Introduction to Data Analysis

R D Parsons, (+ Orel Gueta, Jakob Nordin)

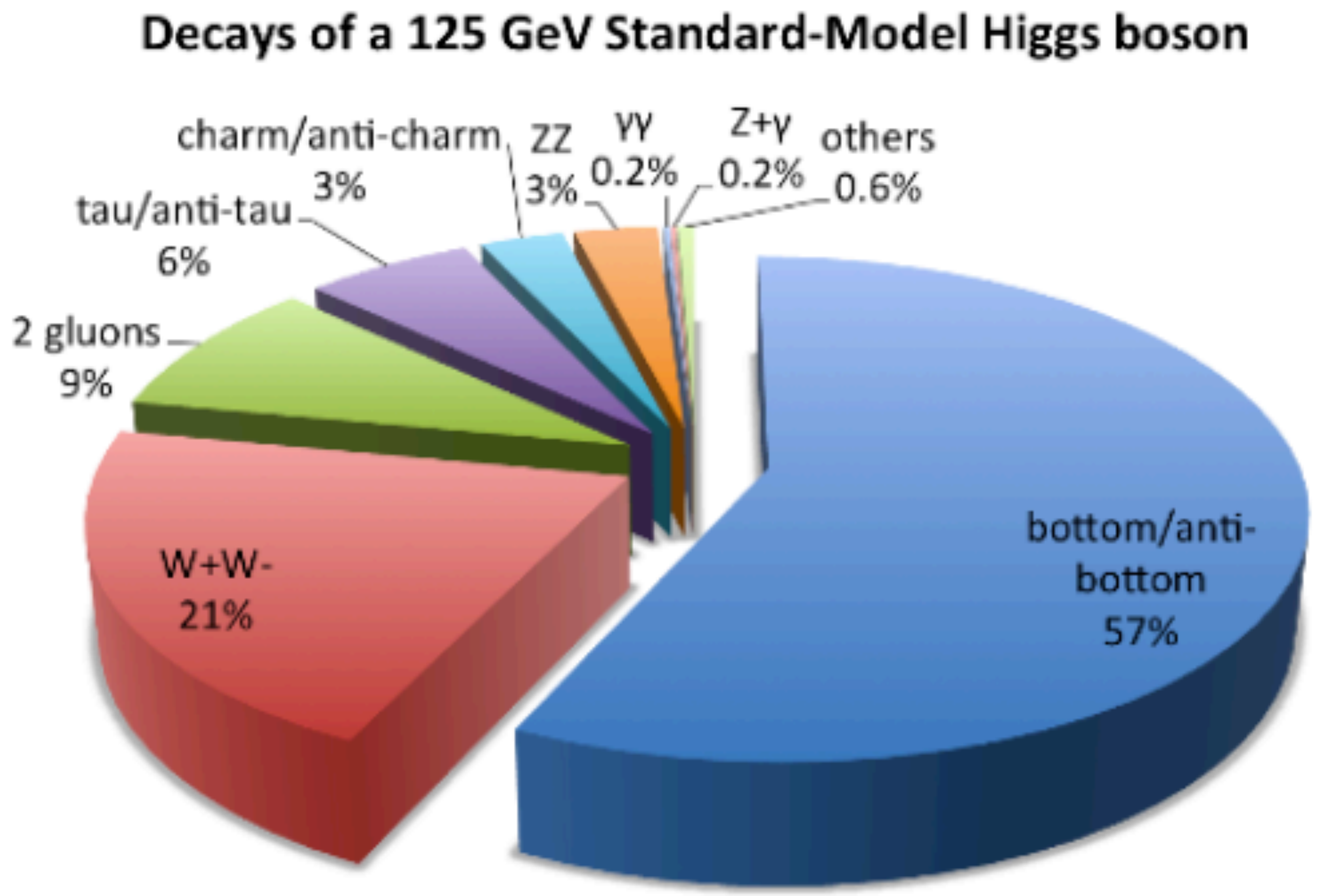
Introduction

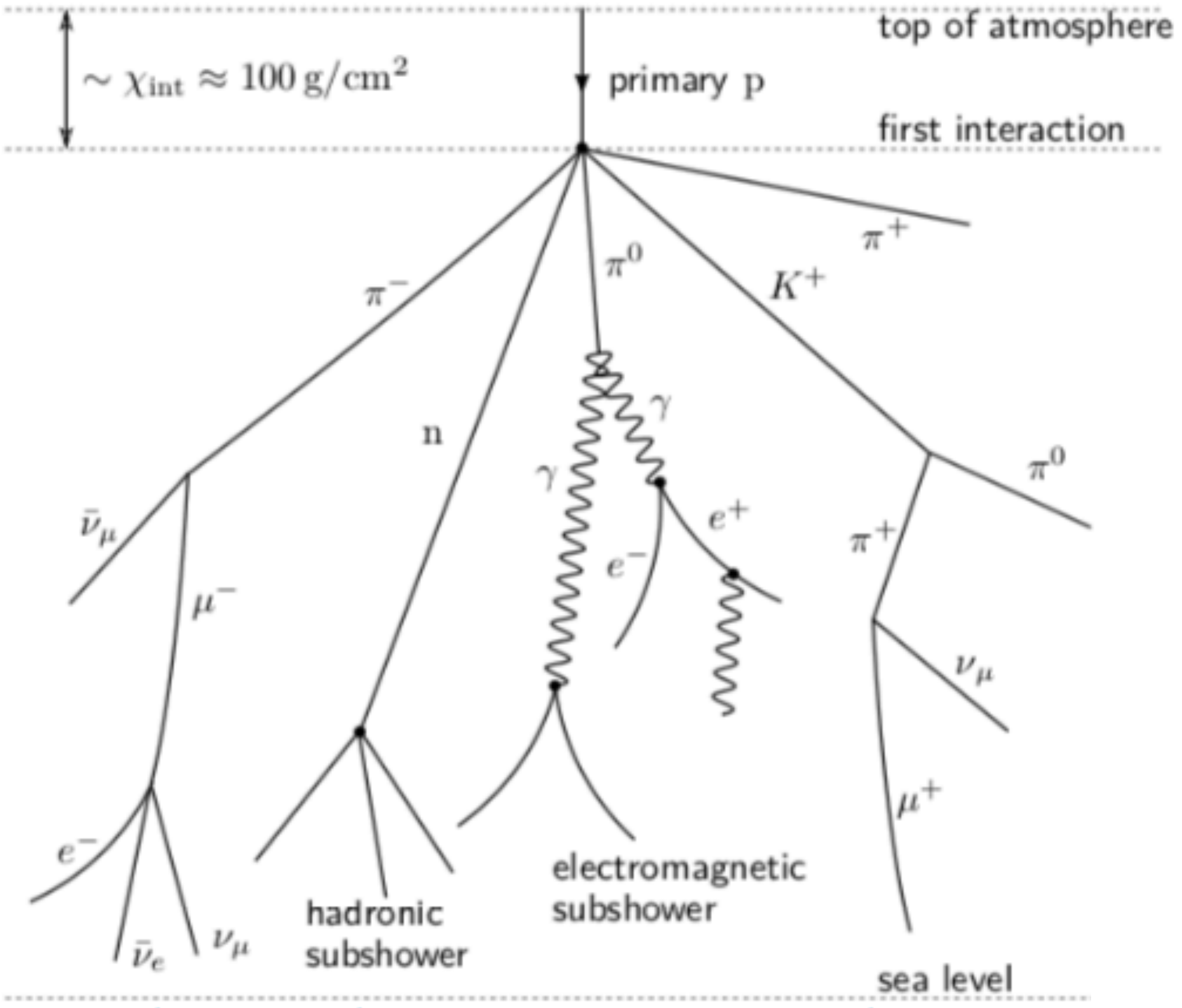
When analysing data randomness appears everywhere

Quantum mechanics, particle interactions etc

We only know probabilities of occurance

Need to use statistics to analyse large volumes of data

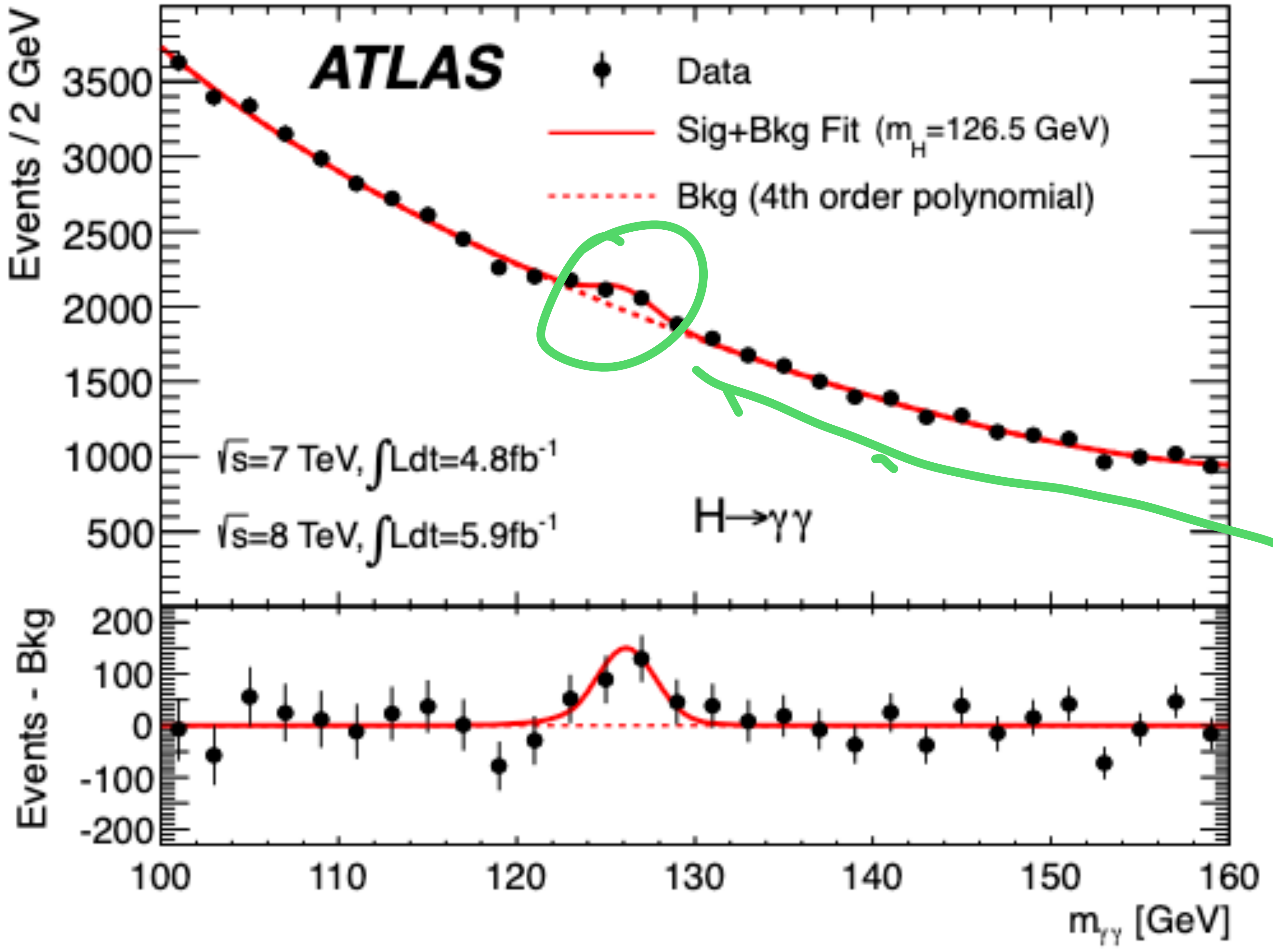


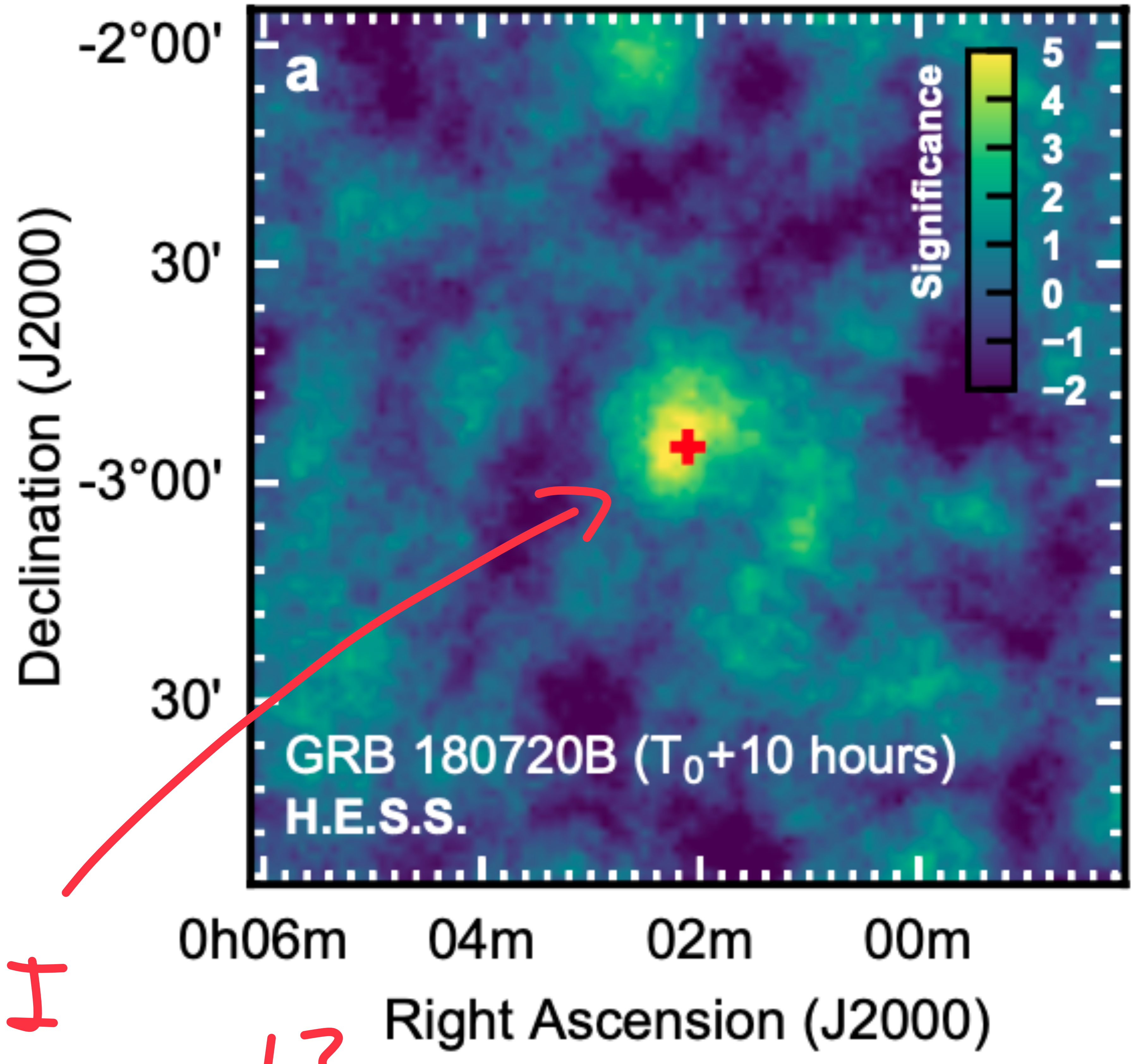


Introduction

Most analysis involves picking out rare "signal" from a stronger background

Due to randomness in my data BG can always fall into my signal region





How many photons before I can trust this is real?

Wiggle in 2 data points, is that a detection?

The Cheating Baker

A restaurant owner orders 30 rolls every day.

The law in the country states that rolls must weigh **75 grams**

After changing suppliers, the owner suspects that the new baker sells underweight rolls

⇒ Investigate! **Weigh the rolls** (1 gram resolution).

```
78, 66, 67, 64, 74, 58, 78, 66, 71, 68, 77, 59, 68, 68, 75, 64, 69, 65, 70, 72
64, 75, 74, 72, 74, 66, 69, 65, 68, 72, 66, 68, 66, 65, 66, 69, 64, 71, 78, 73
69, 65, 66, 78, 70, 66, 70, 80, 70, 73, 71, 68, 64, 68, 68, 72, 74, 74, 71, 74
66, 76, 72, 68, 72, 69, 75, 77, 80, 63, 62, 67, 70, 74, 71, 59, 68, 74, 71, 73
68, 68, 70, 72, 70, 70, 66, 71, 70, 75, 75, 70, 68, 66, 72, 70, 68, 70, 66, 73
67, 76, 72, 72, 64, 70, 73, 65, 68, 70, 63, 71, 74, 65, 71, 63, 69, 61, 75, 72
69, 66, 76, 79, 60, 76, 78, 71, 64, 74, 69, 66, 63, 72, 73, 66, 72, 64, 74, 70
69, 69, 74, 73, 72, 70, 70, 73, 71, 73, 68, 57, 75, 80, 72, 69, 69, 69, 70, 64
67, 65, 71, 68, 72, 69, 74, 71, 80, 60, 66, 74, 82, 68, 70, 68, 76, 68, 73, 68
63, 71, 72, 76, 69, 66, 72, 71, 71, 69, 75, 71, 79, 75, 73, 61, 73, 72, 74, 75
67, 74, 67, 79, 63, 61, 61, 65, 64, 79, 68, 63, 75, 67, 63, 74, 73, 67, 77, 70
78, 72, 77, 70, 63, 82, 67, 69, 66, 68, 71, 71, 73, 78, 69, 63, 64, 66, 61, 74
67, 68, 69, 63, 65, 73, 73, 67, 79, 63, 68, 68, 56, 79, 71, 64, 80, 72, 72, 66
70, 62, 73, 68, 70, 76, 71, 71, 71, 66, 74, 77, 73, 74, 65, 65, 62, 76, 68, 76
66, 67, 70, 74, 70, 71, 70, 70, 64, 70, 69, 69, 72, 66, 69, 68, 72, 73, 65, 72
```

The Cheating Baker

Try combining the data by binning with a resolution

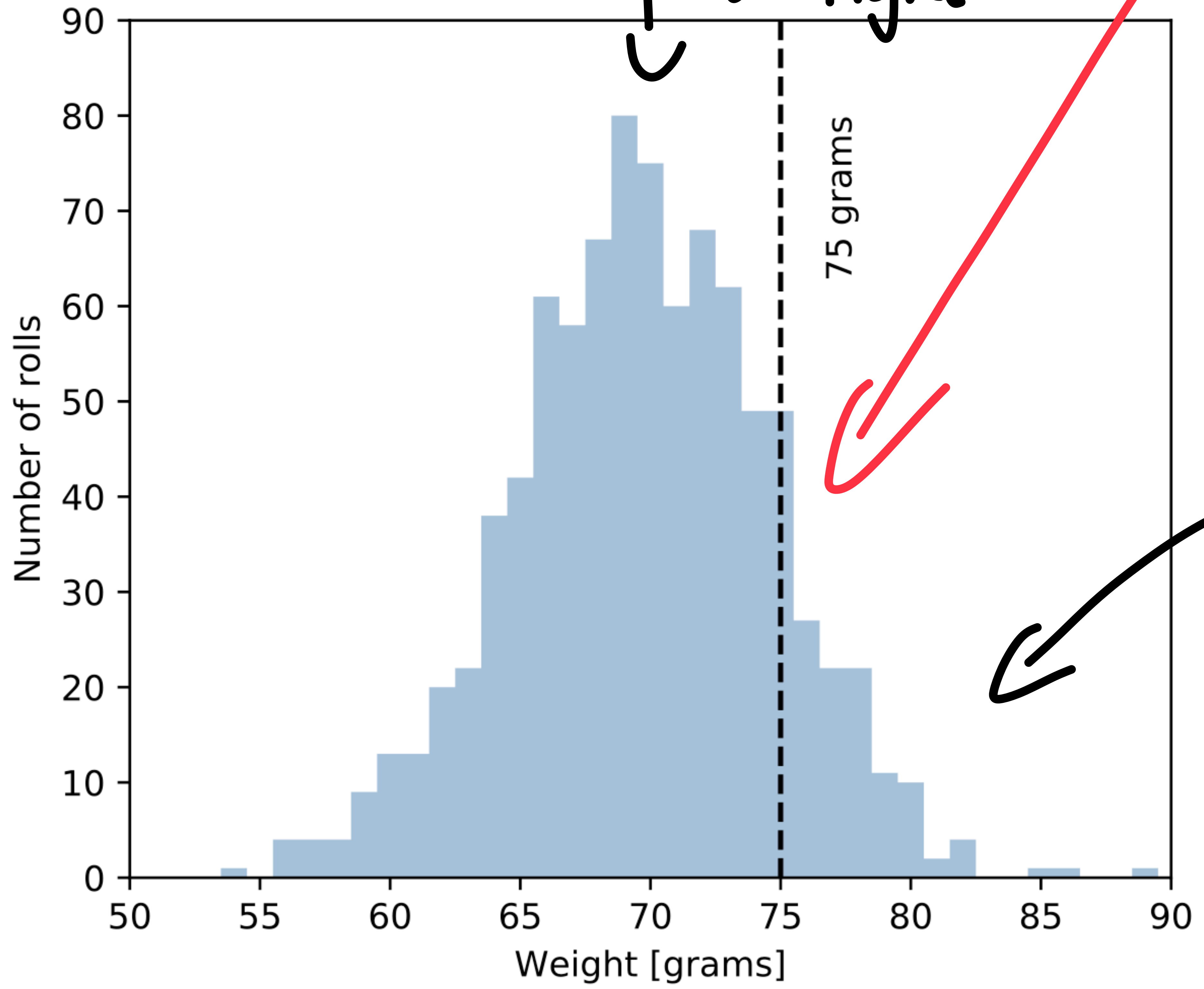
A bit more helpful, but difficult to read

```
Weight[50] = 0  Weight[51] = 0  Weight[52] = 0  Weight[53] = 0  Weight[54] = 1
Weight[55] = 0  Weight[56] = 4  Weight[57] = 4  Weight[58] = 4  Weight[59] = 9
Weight[60] = 13 Weight[61] = 13 Weight[62] = 20 Weight[63] = 22 Weight[64] = 38
Weight[65] = 42 Weight[66] = 61 Weight[67] = 58 Weight[68] = 67 Weight[69] = 80
Weight[70] = 75 Weight[71] = 60 Weight[72] = 68 Weight[73] = 62 Weight[74] = 49
Weight[75] = 49 Weight[76] = 27 Weight[77] = 22 Weight[78] = 22 Weight[79] = 11
Weight[80] = 10 Weight[81] = 2  Weight[82] = 4  Weight[83] = 0  Weight[84] = 0
Weight[85] = 1  Weight[86] = 1  Weight[87] = 0  Weight[88] = 0  Weight[89] = 1
```

The Cheating Baker

Now it is clear, we are being cheated

mean ~ 5g to light



To give the required behaviour we'd need to move the whole distribution to the right

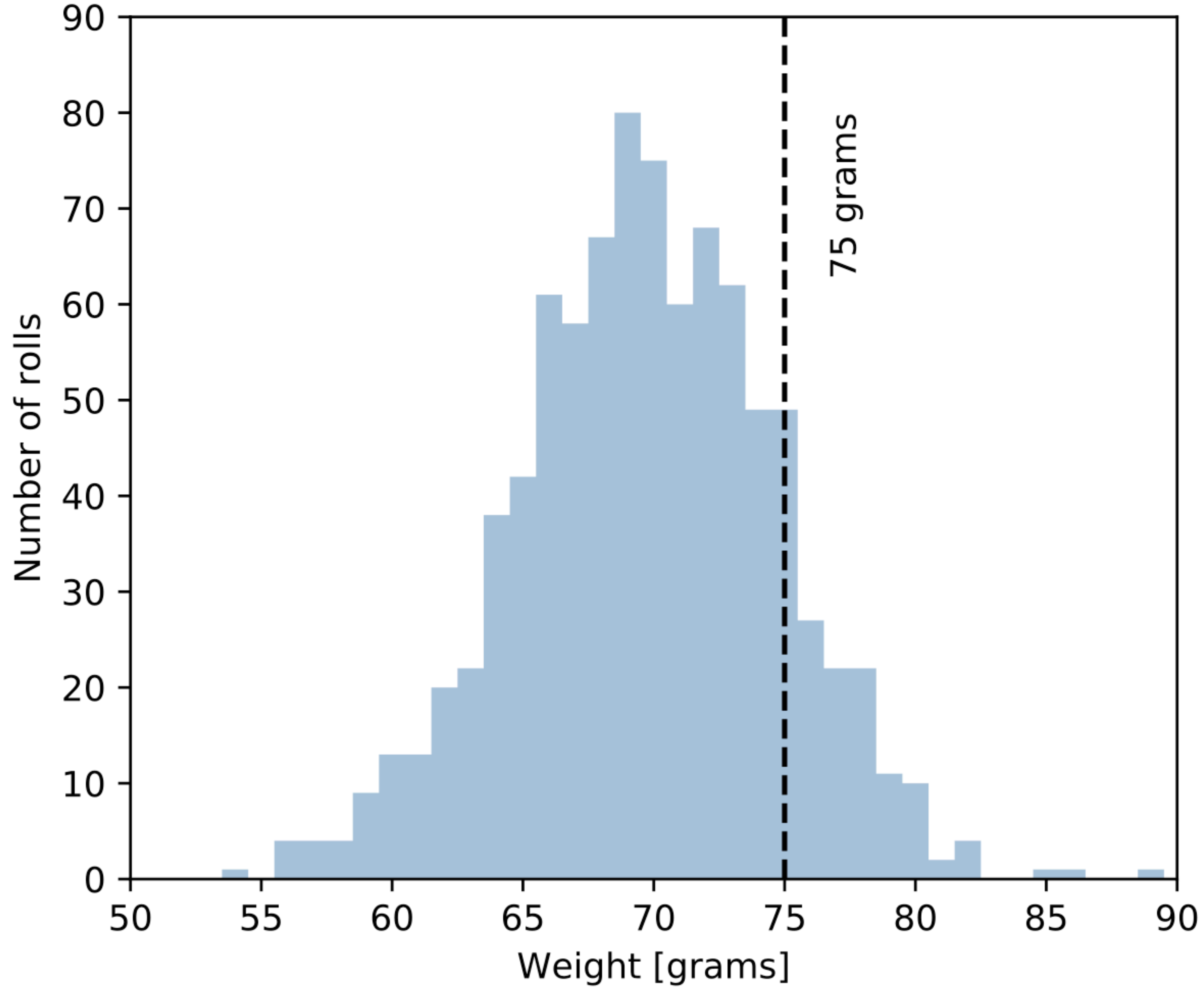
Only a small fraction are above our 75 gram limit

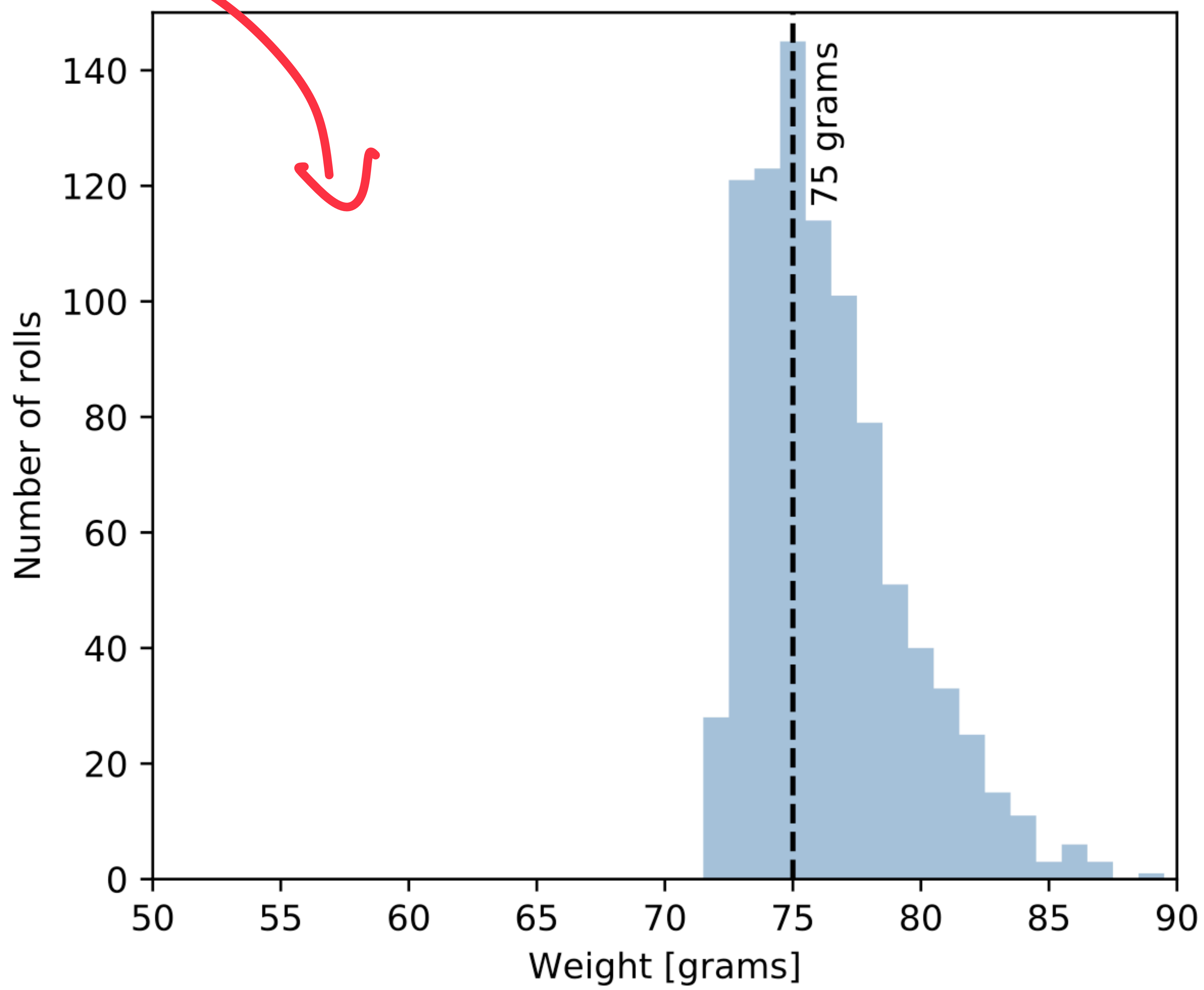
The Cheating Baker

He promises to change his ways

Measure again a few weeks later

well?

