Basic & Advanced Statistical Techniques

CTAO School - 3rd edition

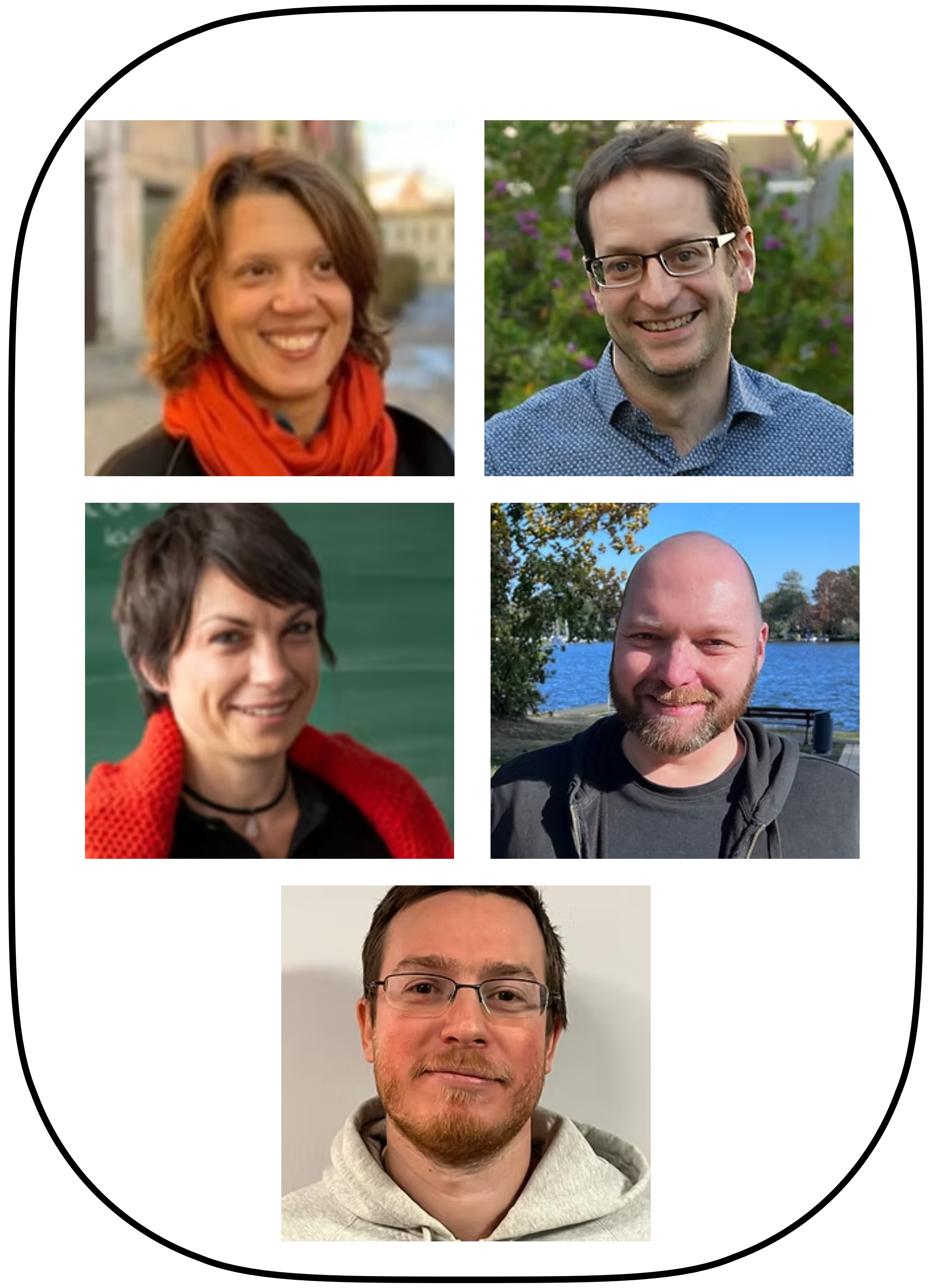
12-22 May 2026

Giacomo D’Amico

Institut de Física d'Altes Energies (IFAE), Barcelona, Spain

Why does it matter?

