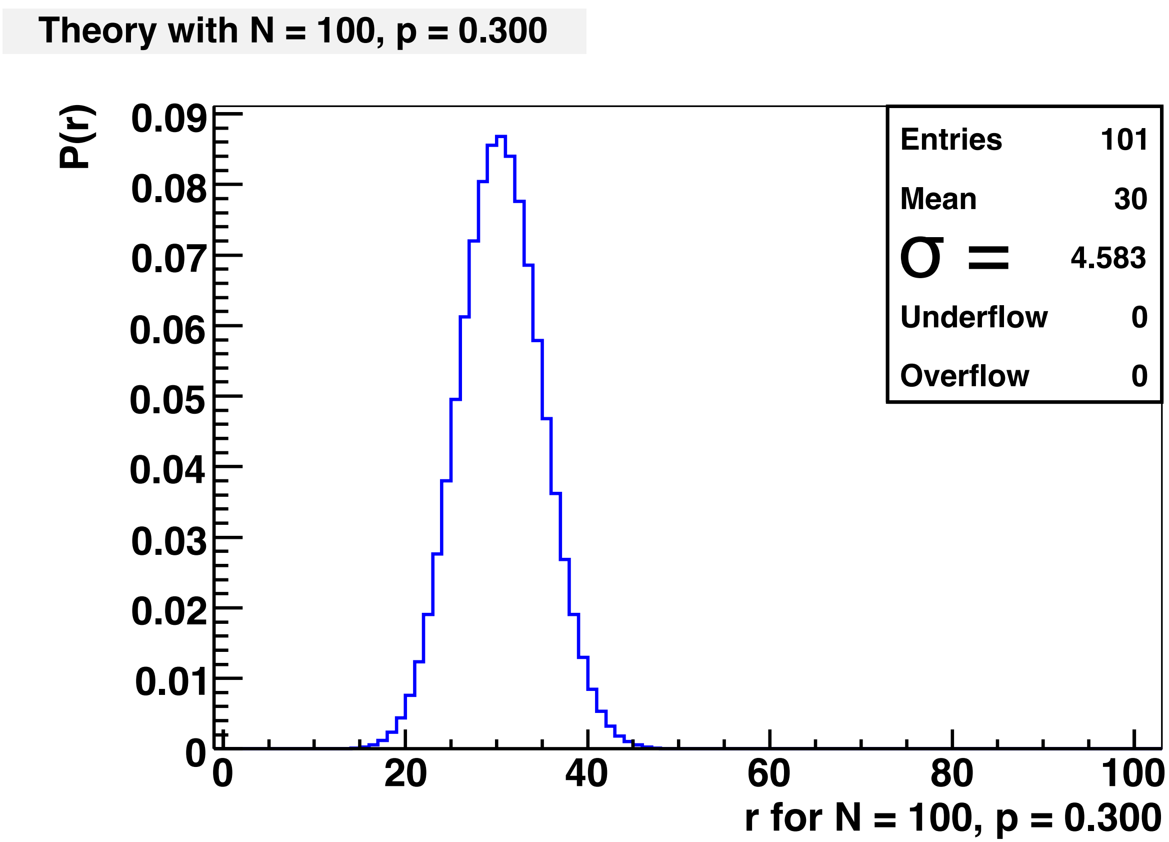
What’s the uncertainty on the mean?



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

N=101

What's the uncertainty on the mean?