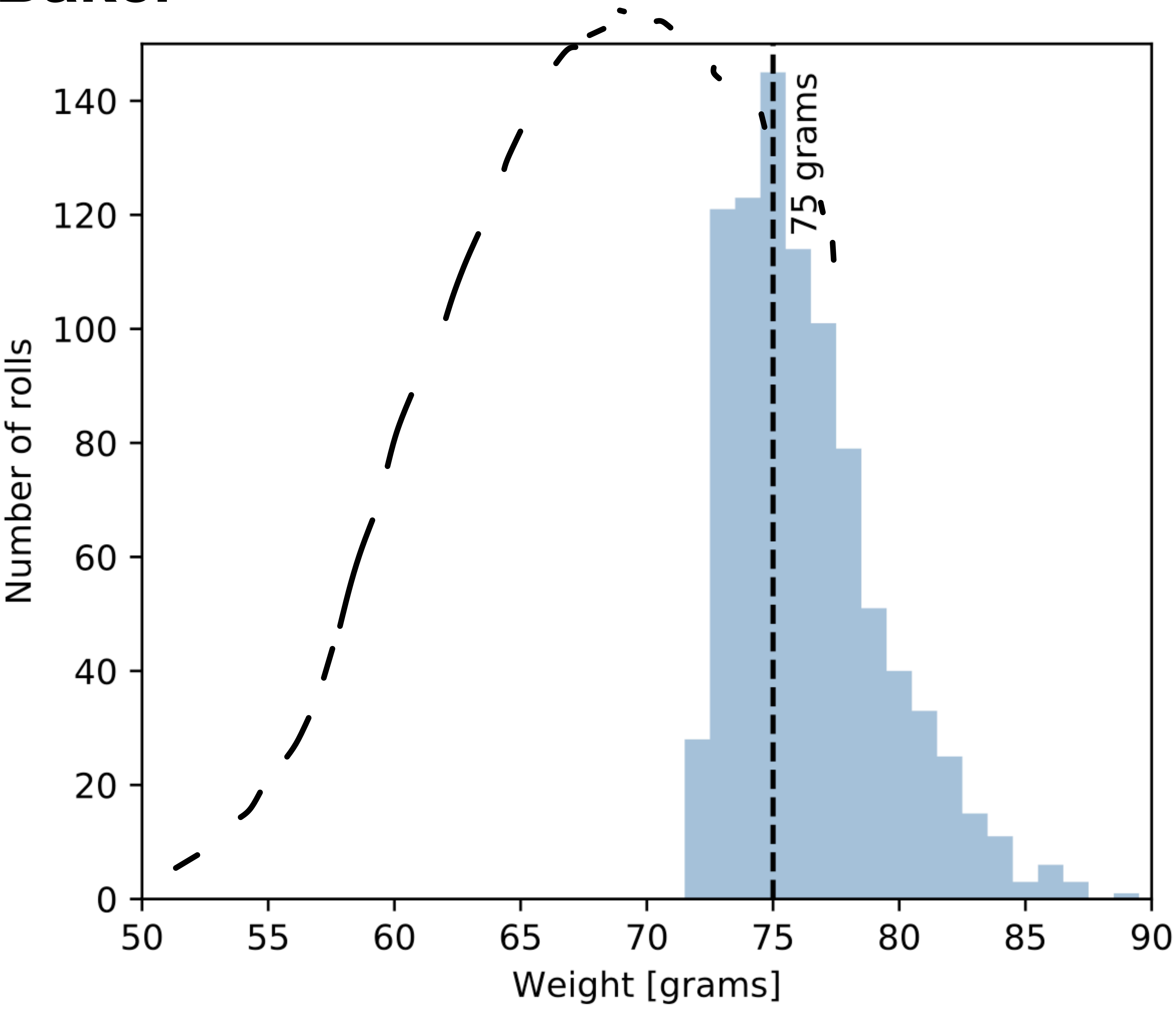




The Cheating Baker

Still cheating...

We get the heavy rolls, light rolls go to someone else



How do we know this?

Probability Distributions

Terminology

An event or random variate is a possible outcome of an experiment governed by a stochastic process.

A population is the set of all possible events. An observation is a realisation of a particular event.

An estimate or a measurement is an attempt to infer properties of the population.

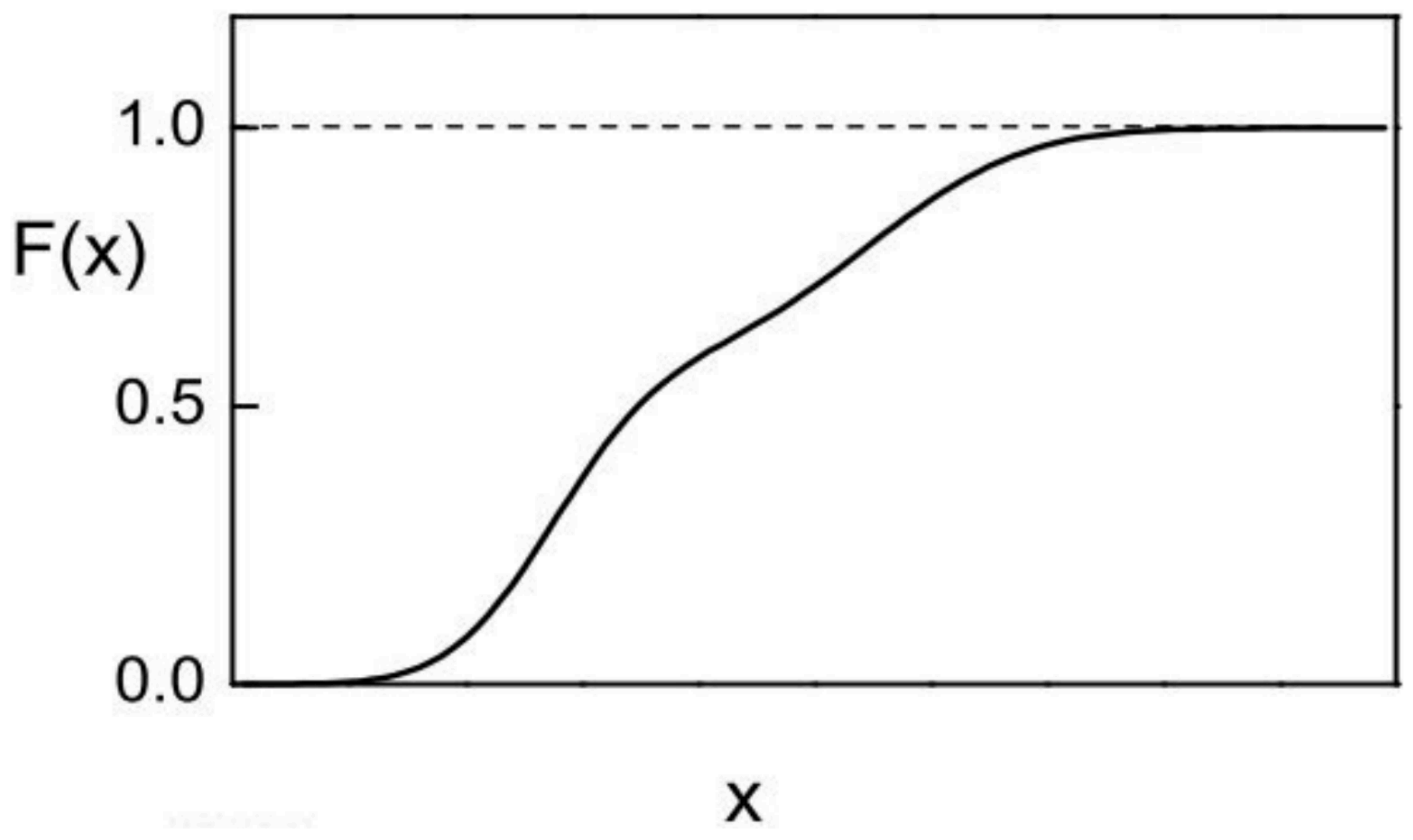
Errors are usually associated with measurements.

A probability distribution function assigns probabilities to events. Can be discrete or continuous.

Cumulative Distribution Function (CDF)

CDF gives us the probability to find a value of x lower than t

$$F(t) = P\{x < t\} \quad , \text{ with } \quad -\infty < t < \infty\,.$$



Axiomatically

$F(t)$ is a non-decreasing function of t,

$F(-\infty) = 0$,

$F(\infty) = 1$.

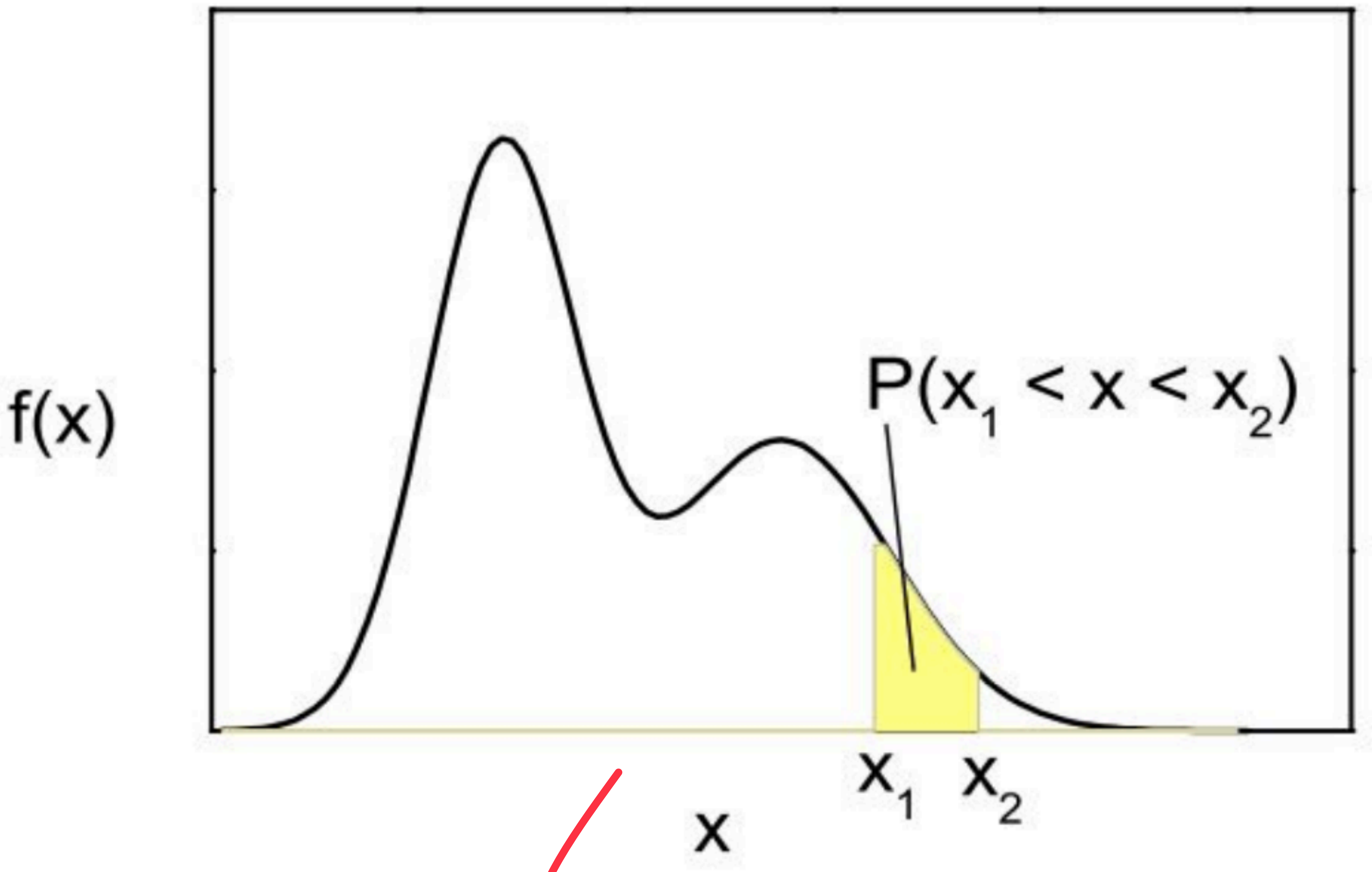
Probability Distribution Function (PDF)

$$f_X(x) = \frac{d}{dx} F_X(x)$$

$f(-\infty) = f(\infty) = 0$

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

$$\Pr[a \le X \le b] = \int_a^b f_X(x)\, dx.$$



Probability of some event over interval dx

Expectation Value

$$E(u(x)) = \sum_{i=1}^{\infty} u(x_i)p(x_i) \quad \text{(discrete distribution)} ,$$

$$E(u(x)) = \int_{-\infty}^{\infty} u(x)f(x)\,\mathrm{d}x \quad \text{(continuous distribution)} .$$

The **expectation value** of some quantity u(x) is denoted **E(x)**

Where **x** is randomly distributed according to **f(x)**,

Can be obtained by taking **the average of an infinite number of samples of u(x)**

Mean Value

$$E(x) \equiv \langle x \rangle = \mu = \textstyle\sum_{i=1}^{\infty} x_i p(x_i) \text{ (discrete distribution)} ,$$
$$E(x) \equiv \langle x \rangle = \mu = \textstyle\int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x \text{ (continuous distribution)} .$$

The (population) mean, μ**=<x>**, is the **expected value of the variate itself**

The sample mean or average based on a finite sample is usually defined as:

$$\overline{x} = \frac{1}{N} \sum_i x_i \; .$$

The sample mean is a random variable, or function of the random variate **x**, which has the expectation value:

$$\langle \overline{x} \rangle = \frac{1}{N} \sum_i \langle x_i \rangle = \langle x \rangle$$

i.e. the mean μ is a property of the unknown, true p.d.f. while the sample mean is a measurement constructed from events or observations. In the limit of large N we expect the sample mean to approach the mean. We can thus use the sample mean to estimate the mean.

Variance

Variance capture the spread (or width of a distribution)

$$\mathrm{var}(x) = \sigma^2 = E\left[(x-\mu)^2\right]$$

$\mathrm{var}(cx) = c^2\,\mathrm{var}(x)$, so if we rescale to μ/σ we get variance of 1

$$\begin{aligned}\sigma^2 &= E(x^2 - 2x\mu + \mu^2)\\ &= E(x^2) - 2\mu^2 + \mu^2\\ &= E(x^2) - \mu^2.\end{aligned}$$

$$\begin{aligned}\sigma^2 &= \langle (x - \langle x \rangle)^2 \rangle\\ &= \langle ((x_1-\mu_1)+(x_2-\mu_2))^2 \rangle\\ &= \langle (x_1-\mu_1)^2 + (x_2-\mu_2)^2 + 2(x_1-\mu_1)(x_2-\mu_2) \rangle\\ &= \langle (x_1-\mu_1)^2 \rangle + \langle (x_2-\mu_2)^2 \rangle + 2\langle x_1-\mu_1 \rangle \langle x_2-\mu_2 \rangle\\ &= \sigma_1^2 + \sigma_2^2\,.\end{aligned}$$

independent variables

Estimating Variance

Easy to estimate variance if you already know the population mean

$$v_\mu^2 = \frac{1}{N}\sum_i (x_i - \mu)^2$$

But usually you don't, so you have to **estimate from the sample mean**

$$\begin{aligned} v^2 &= \frac{1}{N}\sum_i (x_i - \overline{x})^2 \\ &= \frac{1}{N}\sum_i (x_i^2 - 2x_i\overline{x} + \overline{x}^2) \\ &= \frac{1}{N}\sum_i x_i^2 - \overline{x}^2 \ . \end{aligned}$$

Using v to estimate the **population variance** would give a biased estimate.

We rather have to use the (well known) correction:

$$\frac{v^2}{N-1} = \frac{\sum_i (x_i - \overline{x})^2}{N(N-1)}$$

Higher Moments...

$$\gamma_1 = E\left[(x-\mu)^3\right]/\sigma^3$$ Skewness

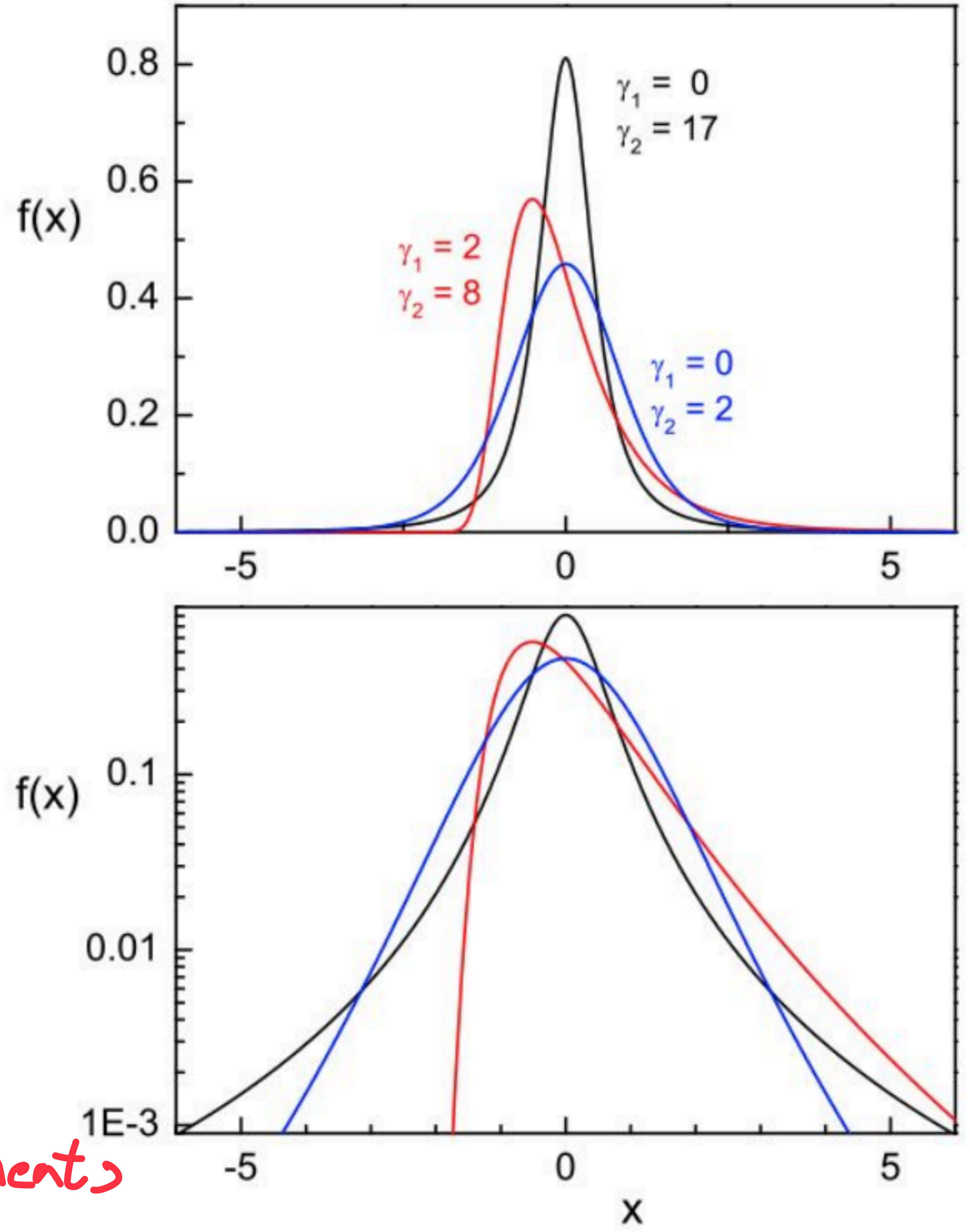
$$\beta_2 = E\left[(x-\mu)^4\right]/\sigma^4$$

$$\gamma_2 = \beta_2 - 3\,,$$ Kurtosis

Central moments

$$\mu_n' = E((x-\mu)^n)$$
$$= \int_{-\infty}^{\infty} (x-\mu)^n f(x)\,dx$$

If distributions have the same moments they are identical



Gaussian (or Normal) Distribution

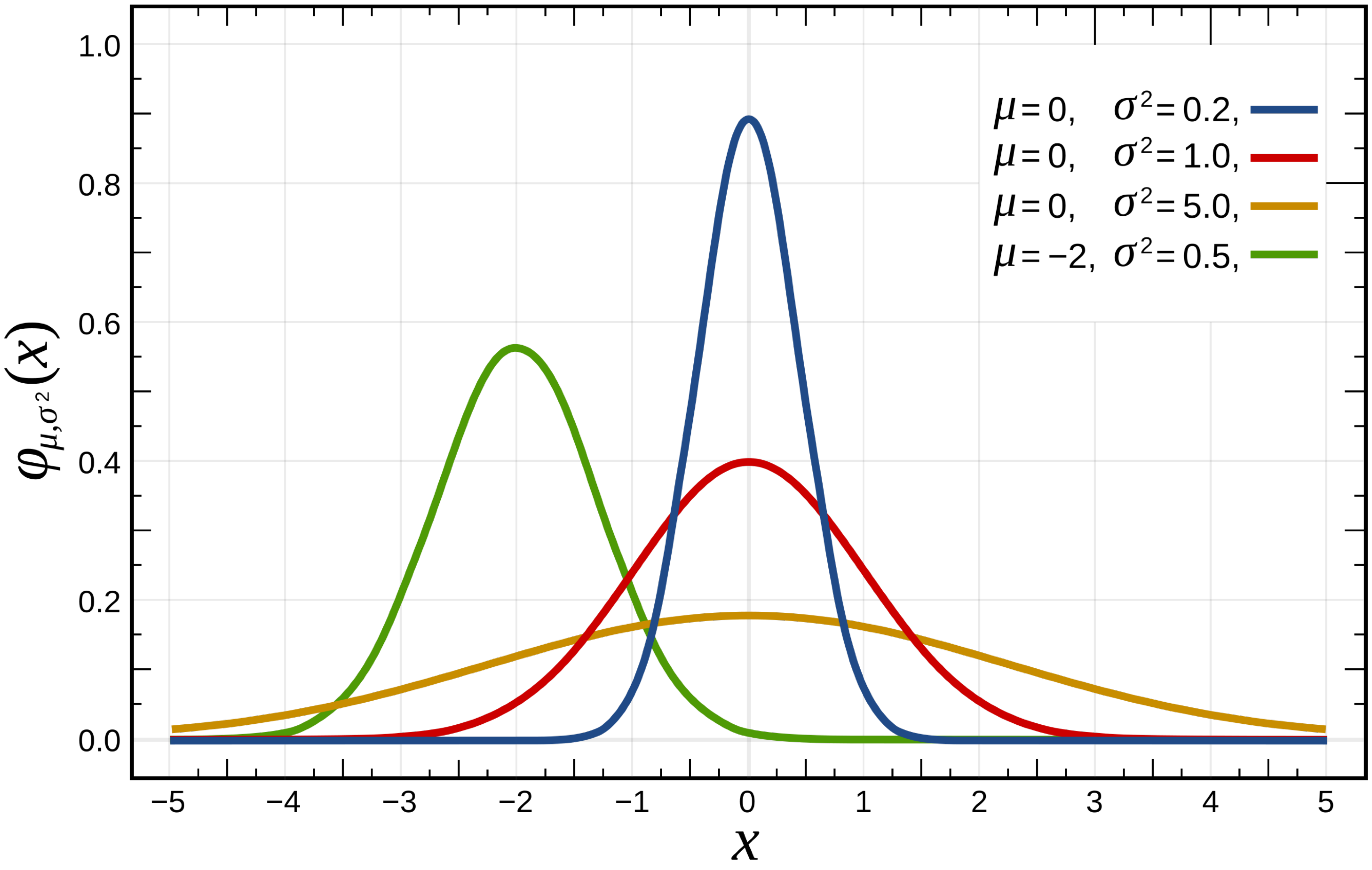
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Probably the best known PDF