INSTRUMENT



DATA ANALYSIS

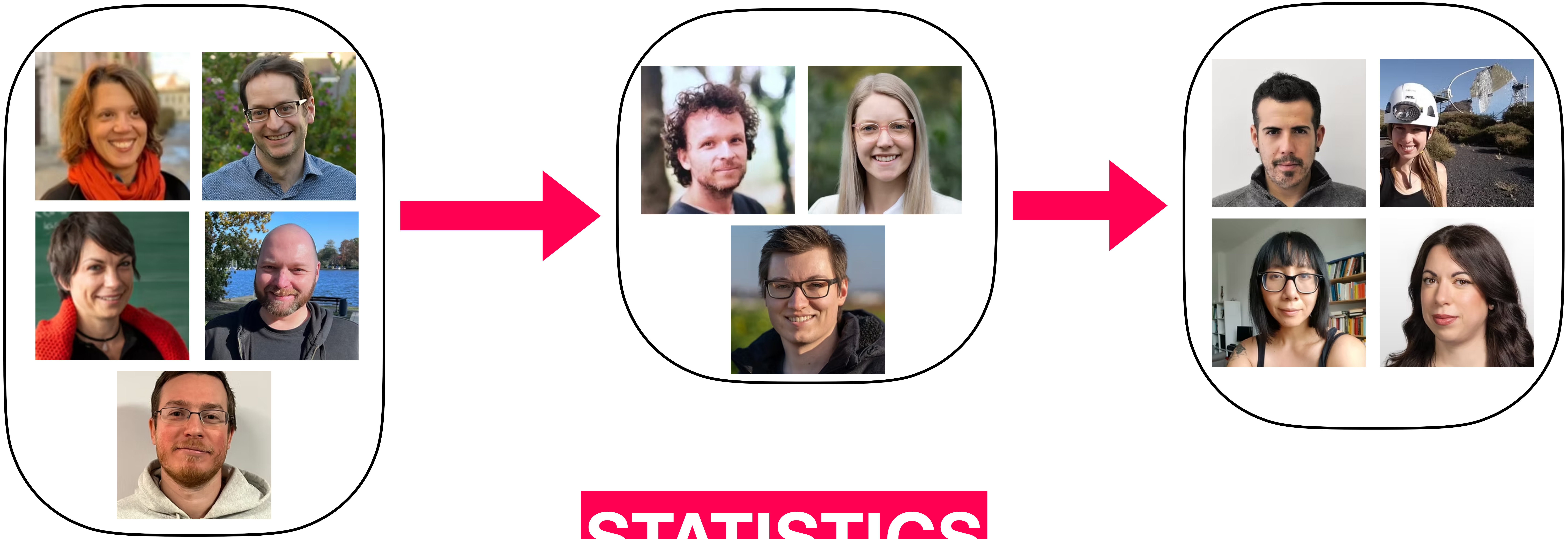
SCIENCE

Why does it matter?

INSTRUMENT

DATA ANALYSIS

SCIENCE



STATISTICS

Outline

- **Part I:** General introduction to statistics
    - Bayesian approach
    - Frequentist approach
    - Likelihood

- **Part II:** Applications in gamma-ray analysis
    - Estimating the excess
    - From the excess to the γ flux

PART I

Outline - Part I

- **The Bayesian approach**
  - priors, posteriors, sensitivity, specificity, etc…
- **The frequentist approach**
  - p-values, confidence levels, “sigmas”, etc…
- **The likelihood**
  - power of the test, the Wilks’ theorem, etc…

The Bayesian approach

- The **Bayesian approach** tries to answer the question:

*Given our **prior** knowledge and the observed **data**, what is the **probability** that the model is true?*

Probability theory

- Marginalised probability

$$p(x) = \int dy\, p(x, y) \qquad p(x) = \sum_i p(x, y_i)$$

- Conditional probability

$$p(x, y) = p(x\,|\,y) \cdot p(y)$$

Probability theory

- Marginalised probability

$$p(x) = \int dy\, p(x, y)$$

- Conditional probability

$$p(x, y) = p(x \,|\, y) \cdot p(y)$$

$$p(x \,|\, y) = \frac{p(y \,|\, x) \cdot p(x)}{\int dy\, p(y \,|\, x) \cdot p(x)}$$