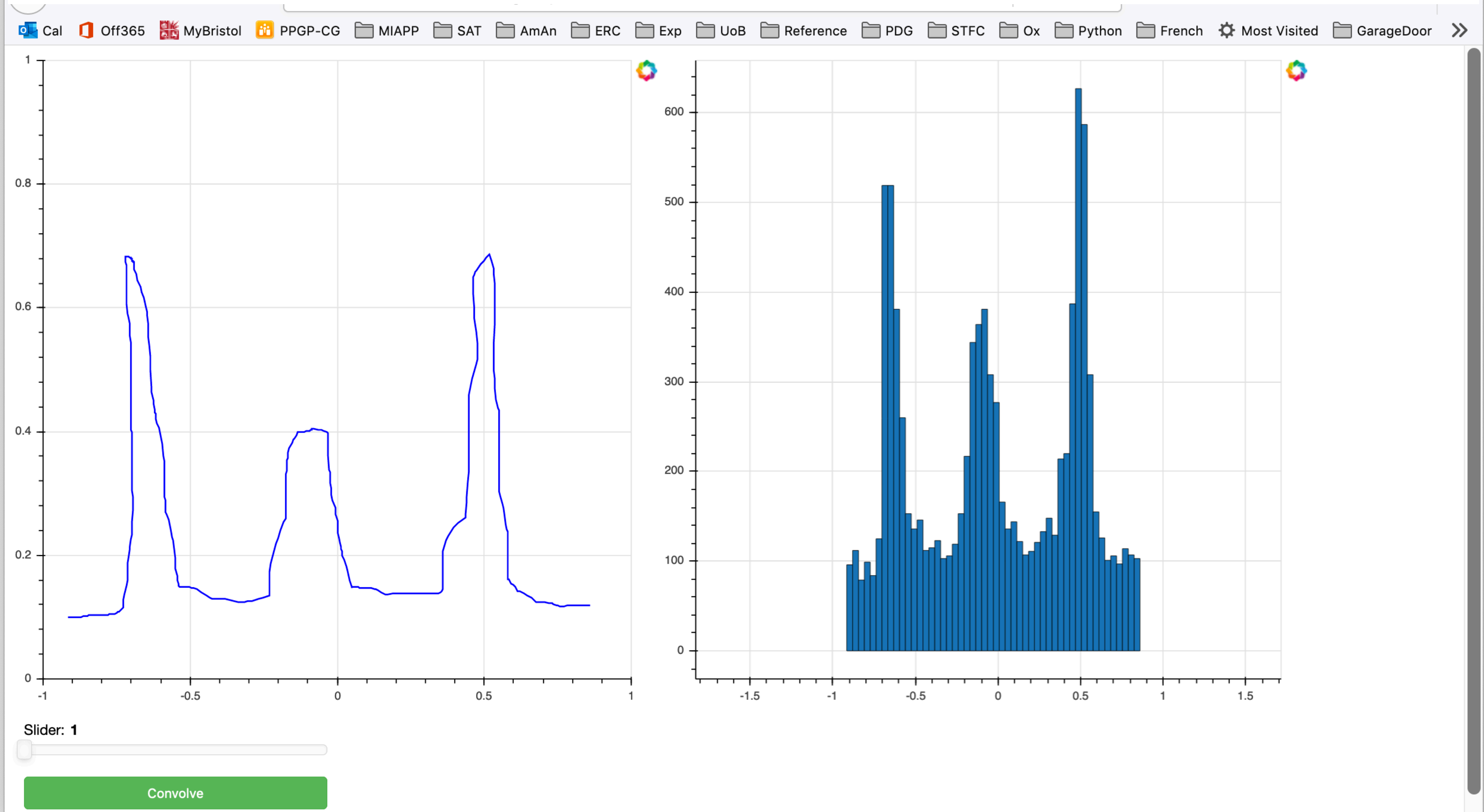
$\sigma_{mean} = \sigma/\sqrt{N}$

N=101

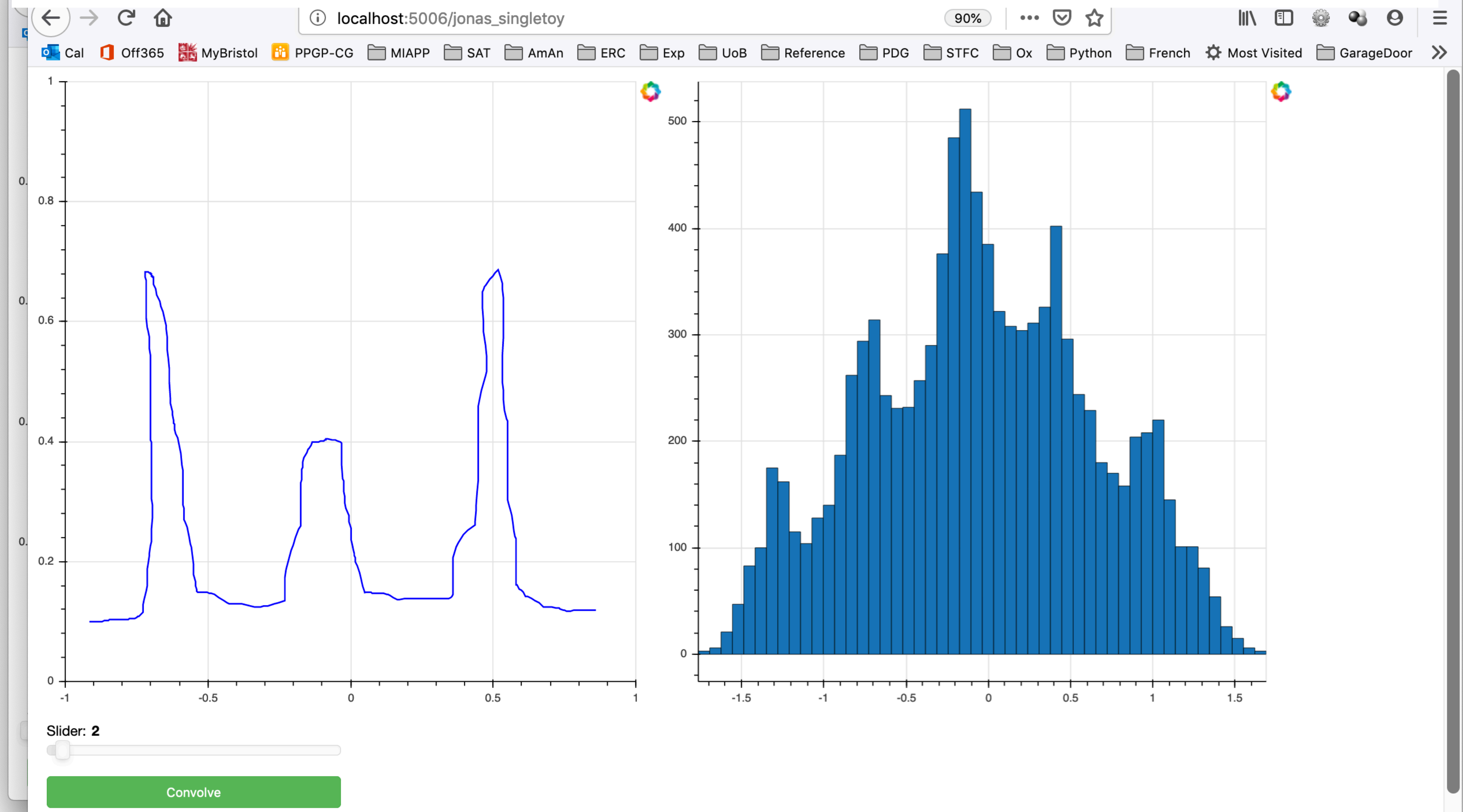
$\sigma_{mean} = 0.46$

The Central Limit Theorem

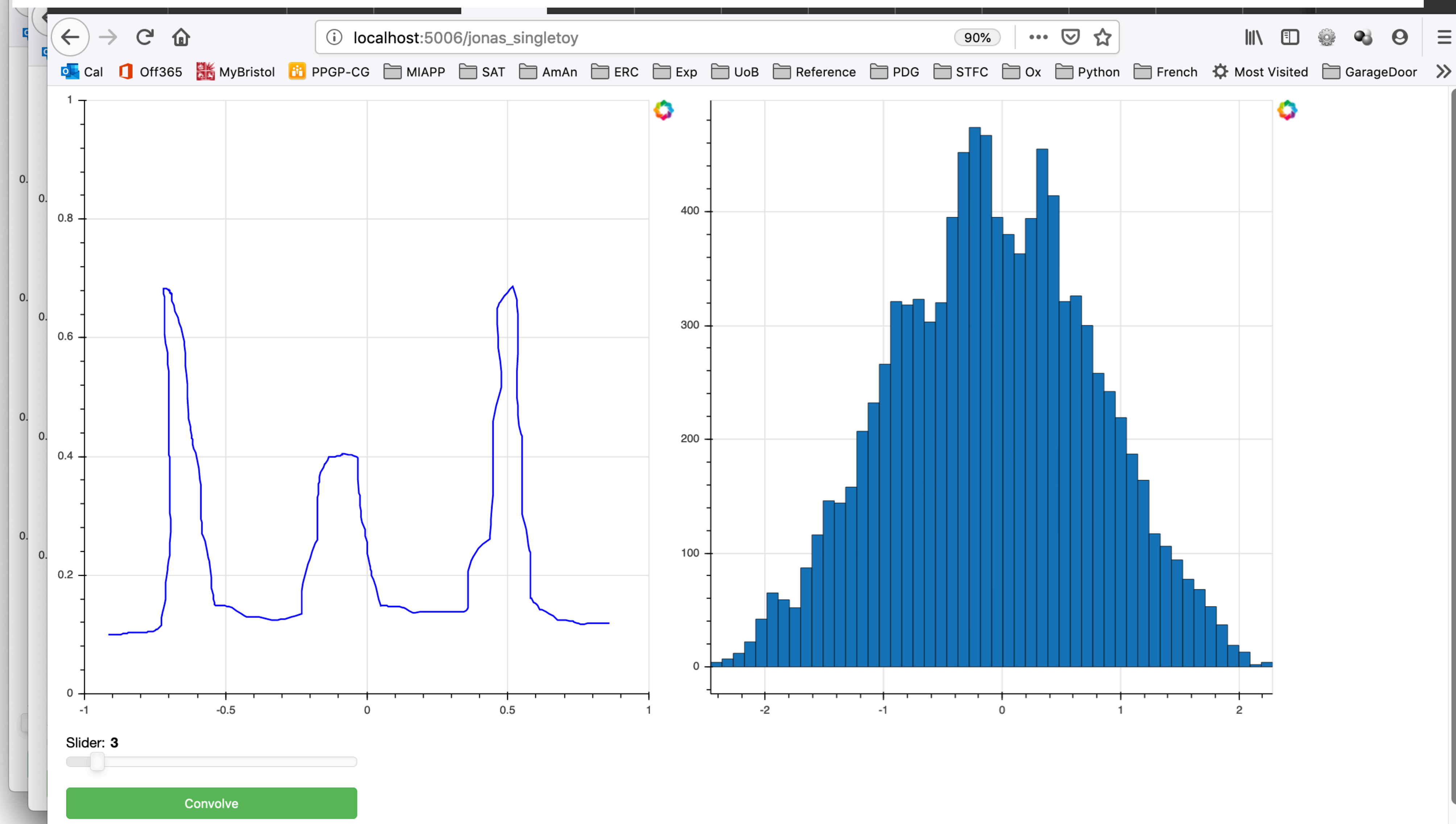
The Central Limit Theorem



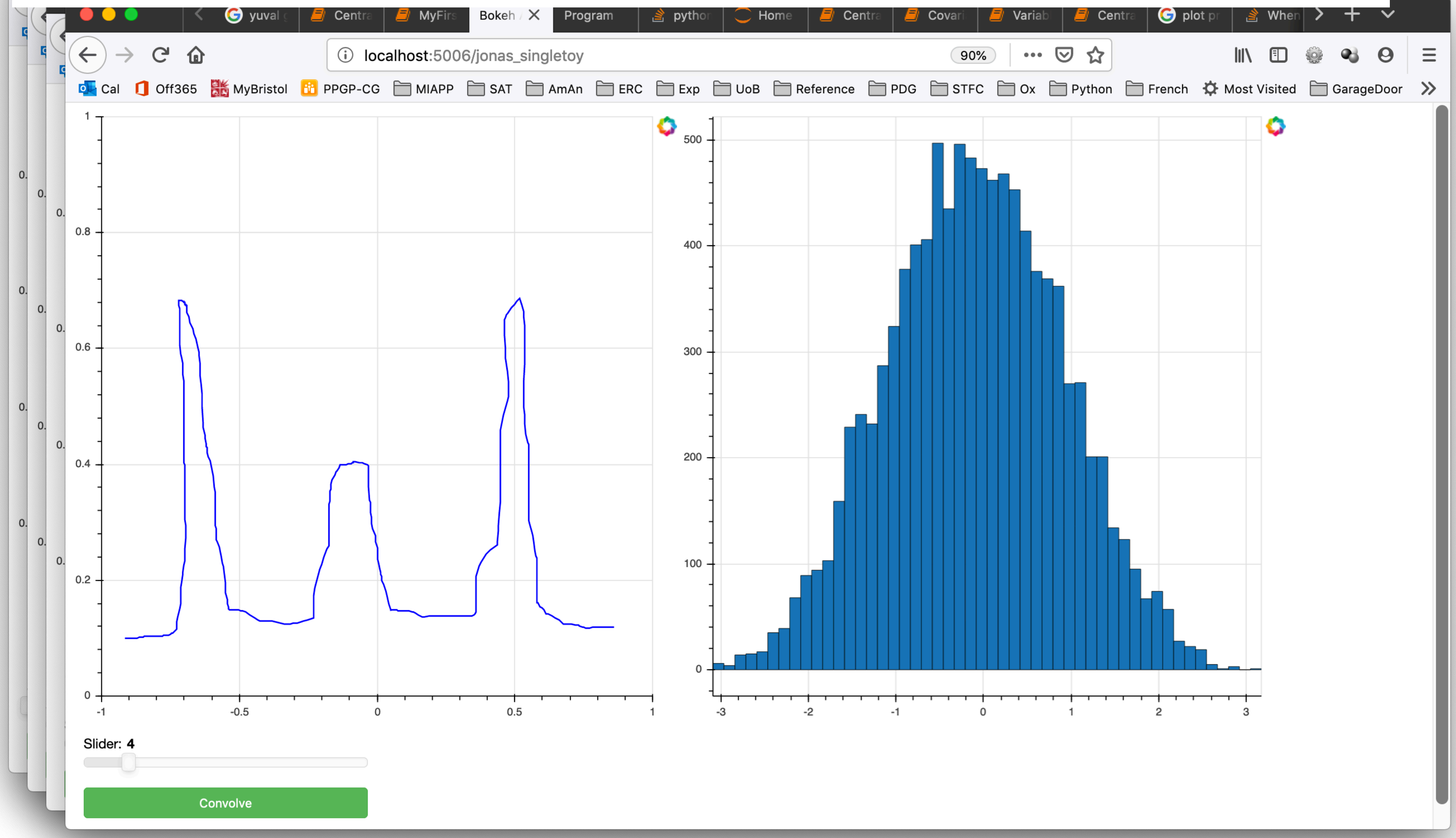
The Central Limit Theorem



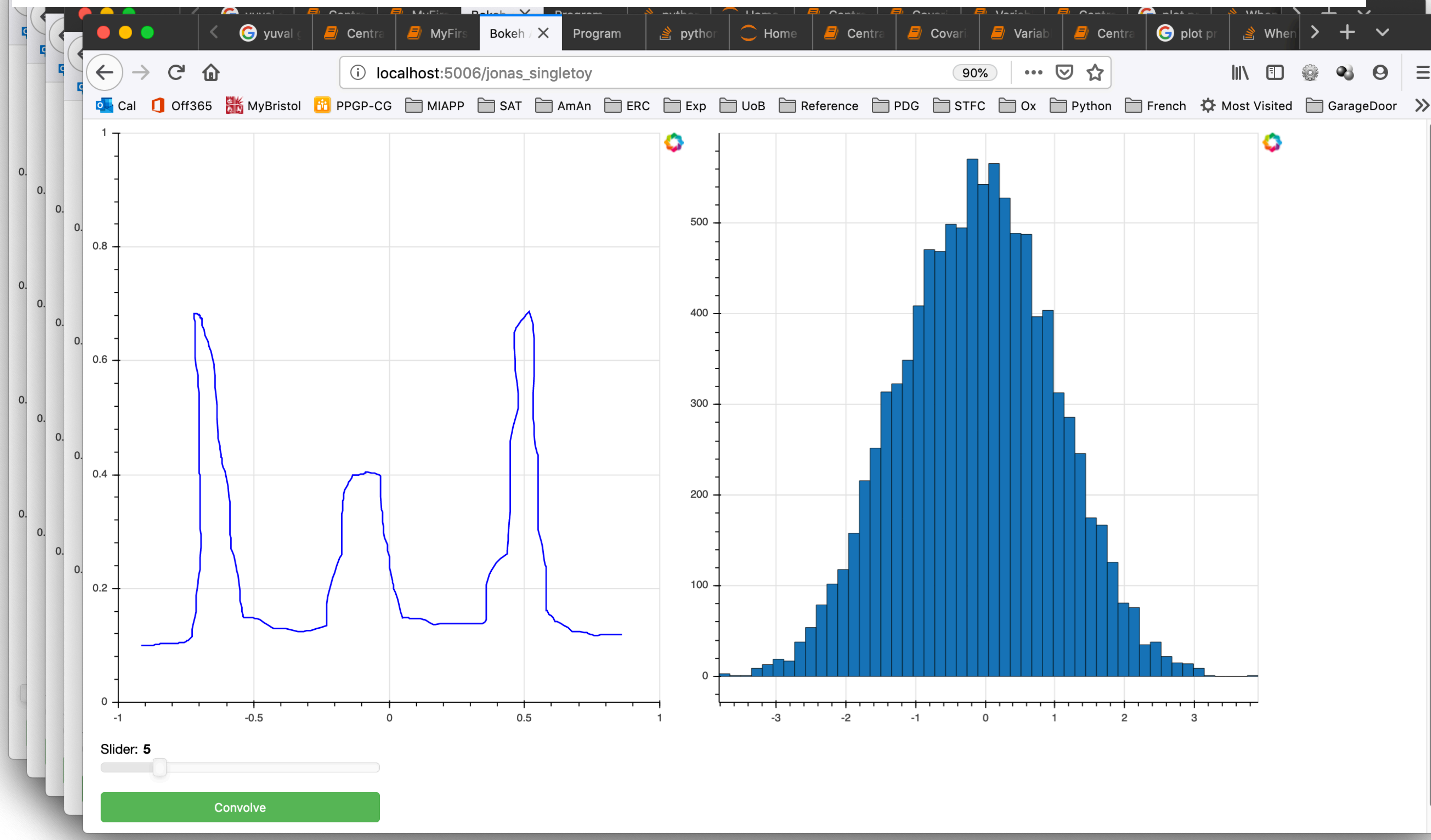
The Central Limit Theorem



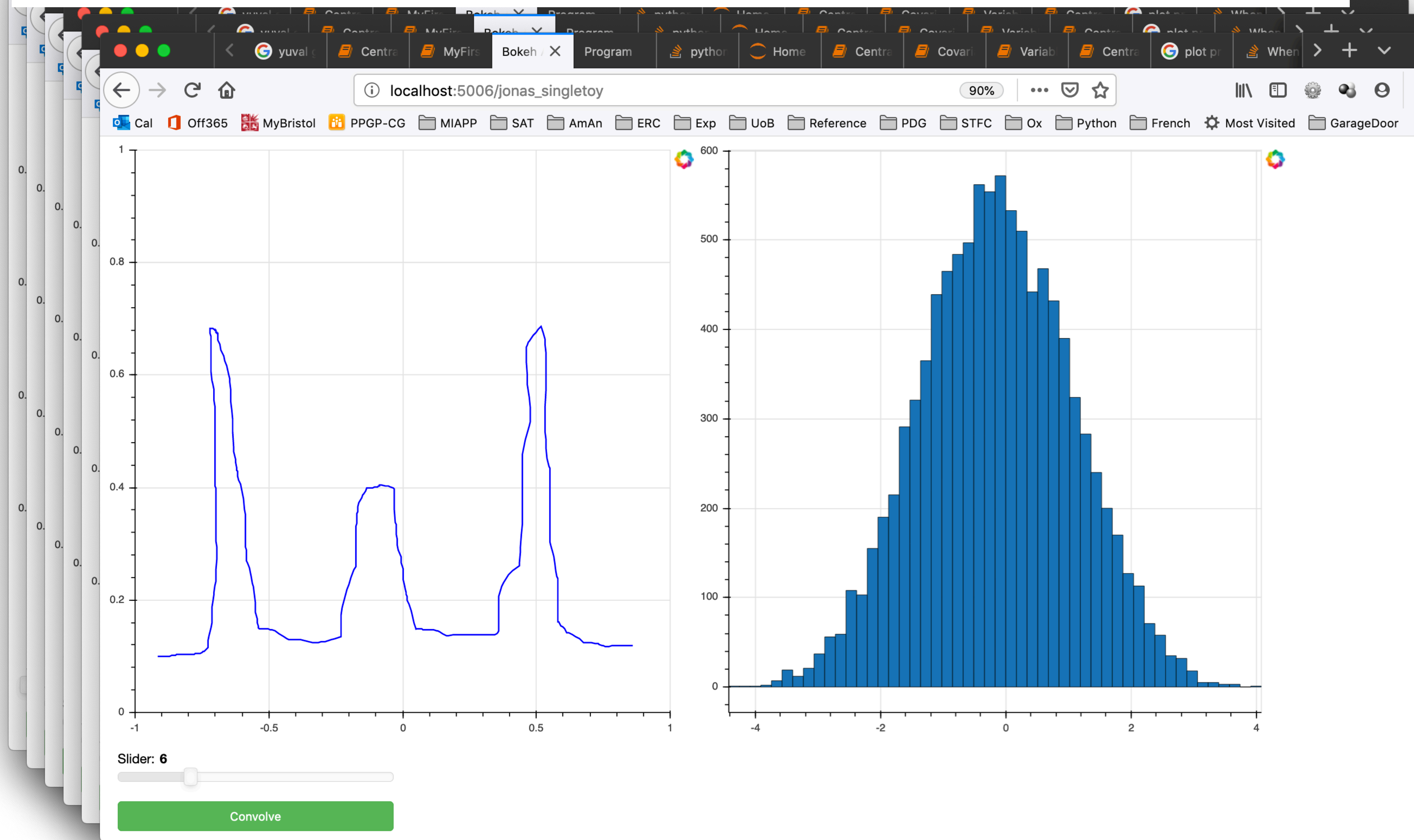
The Central Limit Theorem



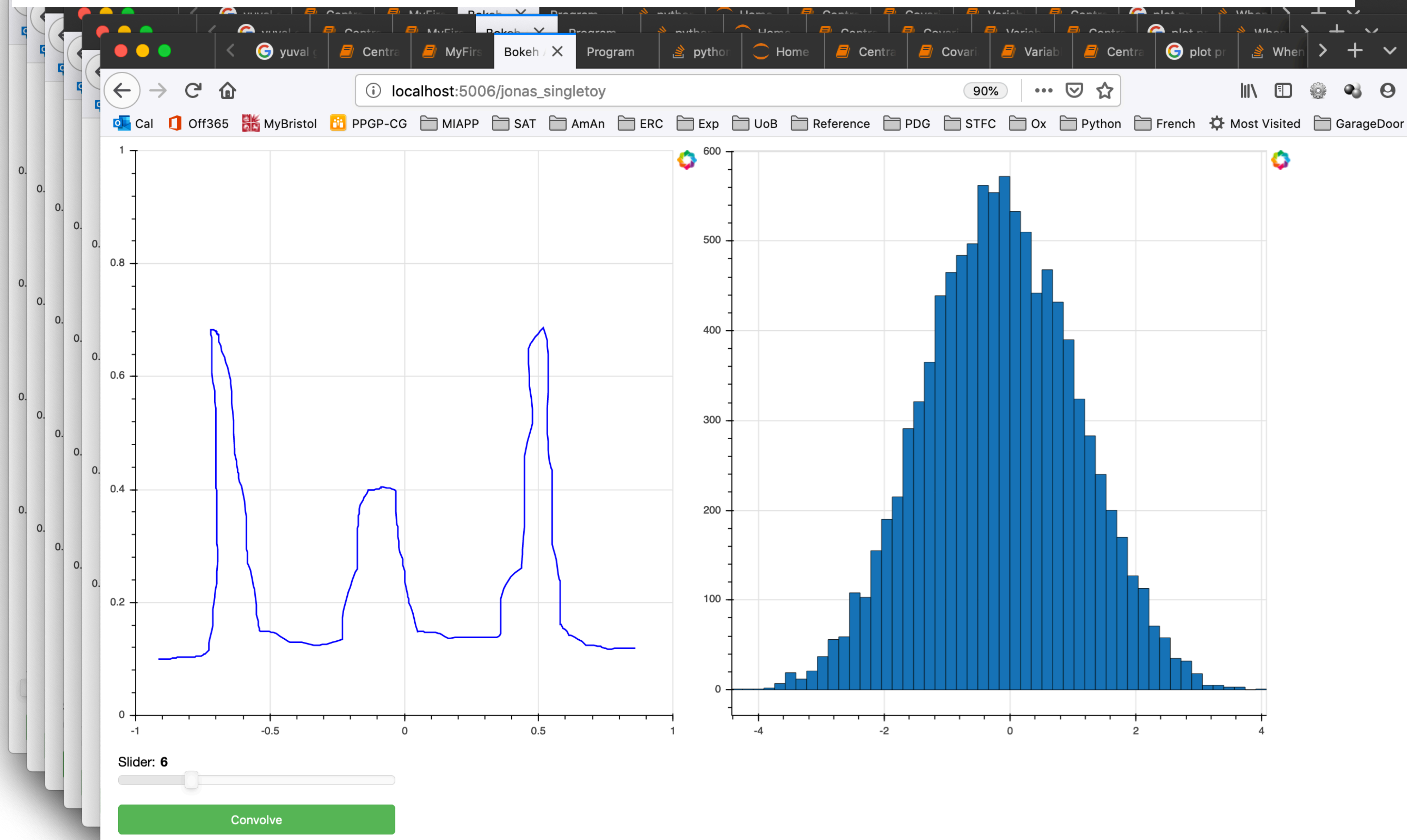
The Central Limit Theorem



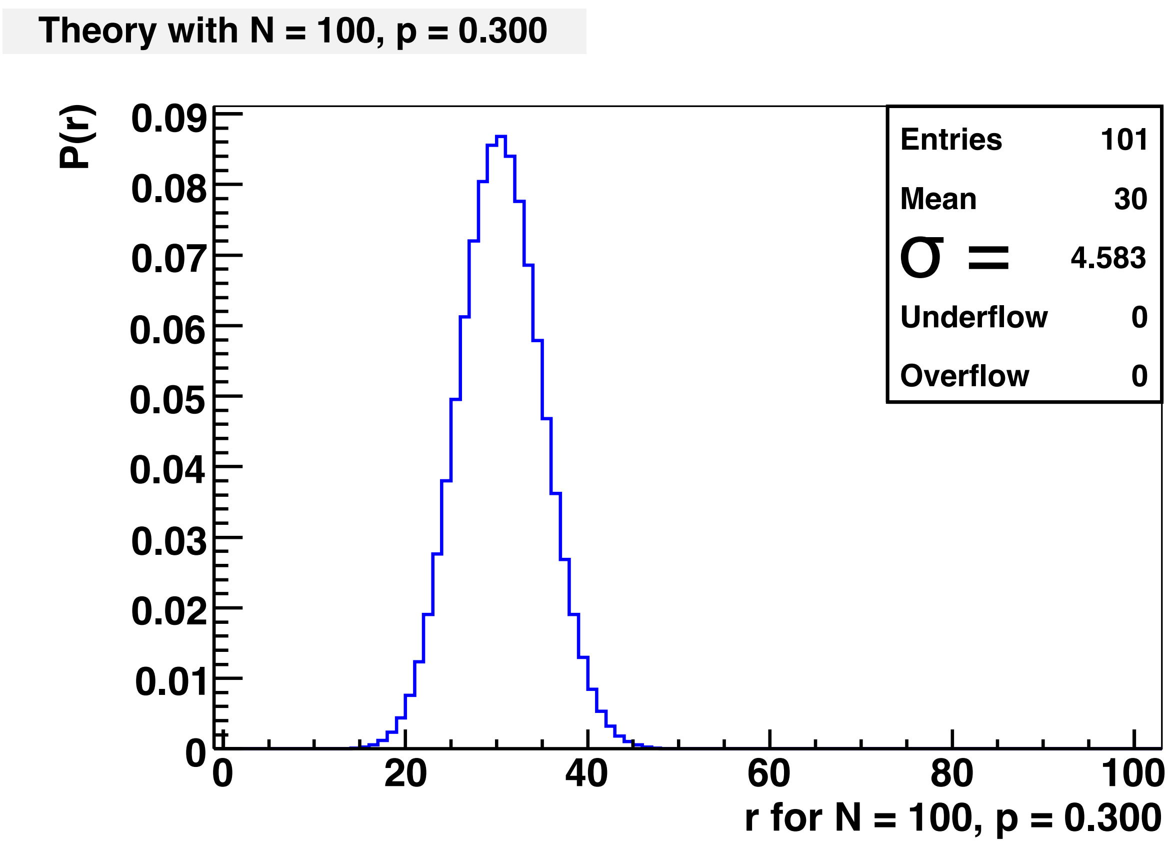
The Central Limit Theorem



The Central Limit Theorem

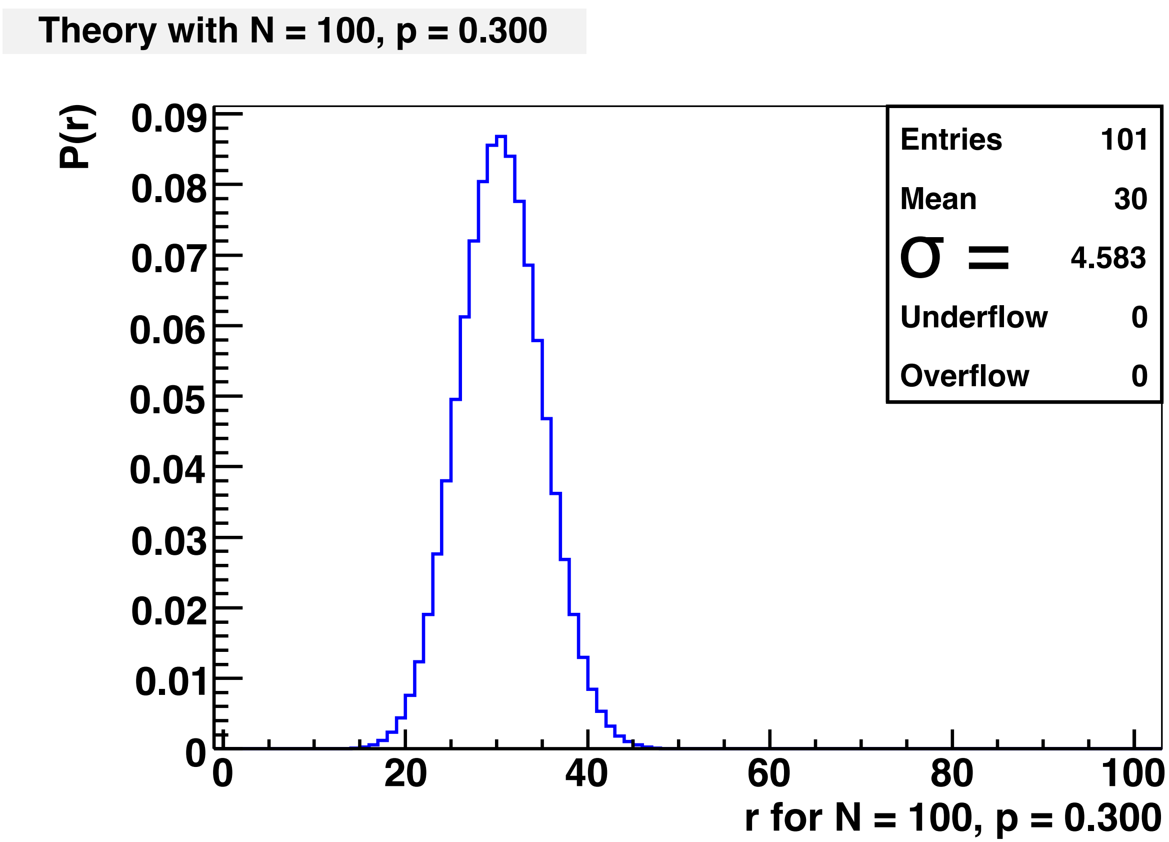


What's the uncertainty on the mean?



The 1st miracle of √N

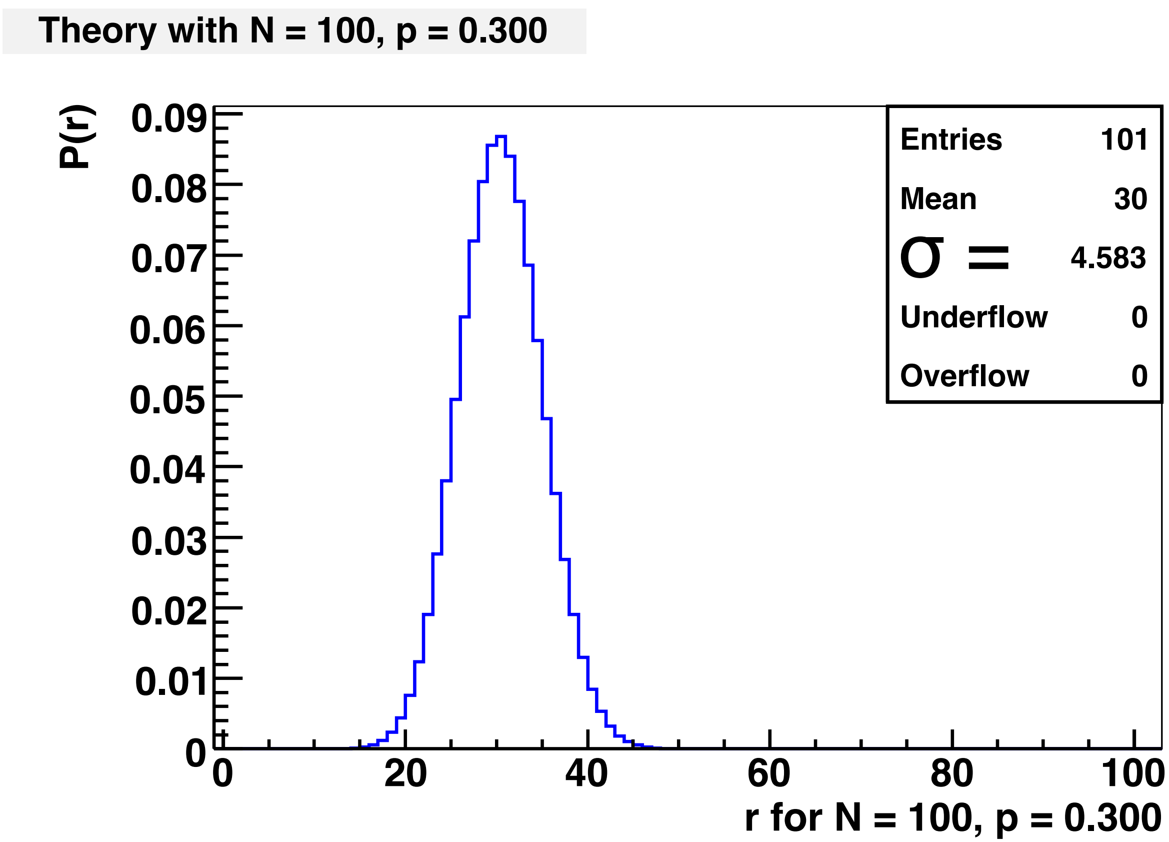
What's the uncertainty on the mean?



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

The 1st miracle of √N

What's the uncertainty on the mean?

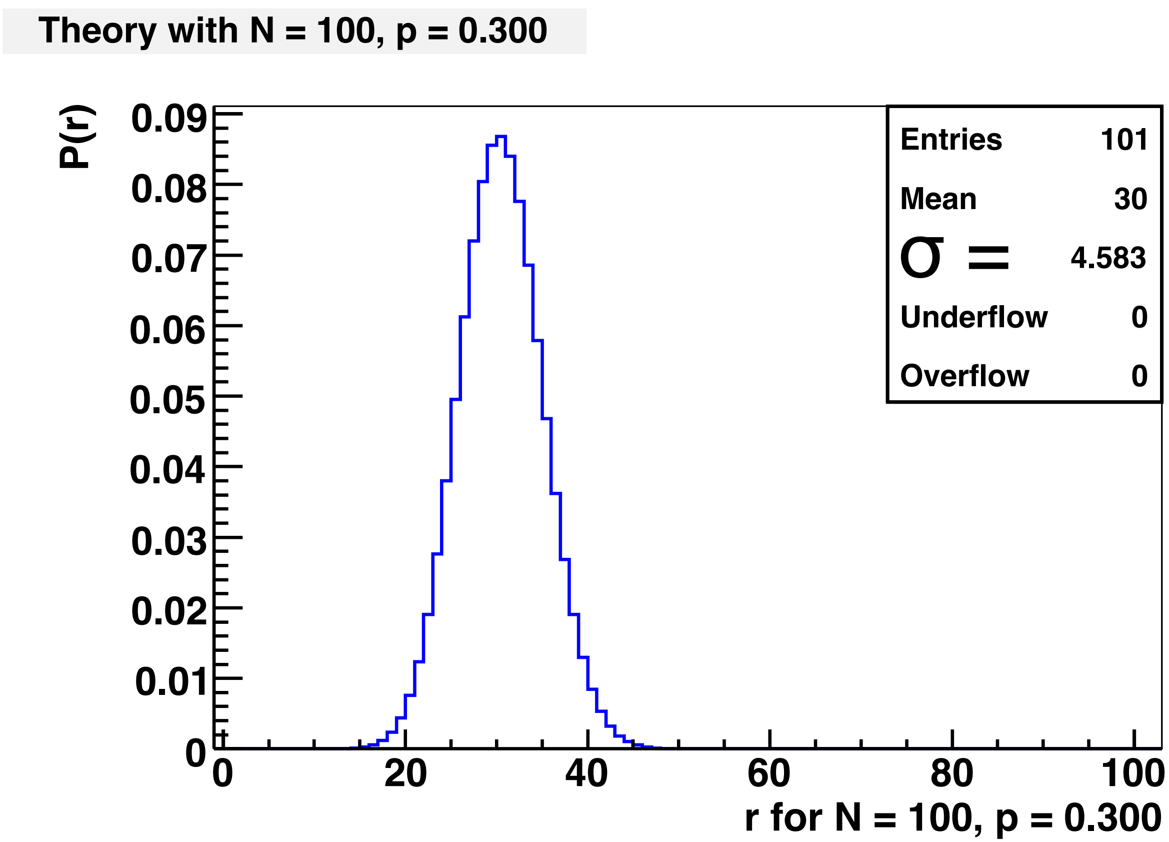


$\sigma_{mean} = \sigma/\sqrt{N}$

N=101

The 1st miracle of √N

What's the uncertainty on the mean?



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

$N=101$

$\sigma_{mean} = 0.46$

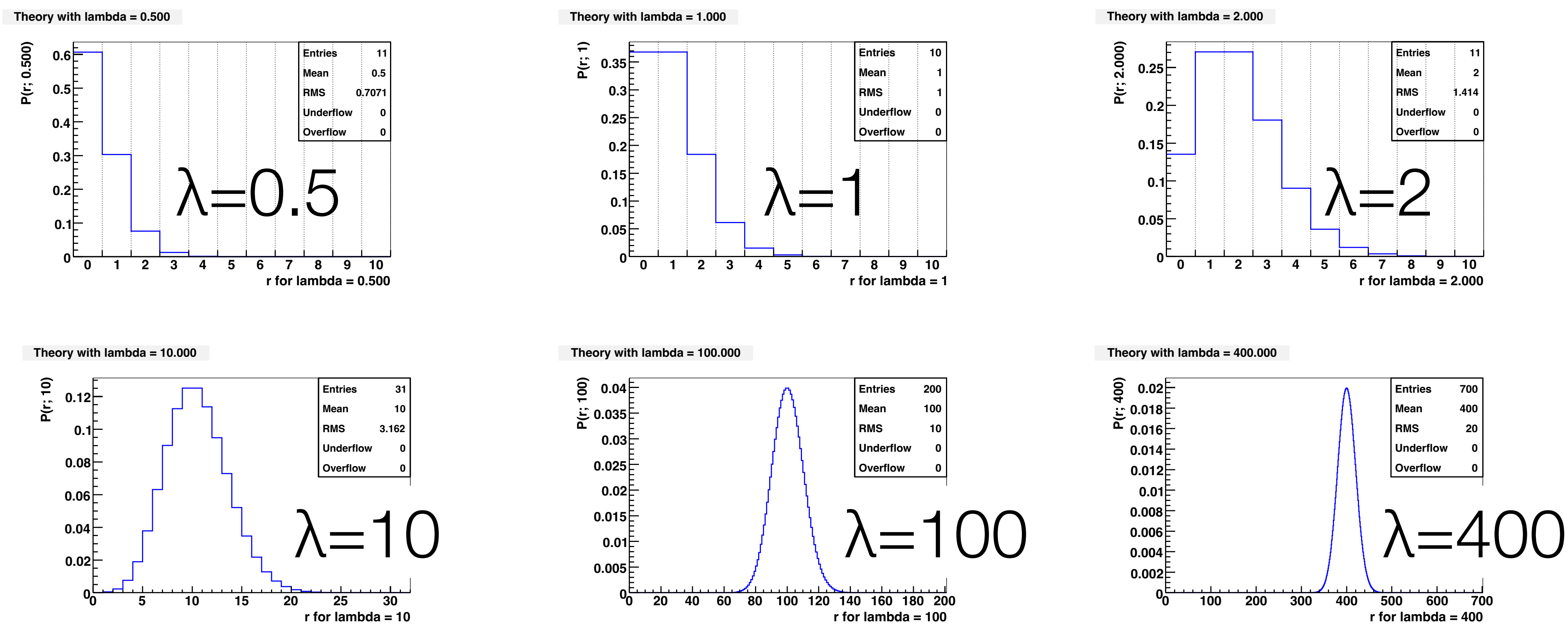
The 1st miracle of √N

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!**

We don't have much time and will do a super-fast version of this on the whiteboard, then continue on slide 87. The more detailed slides will be on indico.