at $x = \mu \pm \sigma$,

$y = y_{max}/\sqrt{e}$

$\sim 0.606\, y_{max}$



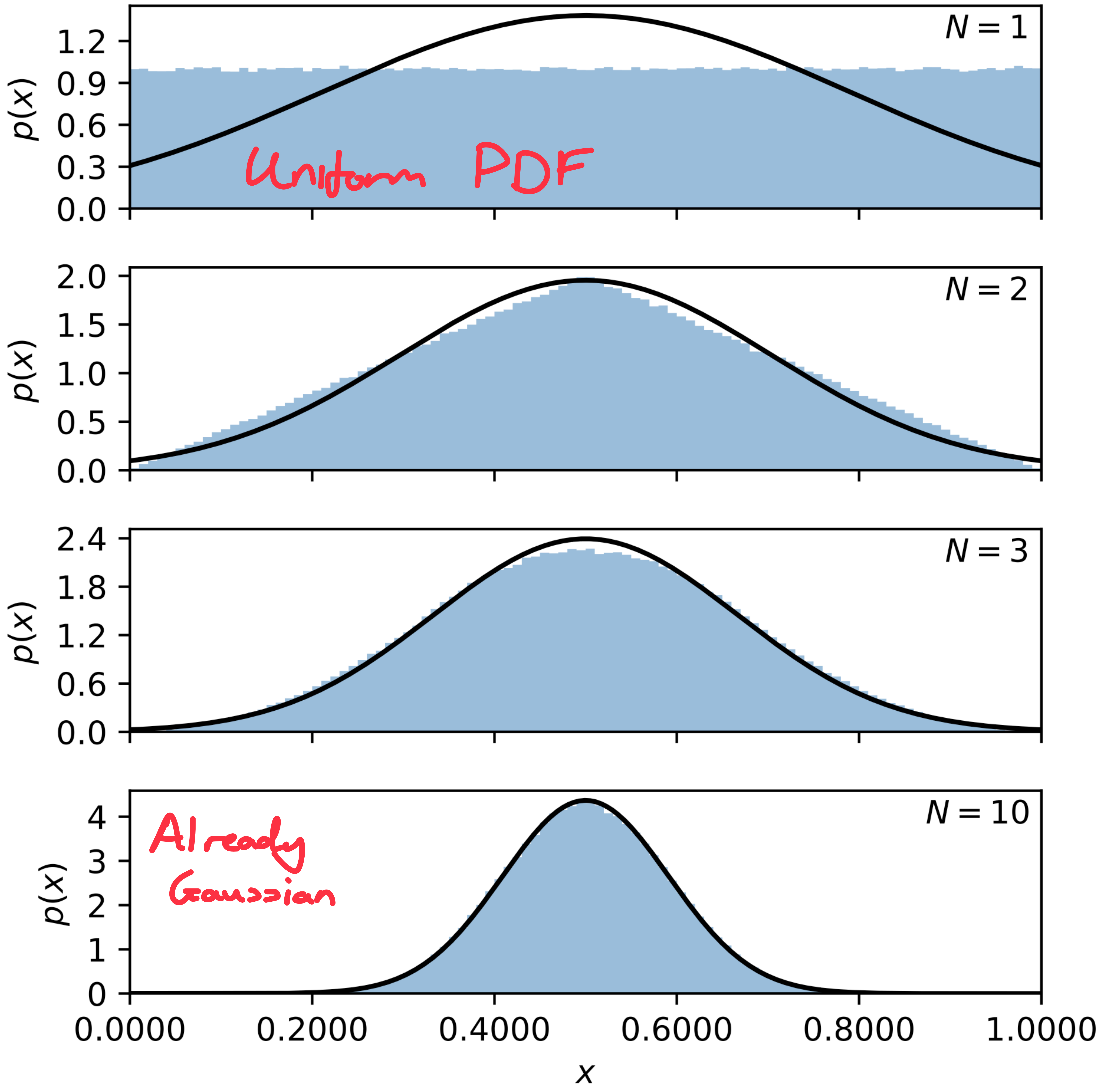
Central Limit Theorem

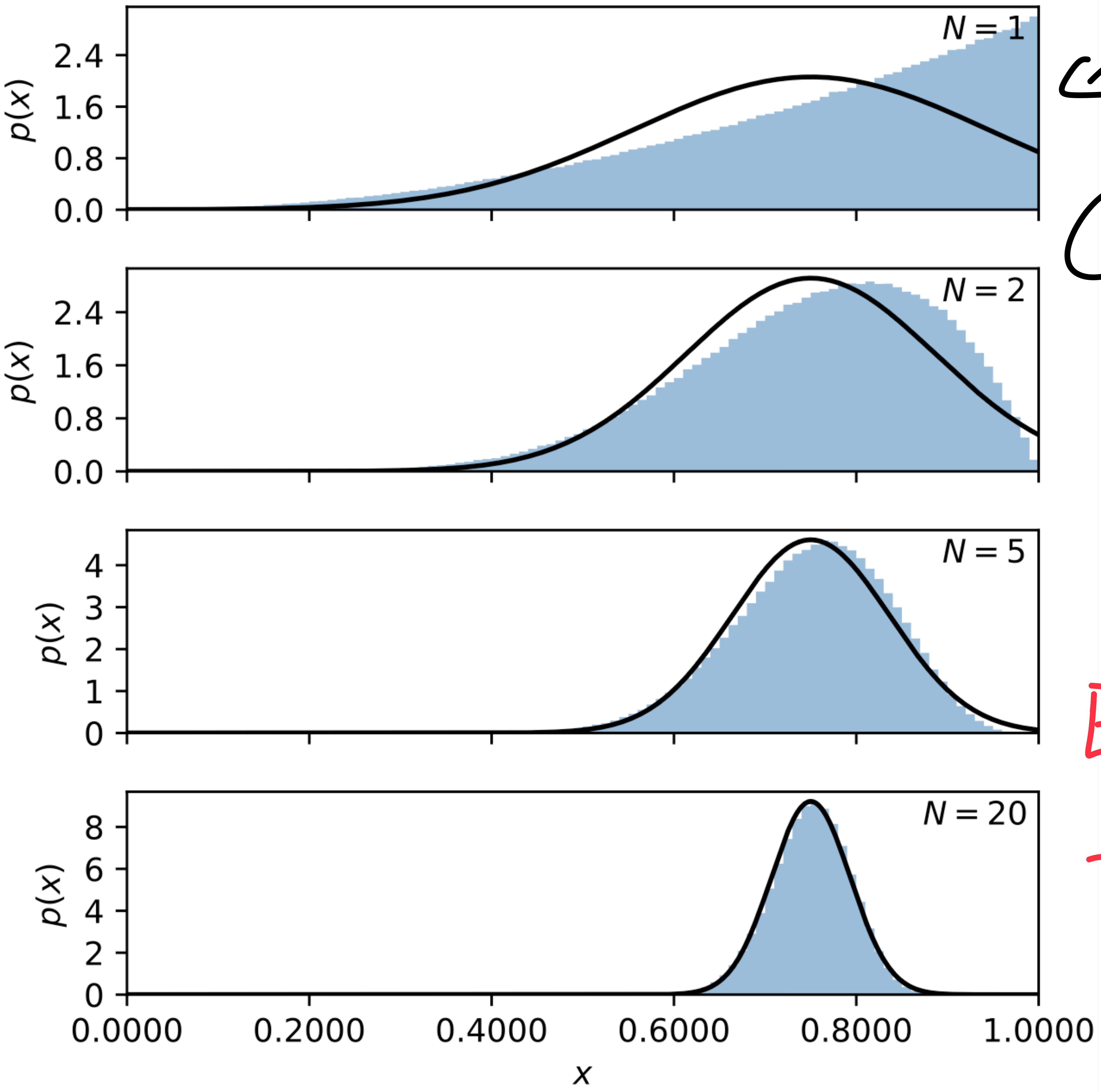
Pick K random variables from any PDF

Repeat N times and take the average (or sum)

Distribution of mean will follow a Gaussian (at large enough N)

Larger N for non-uniform distribution

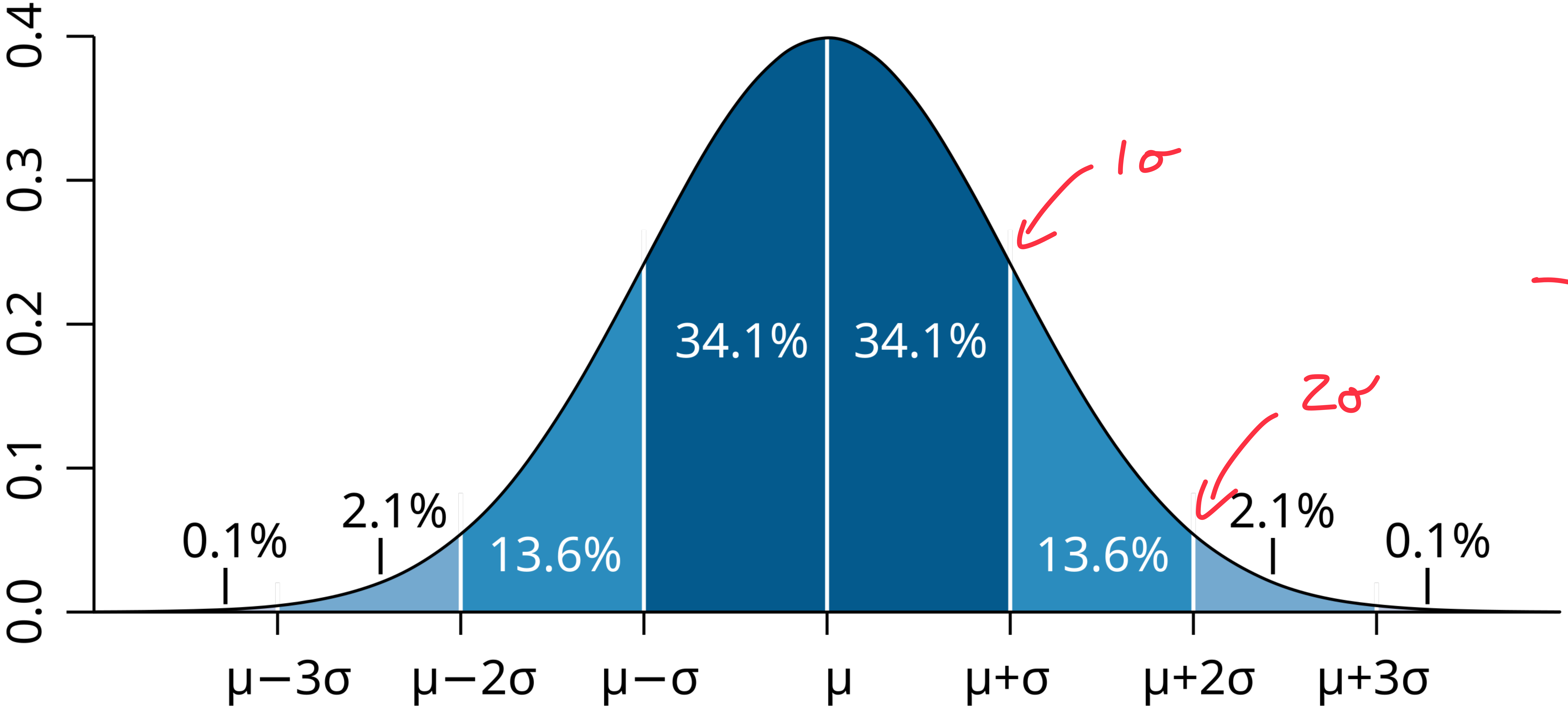




Works with any distribution (Parabola)

Easy to code + try yourself

Gaussian (or Normal) Distribution



Binomial Distribution

Probability of events
with 2 possible outcomes

True or False, Heads or tail

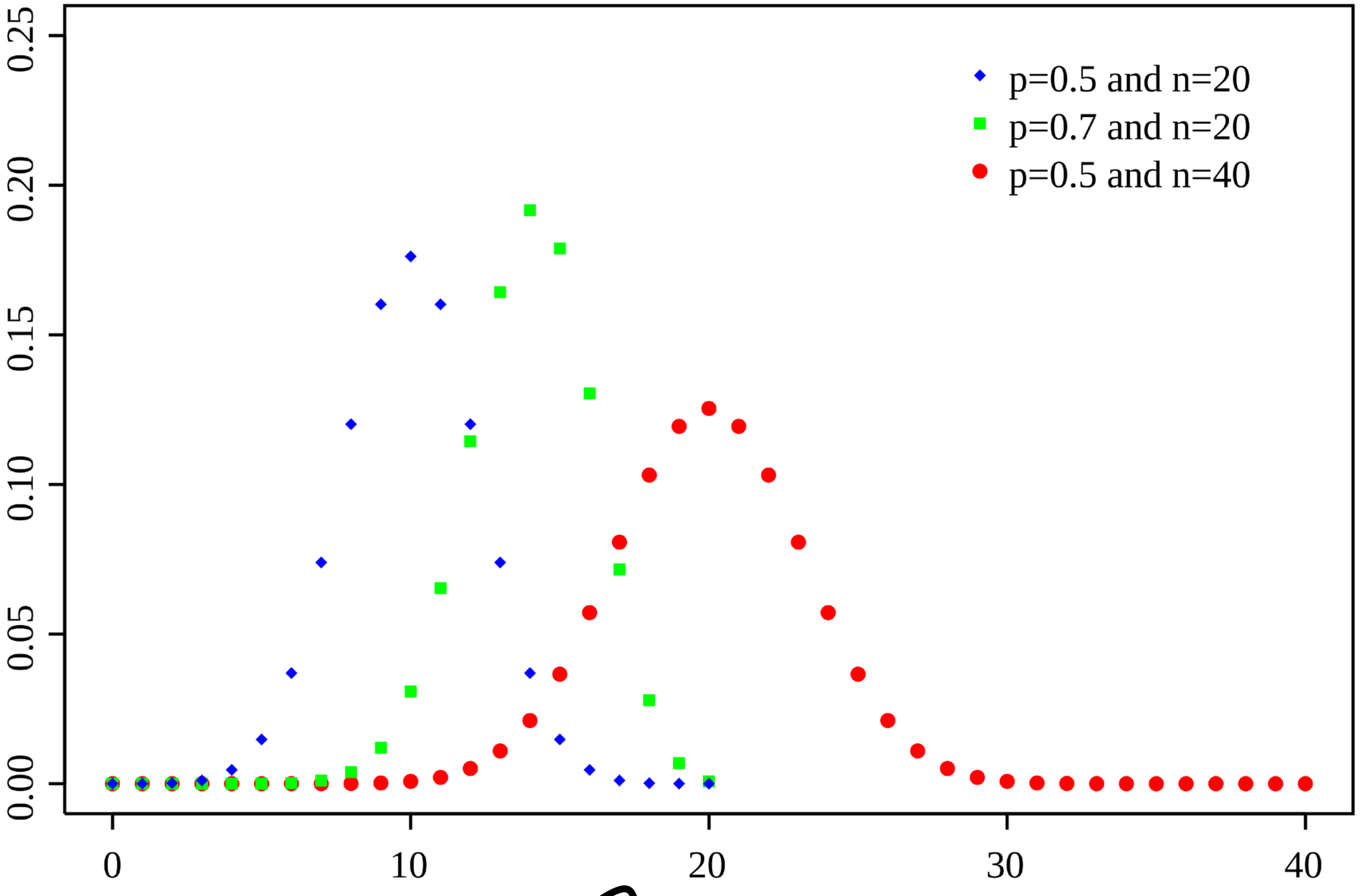
Roll a 6 on a dice 34 times in
100 rolls

$K = 34,\ N = 100,\ p = \frac{1}{6}$

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

$$E[k] = \textstyle\sum_k kP(k) = Np.$$

$$E[(k - \langle k \rangle)^2] = E[k^2] - (E[k])^2 = Np(1-p).$$



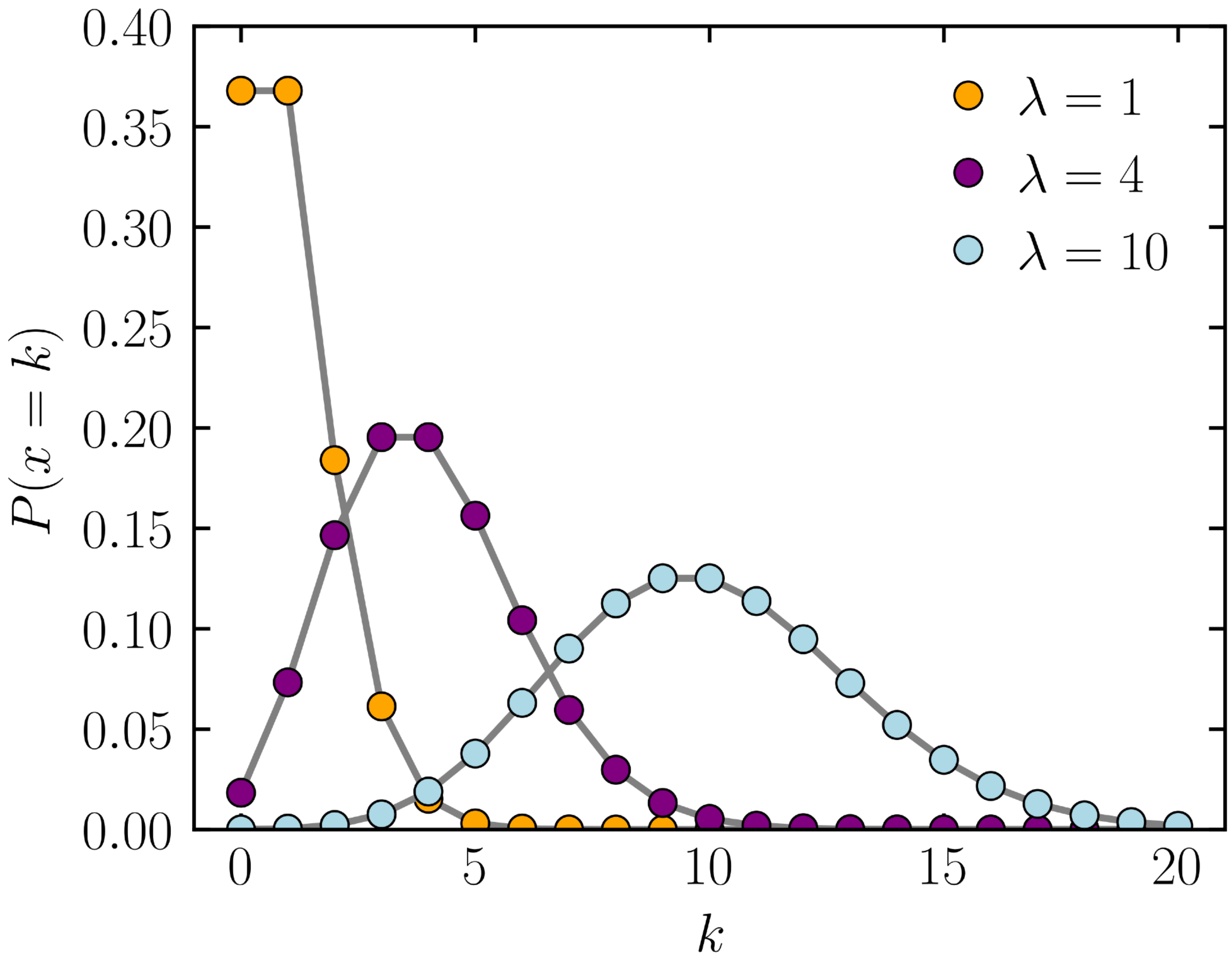
Probability mass function

Poisson Distribution

For N →∞, p →0

Np= const.,

Binomial →**Poisson**



observed number

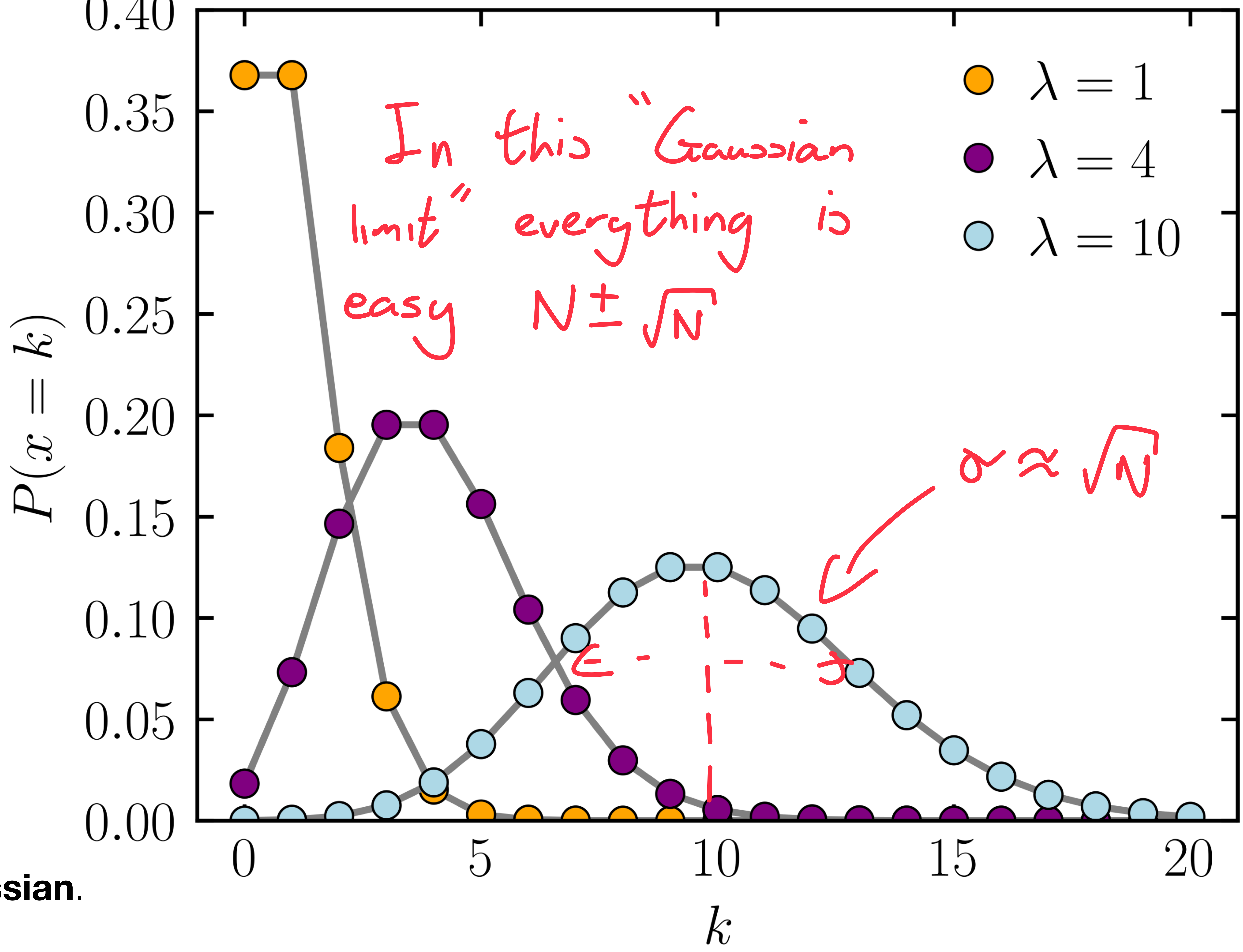
expected number

$$P(N;\mu) = \frac{\mu^N}{N!}e^{-\mu}$$

Probability of a discrete number of events in a fixed time period

$$E[N] = \sum_N NP(N) = \mu.$$

$$\sigma^2 = \mu.$$



For N $\rightarrow \infty$

Poisson $\rightarrow$ **Gaussian**.