Probability theory

- Marginalised probability

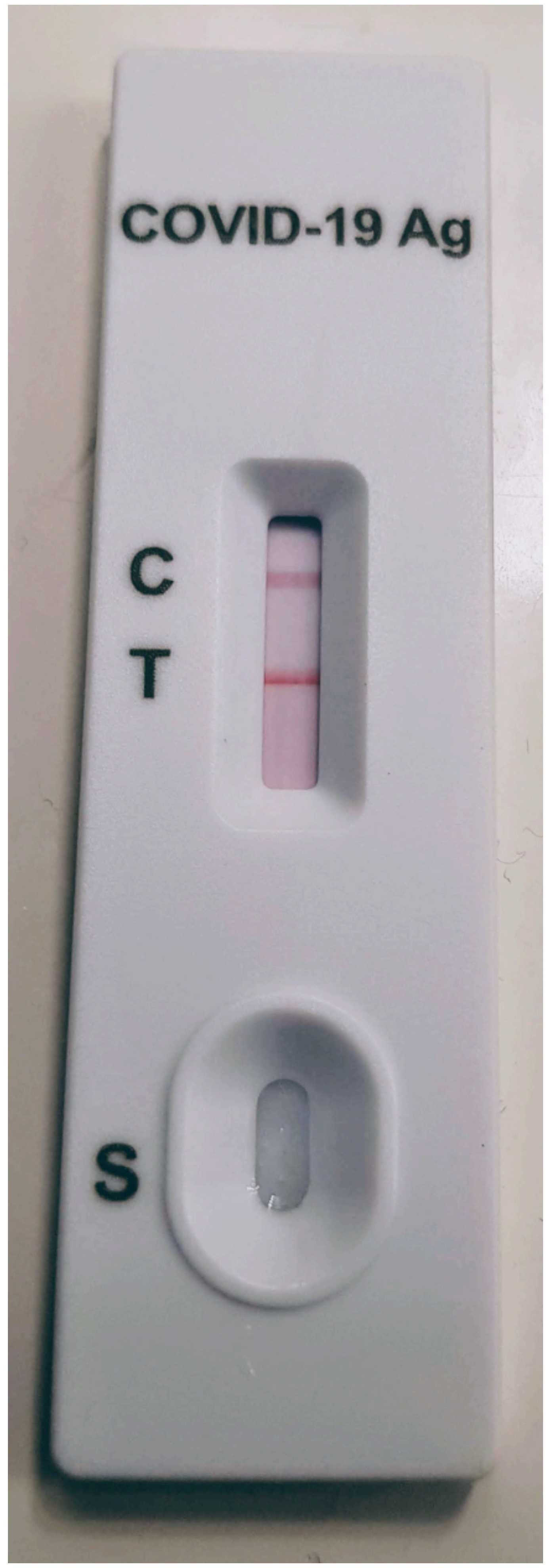
$$p(x) = \int dy\, p(x, y)$$

- Conditional probability

$$p(x, y) = p(x\,|\,y) \cdot p(y)$$

$$\underset{\text{Posterior}}{p(x\,|\,y)} = \frac{\overset{\text{Likelihood}}{p(y\,|\,x)} \cdot \overset{\text{Prior}}{p(x)}}{\underset{\text{Normalisation}}{\int dy\, p(y\,|\,x) \cdot p(x)}}$$