Poisson → Gaussian



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

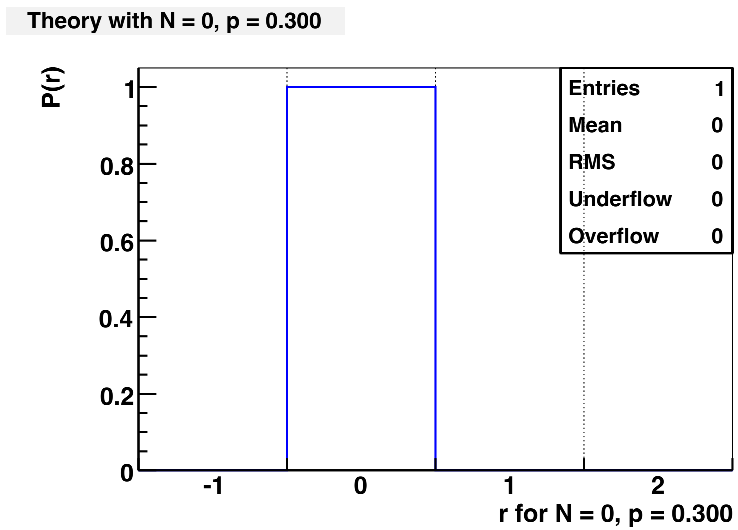
number of "failures"

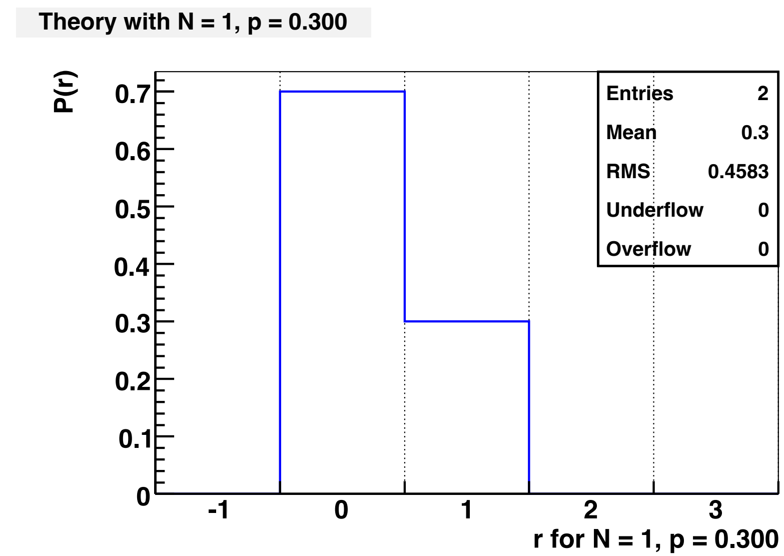
$$P\left(r; N, p\right) = p^r \left(1-p\right)^{N-r} \begin{pmatrix} N \\ r \end{pmatrix}$$

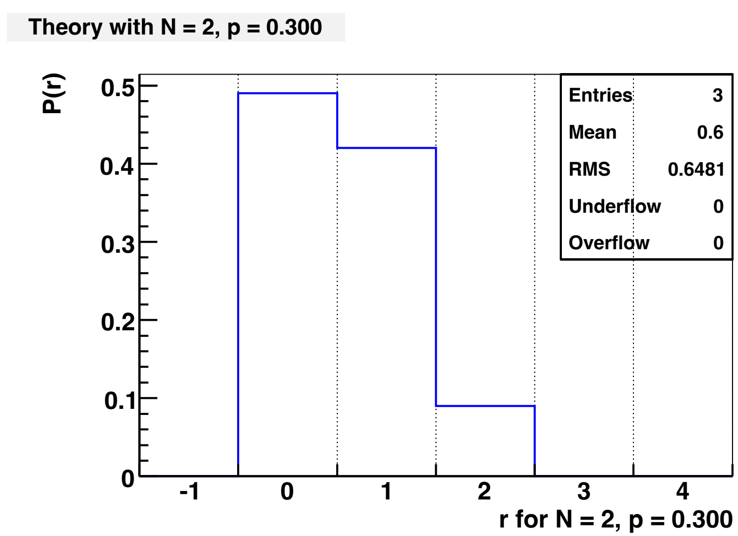
$$= p^r \left(1-p\right)^{N-r} \frac{N!}{r!\ (N-r)!}$$

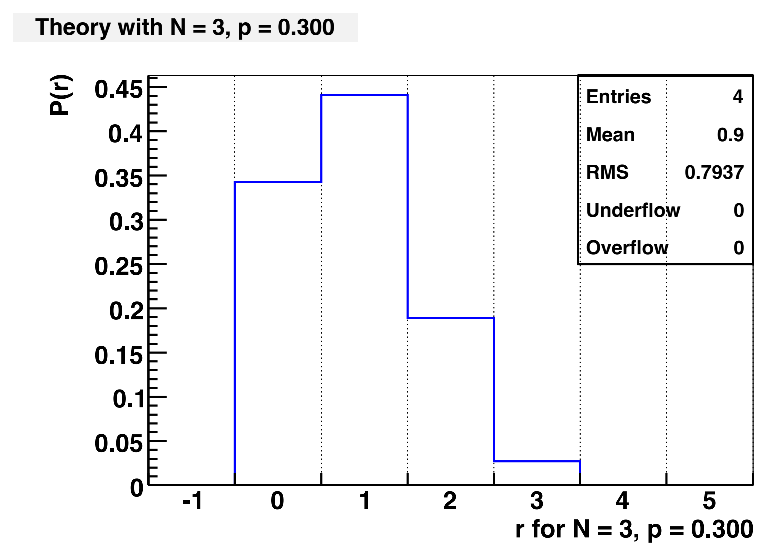
number of different sequences
in which one can have r
successes and N – r failures

Binomi Examples

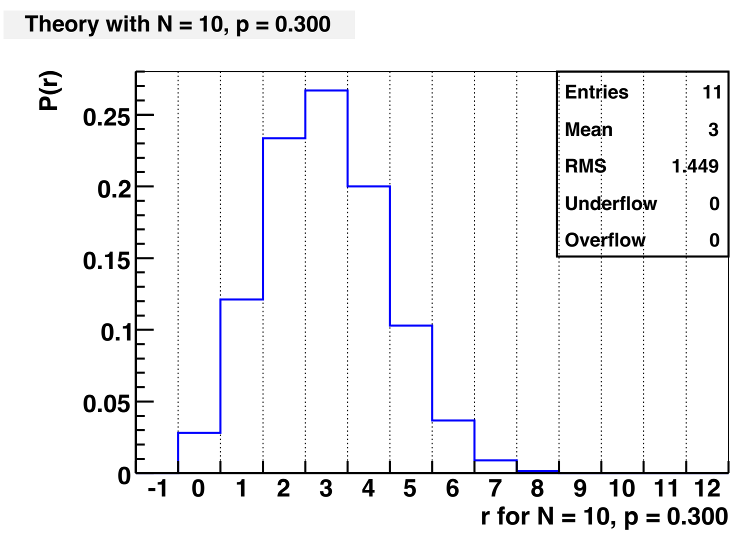


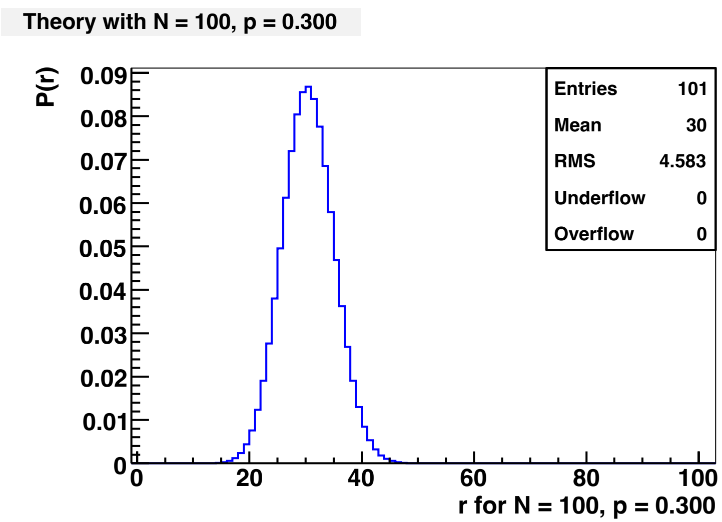


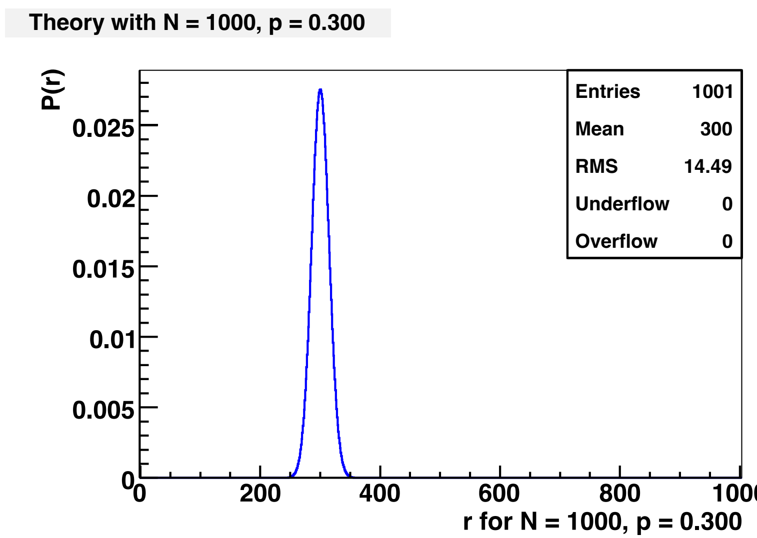


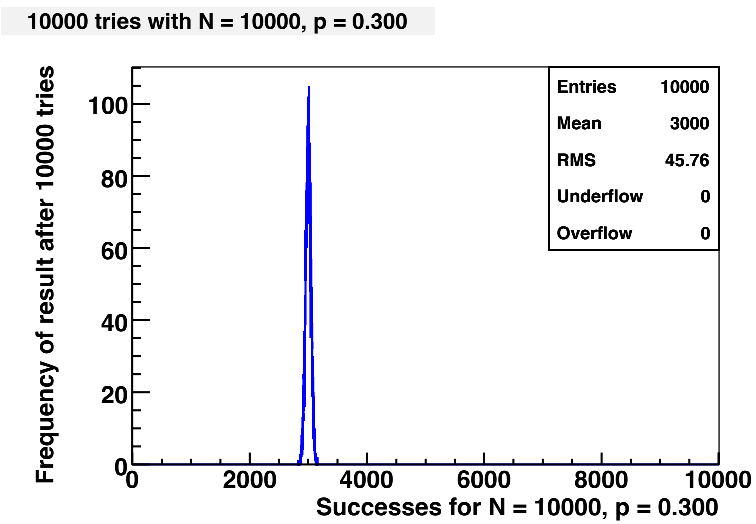


Binomi Examples









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 <u>one single parameter</u> - 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.

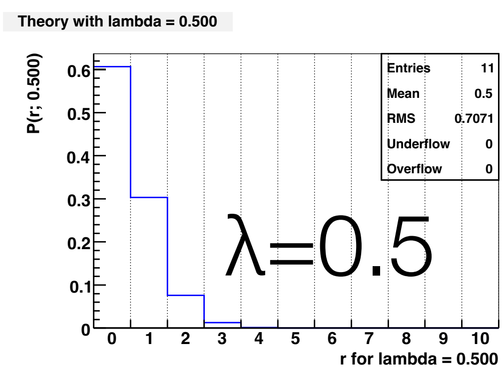
- **The probability to see *r* events, given an expected mean of *λ*, is:**

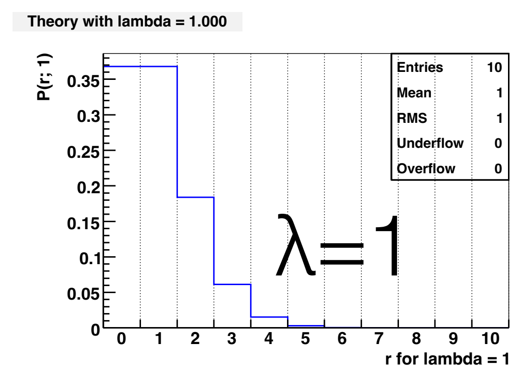
$$P(r;\lambda) = e^{-\lambda}\frac{\lambda^r}{r!}$$

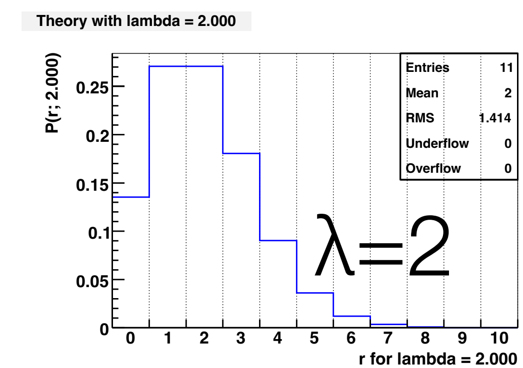
Binomial → Poisson

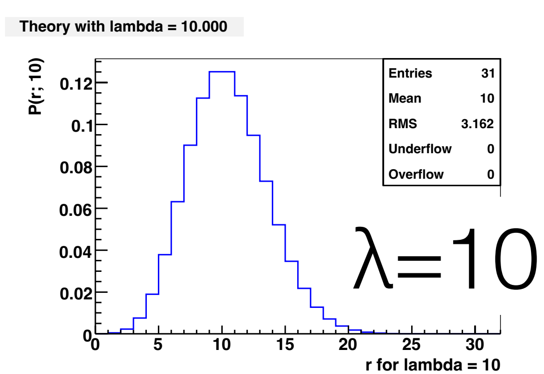
- ... our derivation (if we did it) implies that the Poisson distribution with $\lambda=Np$ is a decent approximation of the Binomial distribution in cases where p is small and N is large.

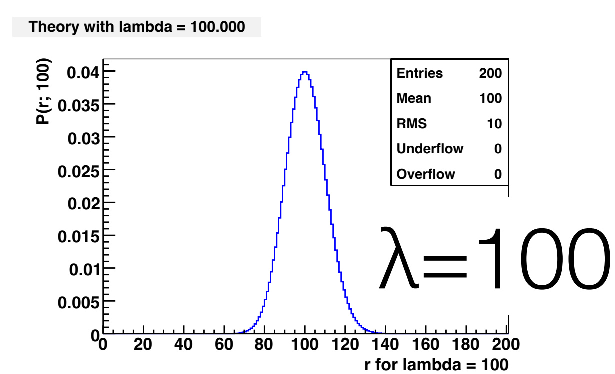
Poisson → Gaussian

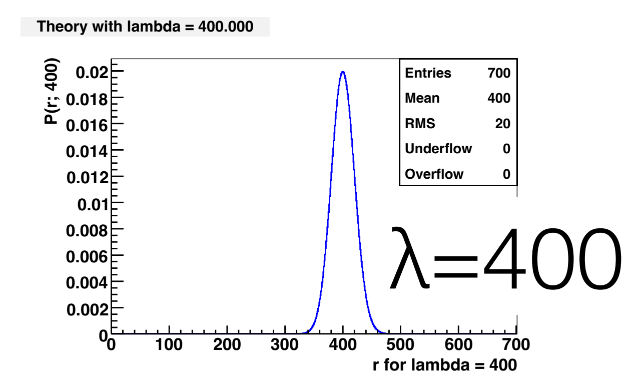












Trinity

$$P\left(r; N, p\right) \quad = \quad p^{r}\ (1-p)^{N-r} \begin{pmatrix} N \\ r \end{pmatrix}$$

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

Binomial

P(r; N,p)

lim N→∞, p→0, N·p=λ

N·p→λ

Poisson

P(r; λ)

lim N→∞

$$\mathrm{N}\cdot\mathrm{p} \quad \rightarrow \quad \mu$$

$$\sqrt{\mathrm{Np}(1-\mathrm{p})} \quad \rightarrow \quad \sigma$$

lim λ→∞

λ→μ,

√λ→σ

Gaussian

P(x; μ,σ)

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

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.

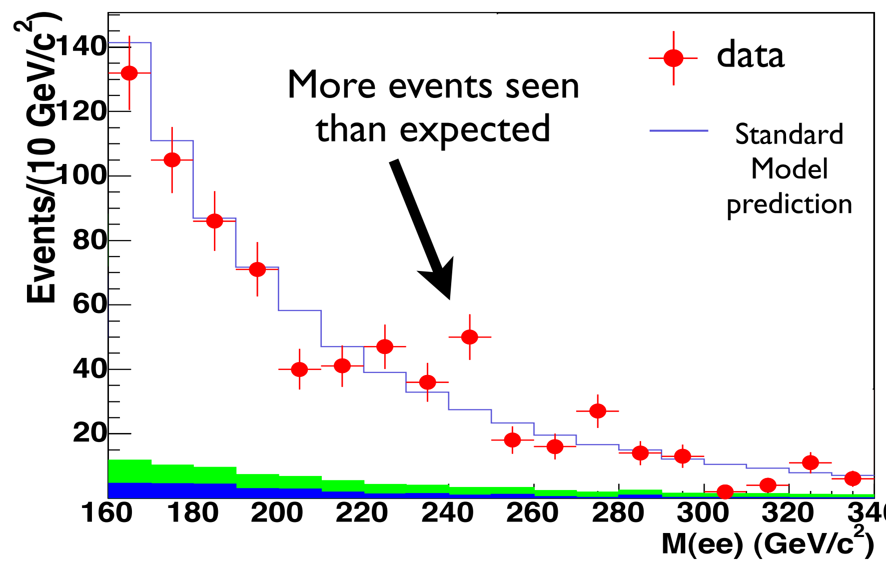
* H1F1 gives you bird flue

https://tinyurl.com/TeshepProblems

More Homework - calculate significances

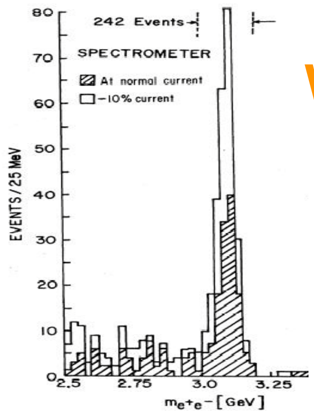
- Estimate the significance of this observation:
  - Step 1: calculate the probability so see an upward fluctuation this big or bigger in the Standard Model, in this one bin
  - Step 2: take into account that they looked in 84 bins (tricky!)
- You should get a fairly small number. Why, do you think, 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:

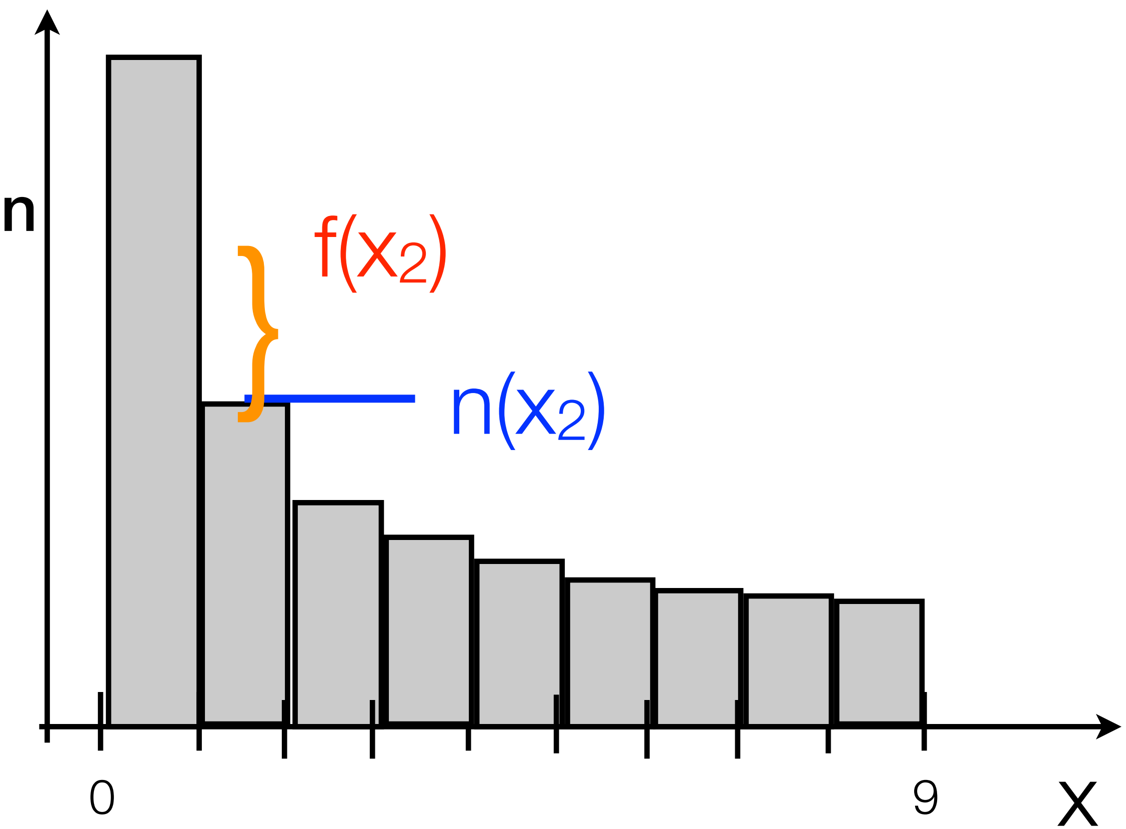
$$P(t) = \frac{1}{\tau} e^{-t/\tau}$$

- 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

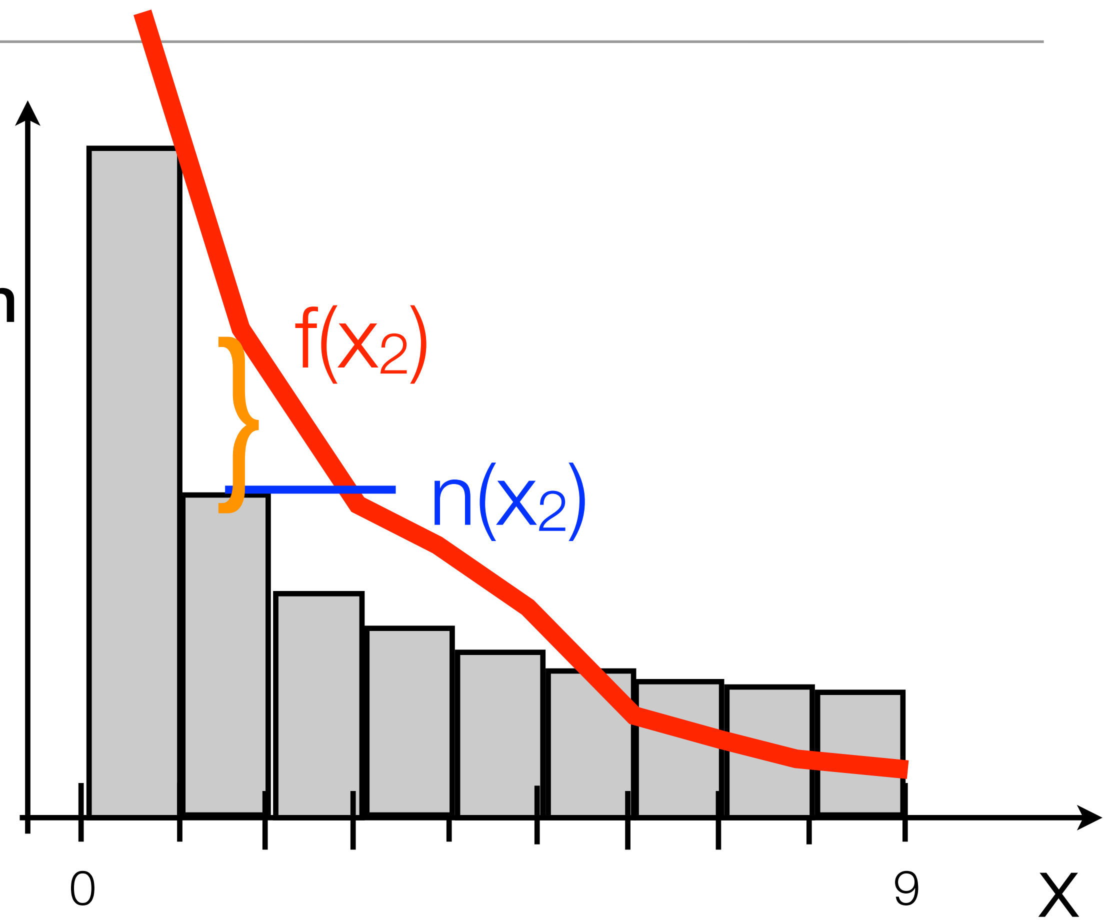
- **Use for binned data**
- **Minimise distance between data and function that describes data.**



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.**

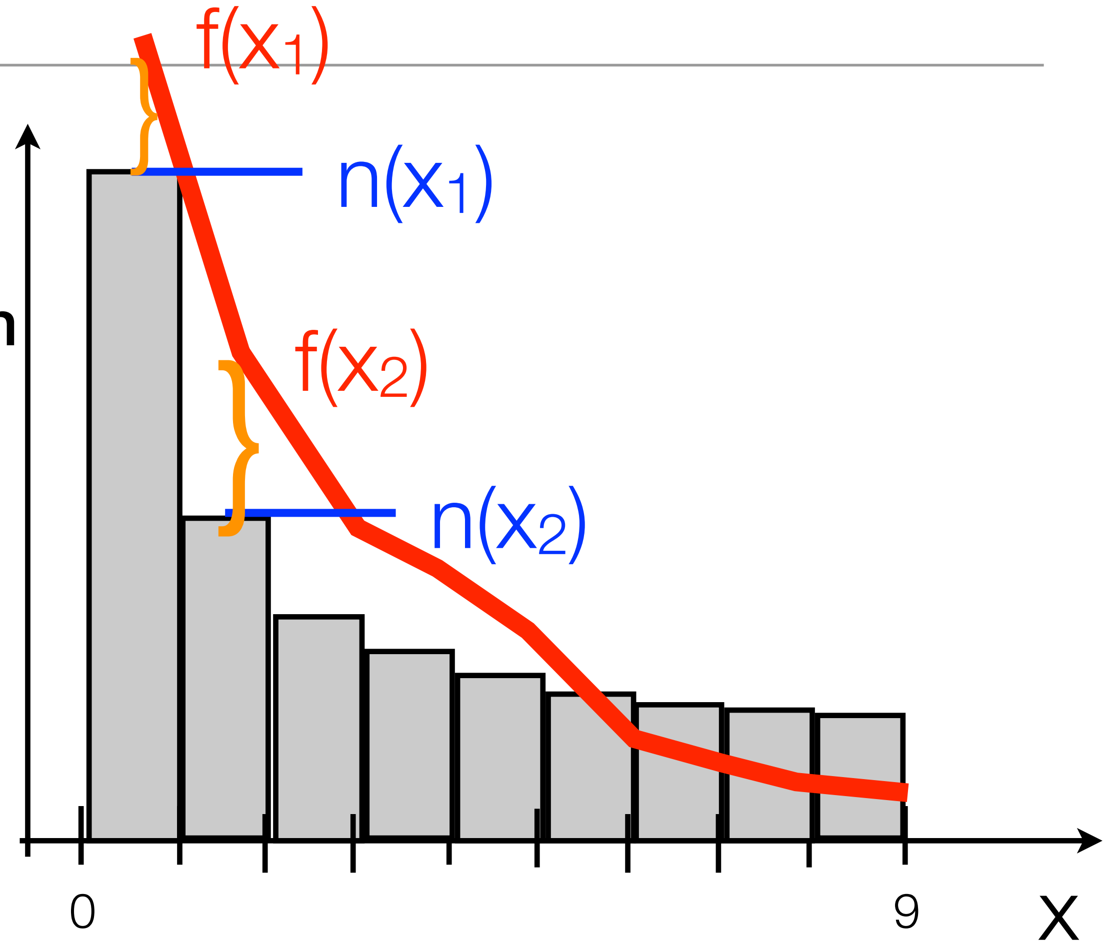


usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

χ^2 Fitting

- **Use for binned data**

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

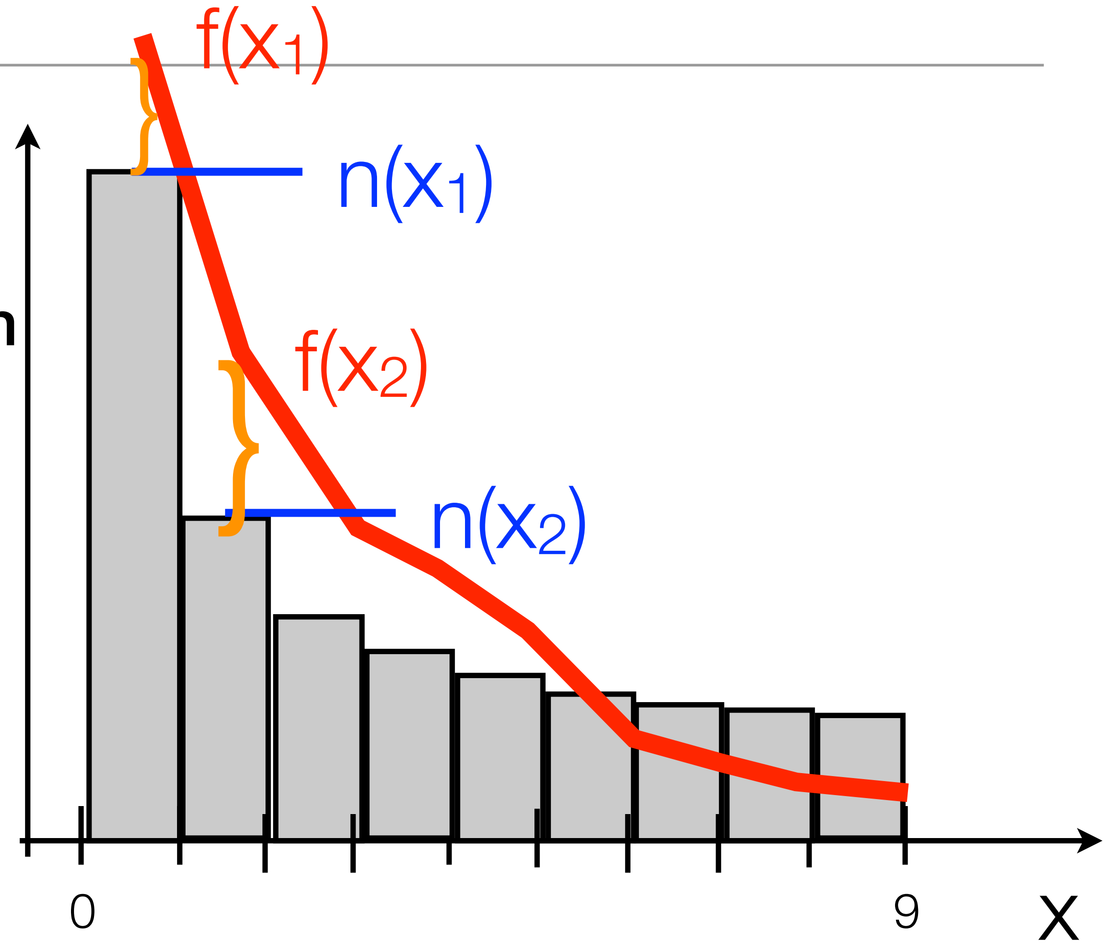


usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

χ^2 Fitting

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

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



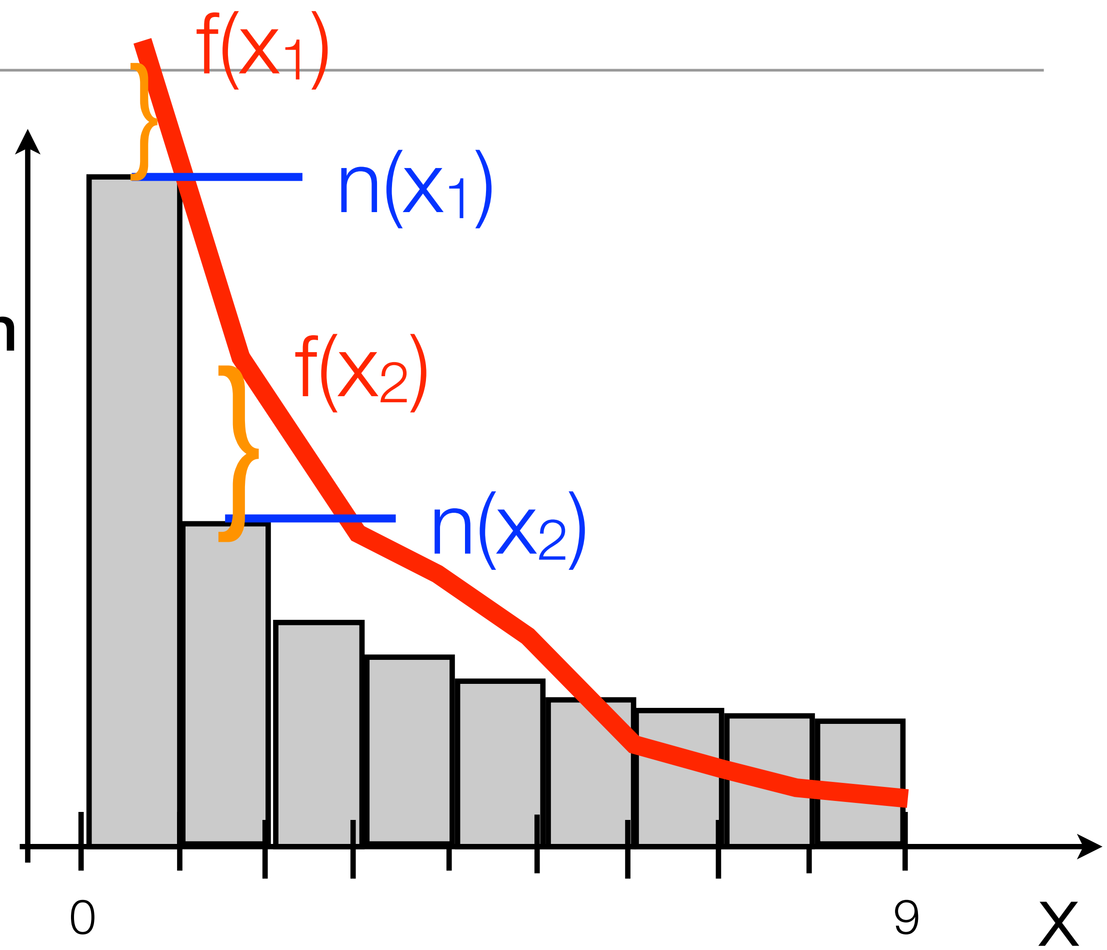
usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

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**



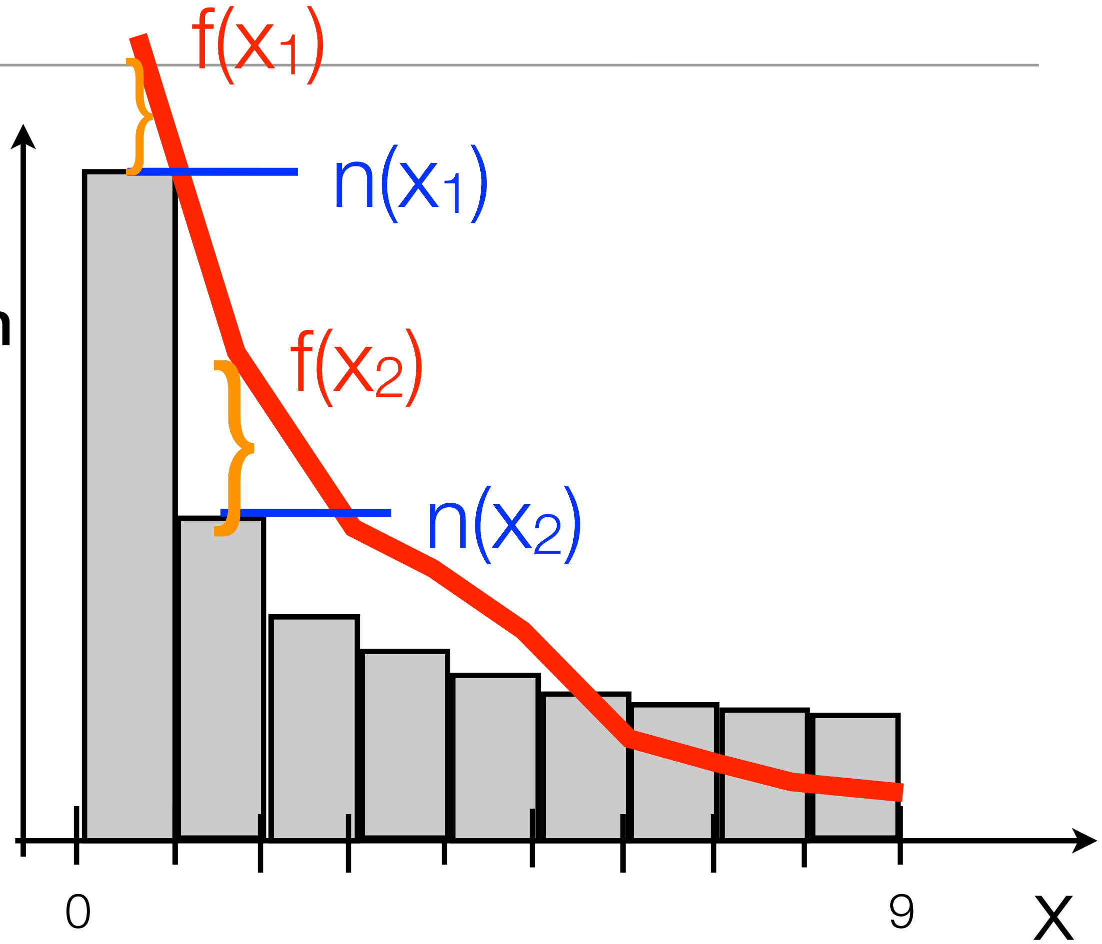
usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

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**

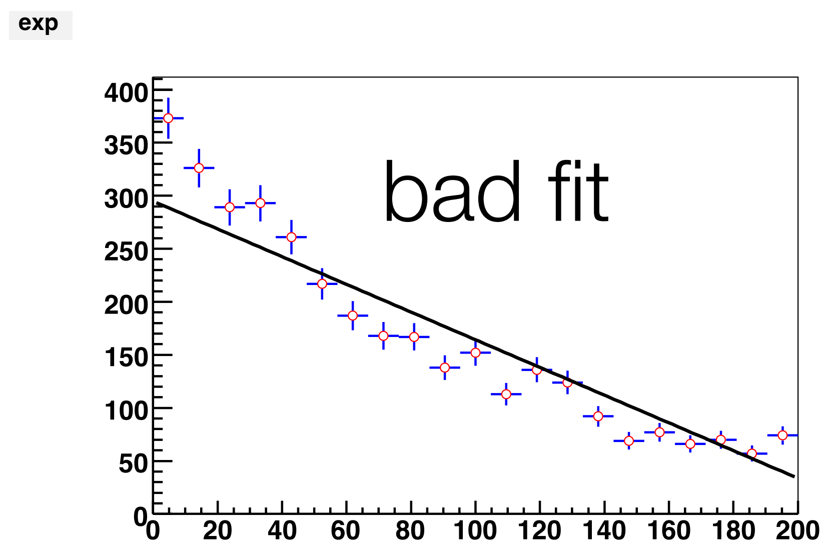


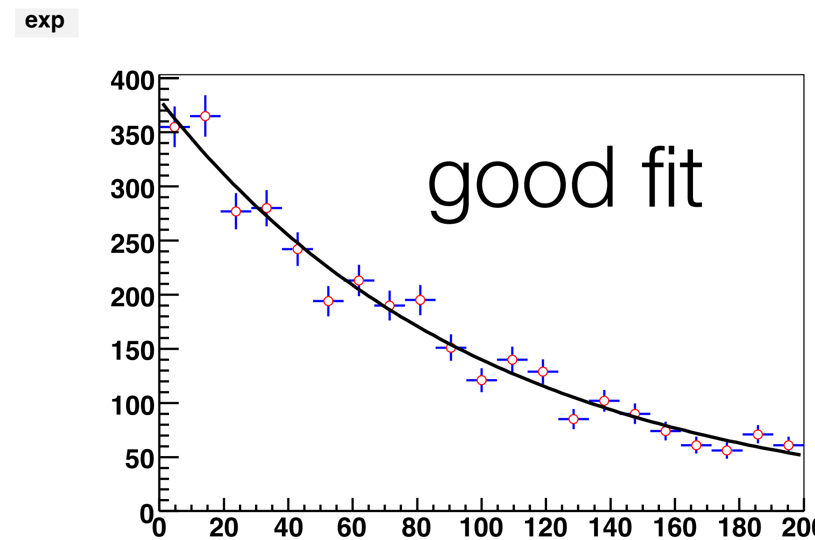
$$\chi^2 \equiv \sum_{\text{all bins}} \frac{\left(n_{\text{meas}}(x_i) - f(x_i)\right)^2}{\sigma^2}$$

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**

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

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

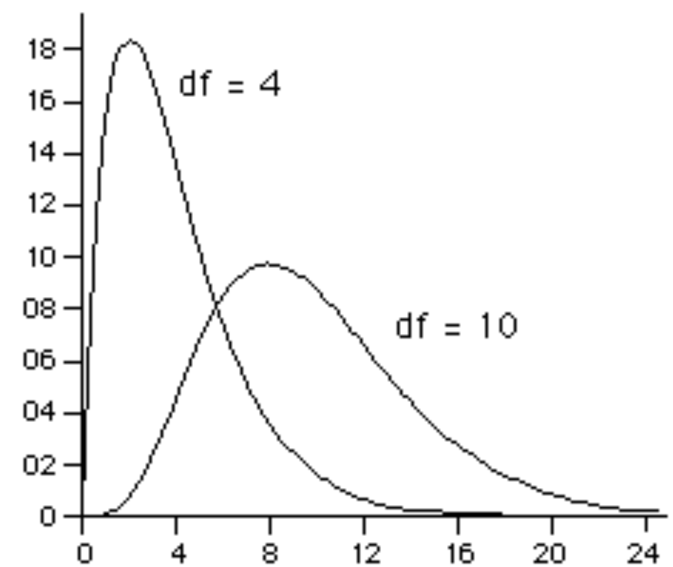
Fit quality as a probability: How likely am I to get a fit that bad or worse if my model is correct?

- **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**

Probabilities, PDFs and likelihood fitting

Skip in TESHEP 2024 lectures
GOTO slide 115.

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?