Exponential Distribution

Used typically in processes that have no "memory"

e.g. radioactive decay

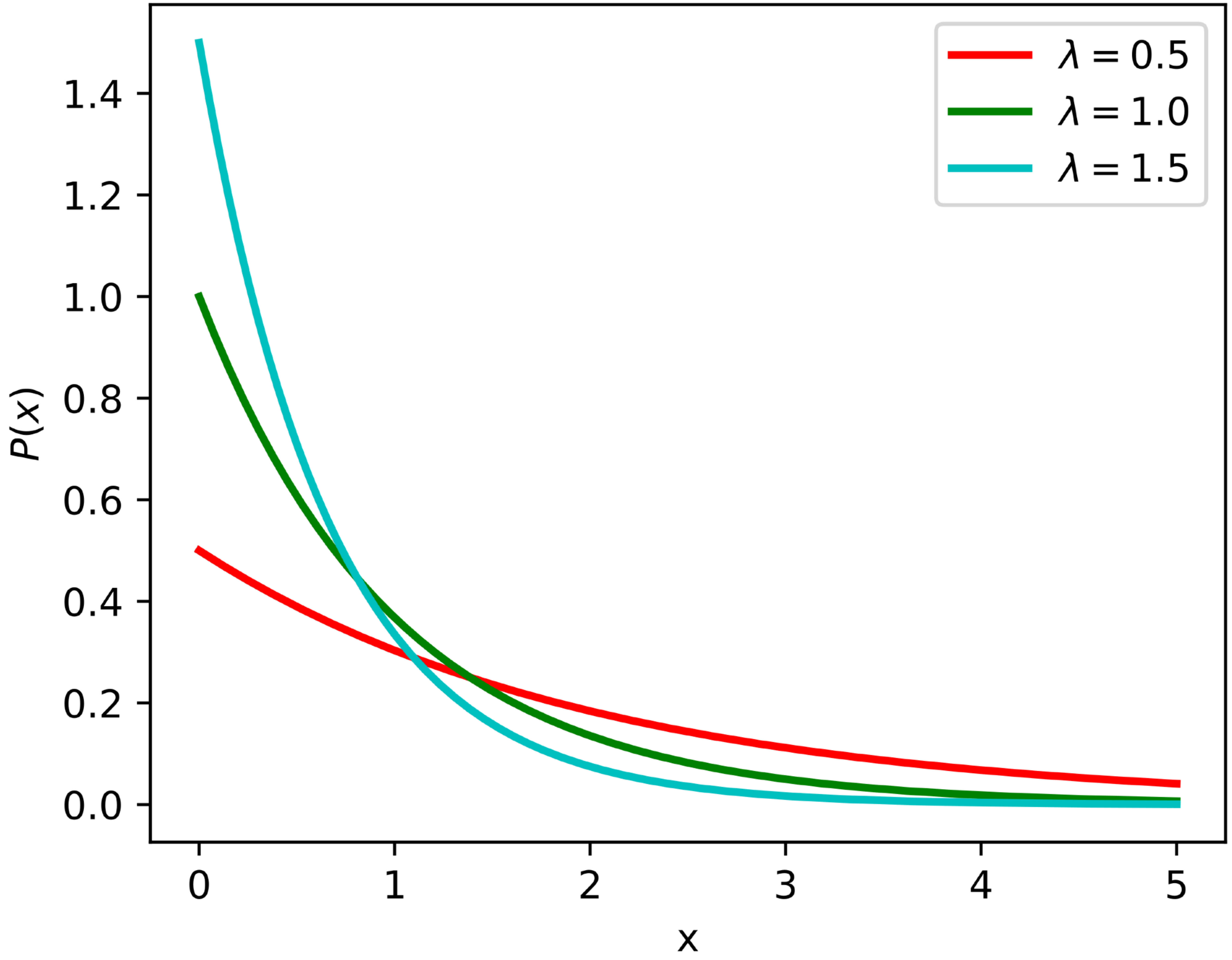
Probability of decay in each time bin **dt** is constant

Lambda>0 and equal to inverse if the lifetime

$$f(t) \equiv f(t|\lambda) = \lambda e^{-\lambda t} \quad \text{for } t \geq 0$$

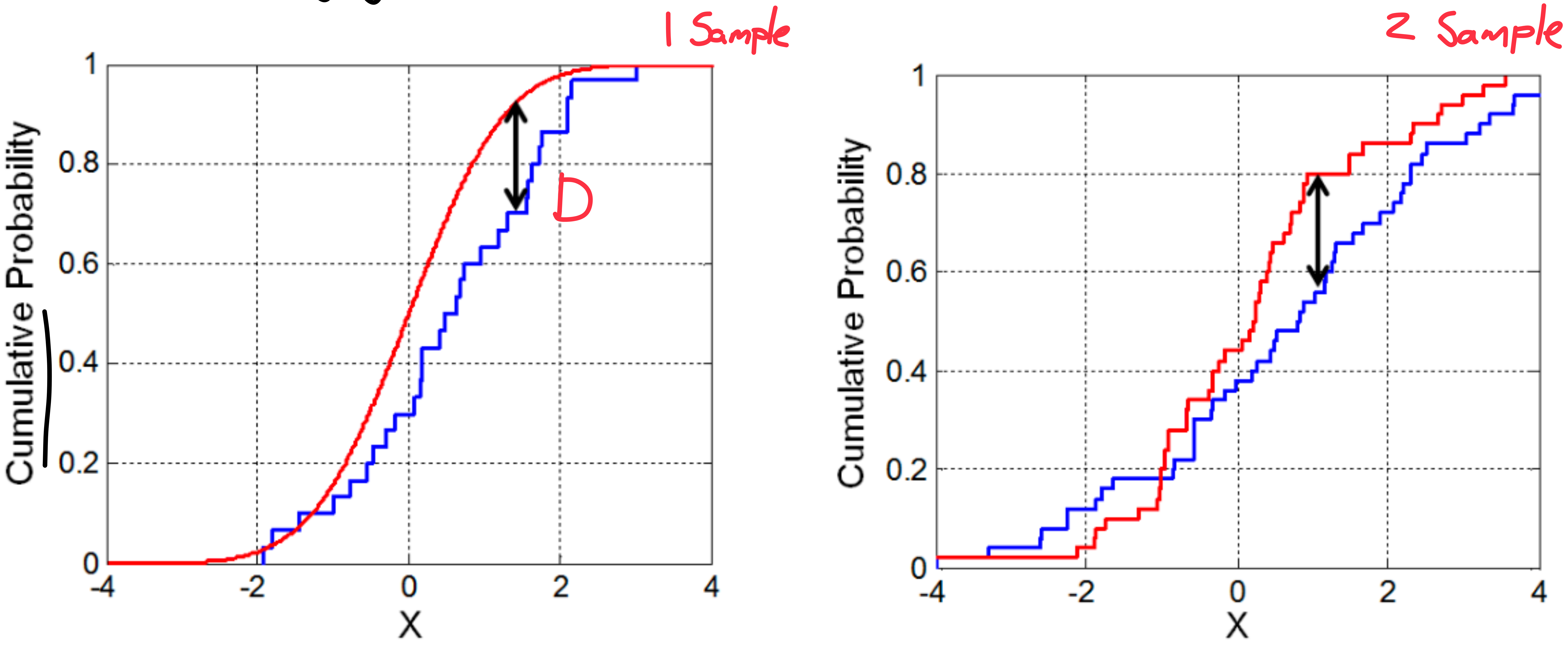
$$F(t) = \int_{-\infty}^{t} f(t')dt' = 1 - e^{-\lambda t}$$

Mean: $\frac{1}{\lambda}$ Variance: $\frac{1}{\lambda^2}$



Kolmogorov Smirnov Test

Compare probability that 2 distributions come from same undlying PDF



Kolmogorov Smirnov Test

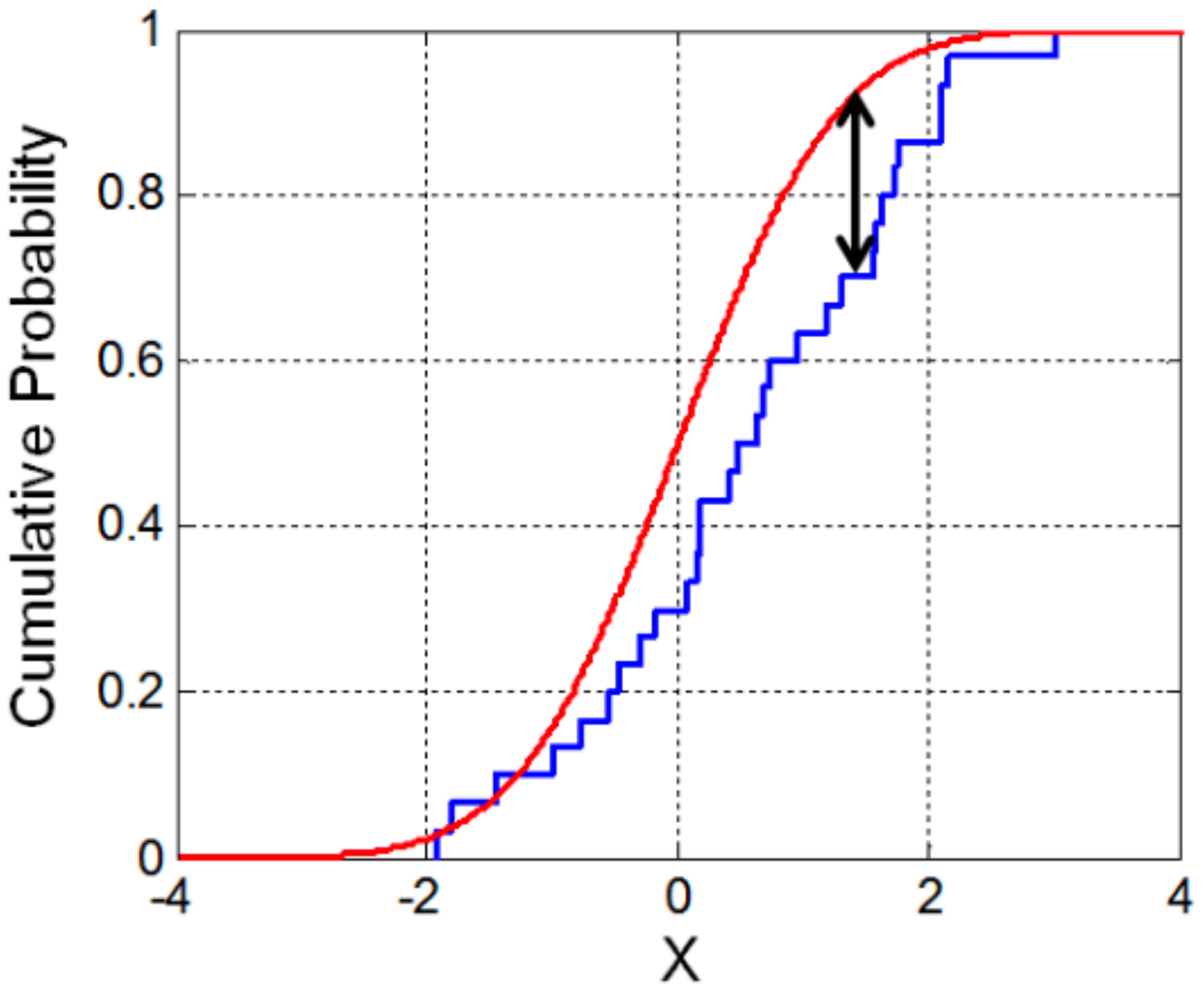
$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{n \cdot m}}.$$

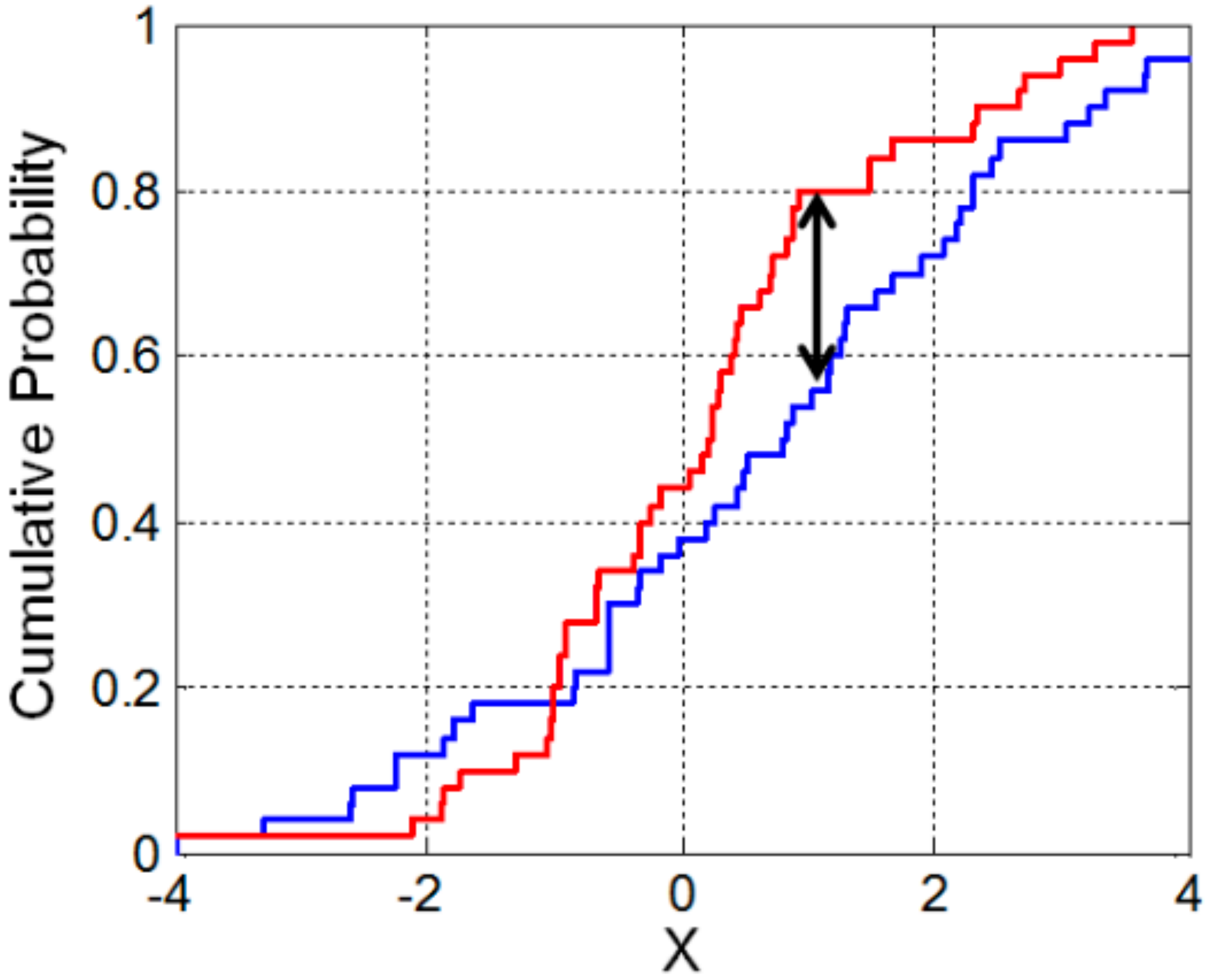
n, m -> number of events in samples

Confidence level

$$D_n = \sup_x |F_n(x) - F(x)|$$

α	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.073	1.138	1.224	1.358	1.48	1.628	1.731	1.949





Covariance and Correlation

Covariance and Correlation

At some point you will come across distributions dependent on several variables

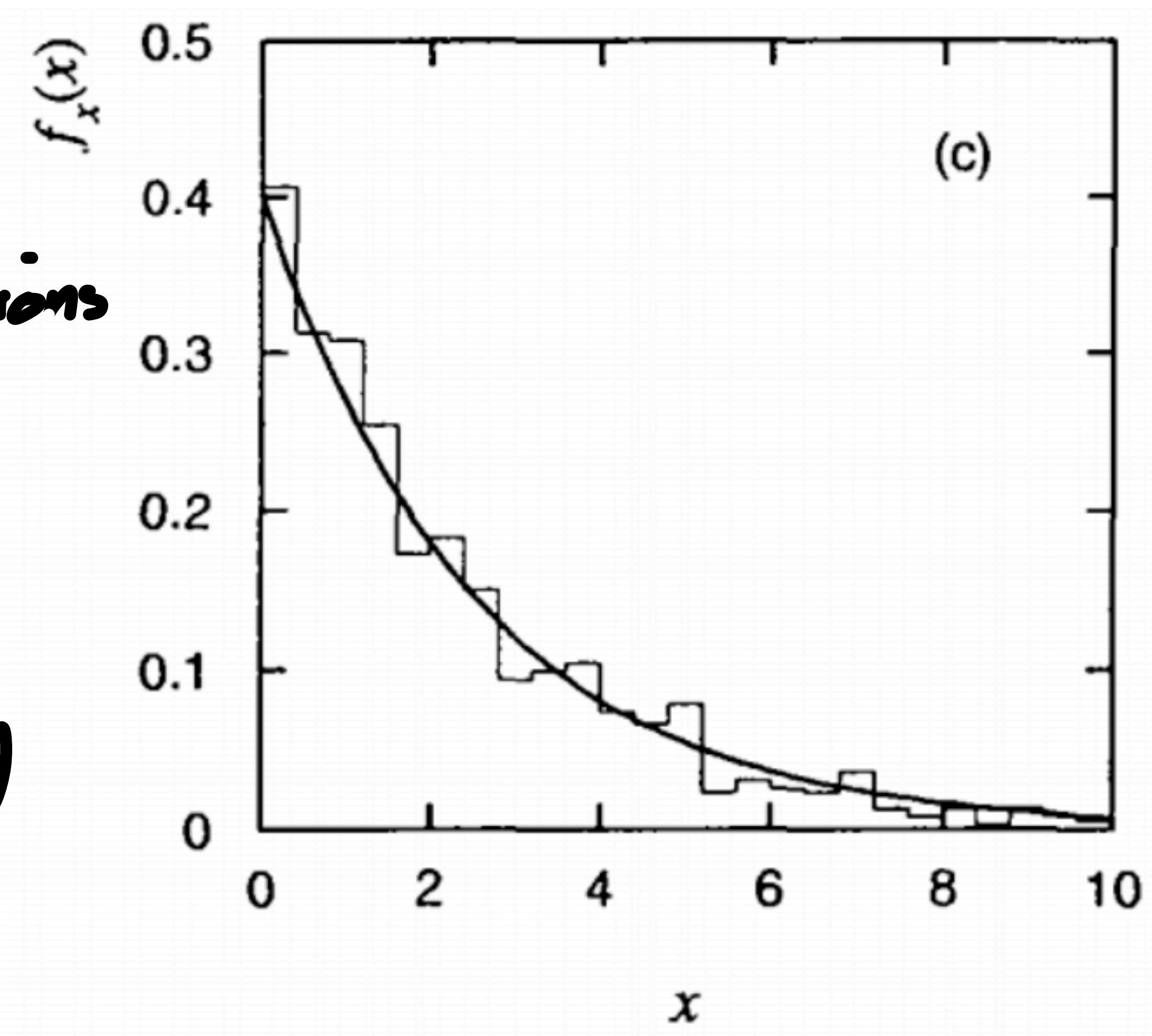
What is the probability for A and B

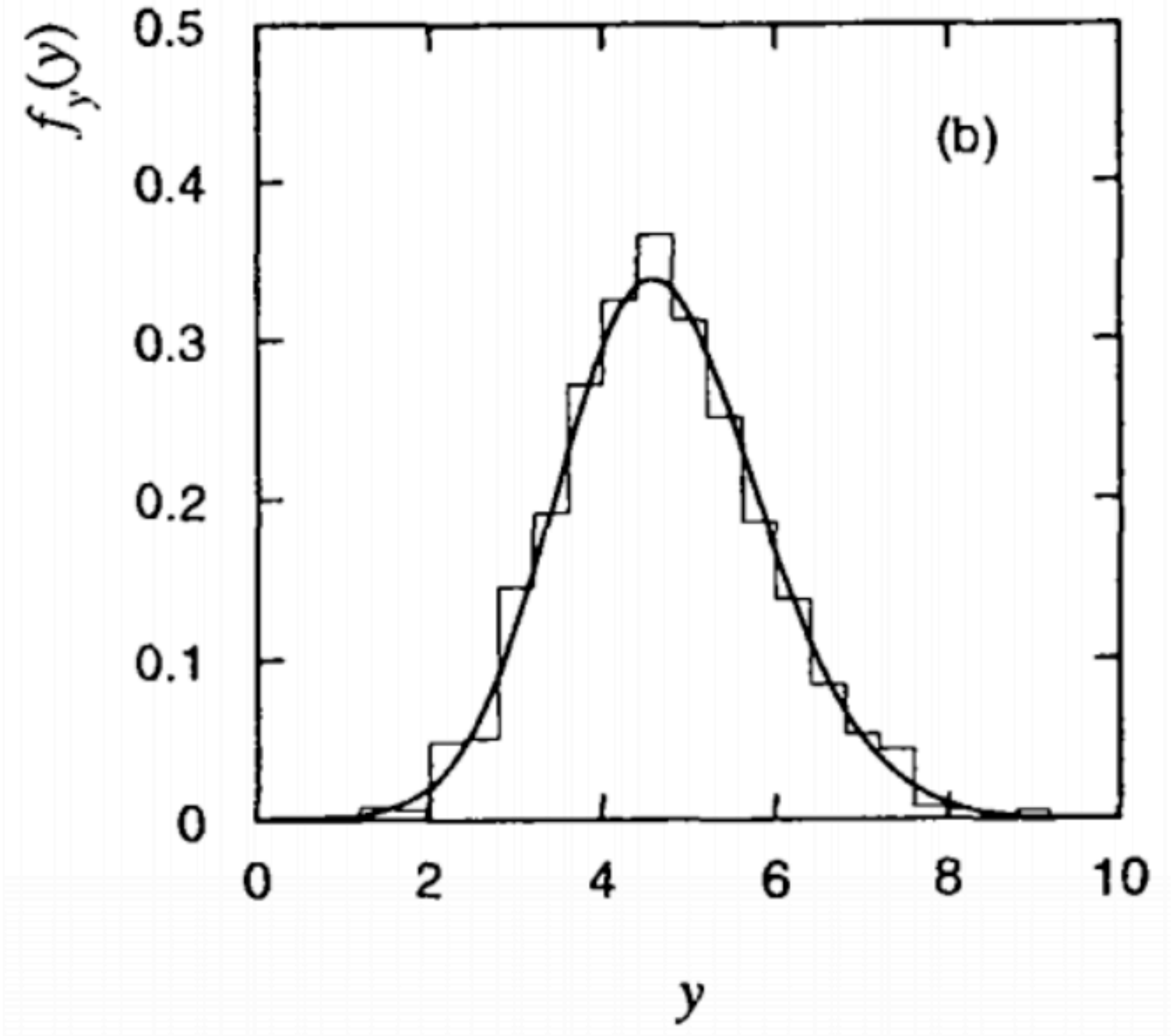
$P(A) = f(x)\,dx$

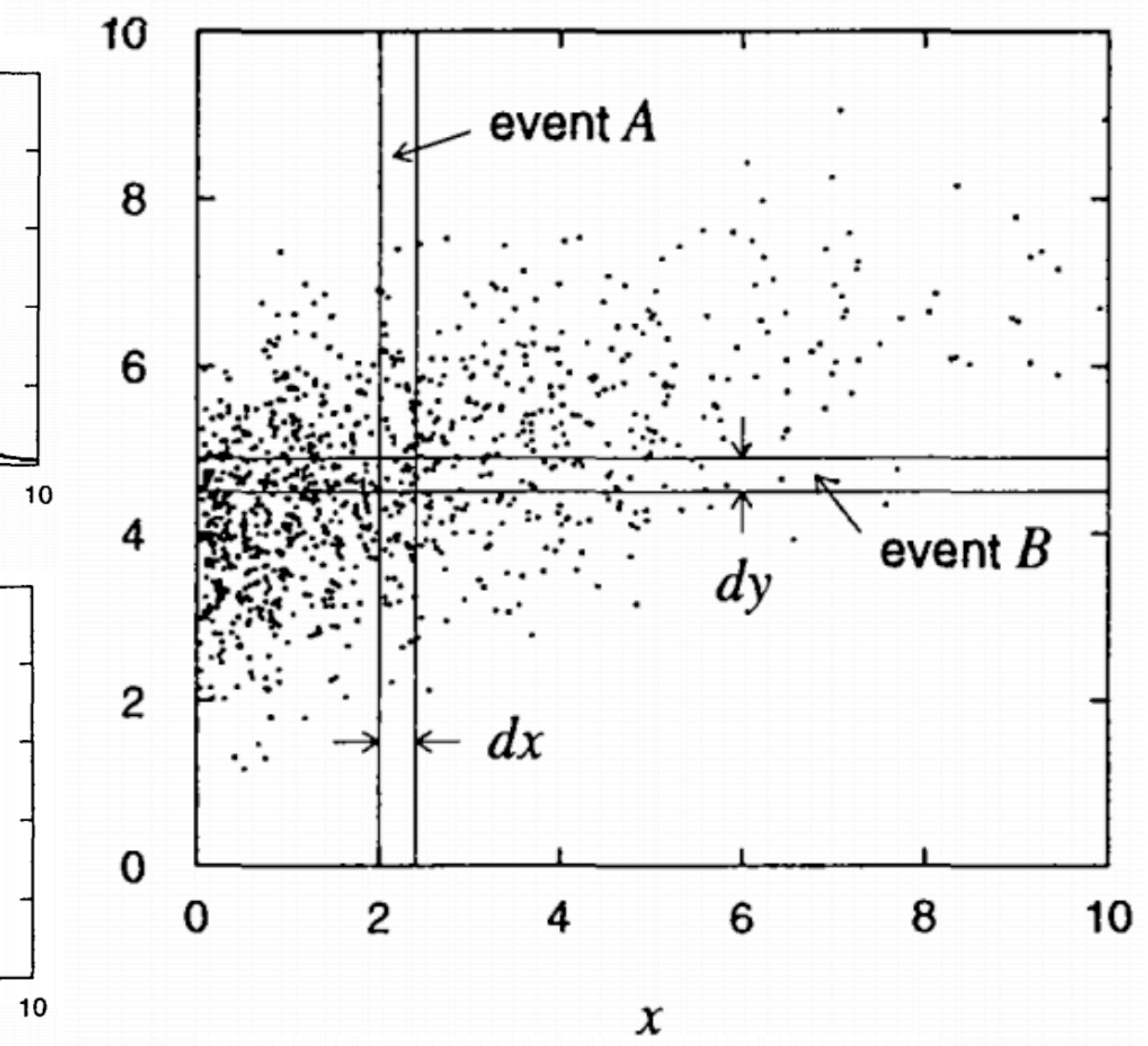
$P(B) = f(y)\,dy$

$\Rightarrow P(A \cap B) = f(x,y)\,dx\,dy$

Can project along one axis if you are care about one variable







Covariance and Correlation

How correlated are x and y?