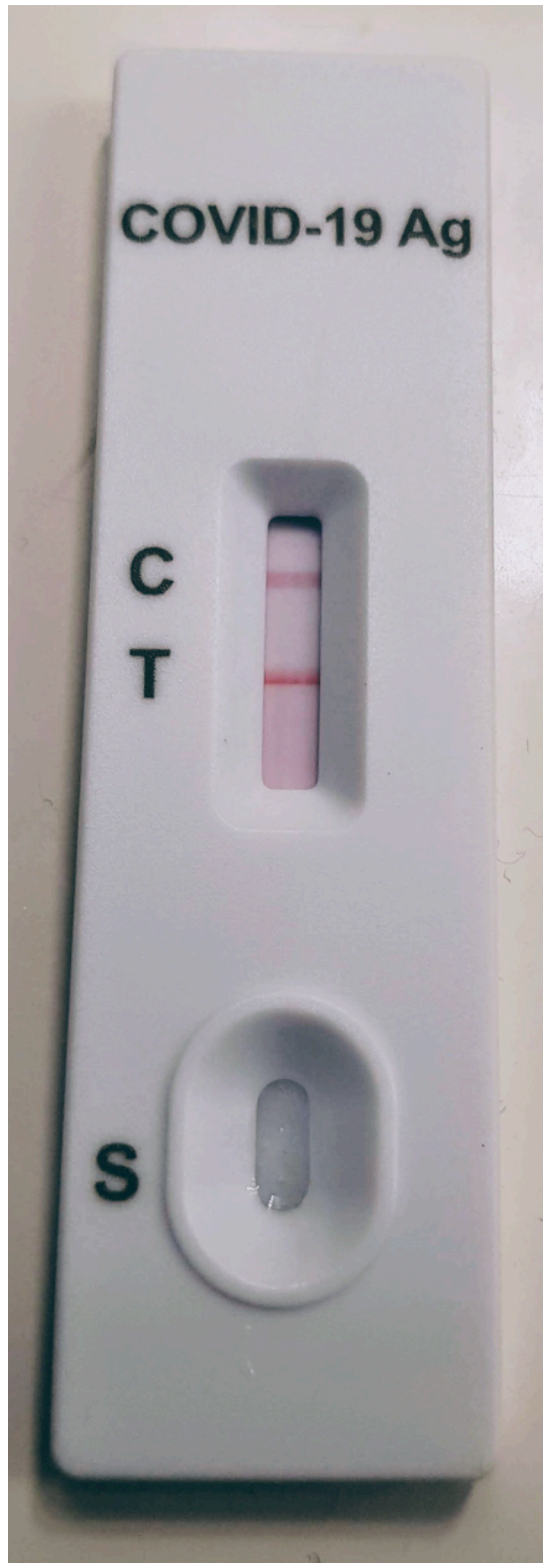
Example 1: How to interpret a positive test?



What's the probability that I am sick (S) ?

$$p(S\,|\,+) = ?$$

Example 1: How to interpret a positive test?



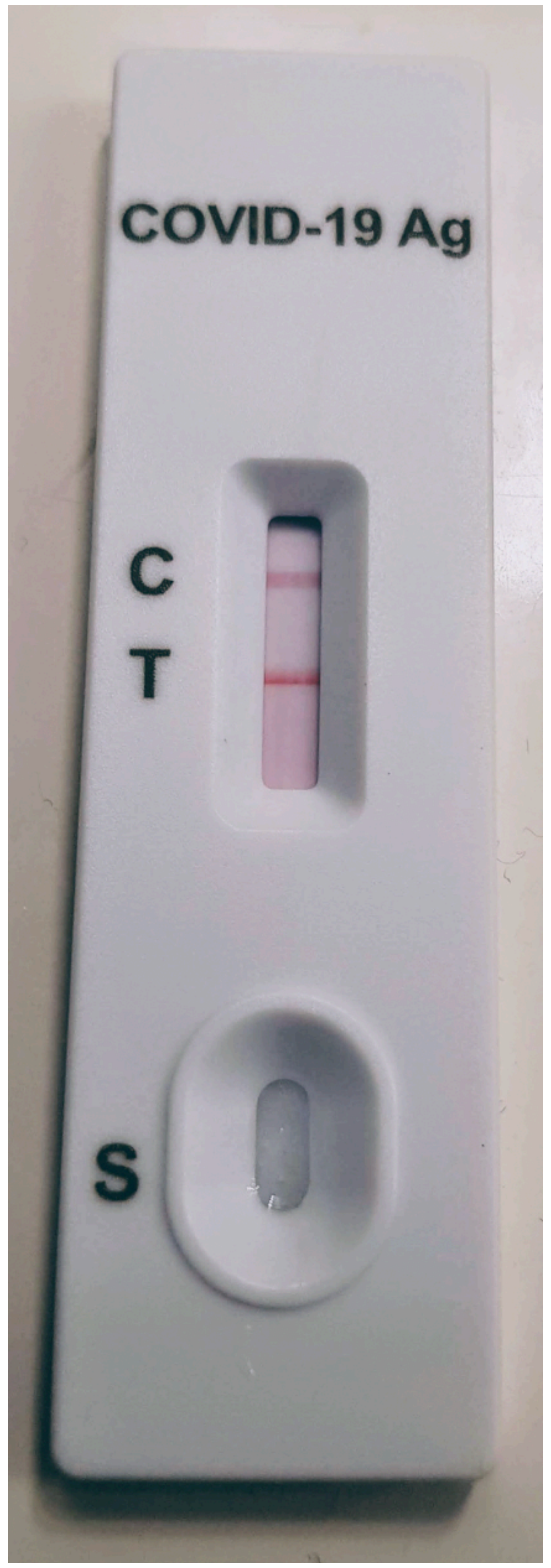
What's the probability that I am sick (S) ?