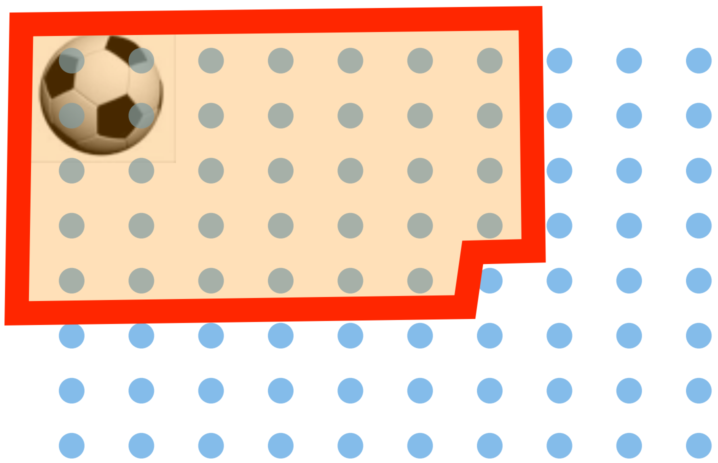
- Bayesian: Probability is a measure for the “degree of belief” that event E happens. One possible definition: I’d bet up to € n_E that E happens, if I get € N if I win: $P(E) = (£\ n_E)/(£\ N)$.
Problem: Subjective (not good for science, but occasionally unavoidable, e.g. for systematics.)

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.**

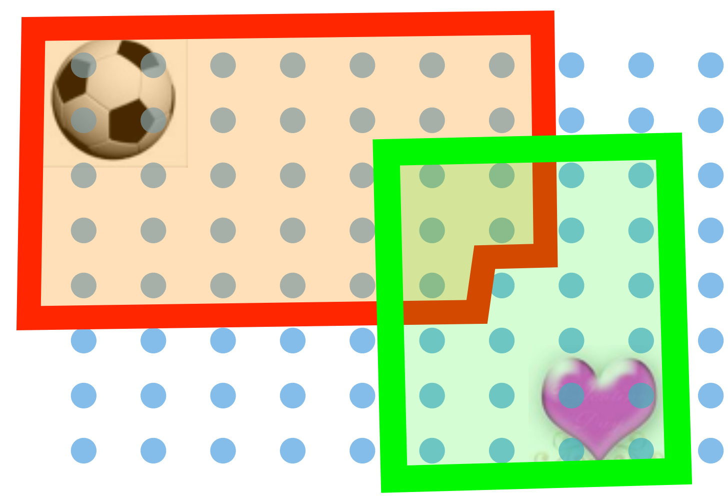


Adding non-exclusive Probabilities

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

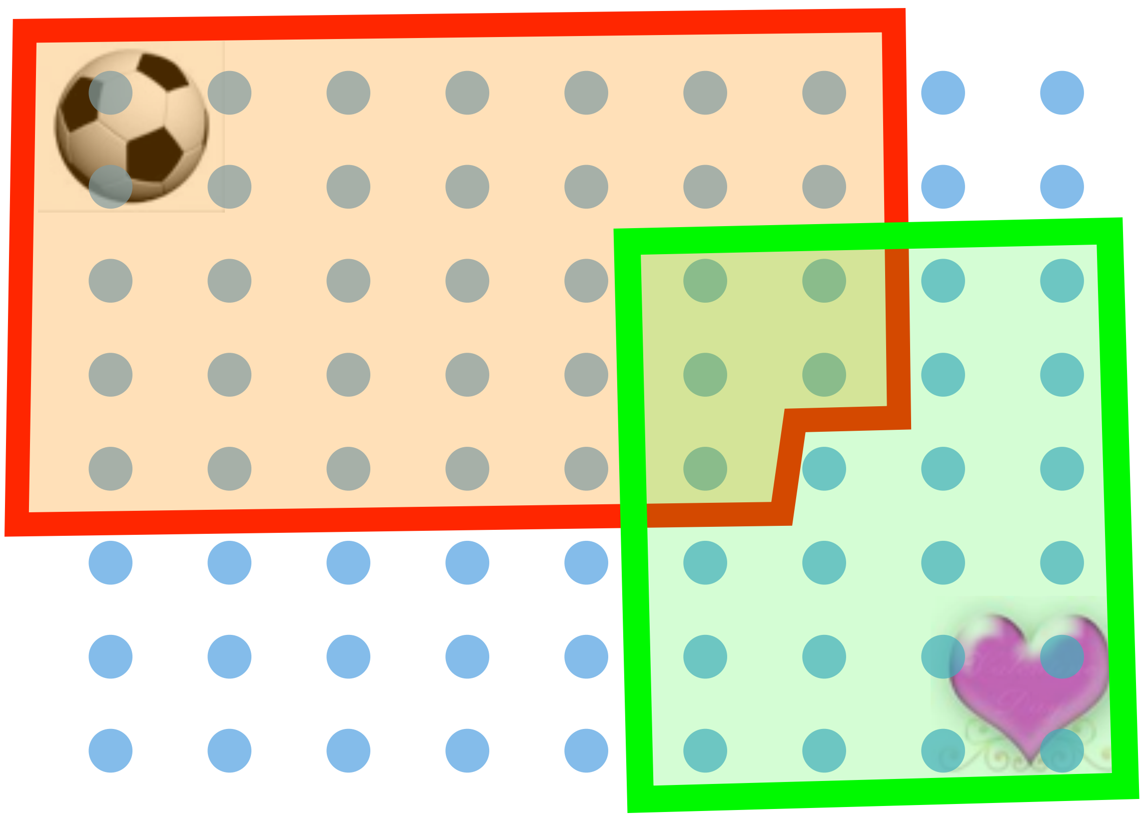
- Not ~~P(A or B) = P(A) + P(B)~~, because we would be double-counting those who like football and the colour pink.

wrong



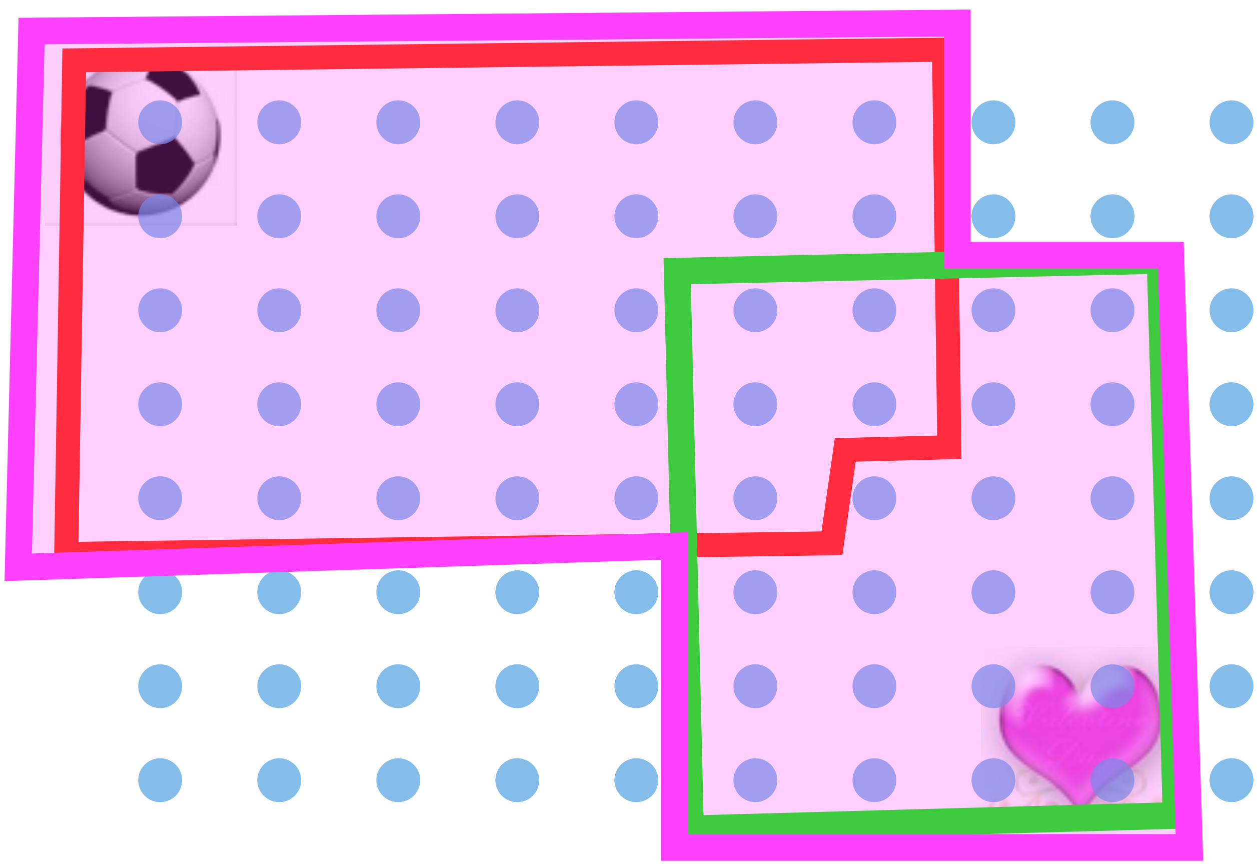
Adding Non-Exclusive Probabilities

- **P(A or B)**



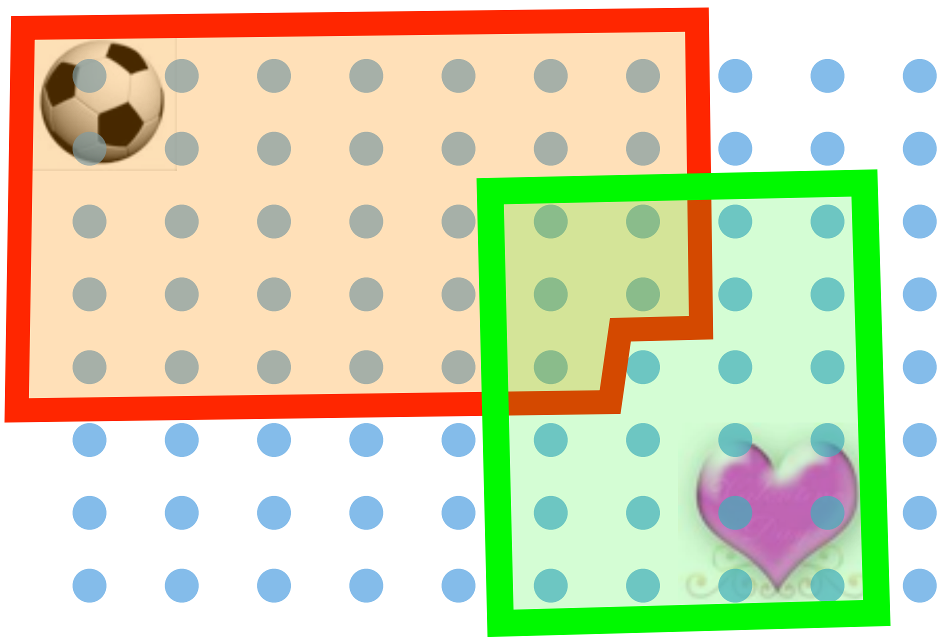
Adding Non-Exclusive Probabilities

- **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

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

- From this follows Bayes' theorem:

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

Bayes' Theorem

- **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)**

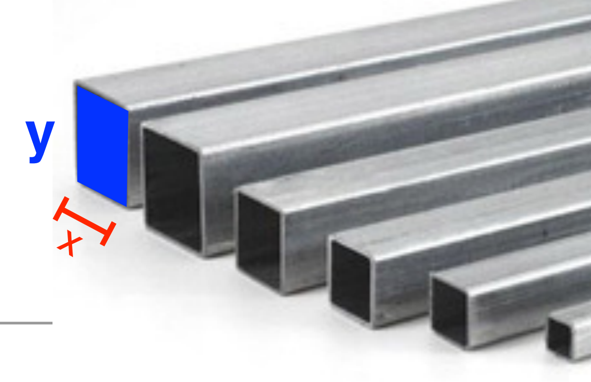
$$\langle x \rangle = \int x' P(x')\,dx' \qquad \langle f(x) \rangle = \int f(x')P(x')\,dx'$$

- **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



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

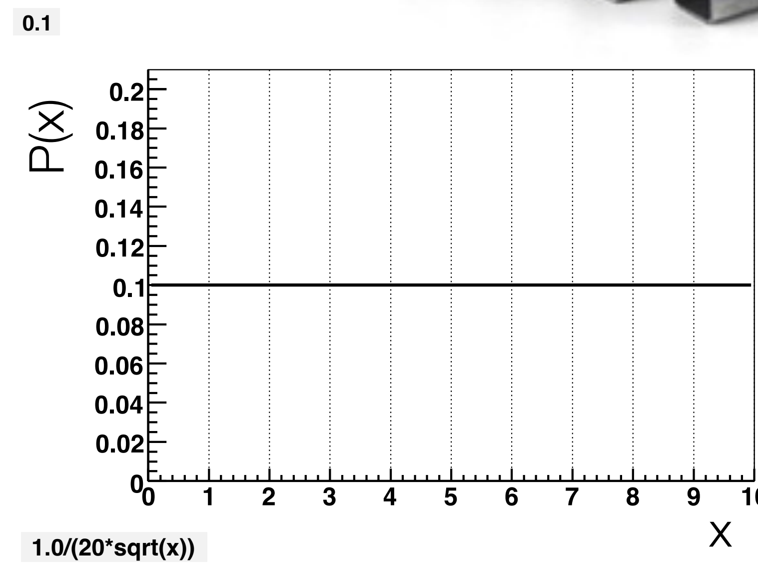
$$y = x^2 \Leftrightarrow x = \sqrt{y} \text{ for } x > 0$$

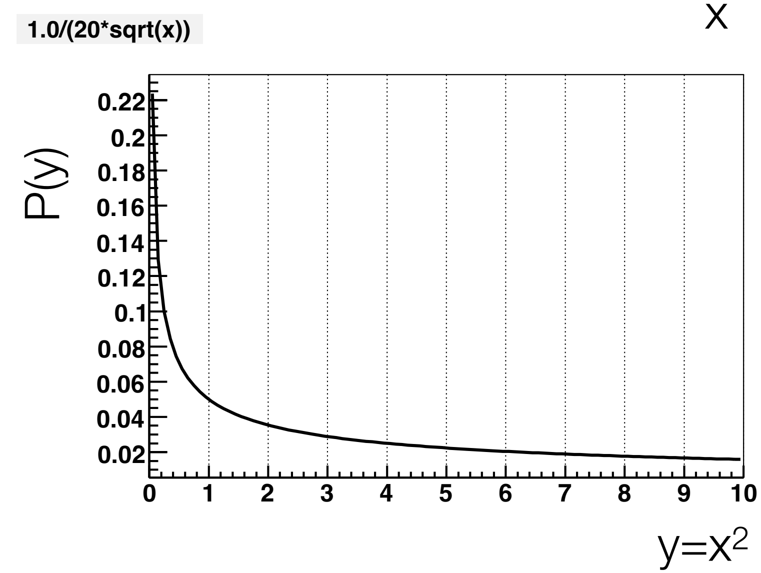
$$P(y)\, dy = P(x)\, dx$$

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

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

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

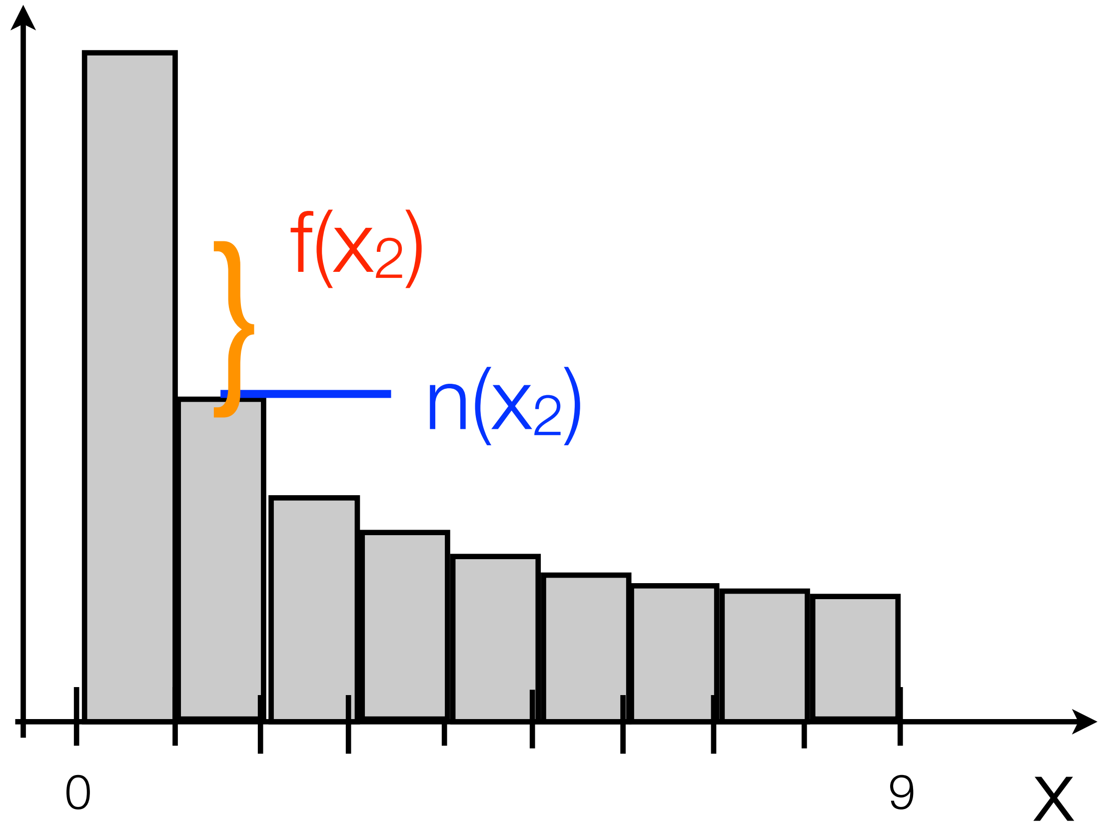




Check out https://tinyurl.com/TeshepVariableTrafo for related python code.