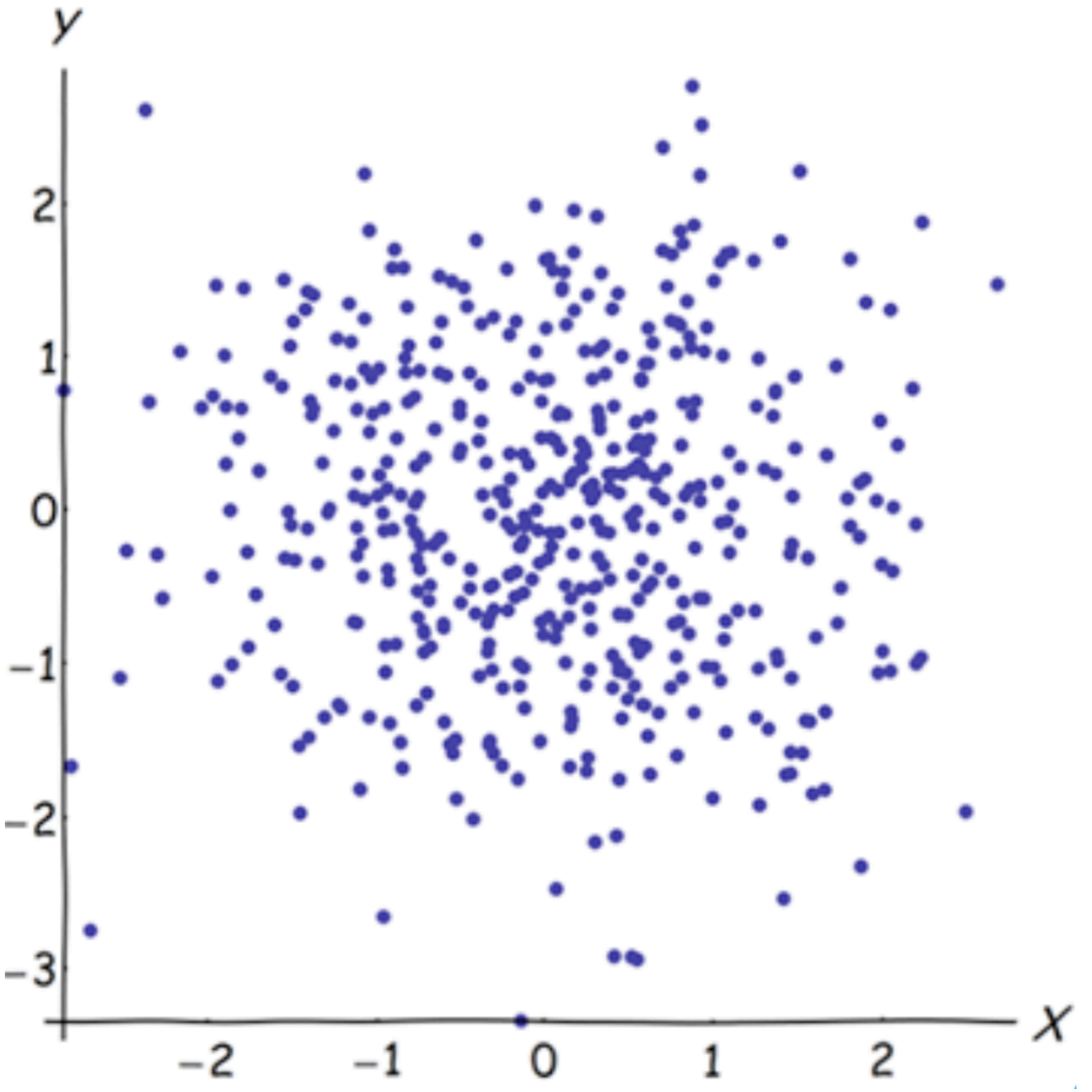
Covariance

$$V(x) = \sigma_x^2 = E[(x - \langle x \rangle)^2]$$
$$= E[x^2] - (E[x])^2$$

$$C(x,y) = V_{x,y} = E[(x - \langle x \rangle)(y - \langle y \rangle)]$$
$$= E[xy] - E[x]E[y]$$

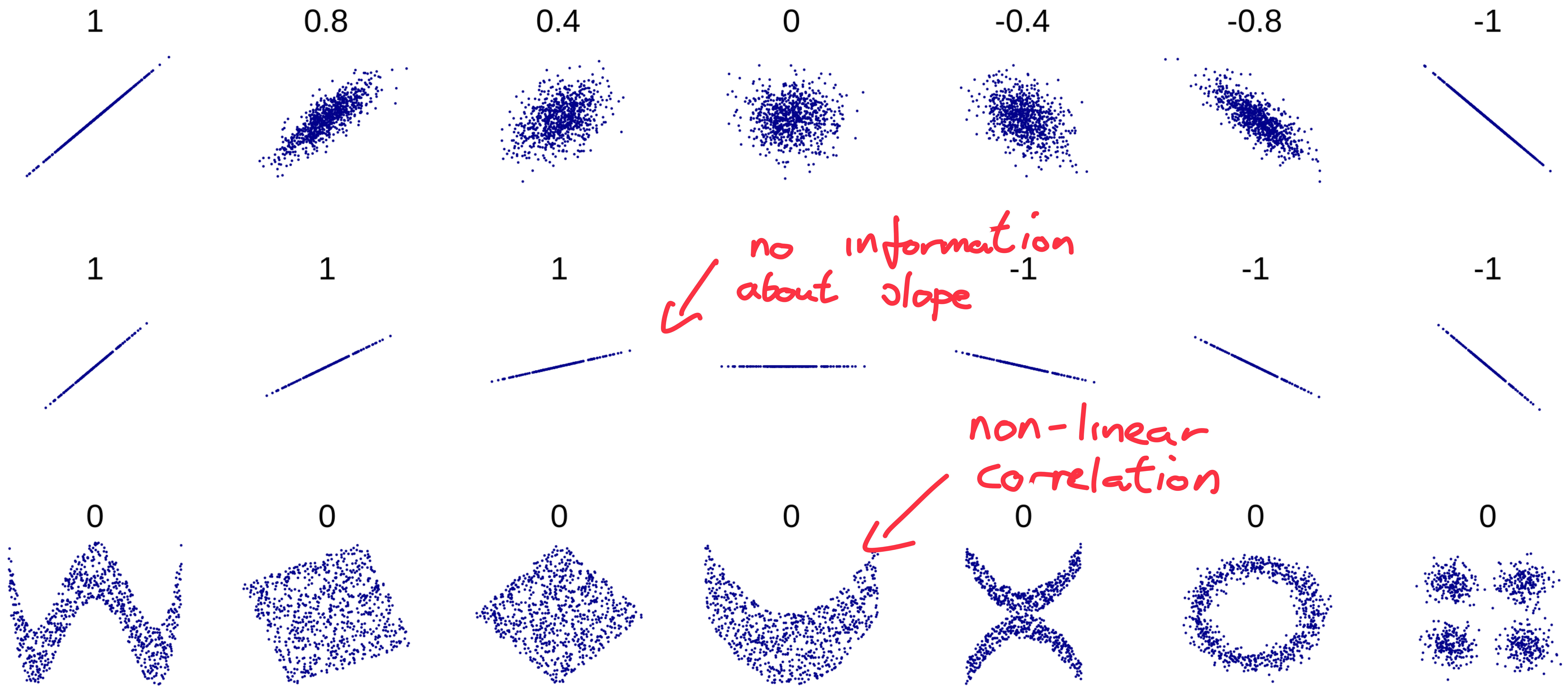
If uncorrelated

$$P(A \cap B) = P(A)P(B)$$



Correlation Coefficient

$$\rho_{x,y} = \frac{C(x,y)}{\sigma_x \sigma_y}$$



no information about slope

non-linear correlation

Covariance and Correlation

$$P(\vec{x}) = \frac{1}{2\pi\sqrt{\det(C)}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T C^{-1}(\vec{x}-\vec{\mu})\right)$$

$$\text{for } \vec{x} = (x, y) \Rightarrow C = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & V_{x,y} \\ V_{x,y} & \sigma_y^2 \end{pmatrix}$$

Variance along one axis

Covariance matrix

Parameter Estimation

Least Squares Fit

Best known fit statistic

$$\chi^2 = \sum_i \left(\frac{(ax_i + b) - y_i}{\sigma} \right)^2$$

Not just good for linear functions

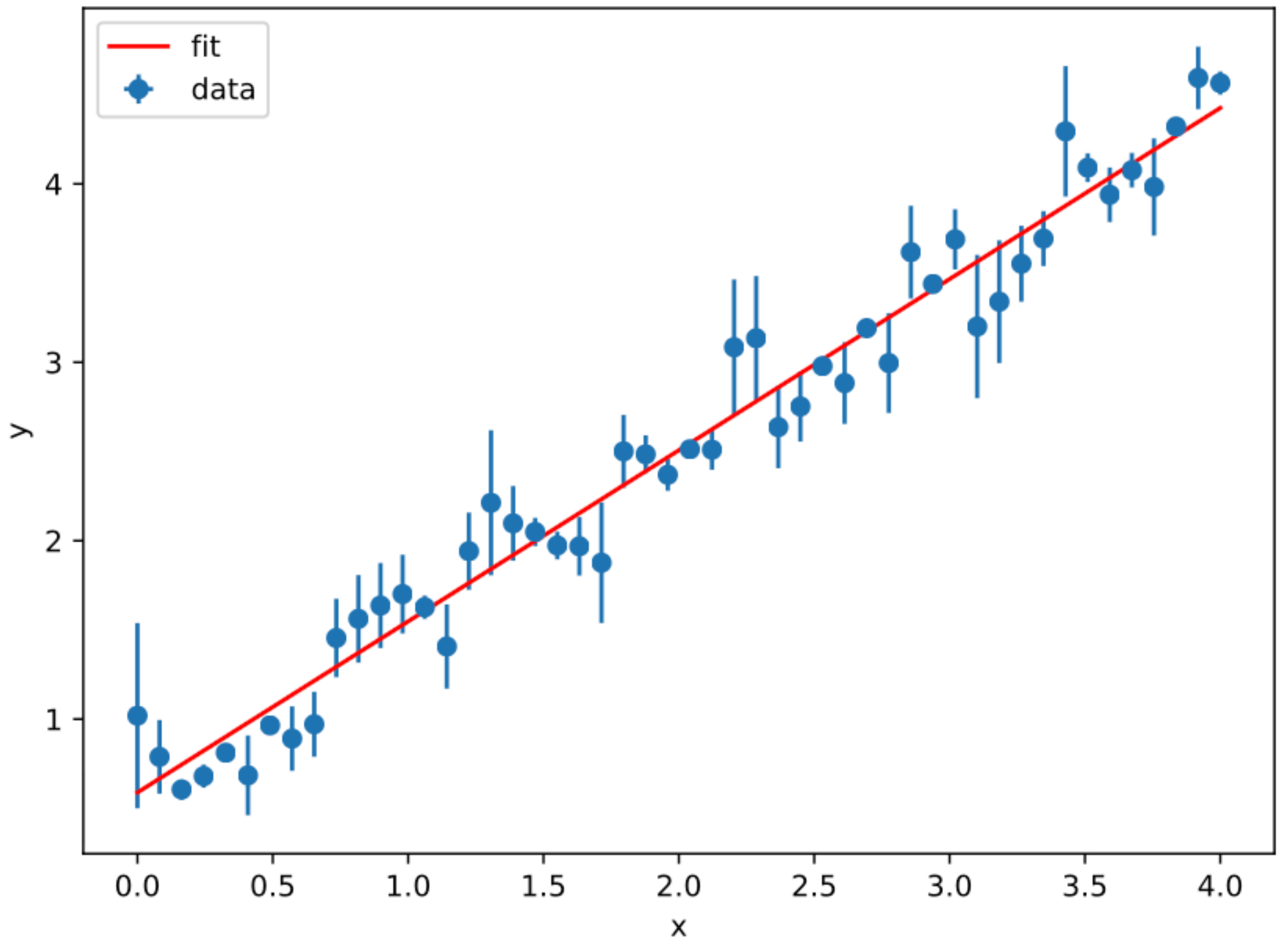
Error bars should be gaussian

Goodness of fit given by reduced χ^2

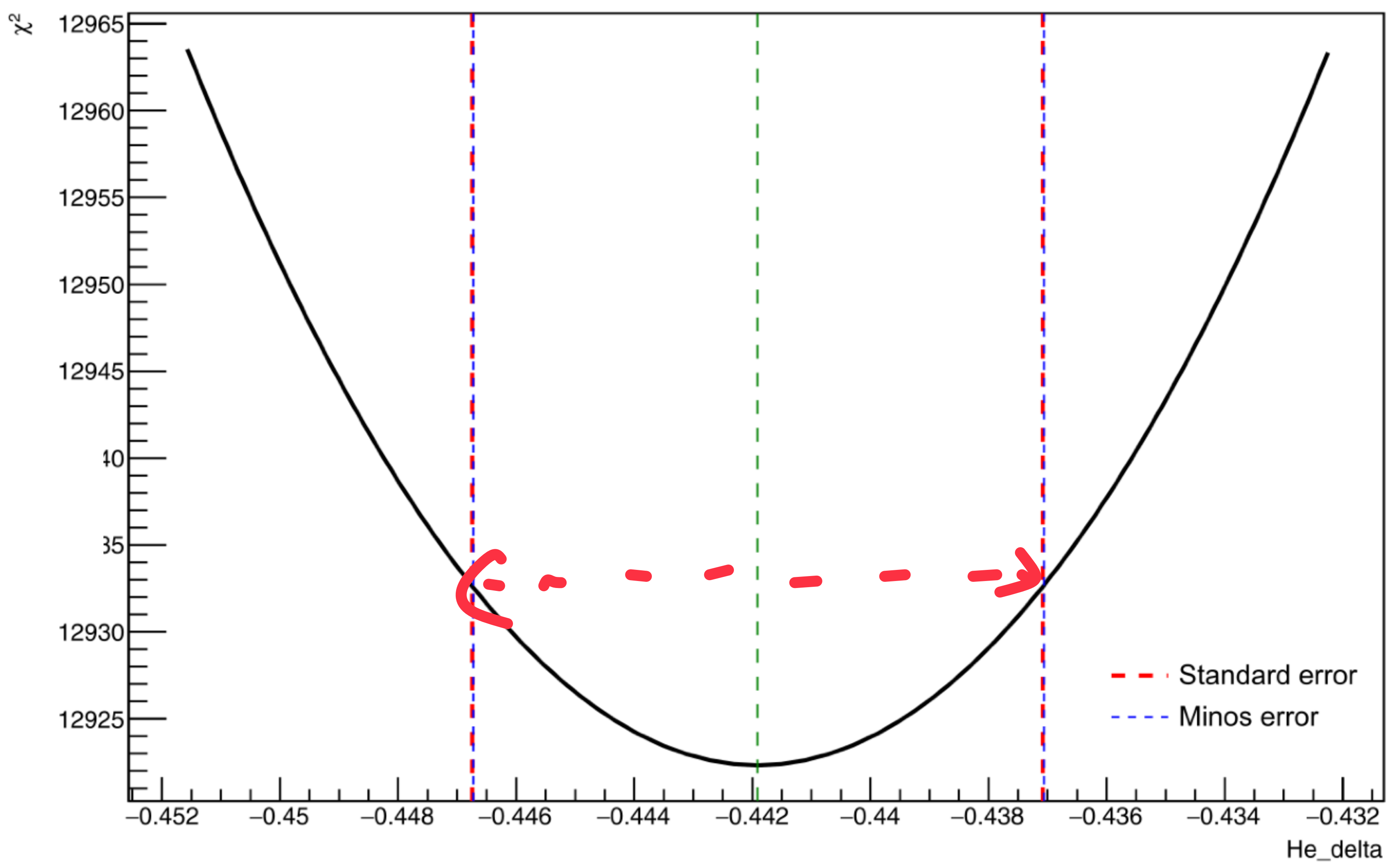
χ^2 / dof

$$\sigma_\theta^2 = 2 / \left(d^2\chi^2 / d\theta^2 \right)$$

$$\chi^2(\theta) = \chi^2_{\text{min}}(\theta_{\text{best}}) + 1;$$

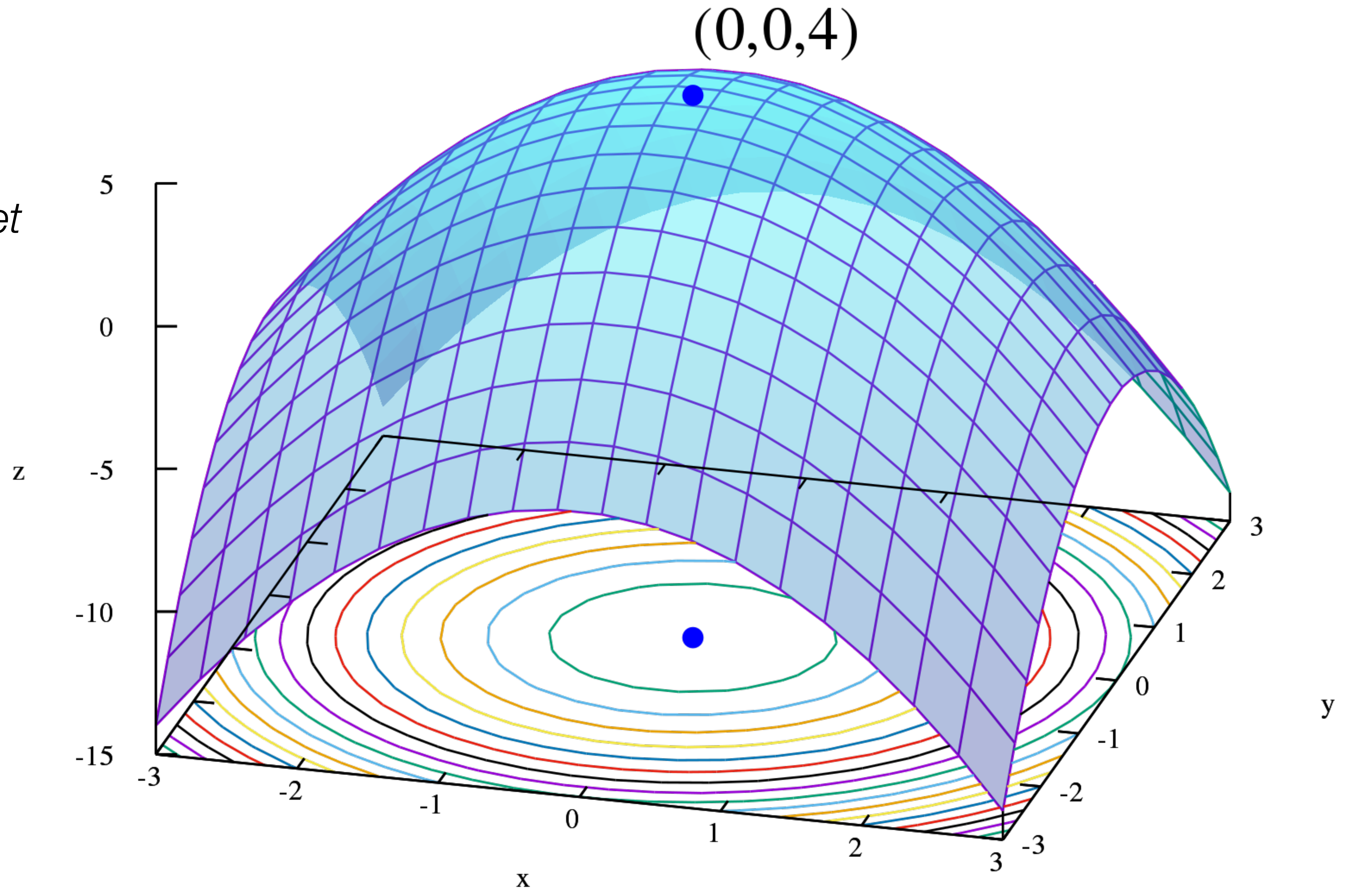


Chi-squared scan



What is Function Minimisation?

"In the more general approach, an optimization problem consists of maximizing or minimizing a real function by systematically choosing input values from within an allowed set and computing the value of the function." - **Wikipedia**



https://en.wikipedia.org/wiki/Mathematical_optimization#/media/File:Max_paraboloid.svg

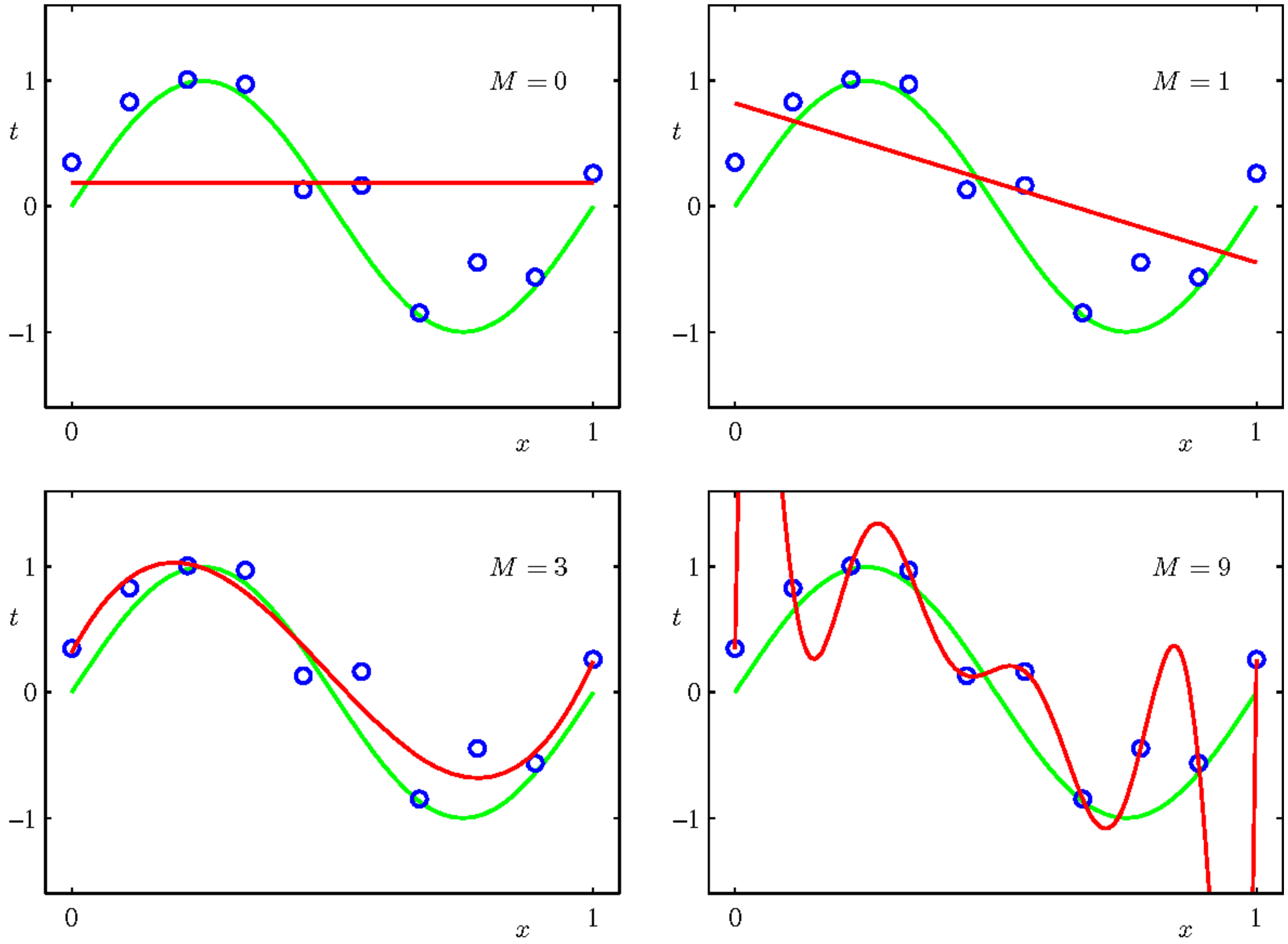
… and why do we want to do it?

Although there are many reasons why you might want to do this

But the most obvious of these is is function fitting

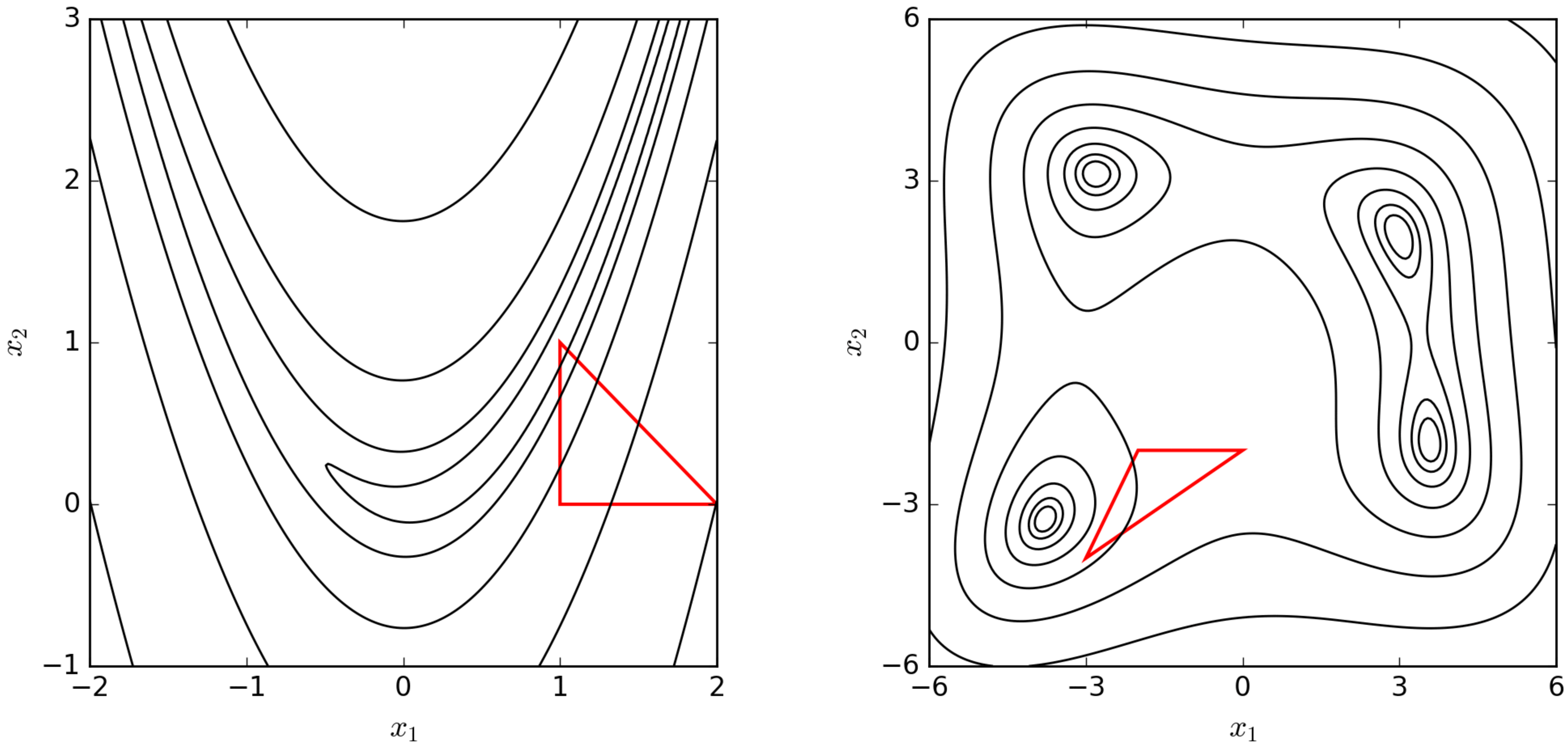
Minimise objective function describing difference between model and data