$$P(S\,|\,+) = \frac{\overbrace{P(+\,|\,S)}\,P(S)}{\underbrace{P(+\,|\,S)}_{\text{Probability of True positive}}\,P(S) + \underbrace{P(-\,|\,S)}_{\text{Probability of False positive}}\,P(\bar{S})}$$

Priors

$$P(S) + P(\bar{S}) = 1$$

Example 1: How to interpret a positive test?



What's the probability that I am sick (S) ?

$$P(S\,|\,+) = \frac{P(+\,|\,S)\,P(S)}{P(+\,|\,S)\,P(S) + P(-\,|\,S)\,P(\bar{S})}$$

Priors

$P(S) + P(\bar{S}) = 1$

Probability of True positive

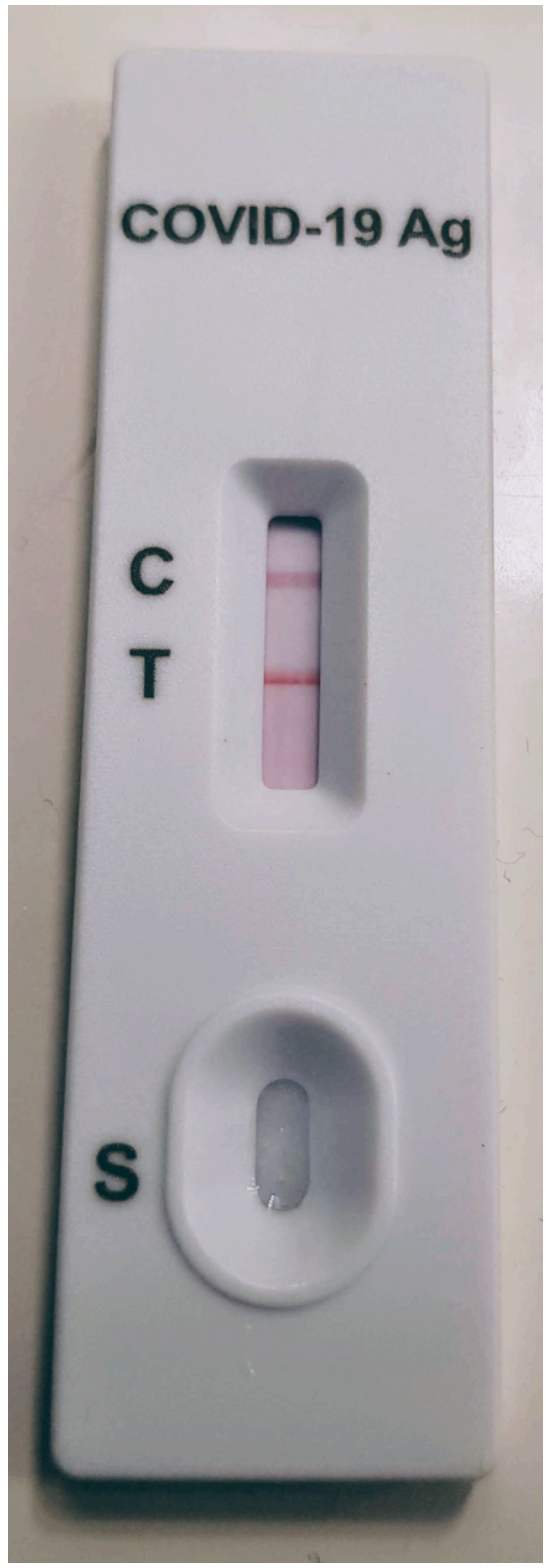
Probability of False positive

Sensitivity of the test $= P(+\,|\,S)$

Specificity of the test $= P(-\,|\,\bar{S}) = 1 - P(+\,|\,\bar{S})$

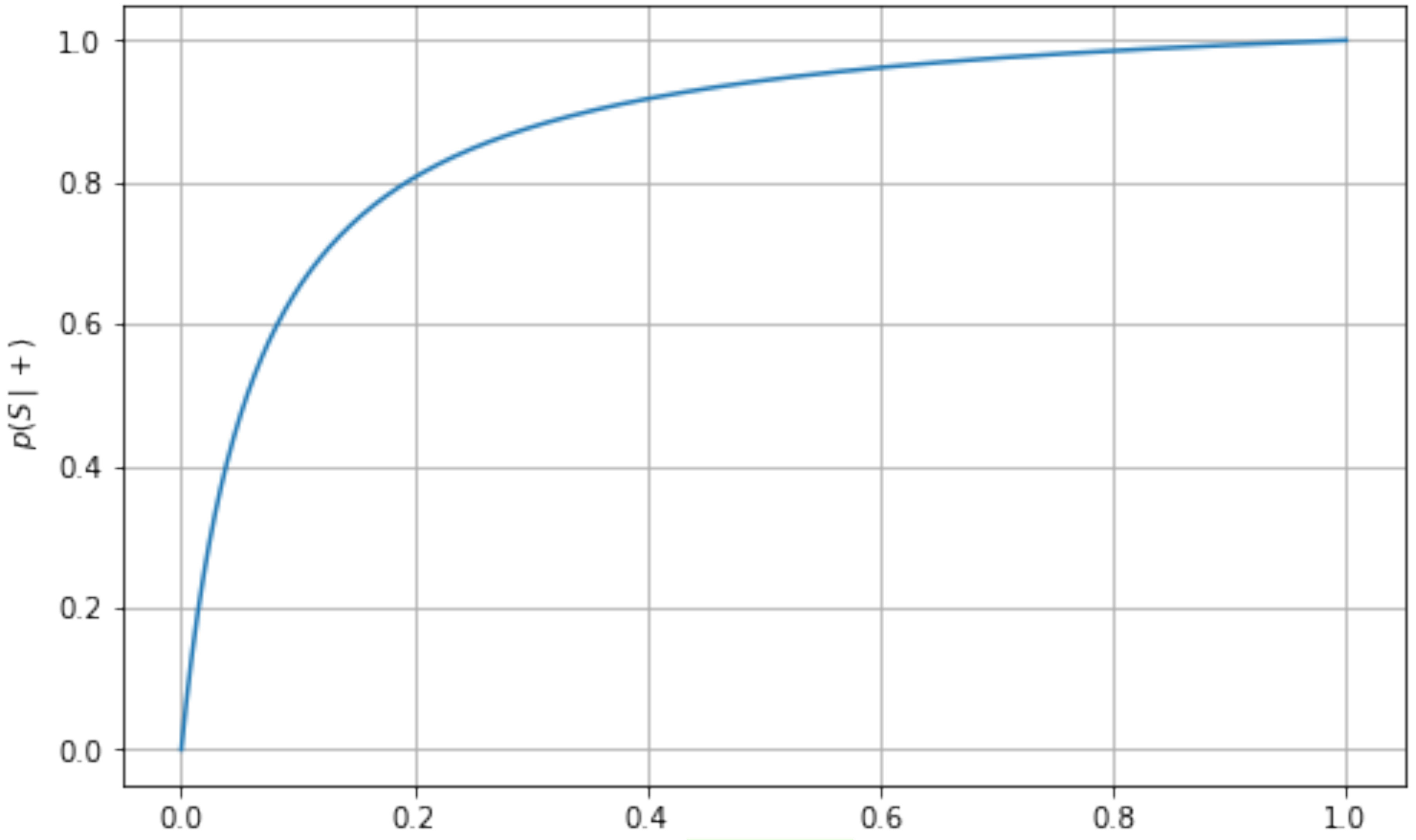
Example 1: How to interpret a positive test?

What’s the probability that I am sick (S) ?



$$P(S\,|\,+) = \left(1 + \frac{1-\mathrm{Sp.}}{\mathrm{Se.}} \cdot \frac{1-P(S)}{P(S)}\right)^{-1}$$

$\mathrm{Se.} = 50\%$ $\mathrm{Sp.} = 97\%$



$P(S)$

Example 2: The Monty Hall problem

In two boxes there is a goat and in the other a car

You have to choose one and only one box

Example 2: The Monty Hall problem

Imagine we randomly pick the first one, but without opening it

Example 2: The Monty Hall problem

Now the host of the game (**who knows where the car is**) shows us the content of the third box, which does not contain the car

Example 2: The Monty Hall problem

We are given the opportunity to change our box (n.1) with the other (n. 2)

What would you do?