Last time: χ^2 Fitting

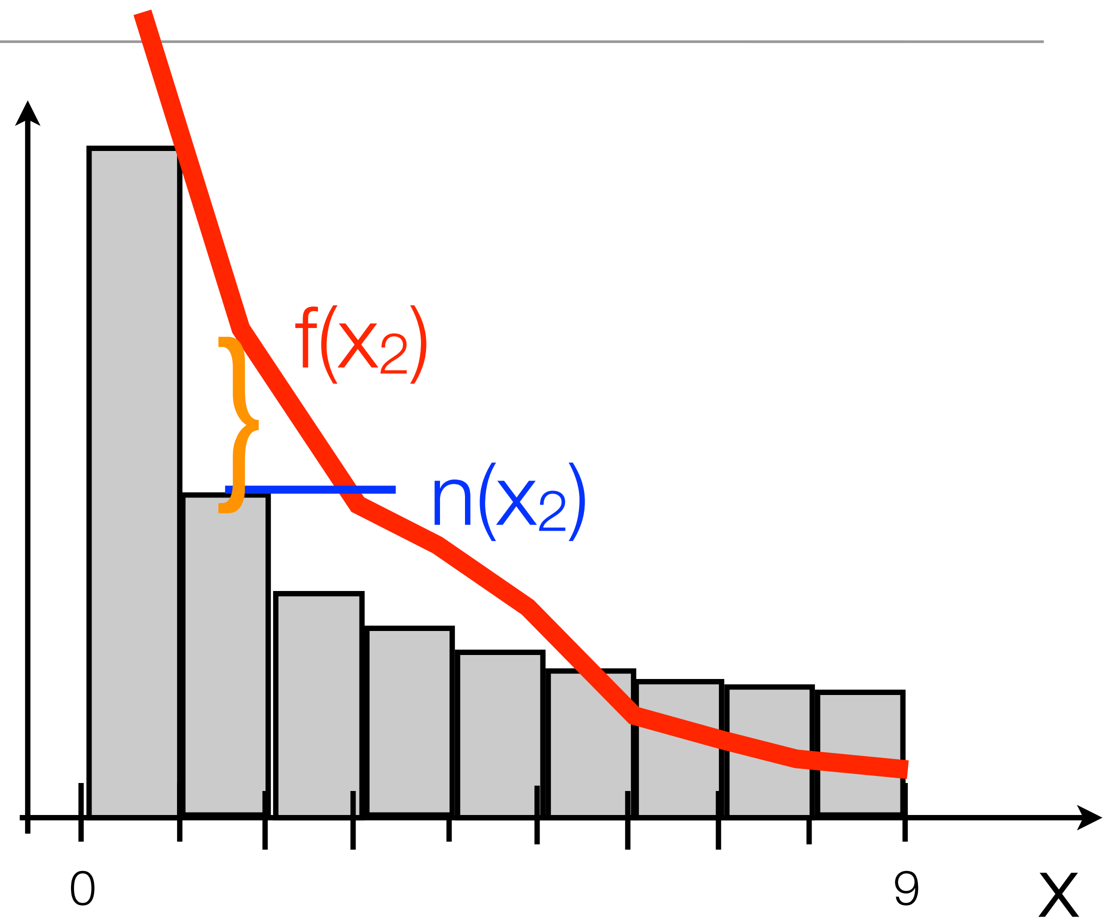
- **Use for binned data**



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

Last time: χ^2 Fitting

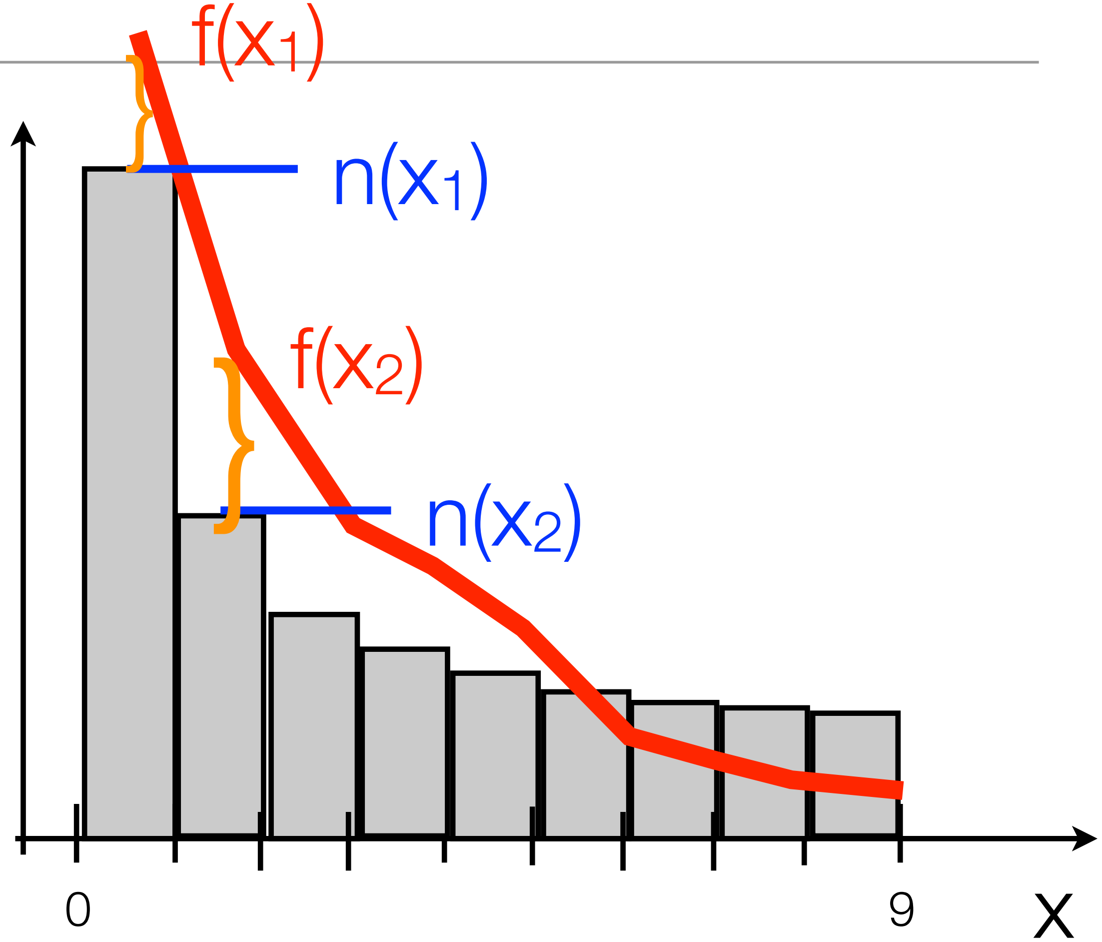
- **Use for binned data**



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

Last time: χ^2 Fitting

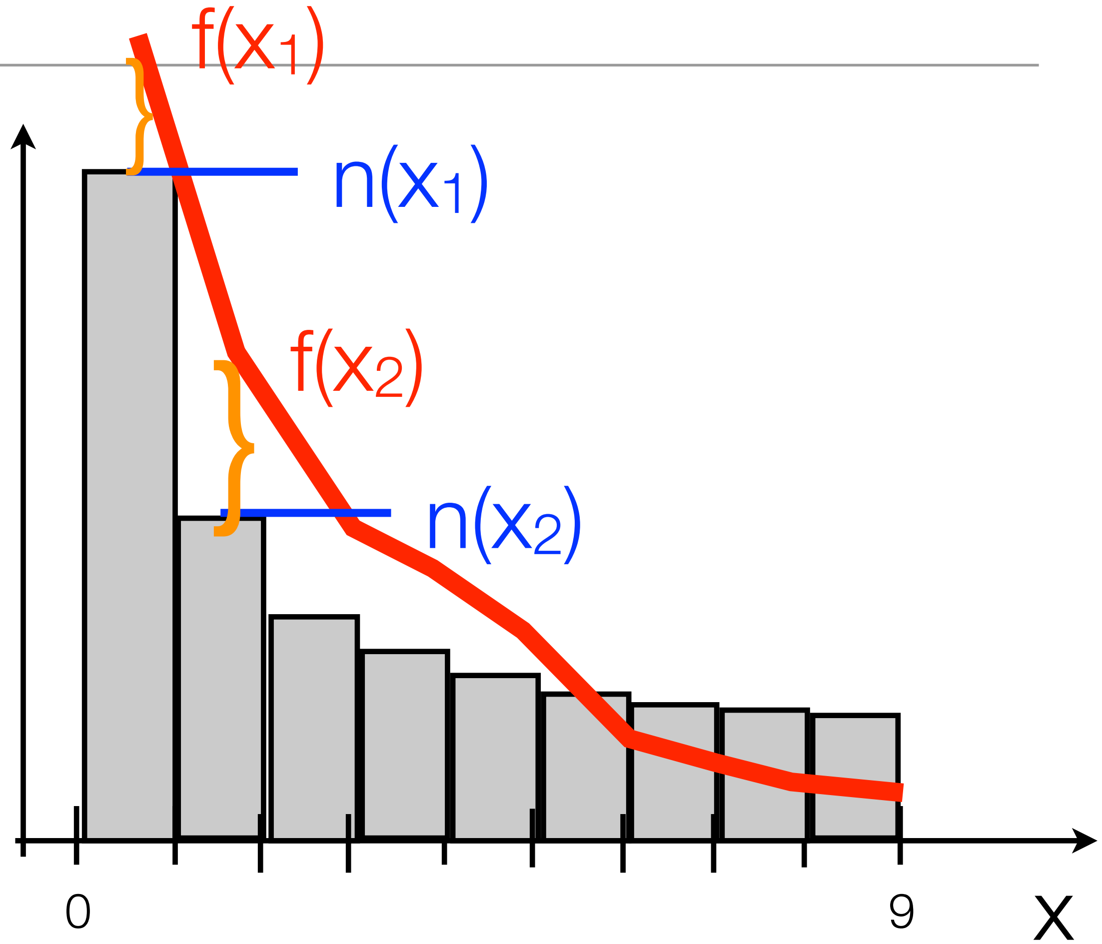
- **Use for binned data**



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

Last time: χ^2 Fitting

- **Use for binned data**

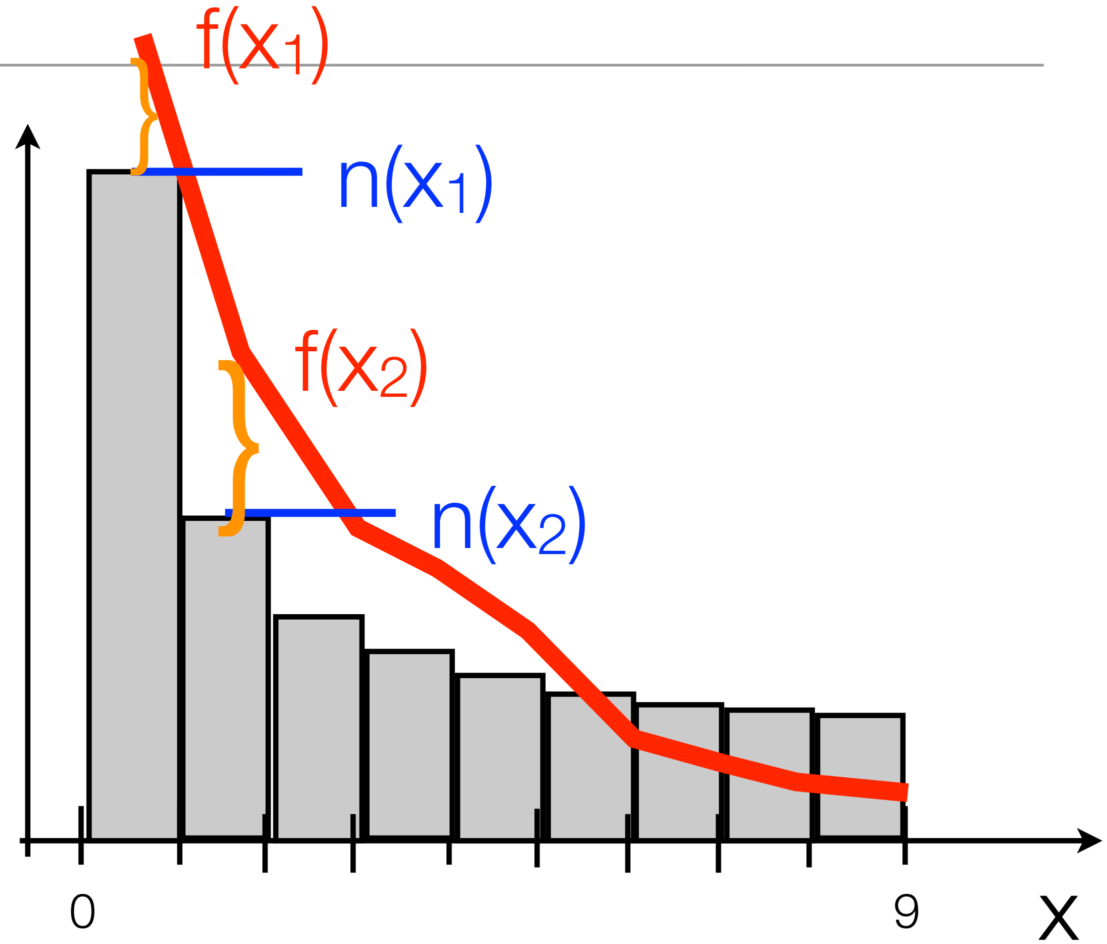


usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

Last time: χ^2 Fitting

- **Use for binned data**

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

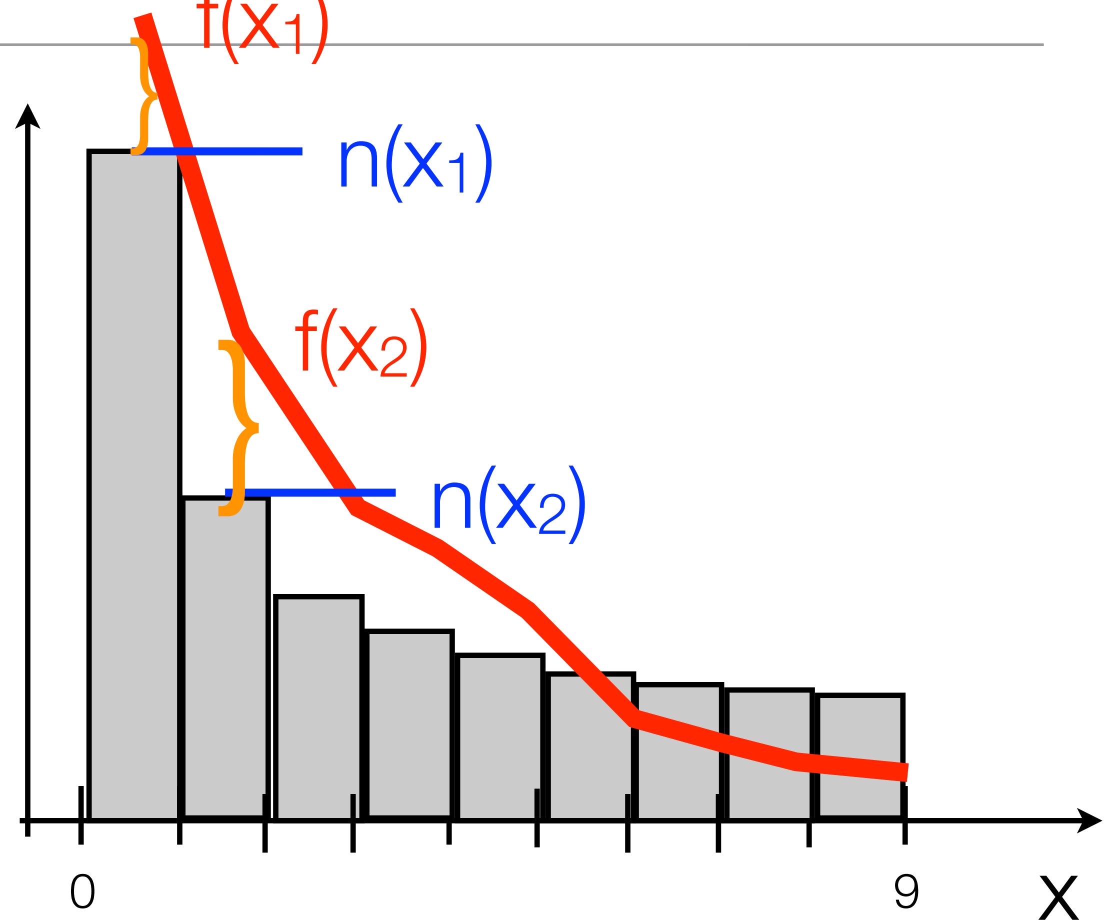


usually $\sigma_i = \sqrt{f(x_i)} \approx \sqrt{n_i}$

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

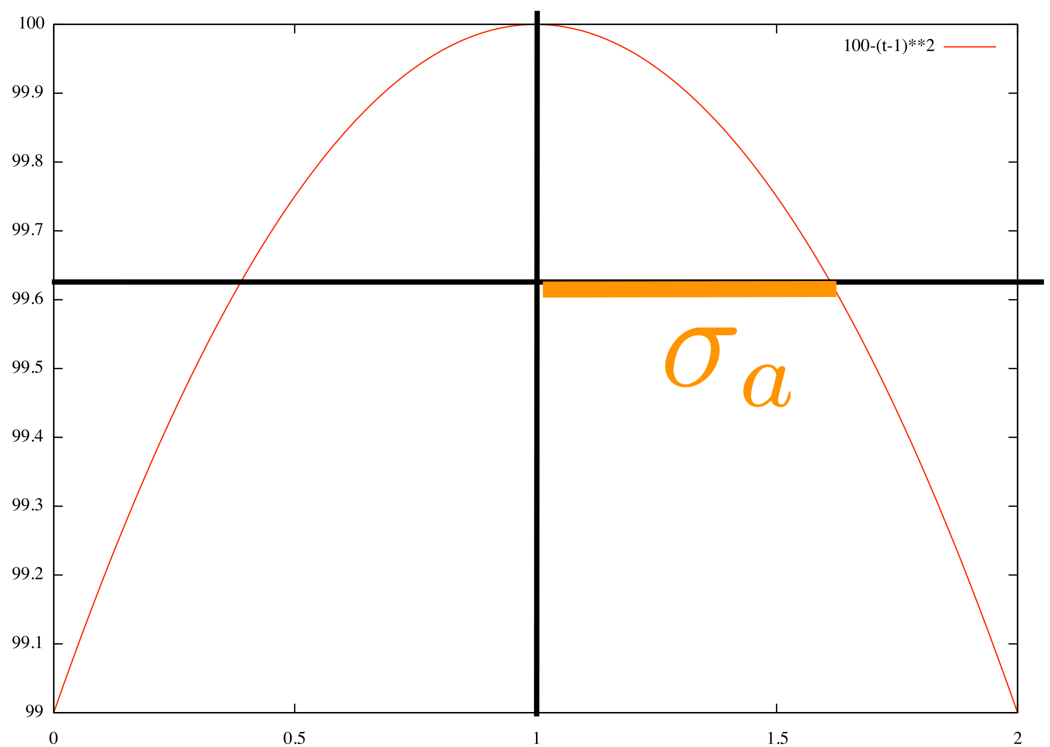
$$P(x) = \frac{f(x)}{\int_{-\infty}^{+\infty} f(x')\, dx'}$$

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

Likelihood Shape

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

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



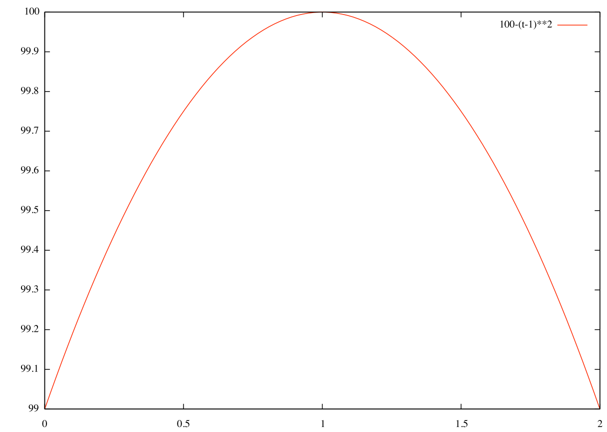
Uncertainty from likelihood "Parabolic Error"

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

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

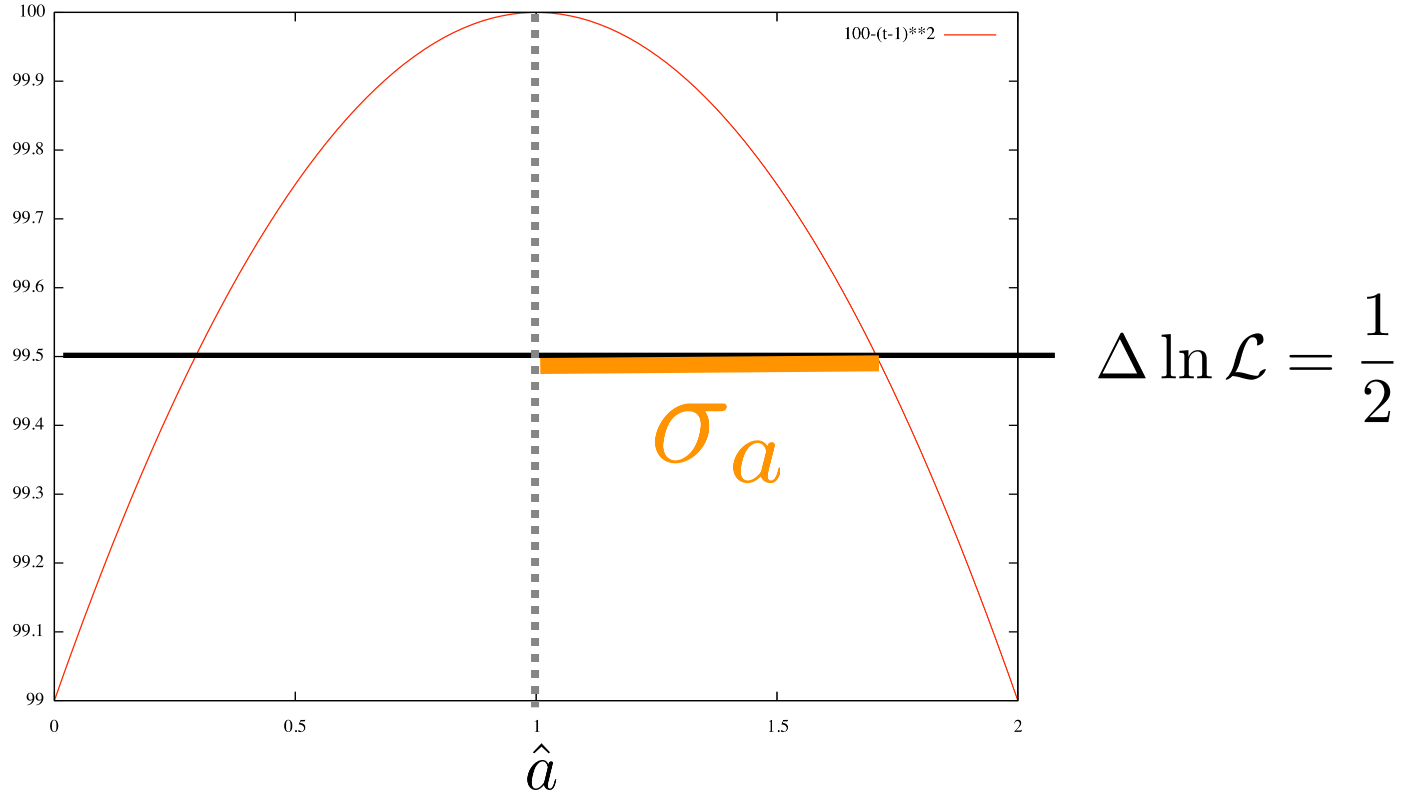
$$\left.\frac{d^2(\ln \mathcal{L})}{d\,a^2}\right|_{\text{at a}=\hat{\text{a}}} = -\frac{1}{\sigma_a^2}$$

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



Error Estimate

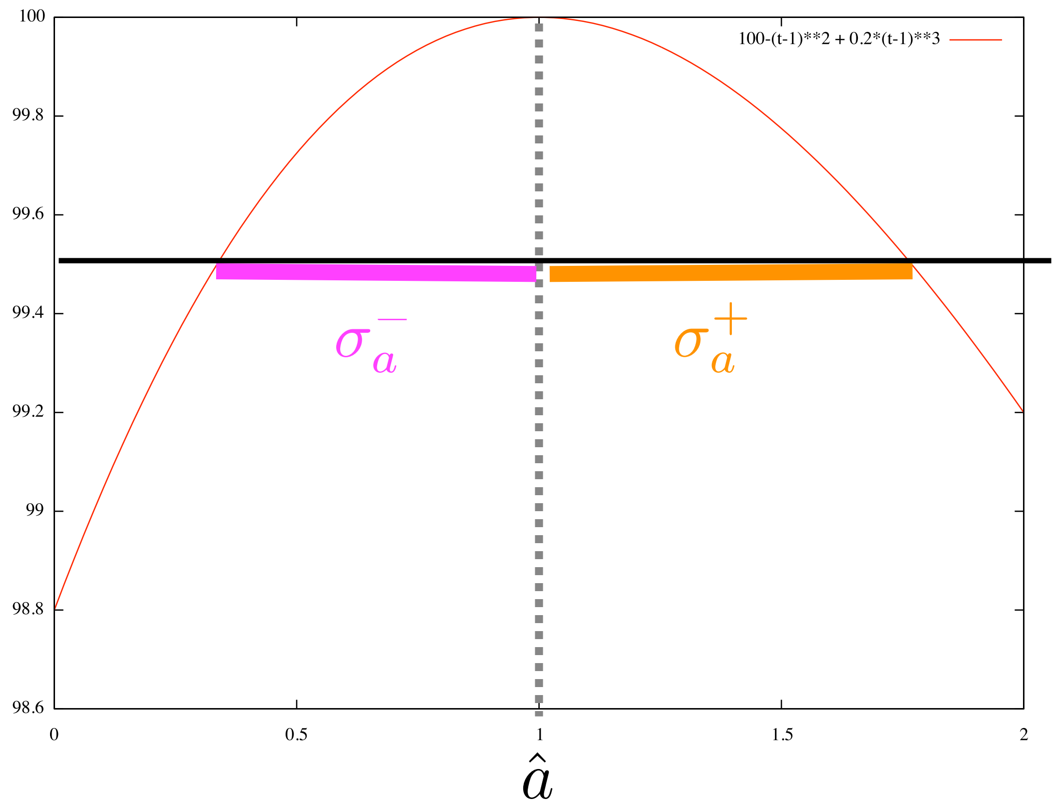
$$\ln \mathcal{L} = -\frac{(a - \hat{a})^2}{2\sigma_a^2} + (\text{meaningless constant})$$