Chi2, likelihood etc

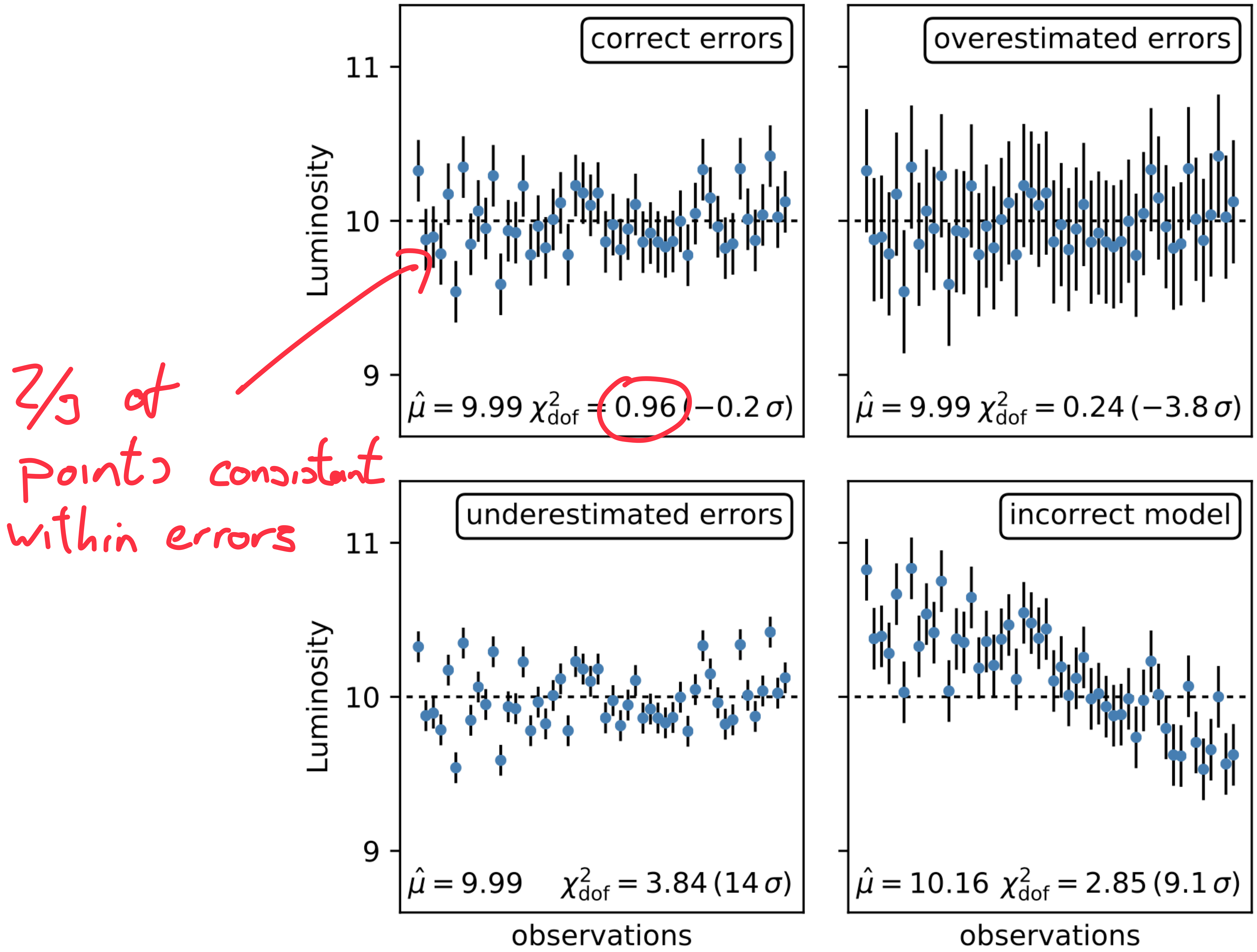


https://www.semanticscholar.org/paper/Introduction-1.1.-Example%3A-Polynomial-Curve-Fitting/001cc52a63e9ed49b7a3c29dc5c1a0dc16406049

Let's start Simple(x)



https://en.wikipedia.org/wiki/Nelder%E2%80%93Mead_method#/media/File:Nelder-Mead_Himmelblau.gif
https://en.wikipedia.org/wiki/Nelder%E2%80%93Mead_method#/media/File:Nelder-Mead_Rosenbrock.gif


correct errors
overestimated errors
underestimated errors
incorrect model
Luminosity
11
10
9
observations
$\hat{\mu} = 9.99\ \chi^2_{\text{dof}} = 0.96\ (-0.2\,\sigma)$
$\hat{\mu} = 9.99\ \chi^2_{\text{dof}} = 0.24\ (-3.8\,\sigma)$
$\hat{\mu} = 9.99\quad \chi^2_{\text{dof}} = 3.84\ (14\,\sigma)$
$\hat{\mu} = 10.16\ \chi^2_{\text{dof}} = 2.85\ (9.1\,\sigma)$
2/3 of points consistent within errors

Likelihood

PDF constructed with x as variable

$$x \mapsto f(x \mid \theta),$$

$$\theta \mapsto f(x \mid \theta),$$

Likelihood asks question, how likely is my data given this model

We can create a likelihood function by assuming a PDF then comparing model to datapoints

$$L(\theta) = \prod_i^N p(x_i, \theta)$$

or $$\ln L(\theta) = \sum_i^N \ln p(x_i, \theta)$$

Likelihood

Assume we flip a coin 3 times, what is the $L(HHT, p_H = 0.5)$

$$= 0.5^3 = 0.125$$

But we can test other assumptions about the coin

$$L(HHT, p_H = 1) = 1 \times 1 \times 0 = 0$$

$$L(HHT, p_H = 0.25) = 0.25 \times 0.25 \times 0.75 = 0.045$$

We can learn something about coin!

Most time you will see this as $-2 \log L = TS$

Maximum Likelihood

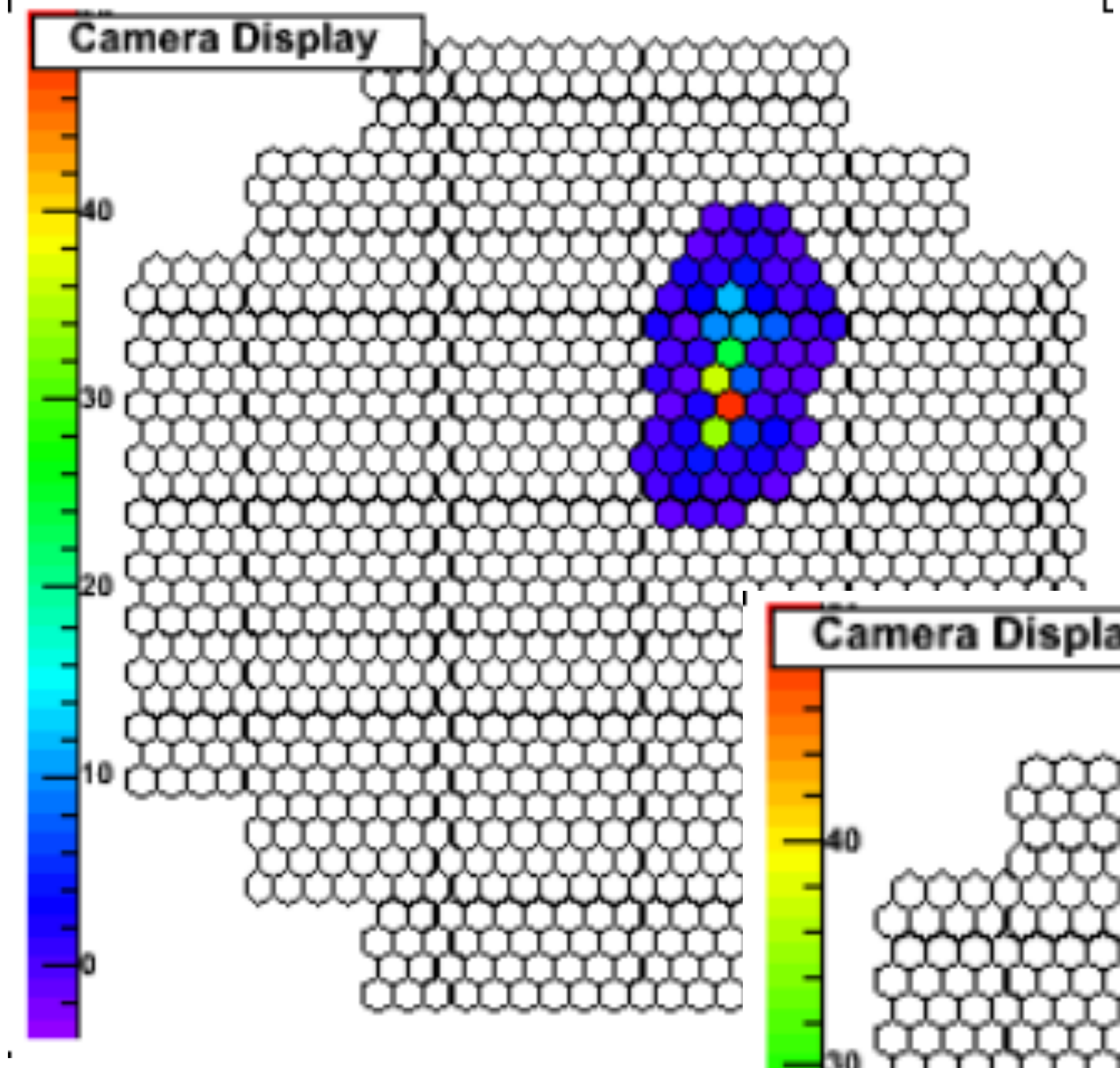
If I can construct a model + likelihood function I can test find which model parameters best fit my data

i.e. Maximise Likelihood

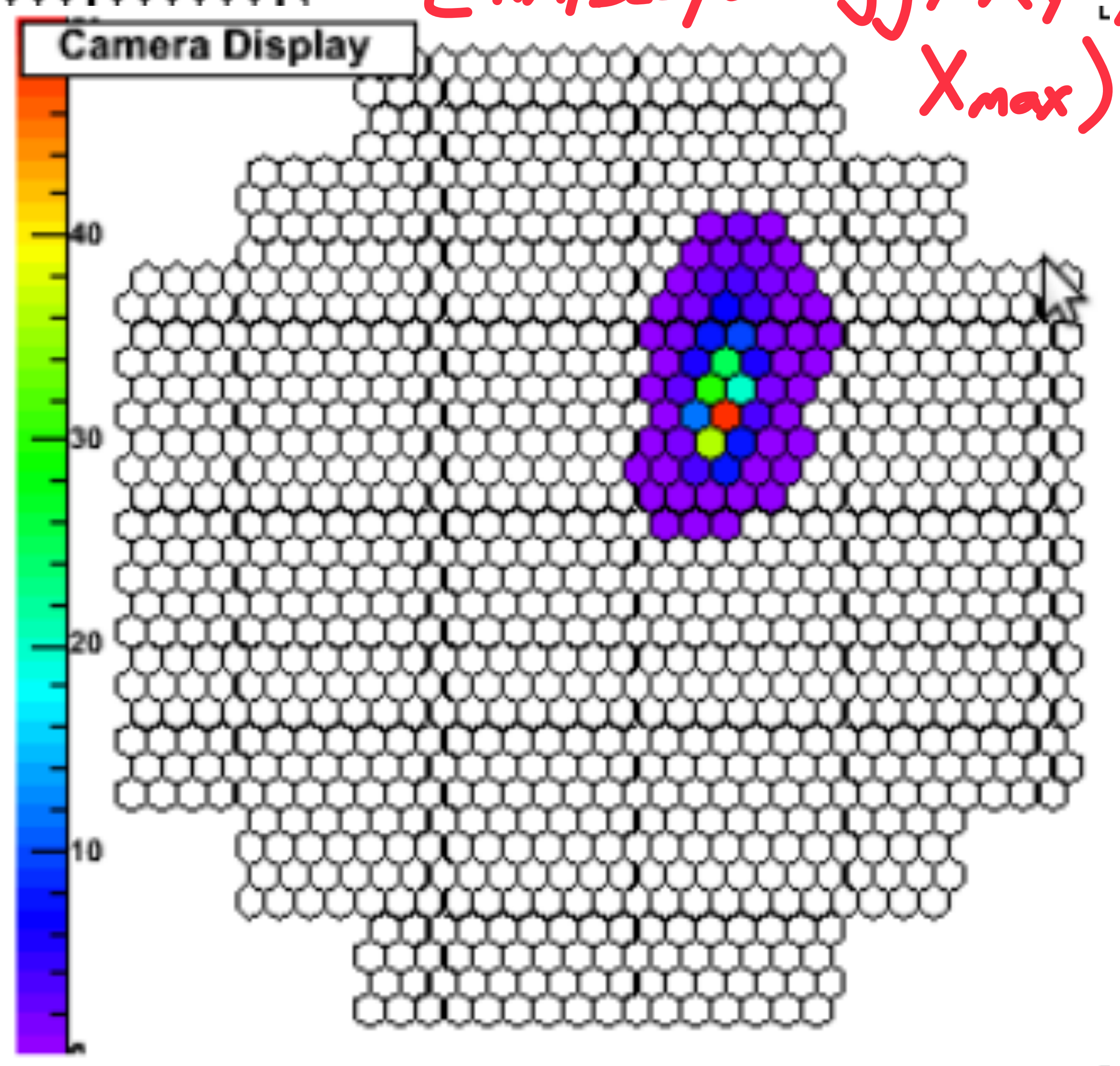
Usually you instead minimise

$-2 \ln L$

Image



Model expectation (RA, Dec, Energy, X, Y, X_{max})



Likelihood

$$P(s|\mu,\sigma_p,\sigma_\gamma) = \sum_n \frac{\mu^n e^{-\mu}}{n!\sqrt{2\pi(\sigma_p^2 + n\sigma_\gamma^2)}} \exp\left(-\frac{(s-n)^2}{2(\sigma_p^2 + n\sigma_\gamma^2)}\right)$$

Maximum Likelihood

We can then search model space using function minimisation like χ^2

In fact χ^2 is a special case of MLE

Calculate parameter errors the same, increase of $-2 \ln L$ of 1

Hypothesis Testing

Uncertainties

Now I have best fit values + uncert, how to compare to data

Lets say I get a ratio of Parameter/expectation of 0.97, is this compatible with theory

0.97 ± 0.005 $\quad$ 0.97 ± 0.05 $\quad$ 0.97 ± 0.5

✗ $\quad$ ✓ $\quad$?

Uncertainties

Statistical uncertainties

- Random in nature.
- Fluctuates independently per measurement.
- Unavoidable.
- Usually, more data $\to$ lower uncertainty ($\propto \sqrt{N}$).
- e.g., counting statistics, electronic noise, etc.

Systematic uncertainties

- Usually originate in the instrument.
- Bias the data by unknown $\sim$constant offset.
- Hard to detect, correct for, estimate.
- e.g., miscalibration, diff. between data and simulation, simulation statistics, etc.

can occur on data + theoretical expectation

Normally report these separately

i.e. $0.97 \pm 0.005_{stat} \pm 0.03_{sys}$

Uncertainties

$z = ax \pm b$	$\delta z = a \cdot \delta x$
$z = x \pm y$	$\delta z = [(\delta x)^2 + (\delta y)^2]^{\frac{1}{2}}$
$z = cxy$	$\frac{\delta z}{z} = \left[\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2\right]^{\frac{1}{2}}$
$z = c\frac{x}{y}$	$\frac{\delta z}{z} = \left[\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2\right]^{\frac{1}{2}}$
$z = cx^a$	$\frac{\delta z}{z} = a\frac{\delta x}{x}$
$z = cx^a y^b$	$\frac{\delta z}{z} = \left[\left(a\frac{\delta x}{x}\right)^2 + \left(b\frac{\delta y}{y}\right)^2\right]^{\frac{1}{2}}$

Errors can be propagated in the well known way

Only works if independent and gaussian

If not independent then covariance matrix needed

Hypothesis Testing

An example, measuring Higgs properties

Null model H_0, SM only

New physics H_1

Simple hypothesis

Calculate $\mathcal{L}(H_0(\theta))$

- > decide if data is likely for H_0 (p-value).
- > If not, claim discovery (of what?)
- → Existence of a particle (Higgs, new particle)
- → A new γ-ray source.

Composite hypothesis

compare $\mathcal{L}(H_0(\theta))$ and $\mathcal{L}(H_1(\theta))$.

- > Usually likelihood ratio is used.
- > More sensitive to H_1.
- > Based on p-values, which H_i is more likely.
- → Particle with certain mass, width, coupling constants.
- → Position and spectra of γ-ray source.

Hypothesis Testing

		True State of Nature	
		H_0 is true	H_0 is false
Our Decision	Do not reject H_0	Correct decision	Type II error
	Reject H_0	Type I error	Correct decision

Type I error rate (significance α)

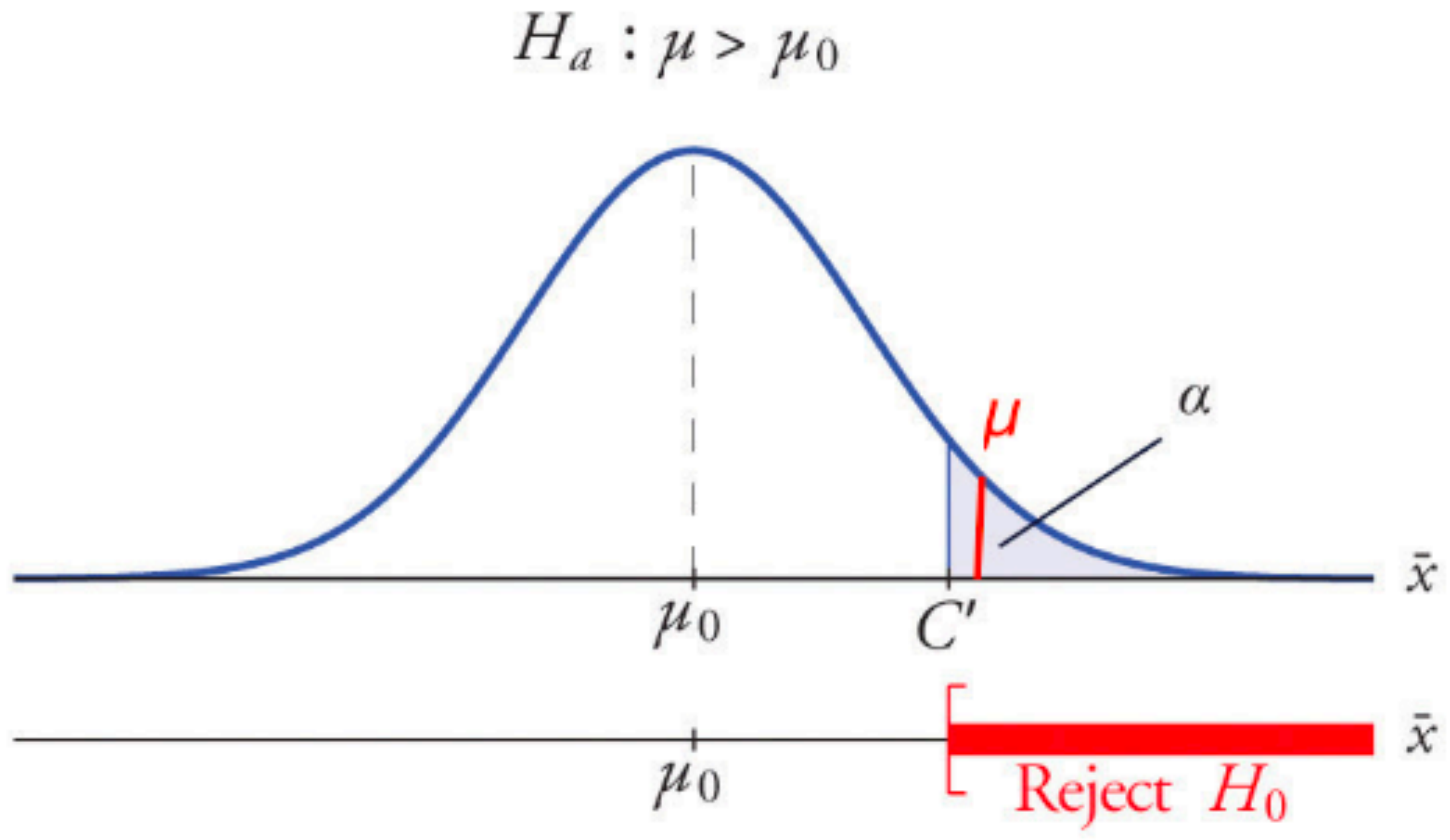
$$\alpha = \int_{x \geq x_0} p(x \mid H_0)\, dx$$

P-value

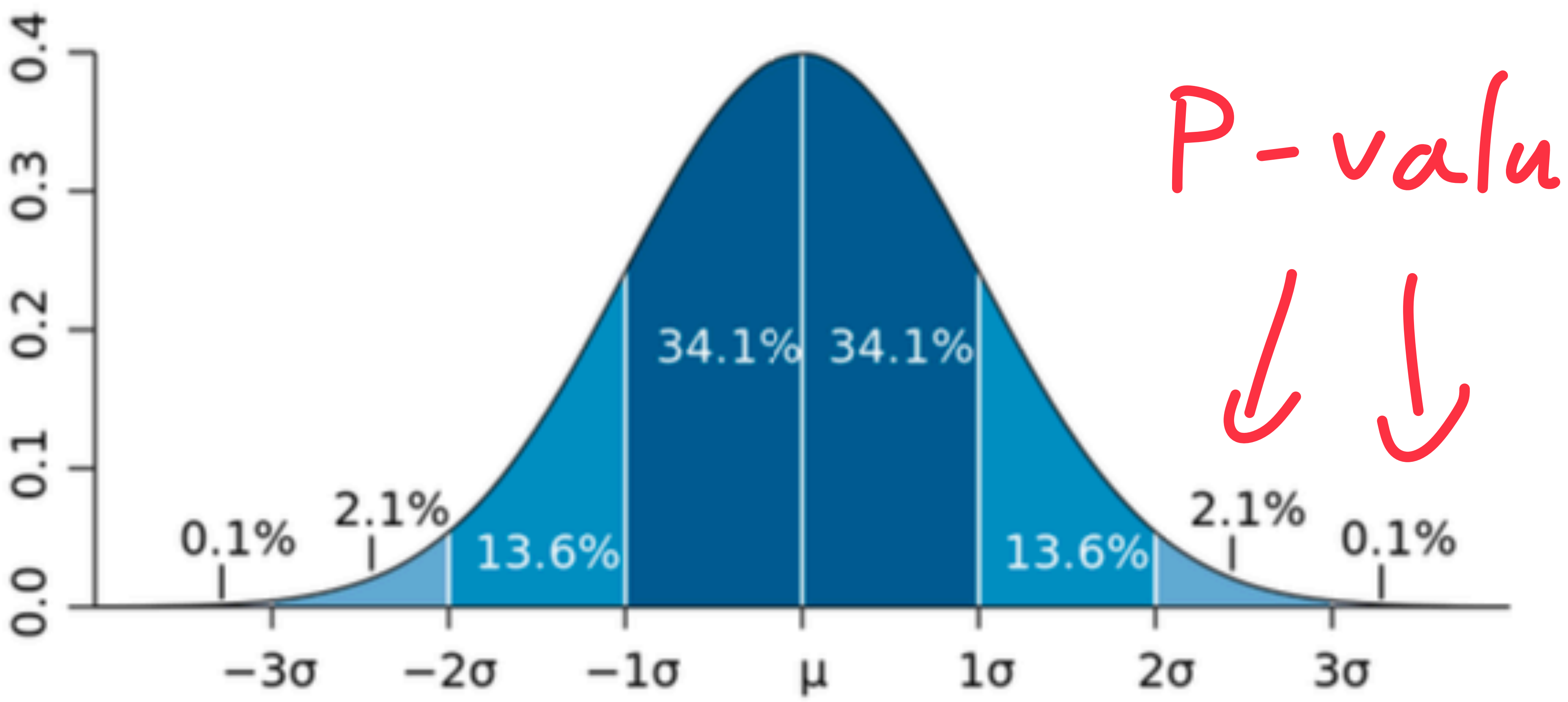
Hypothesis Testing

Compare my result to the expected distribution of H_0

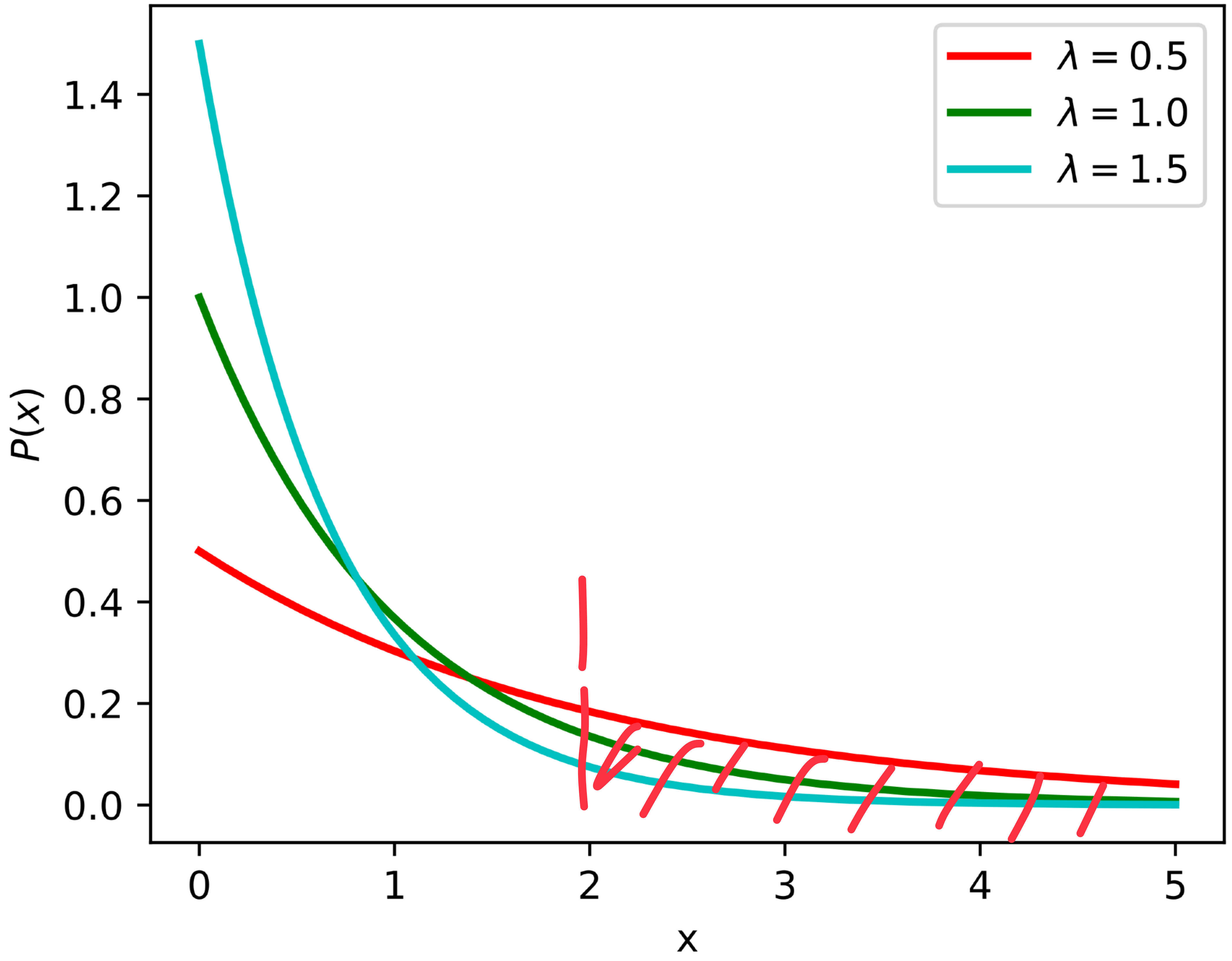
If it lies far from expectation we can integrate beyond this to get p-value