Example 2: The Monty Hall problem

(1) Set up the problem

- H_i The hypothesis “the car is in the i-th box”
- E The **event** “the host shows use the content of the third box”
- I Our prior knowledge:

  “3 boxes and 1 car” $\oplus$ “the host knows where the car is”

We need to compute the **posterior** $P(H_i \,|\, E, I)$

Example 2: The Monty Hall problem

(2) Bayes theorem for each hypothesis

$$P(H_1 \mid E, I) = \frac{P(E \mid H_1, I)\, P(H_1 \mid I)}{P(E \mid I)} = \ldots$$

$$P(H_2 \mid E, I) = \frac{P(E \mid H_2, I)\, P(H_2 \mid I)}{P(E \mid I)} = \ldots$$

$$P(H_3 \mid E, I) = \frac{P(E \mid H_3, I)\, P(H_3 \mid I)}{P(E \mid I)} = \ldots$$

Example 2: The Monty Hall problem

(3) Add the likelihoods

$$P(H_1 \mid E, I) = \frac{P(E \mid H_1, I)\, P(H_1 \mid I)}{P(E \mid I)} = \frac{1/2 \cdot \qquad}{\qquad}$$

$$P(H_2 \mid E, I) = \frac{P(E \mid H_2, I)\, P(H_2 \mid I)}{P(E \mid I)} = \frac{1 \quad \cdot \qquad}{\qquad}$$

$$P(H_3 \mid E, I) = \frac{P(E \mid H_3, I)\, P(H_3 \mid I)}{P(E \mid I)} = \frac{0 \quad \cdot \qquad}{\qquad}$$

Example 2: The Monty Hall problem

(4) Add the priors $P(H_1|I) = P(H_2|I) = P(H_3|I) = 1/3$

$$P(H_1 \mid E, I) = \frac{P(E \mid H_1, I)\, P(H_1 \mid I)}{P(E \mid I)} = \frac{1/2 \cdot 1/3}{}$$

$$P(H_2 \mid E, I) = \frac{P(E \mid H_2, I)\, P(H_2 \mid I)}{P(E \mid I)} = \frac{1 \quad \cdot 1/3}{}$$

$$P(H_3 \mid E, I) = \frac{P(E \mid H_3, I)\, P(H_3 \mid I)}{P(E \mid I)} = \frac{0 \quad \cdot 1/3}{}$$

Example 2: The Monty Hall problem

(5) Add the normalization $P(E \mid I) = \sum_i P(H_i \mid E, I) P(E \mid I) = 1/2$

$$P(H_1 \mid E, I) = \frac{P(E \mid H_1, I)\, P(H_1 \mid I)}{P(E \mid I)} = \frac{1/2 \cdot 1/3}{1/2}$$

$$P(H_2 \mid E, I) = \frac{P(E \mid H_2, I)\, P(H_2 \mid I)}{P(E \mid I)} = \frac{1 \quad \cdot 1/3}{1/2}$$

$$P(H_3 \mid E, I) = \frac{P(E \mid H_3, I)\, P(H_3 \mid I)}{P(E \mid I)} = \frac{0 \quad \cdot 1/3}{1/2}$$