Error Estimate for low N

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

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

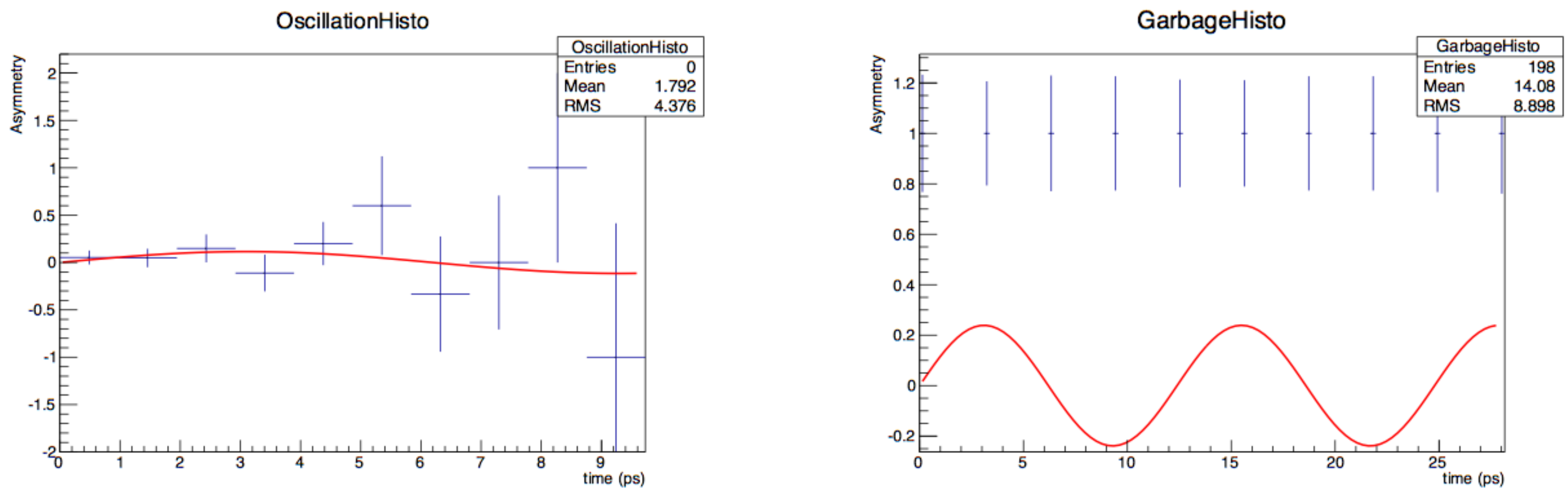


$$\Delta \ln \mathcal{L} = \frac{1}{2}$$

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

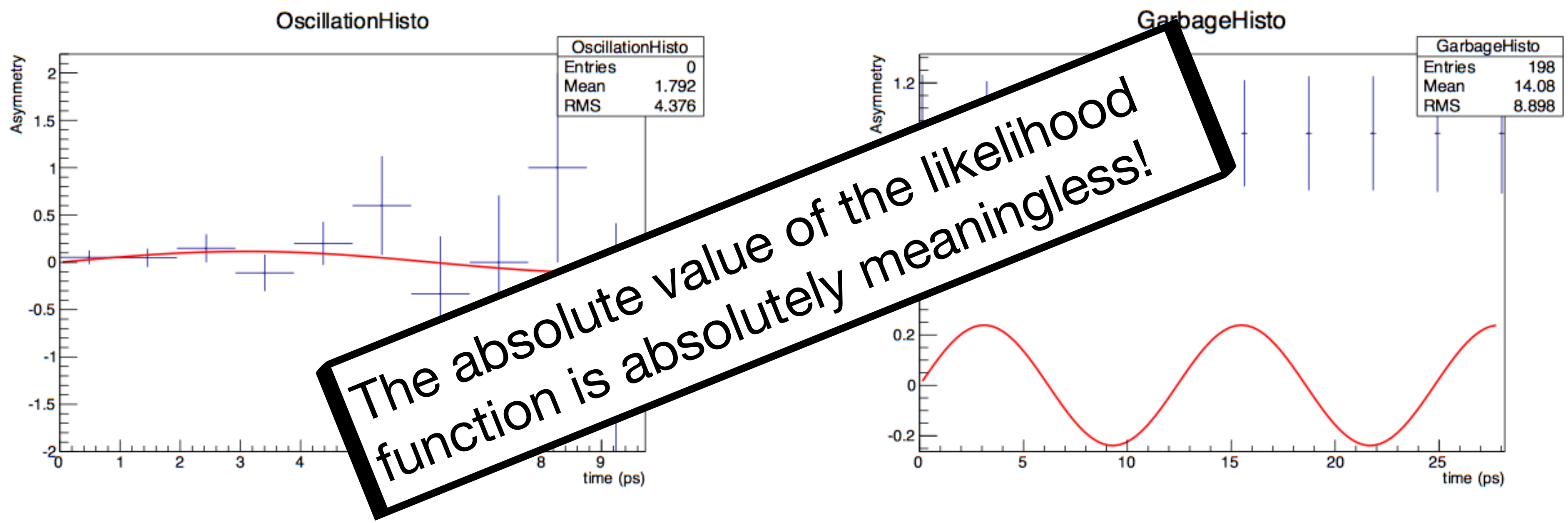


- **One solution: After doing an un-binned likelihood fit, bin the data and calculate the χ^2 between data and fit.**

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



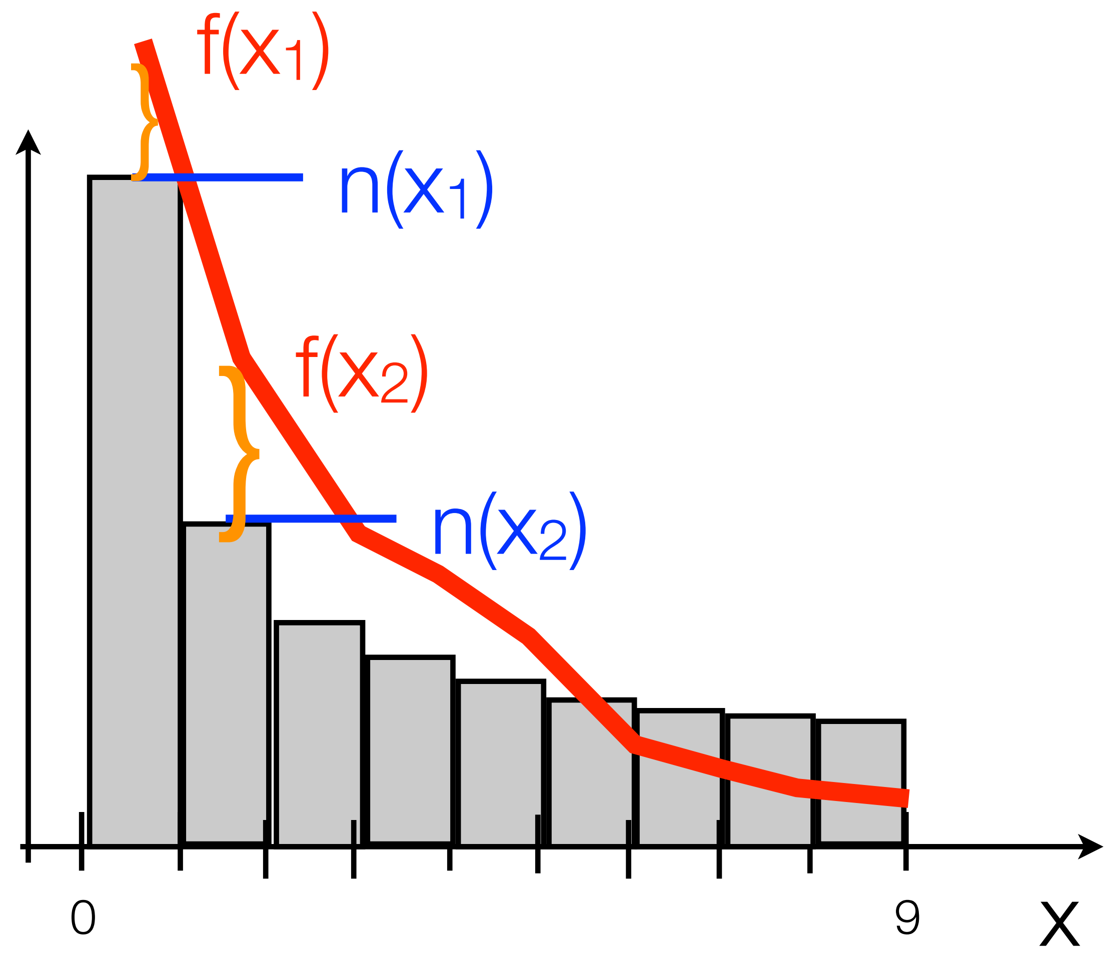
- **One solution: After doing an un-binned likelihood fit, bin the data and calculate the χ^2 between data and fit.**

χ^2 Fitting and likelihood.

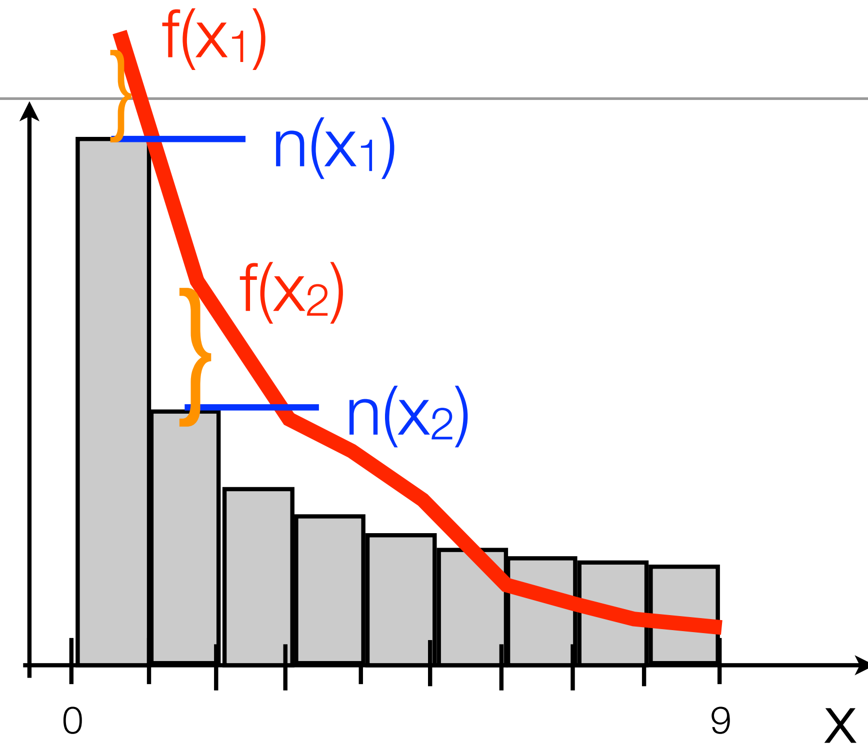
- **Let's do a binned likelihood fit. Our model predicts *f*(x1) events for bin centred at x1.**

- **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**

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

Gaussian that inherits from Poisson with $\lambda \equiv f(x_i) = \mu_i = \sigma_i^2$

- **log-likelihood**

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

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

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**

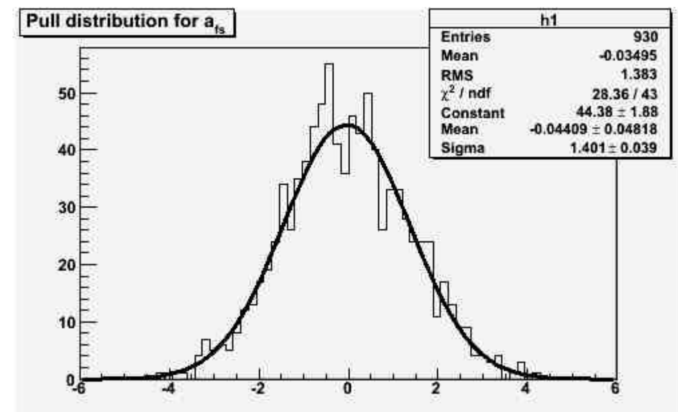
pull = $$\frac{(\text{fit result}) - (\text{true value})}{(\text{error estimate})}$$

and put it in a histogram.

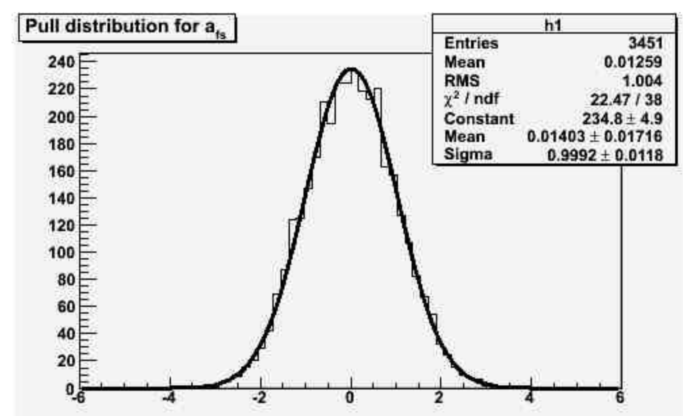
- **For a good, unbiased fitter, you get:**

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

σ=1.4 for 1k events ⇒ wrong errors



σ=1.0 for 1k events ⇒ correct errors



Monte Carlo

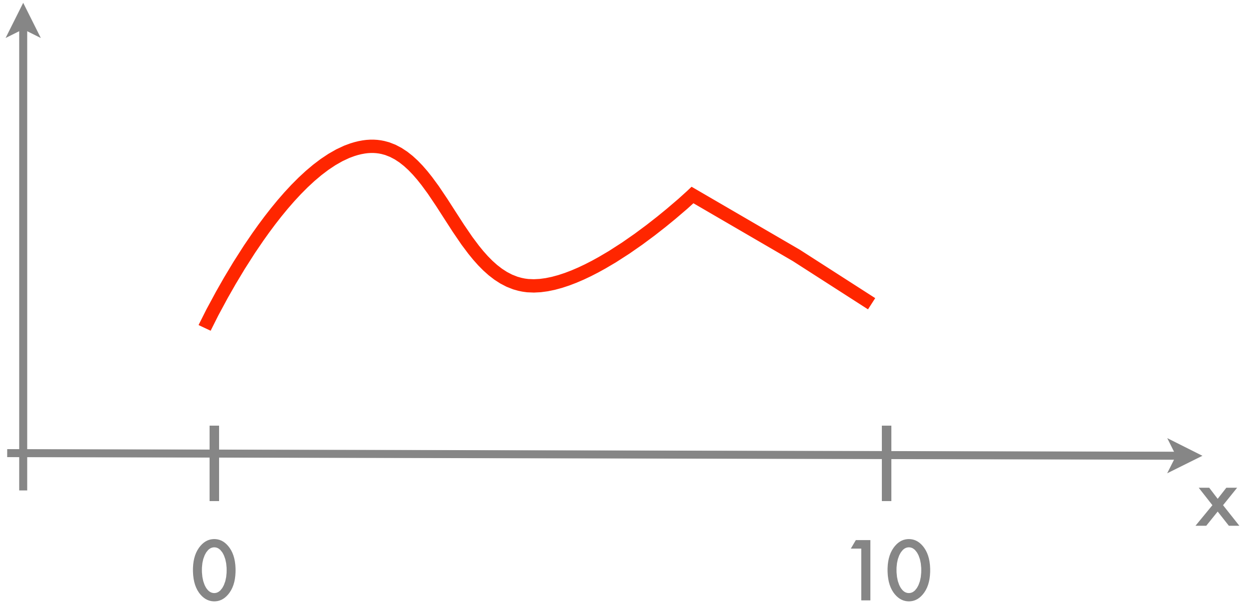
Gamblers in the casino at Monte-Carlo. c.1910

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

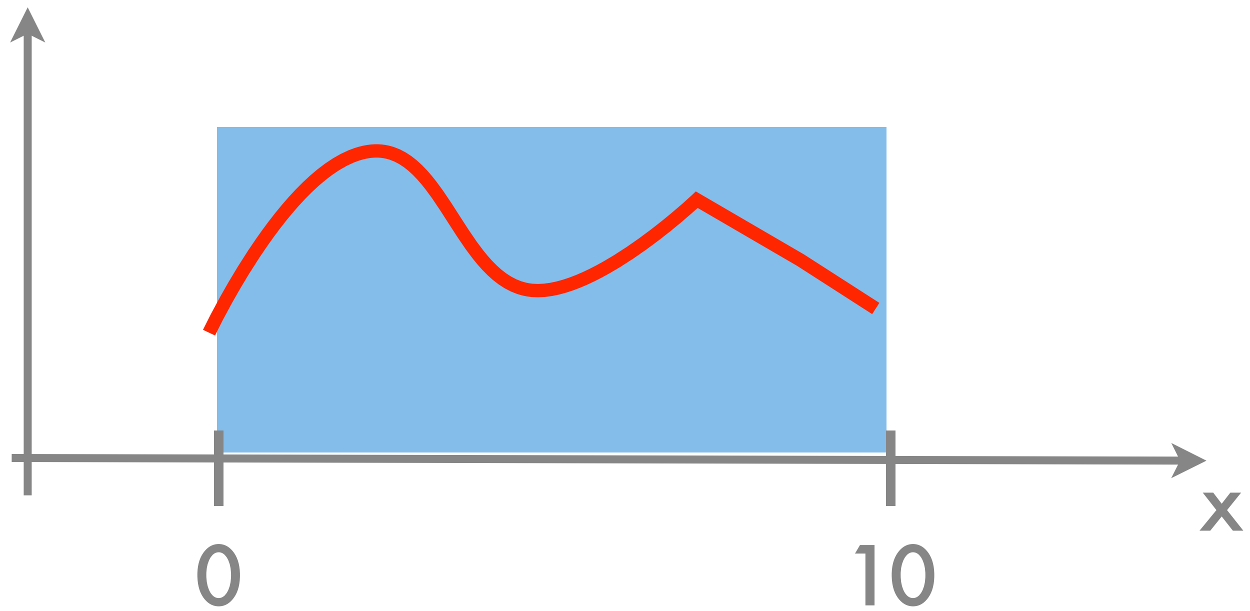
Von Neumann Accept-Reject

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



Von Neumann Accept-Reject

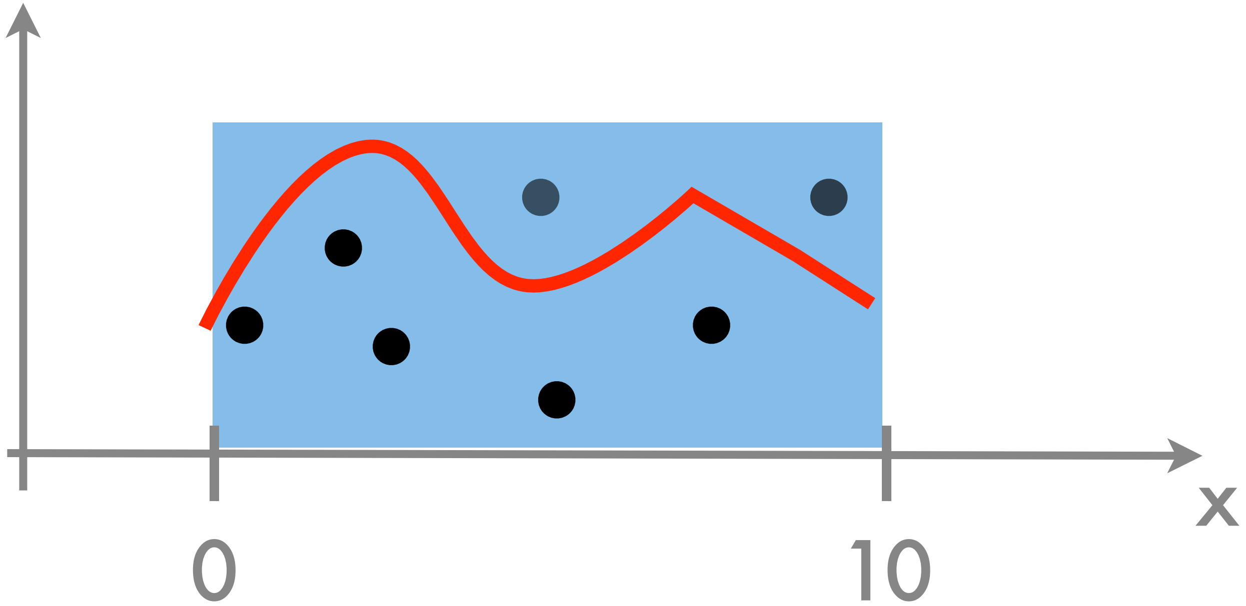
- Aim: Generate f(x) between 0 and 10



- 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).