Can be one-sided or two-sided depending on the question I ask

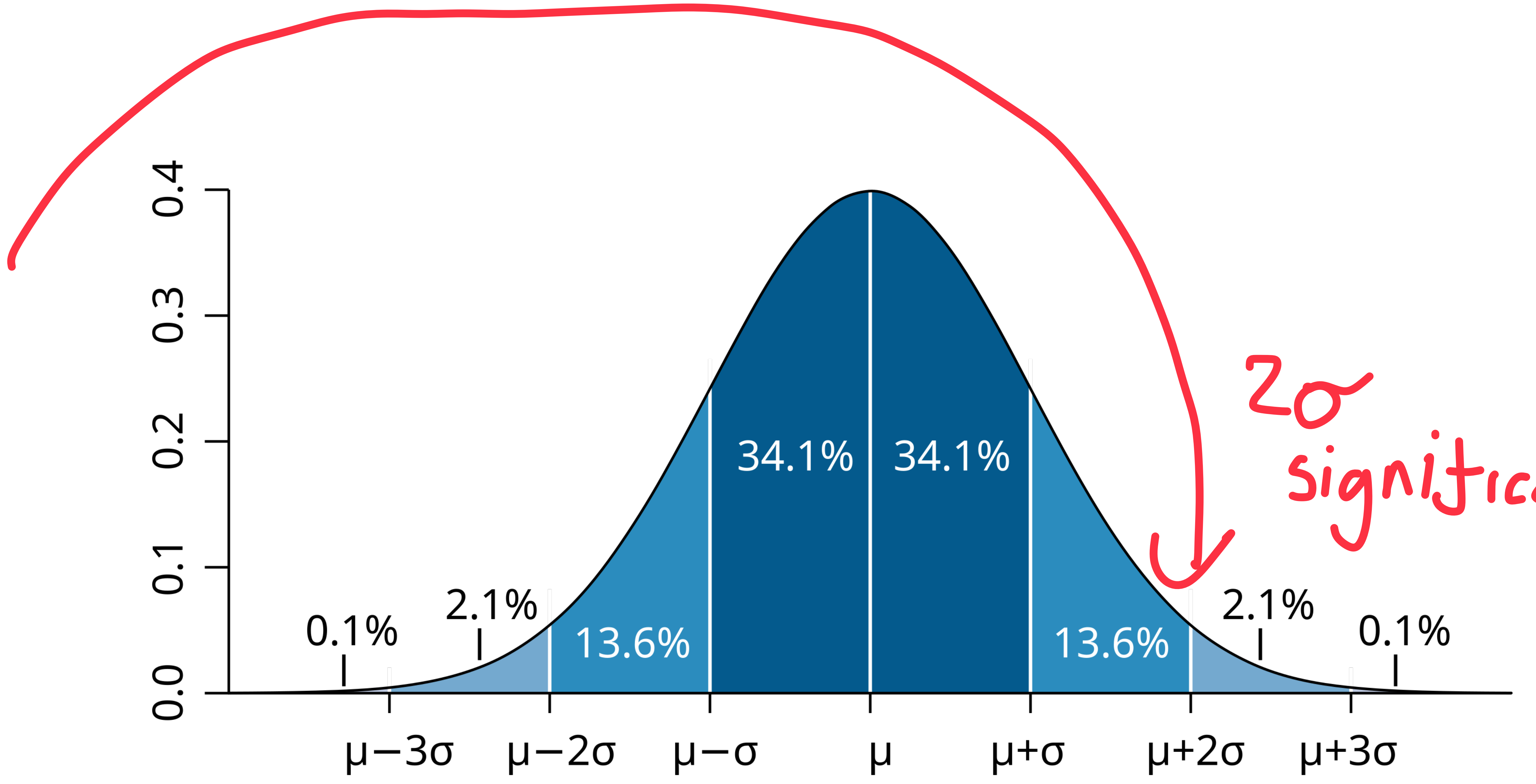


Hypothesis Testing



P-values not just for gaussians

We can calculate for any distribution and convert to equivalent gaussian number of "sigmas"



2σ significance

Hypothesis Testing

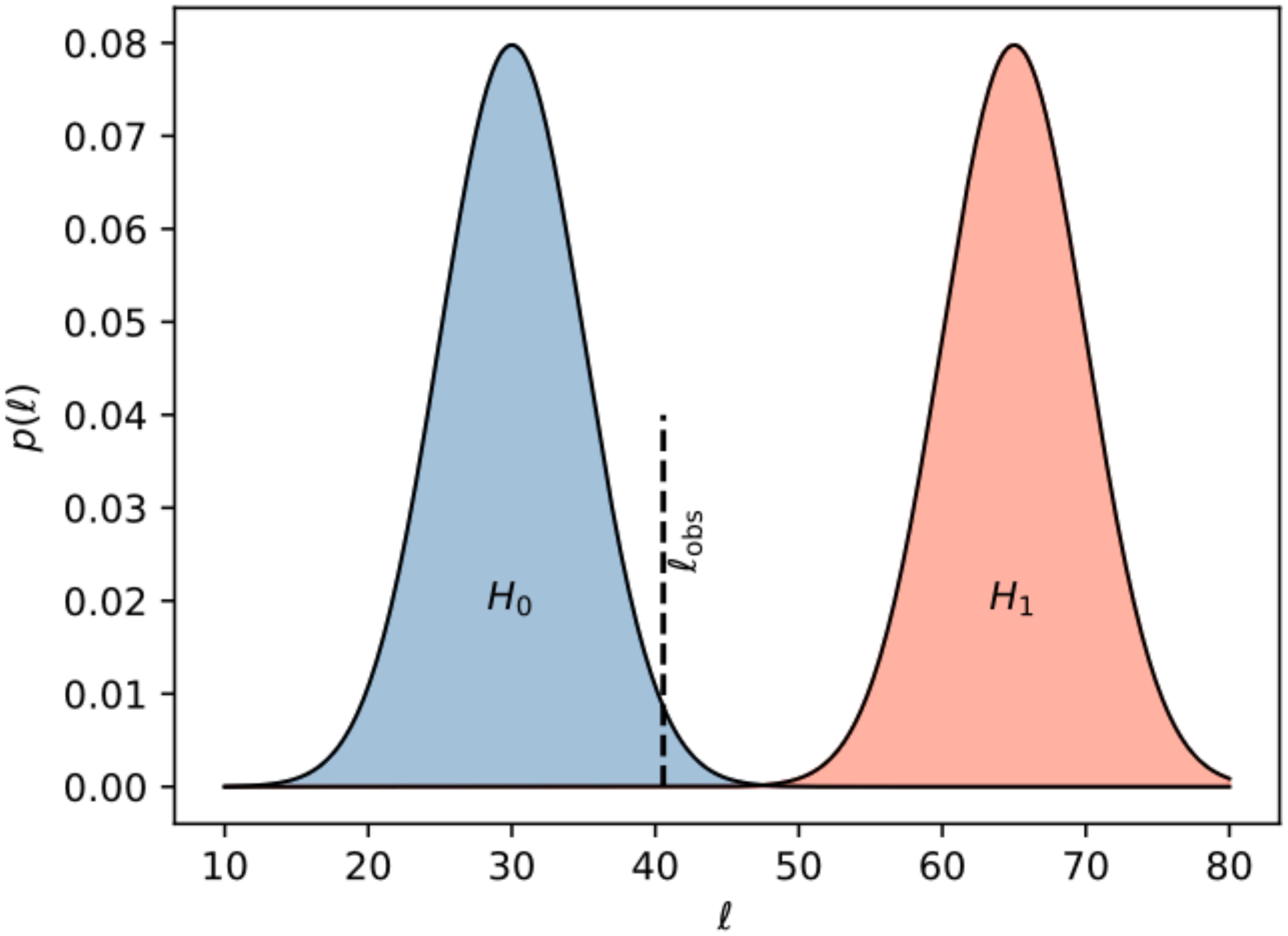
It's often more powerful to compare one hypothesis to another

i.e. which is more likely H_0 or H_1 ?

Liklihood ratio test

$$\ell = \Delta TS = -2 \ln \left(\frac{L_{H_1}}{L_{H_0}} \right)$$

$$= -2 \left(\ln L_{H_0} - \ln L_{H_1} \right)$$

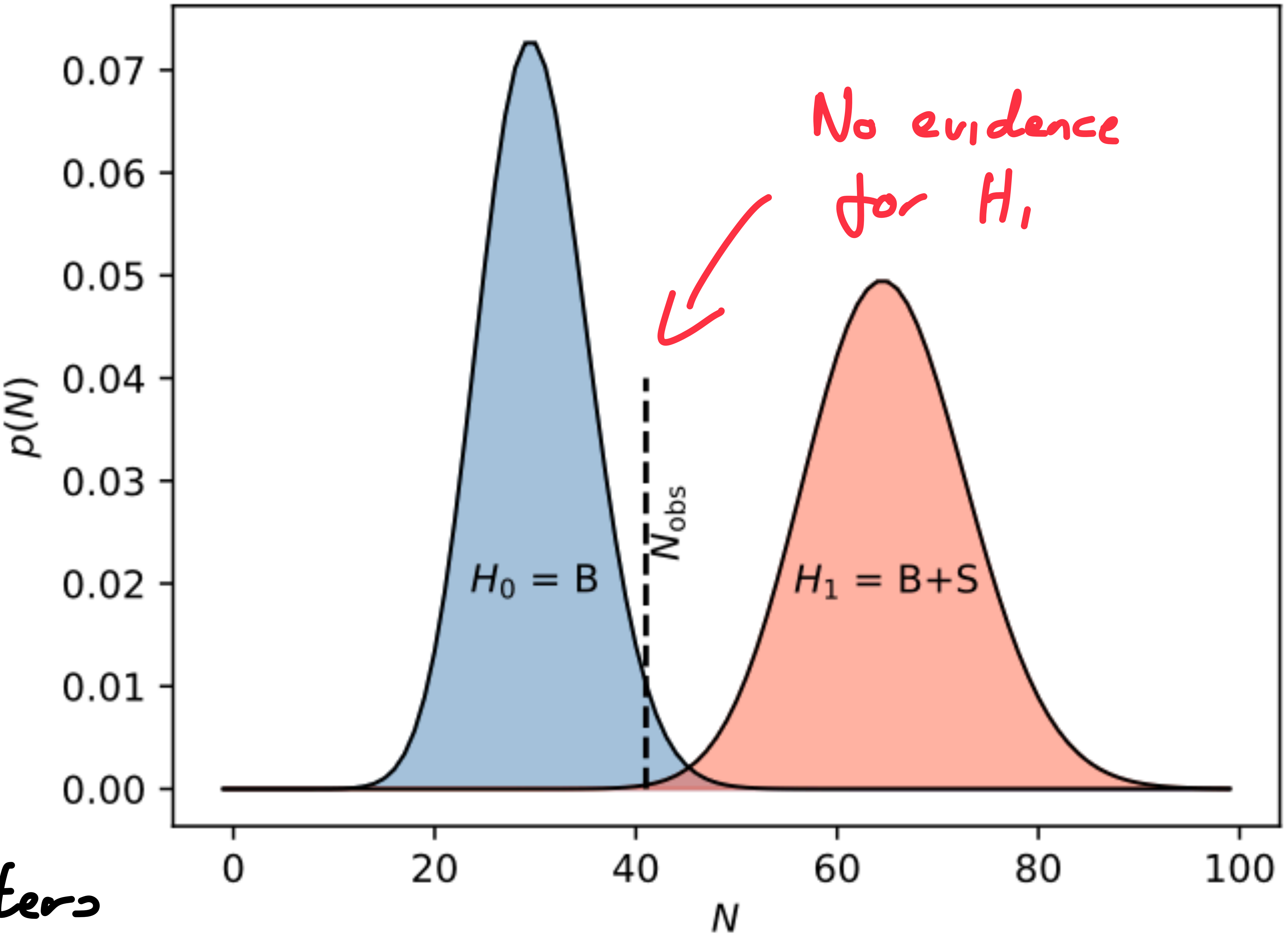


Hypothesis Testing

H_0 is our background, H_1 is signal + background

Perform ML fit for both models calculate likelihood ratio

ℓ should be distributed as a chi-squared distribution with n dof of number of free parameters different



No evidence for H_1

e.g. $P(\theta_1, \ldots, \theta_n)$

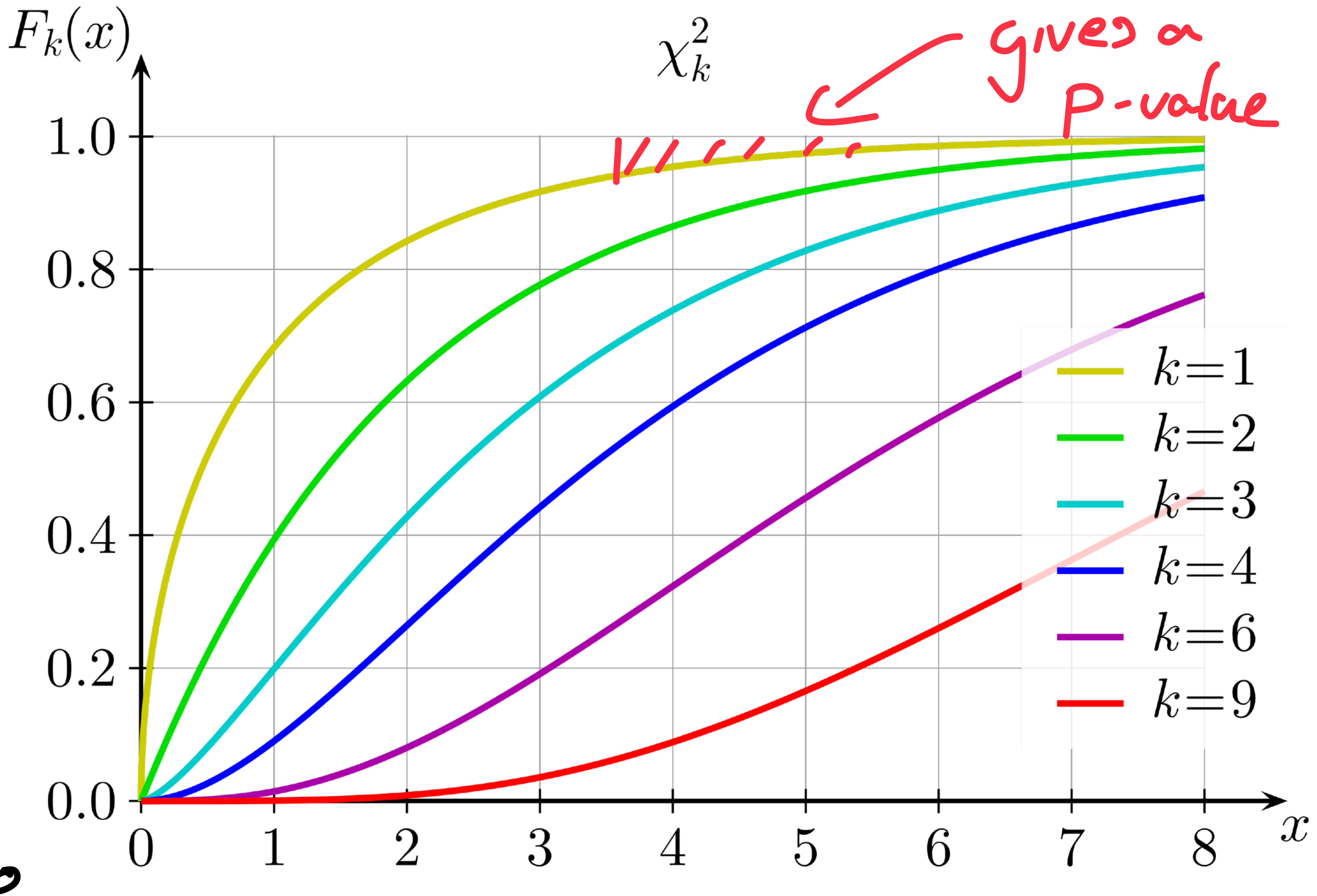
vs $P(\theta_1 \ldots \theta_{n+1})$

Only works on nested models!

Hypothesis Testing

$$l = -2 \ln \left(\frac{L_{H_1}}{L_{H_0}} \right)$$

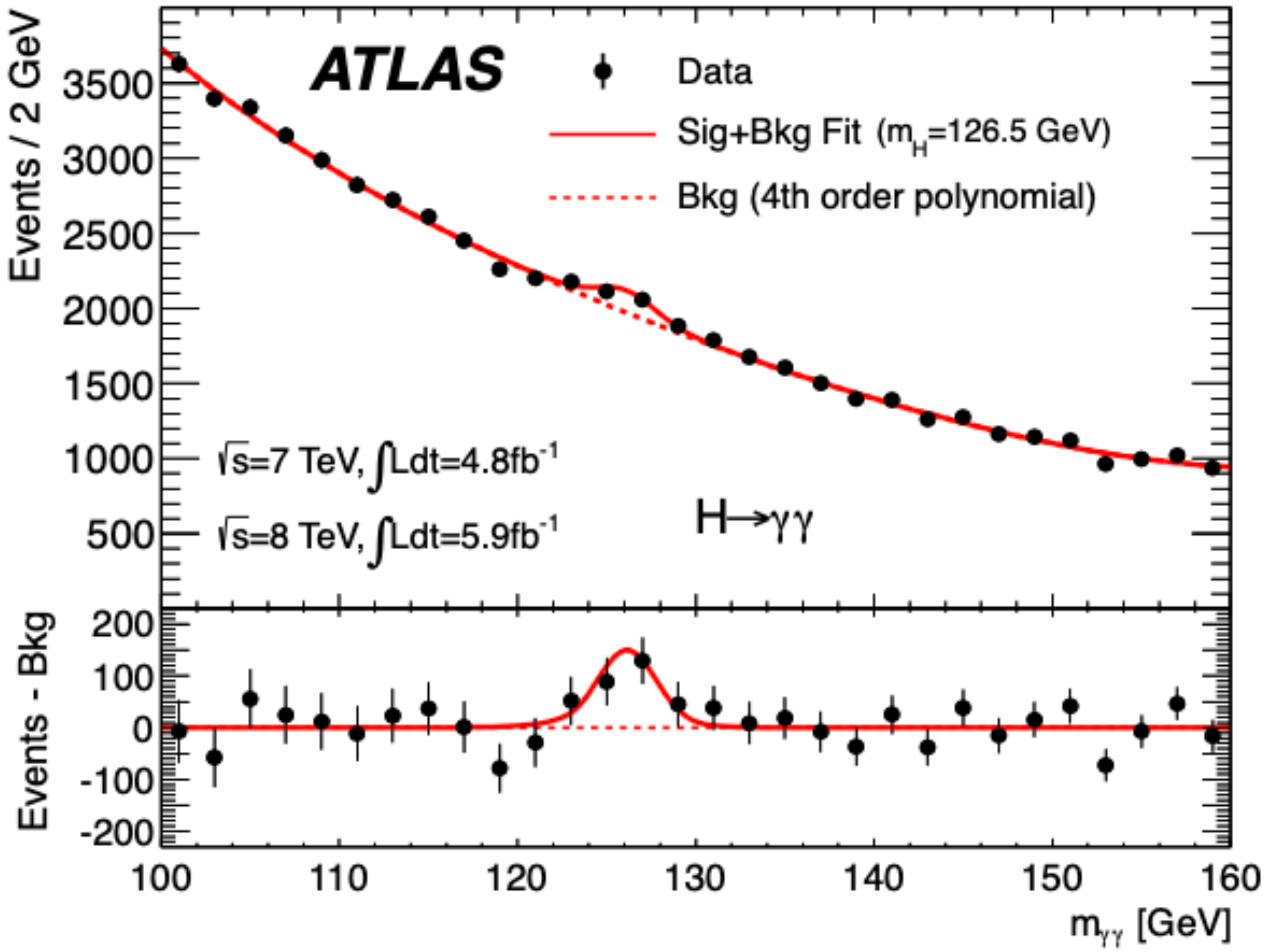
Distributed as χ^2, K difference in free parameters