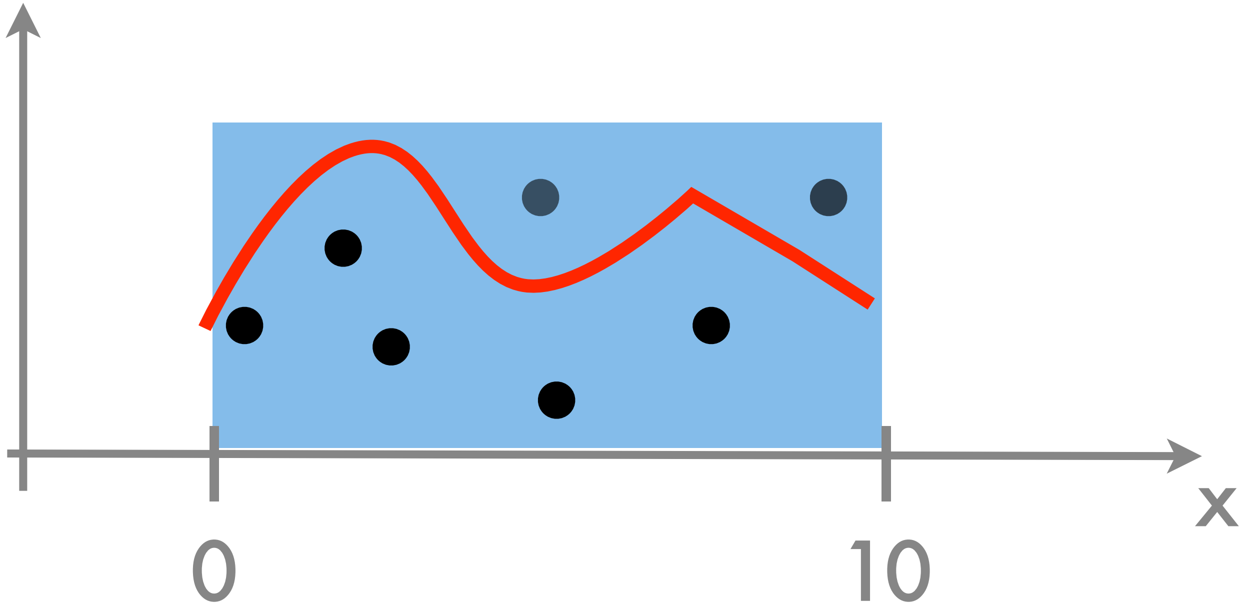
Von Neumann Accept-Reject

- 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

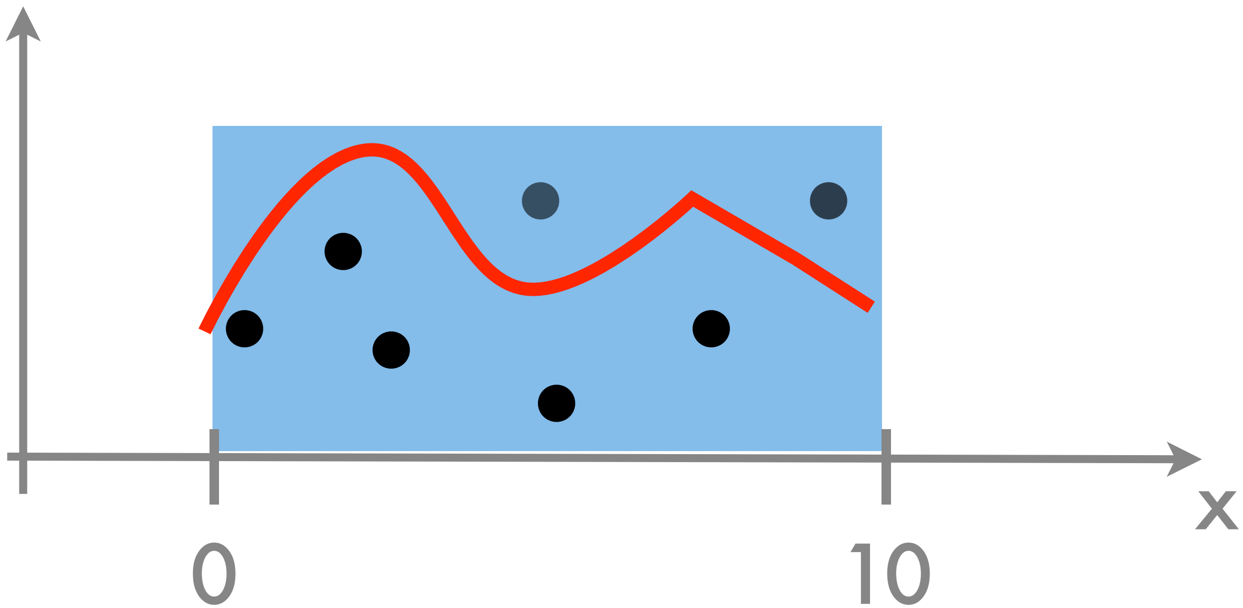


- x = rnd->Rndm() · 10;

  y = rnd->Rndm() · fmax;

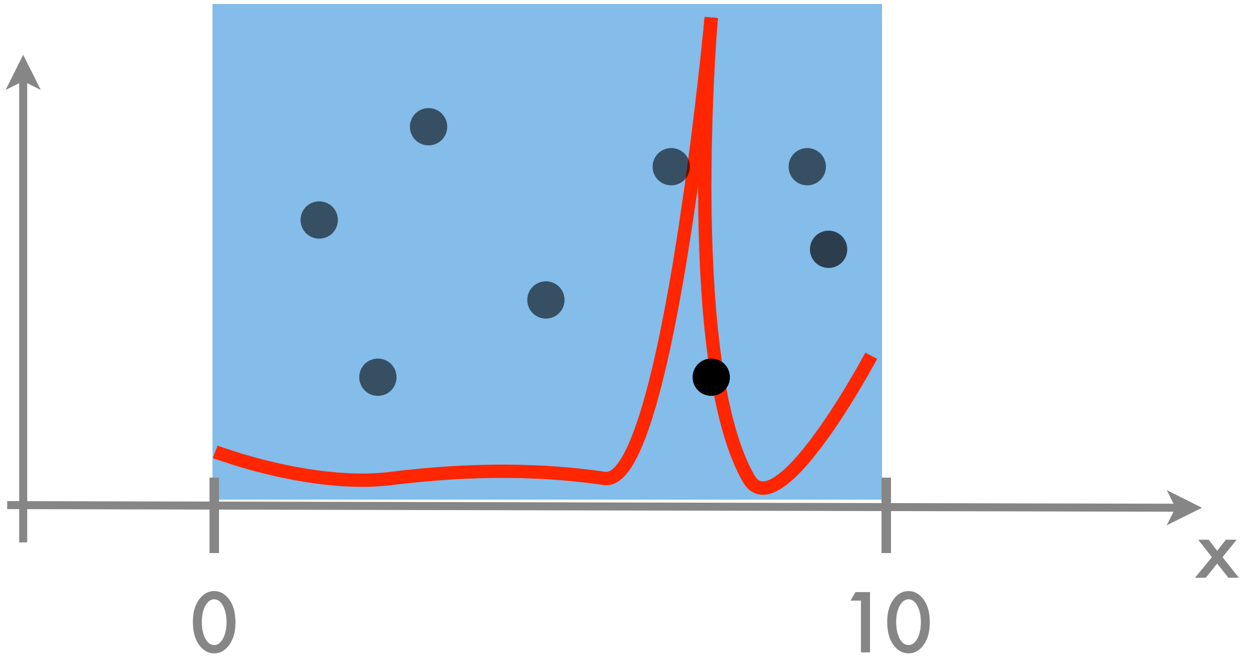
  if(y < f(x)) acceptEvent(x,y)

MC-integration

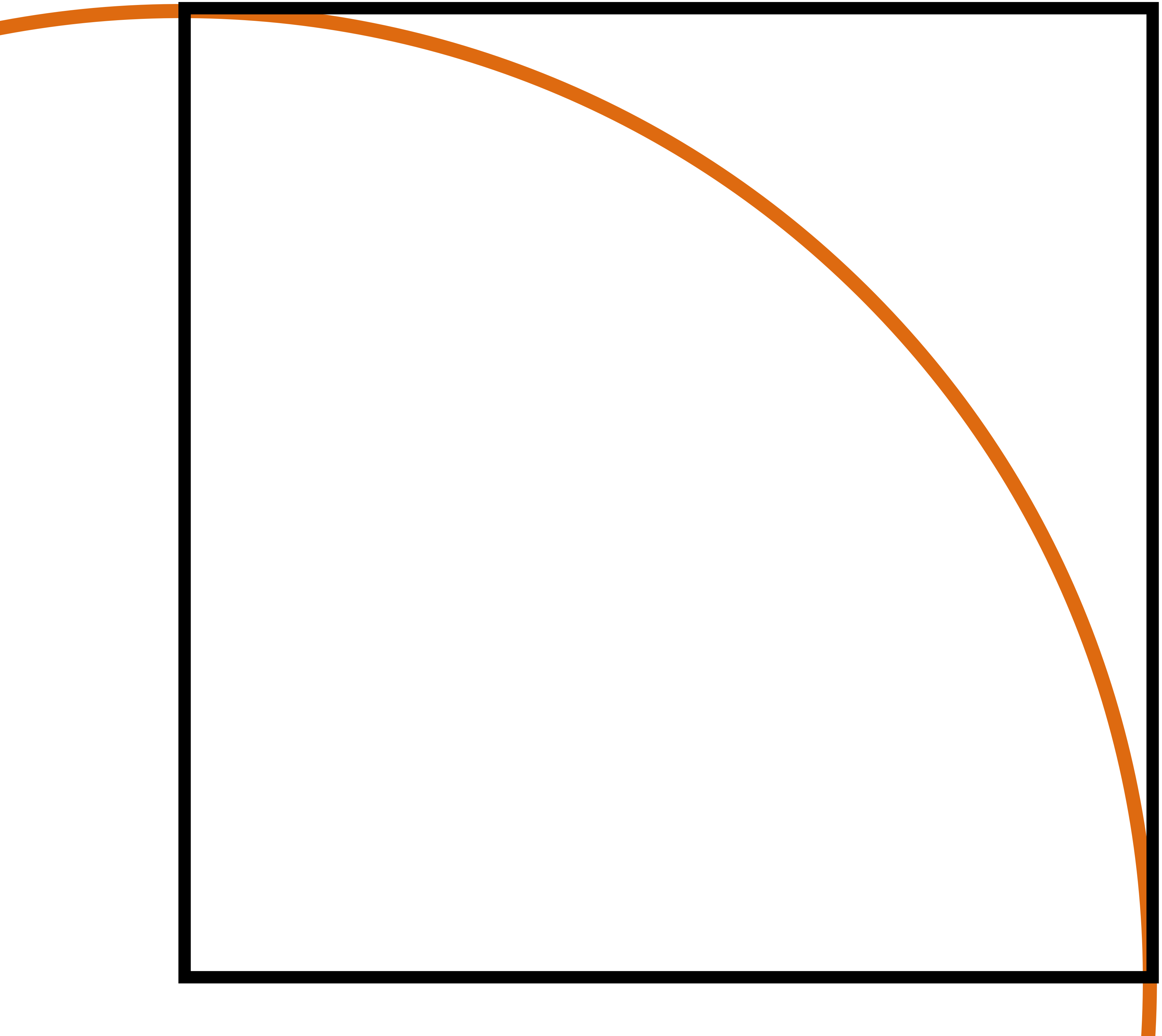


- This can be used for MC integration - the fraction of points accepted is $\propto$ to the area under the curve.
- This is the most efficient method of numerical integration in many dimensions (say more than 3).

Von Neumann Accept-Reject



- Can be very inefficient for peaky distributions



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
Jupyter Workbook for Chi2 fit à la TESHEP	https://tinyurl.com/TeshepFit
Solutions:	https://tinyurl.com/TeshepFitSolved

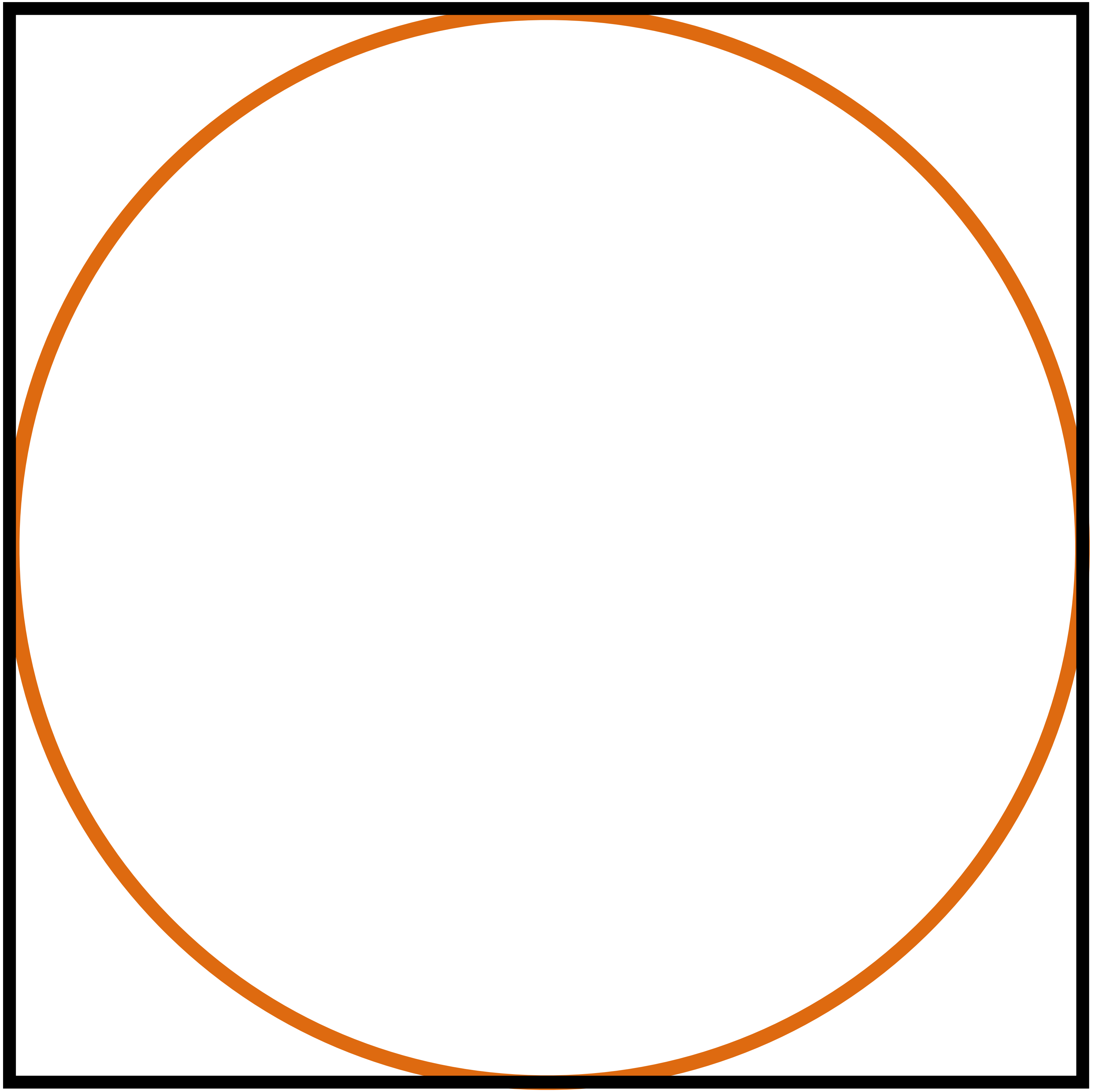
Additional Jupyter notebooks to play around with:

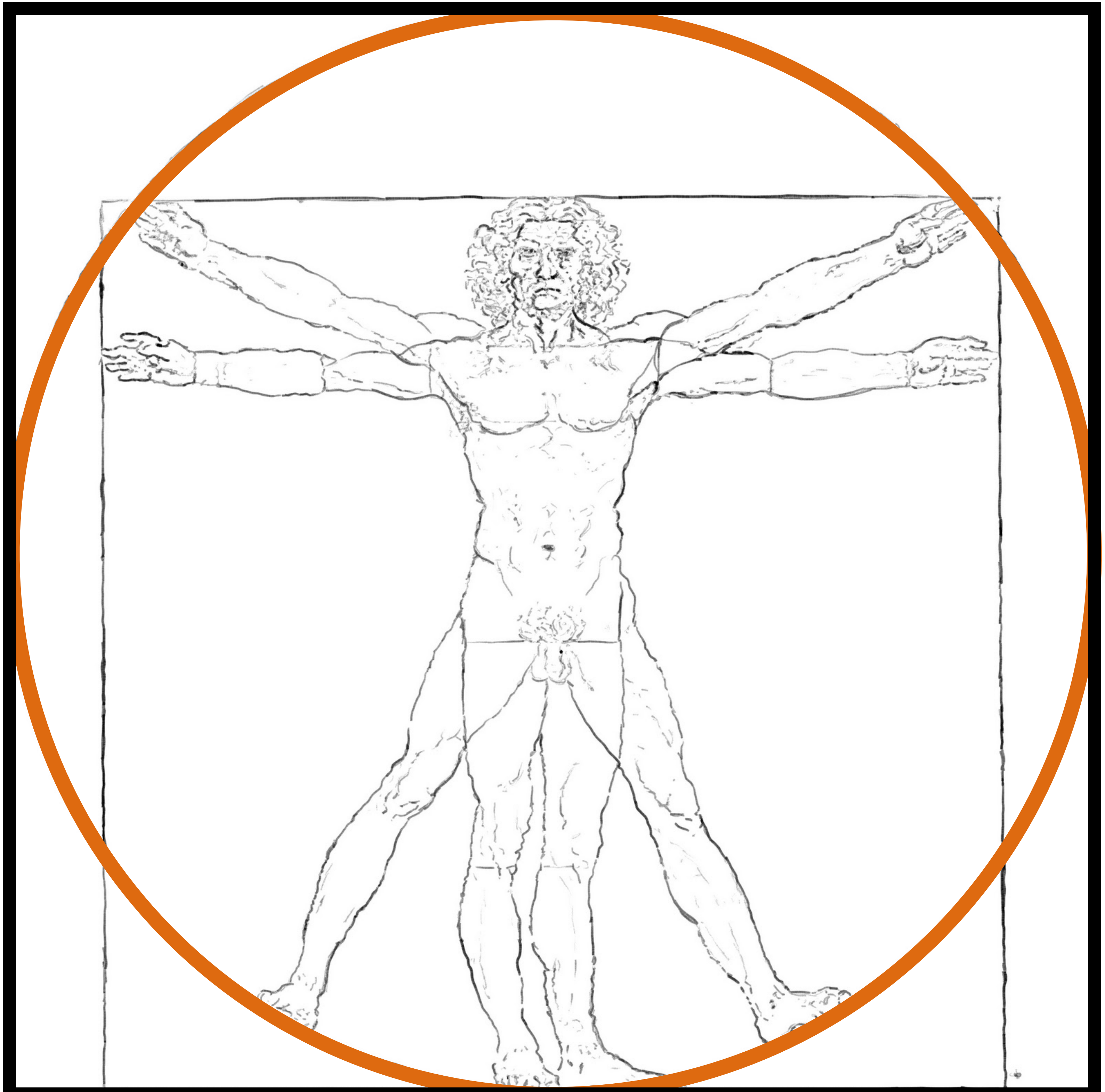
https://tinyurl.com/TeshepStatCode

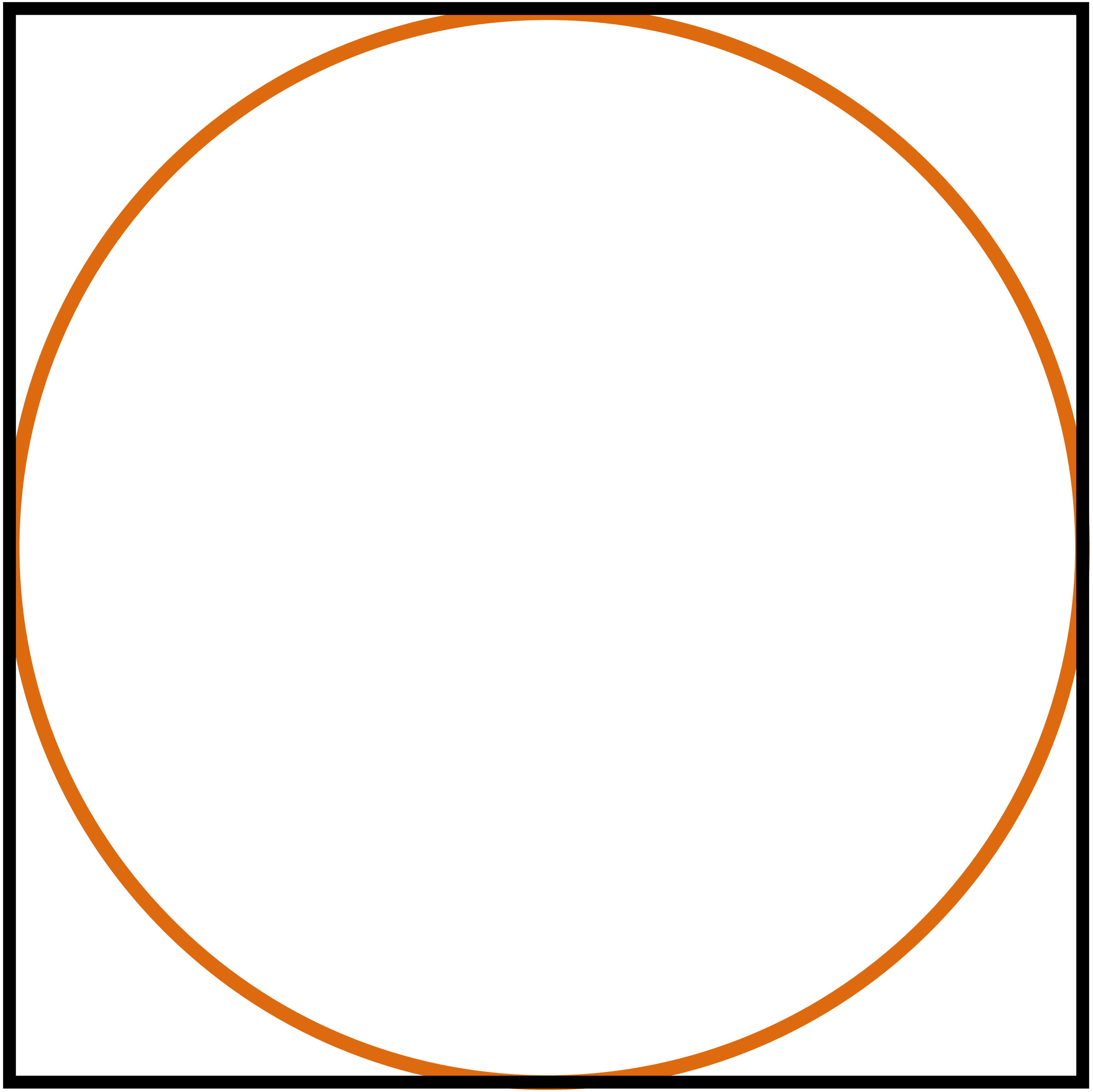
Links for installing jupyter and anaconda:

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

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







The End