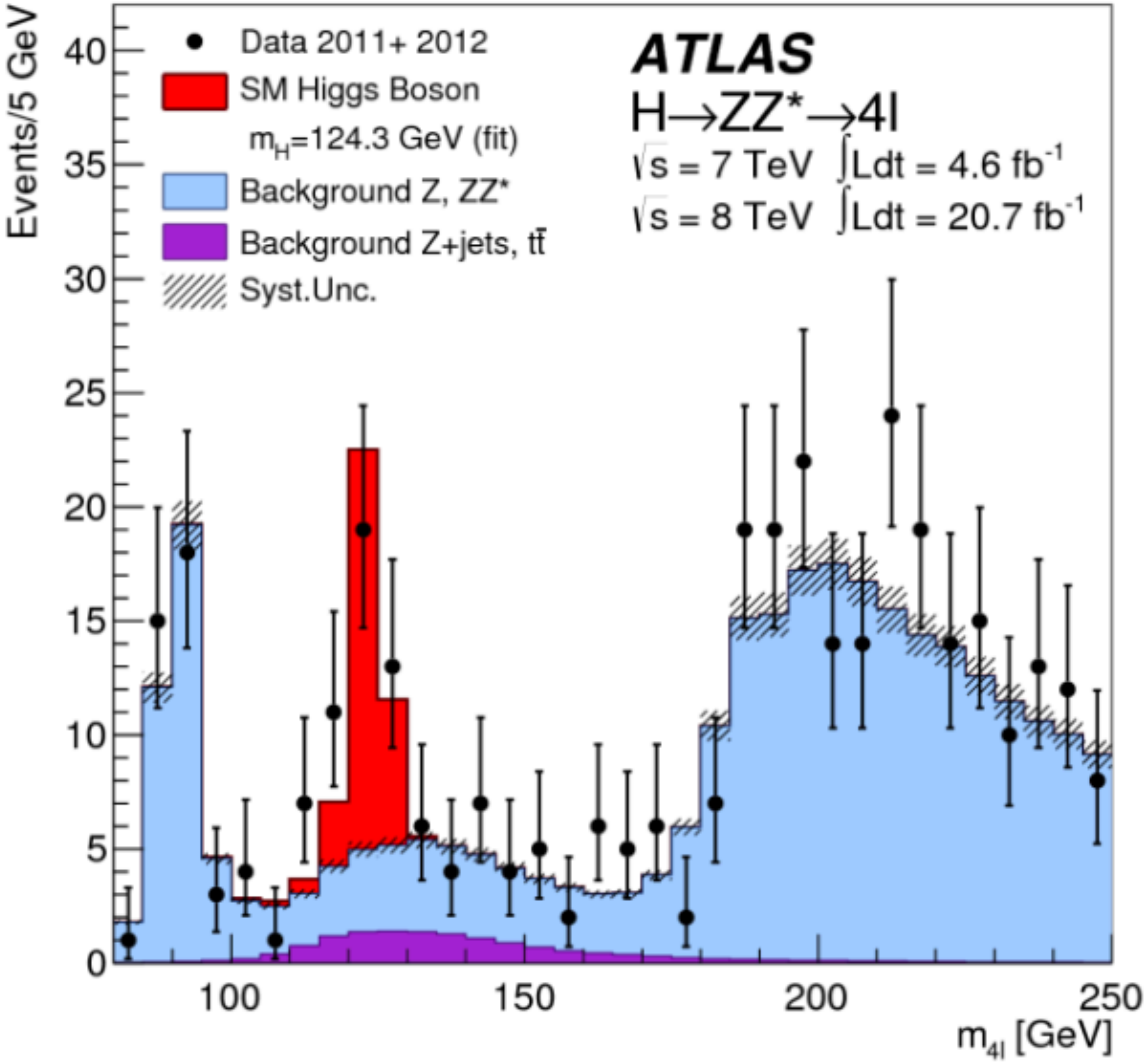


P-value can be converted to gaussian σ

If $K=1$ significance $\sim \sqrt{l}$

Hypothesis Testing

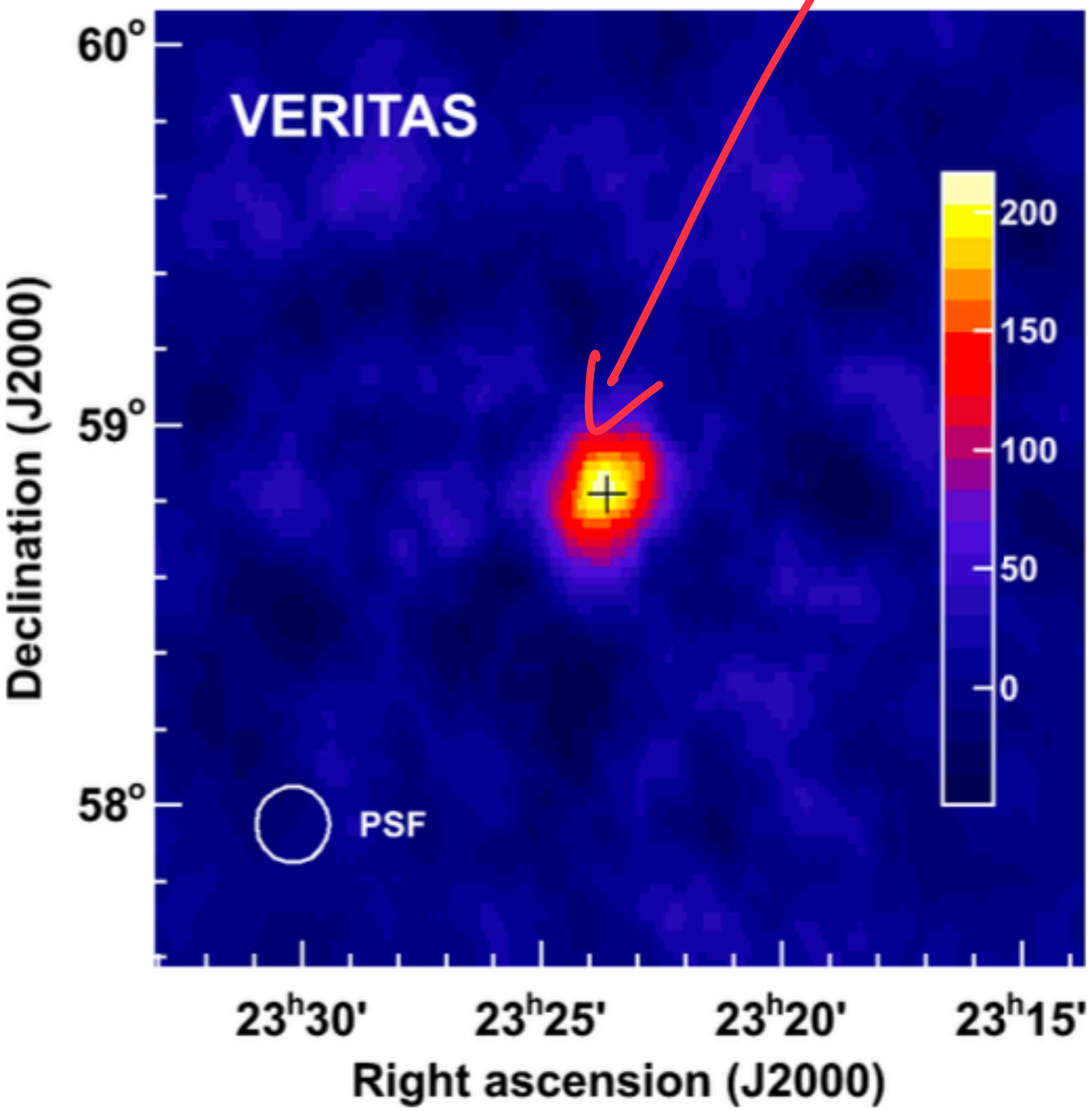


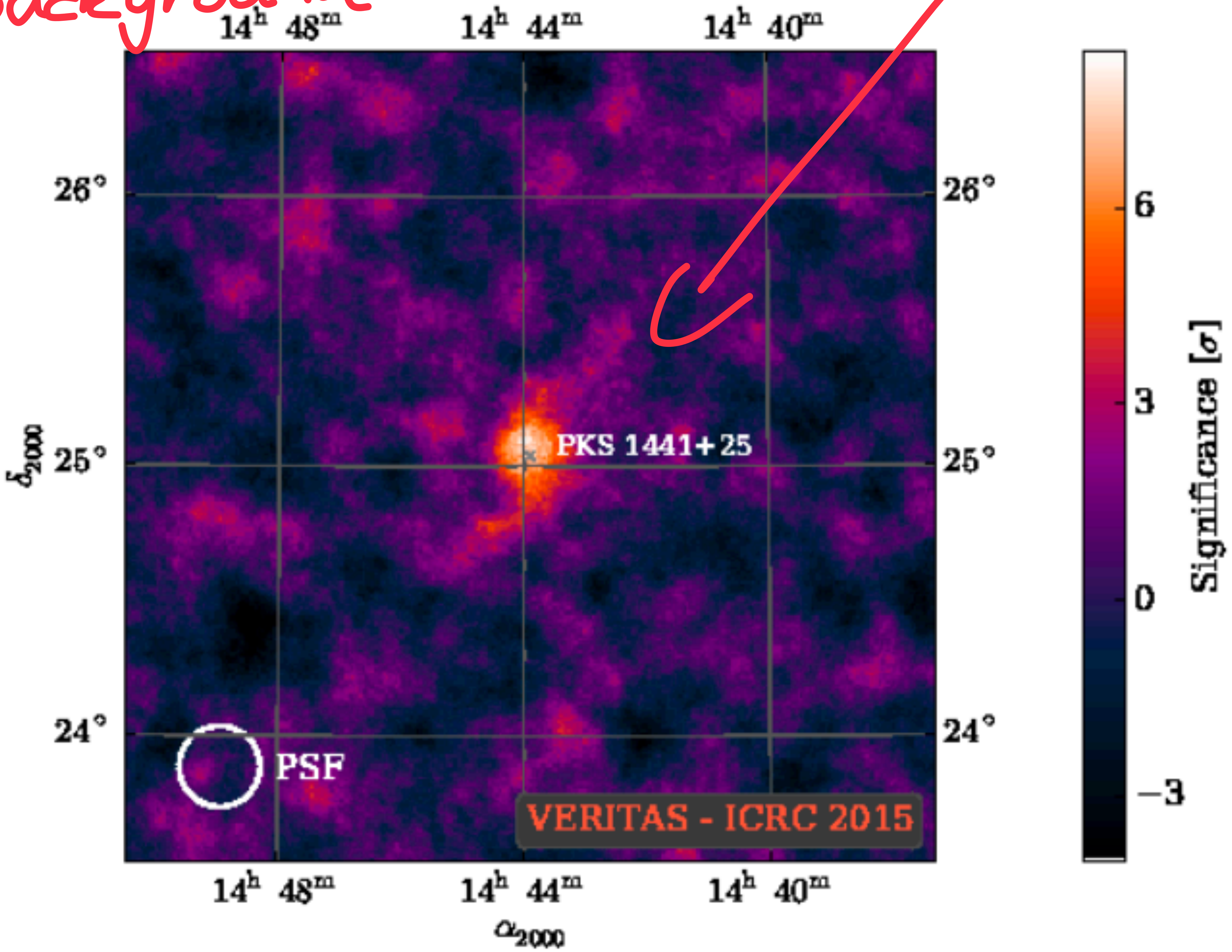


Hypothesis Testing

How likely is this to be a fluctuation of background

7 sigma not very...

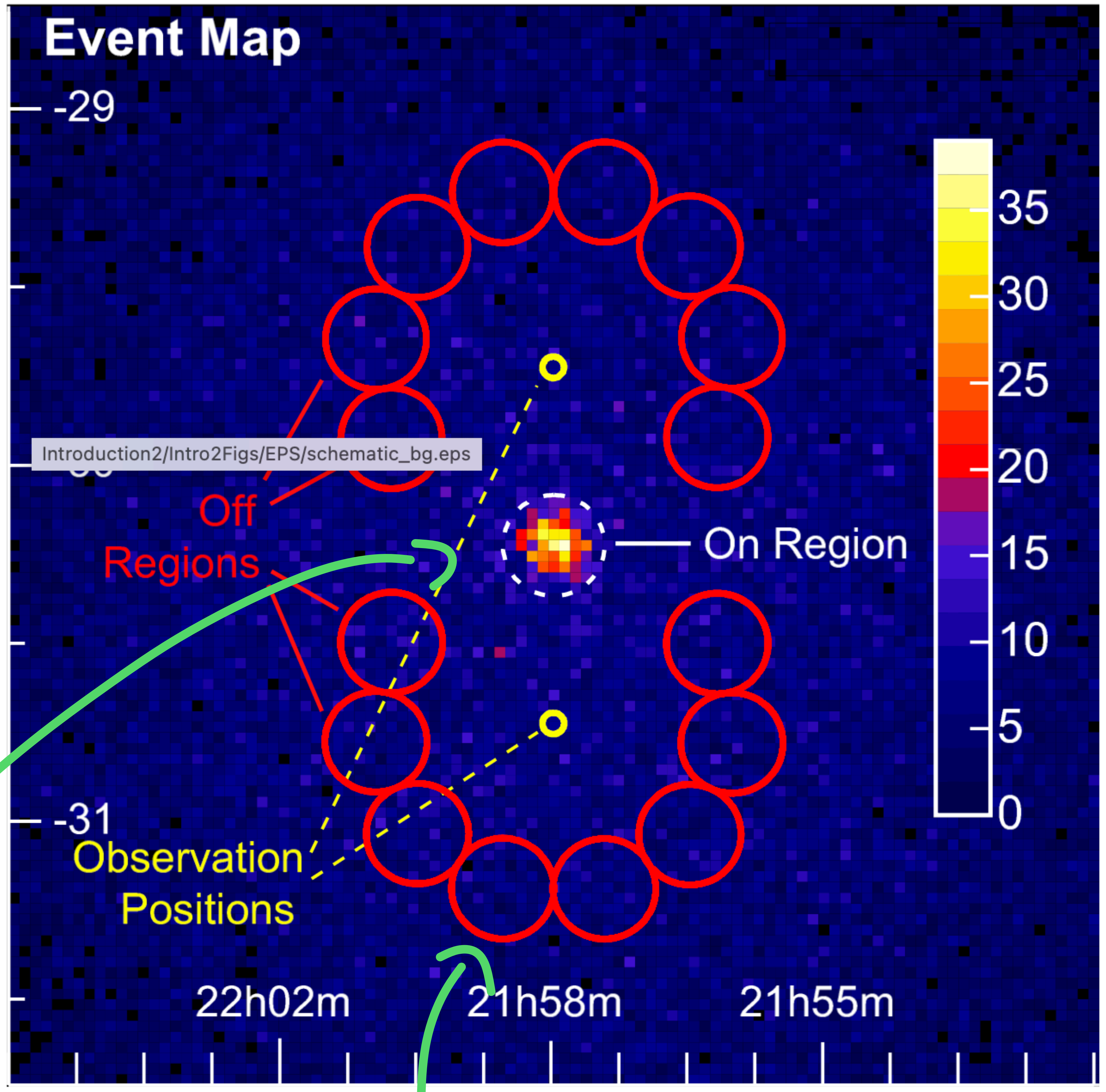




Hypothesis Testing

Can estimate source significance by Li and Ma formula

Likelihood ratio test of BG-only hypothesis vs BG + Signal



Signal "on" region

BG level from "off" regions

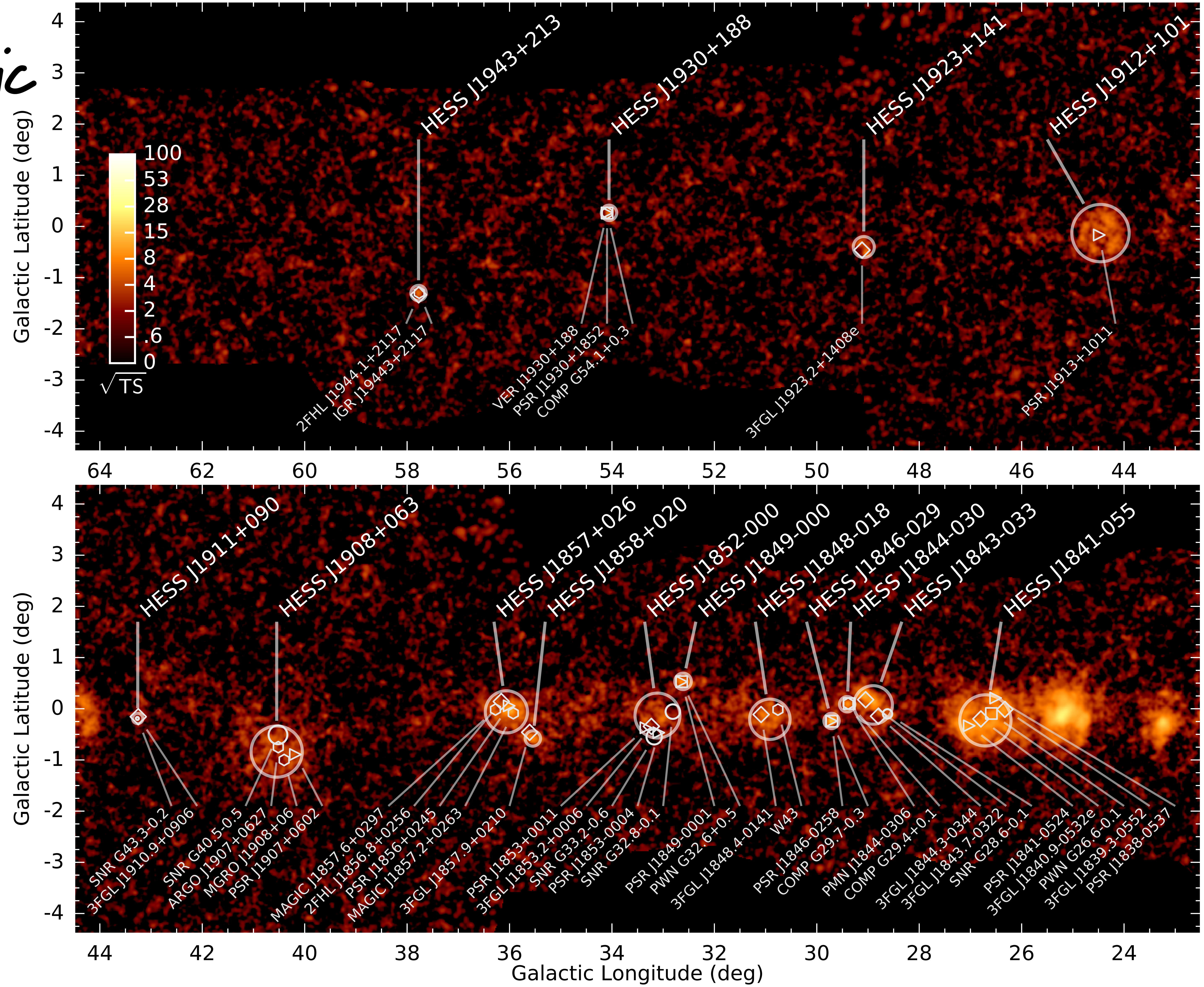
$$S = \sqrt{2}\left(N_{\text{on}}\ln\left[\frac{1+\alpha}{\alpha}\left(\frac{N_{\text{on}}}{N_{\text{off}}+N_{\text{on}}}\right)\right] + N_{\text{off}}\ln\left[(1+\alpha)\left(\frac{N_{\text{off}}}{N_{\text{off}}+N_{\text{on}}}\right)\right]\right)^{\frac{1}{2}}$$

Look elsewhere effect (trials factor)

HESS galactic plane scan

Test for a source at every pixel in map

Highly unlikely fluctuations become possible



10^5 trials?

Need to modify significance

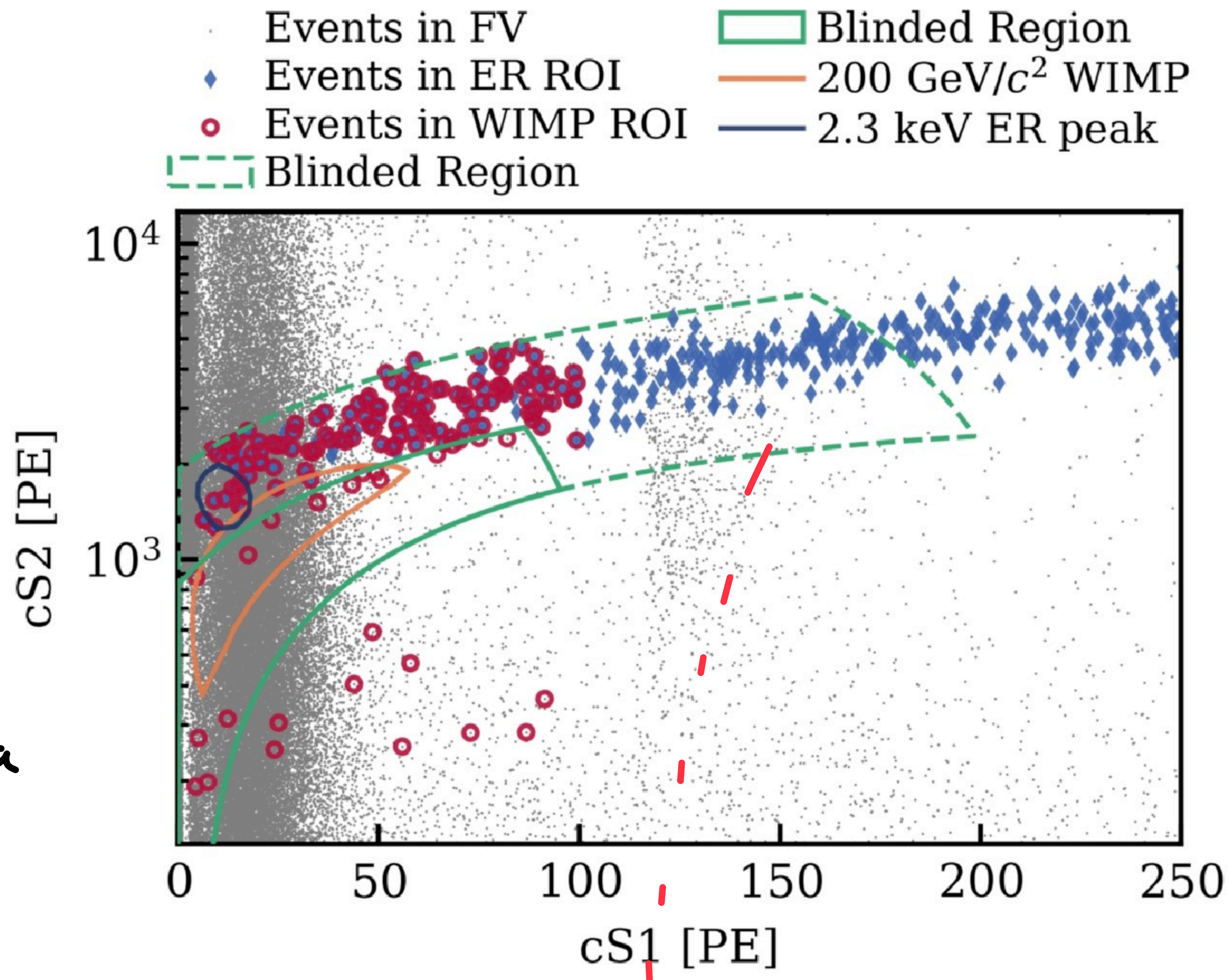
Blinding of analysis

In order to reduce the number of trials and increase confidence data can be "blinded"

Do not allow ourselves to observe expected signal region until analysis is fixed

Could optimise on small fraction of data

That data must then be discarded!



Don't look here!

Why 5 Sigma?

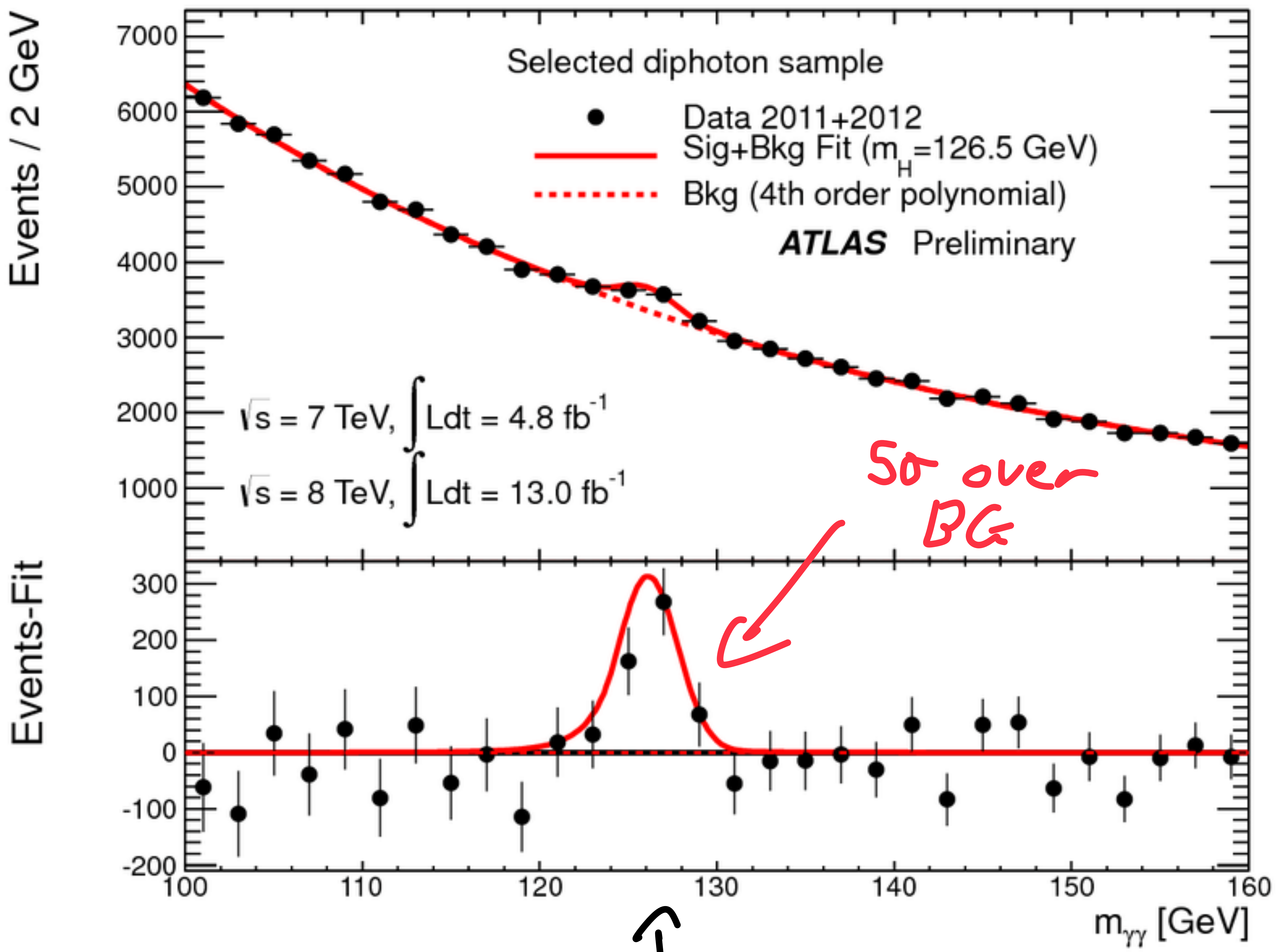
In the world of particle physics (+ astroparticle) 5 σ is the "gold standard"

p value < 1e^{-6}, very unlikely

NOT universal

- Astronomy 3σ
- Biology ~ 2σ etc

If its new physics we better be certain



If a few trials missed it doesn't change much

Maybe some systematics?

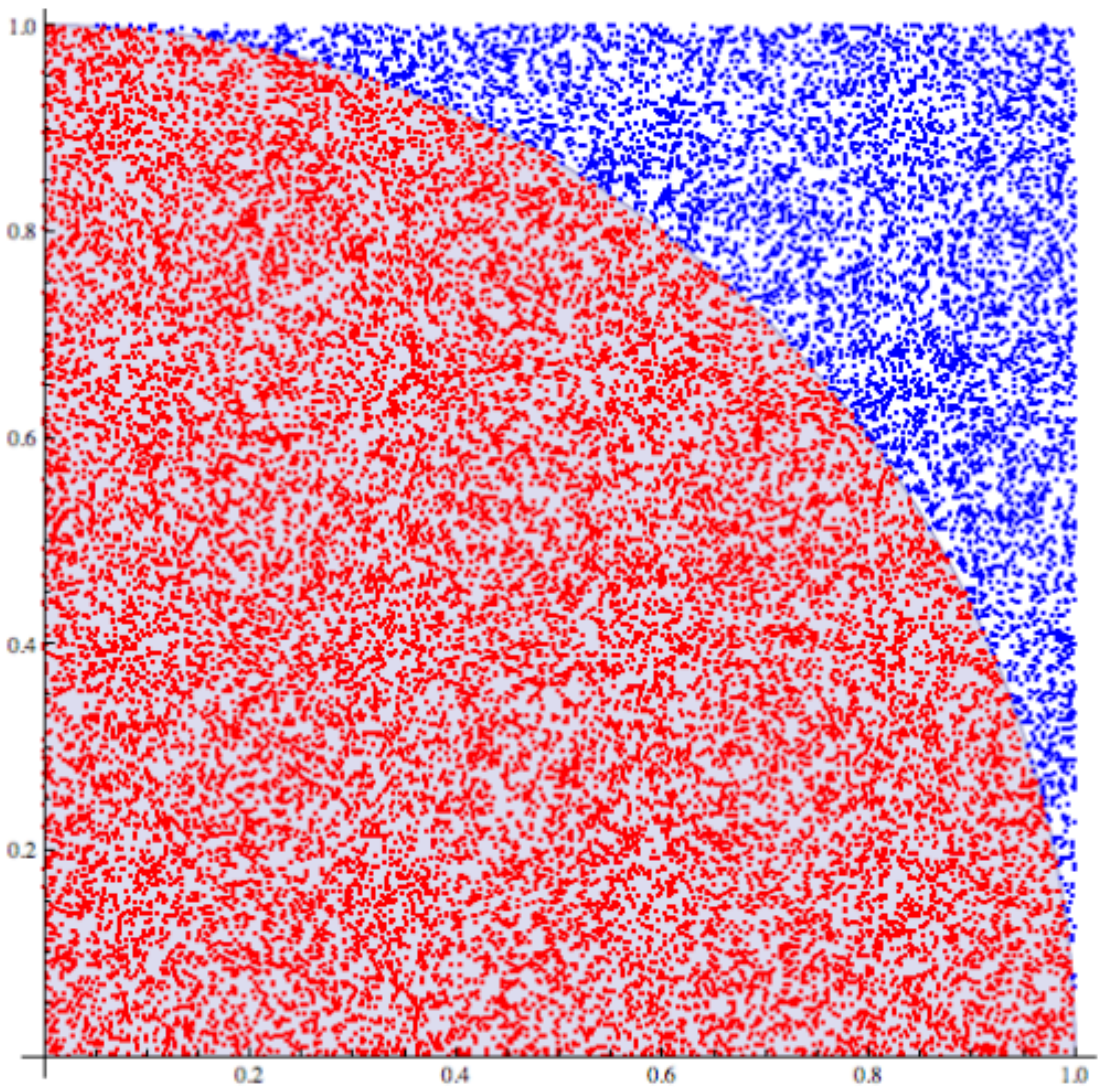
Monte Carlo Techniques

Monte Carlo methods

Algorithm that relies on repeated random sampling

Often used with complex non-linear systems

- Numerical integration
- Particle interactions + decay
- Uncertainty Estimation




```python
from random import random
from math import sqrt, pi
inside=0
n=10000000
i_print = [1, 10, 100, 1000, 10000, 100000, 1000000]
for i in range(0,n+1):
    x=random()
    y=random()
    if sqrt(x*x+y*y)<=1:
        inside+=1
    if i in i_print:
        piNow=4*inside/i
        print ('pi(i=%d) = %.4f, error = %.4f' % (i, piNow, abs(piNow - pi)))
```

```
pi(i=1) = 4.0000, error = 0.8584
pi(i=10) = 3.6000, error = 0.4584
pi(i=100) = 3.3600, error = 0.2184
pi(i=1000) = 3.1240, error = 0.0176
pi(i=10000) = 3.1264, error = 0.0152
pi(i=100000) = 3.1433, error = 0.0017
pi(i=1000000) = 3.1402, error = 0.0014
```

Monte Carlo integration

Standard computation -> split into sub-shapes + sum area

Trapezium rule , $N_{calc} = n^d$ ← dimensions

Can instead randomly sample functions

Integration error $\propto 1/\sqrt{N}$ → faster at large d