Example 2: The Monty Hall problem

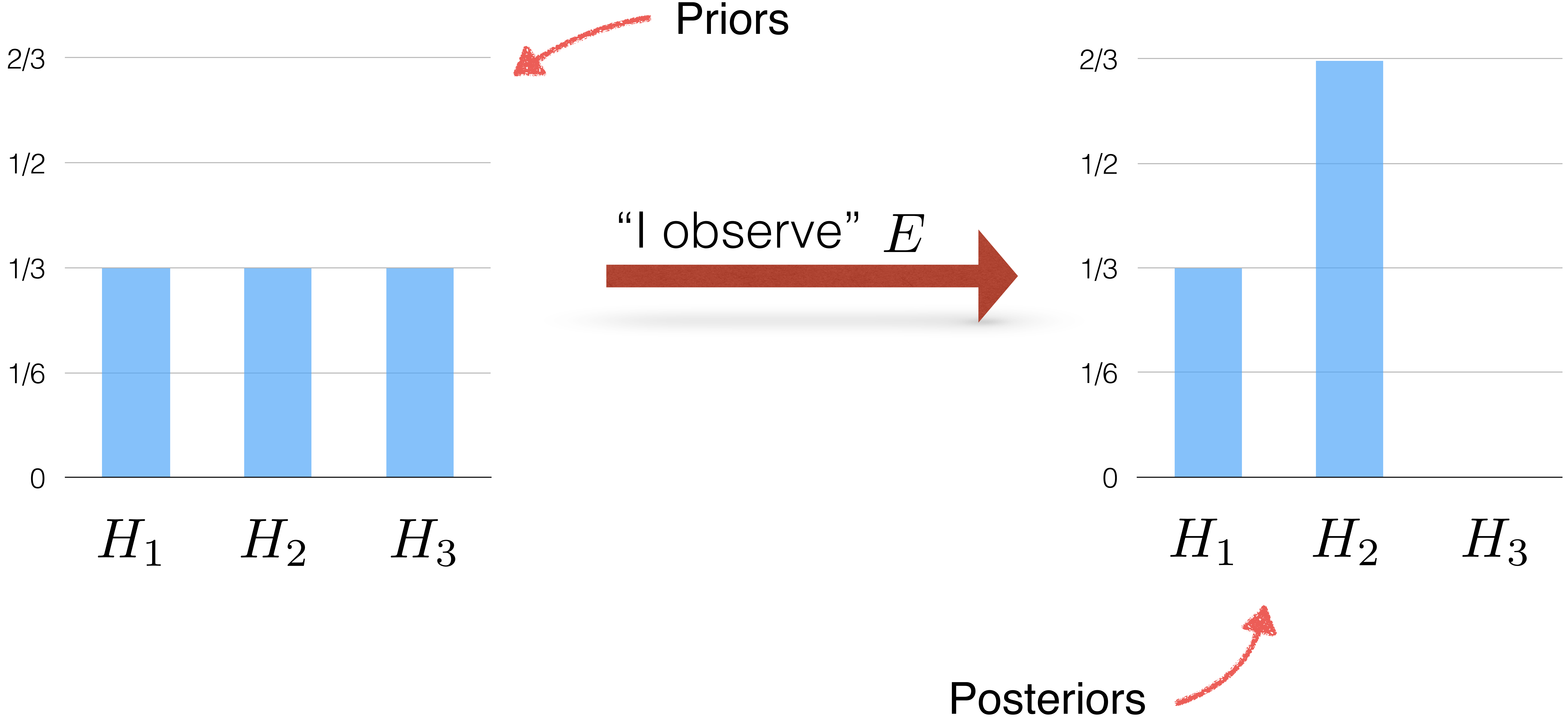
Result:

$$P(H_1 \mid E, I) = \frac{P(E \mid H_1, I)\, P(H_1 \mid I)}{P(E \mid I)} = \frac{1/2 \cdot 1/3}{1/2} = \frac{1}{3}$$

$$P(H_2 \mid E, I) = \frac{P(E \mid H_2, I)\, P(H_2 \mid I)}{P(E \mid I)} = \frac{1 \quad \cdot 1/3}{1/2} = \frac{2}{3}$$

$$P(H_3 \mid E, I) = \frac{P(E \mid H_3, I)\, P(H_3 \mid I)}{P(E \mid I)} = \frac{0 \quad \cdot 1/3}{1/2} = 0$$

Example 2: The Monty Hall problem



Priors

"I observe" E

Posteriors

Example 3: Counting experiment

- Assumptions:
  - No background
  - Expected signal counts: $s = r \cdot T_{obs}$

    Rate

    Observation time
  - Observed counts n

Example 3: Counting experiment

- Assumptions:
  - No background
  - Expected signal counts: $s = \overset{\text{Rate}}{r} \cdot \overset{\text{Observation time}}{T_{obs}}$
  - Observed counts n

Likelihood: $p(n \,|\, s) = \dfrac{s^n}{n!} e^{-s}$

Bayes theorem: $p(s \,|\, n) = \dfrac{p(n \,|\, s) p(s)}{p(n)} = \dfrac{p(n \,|\, s) p(s)}{\int ds\, p(n \,|\, s) p(s)} = \dfrac{s^n e^{-s} p(s)}{\int ds\, s^n e^{-s} p(s)}$

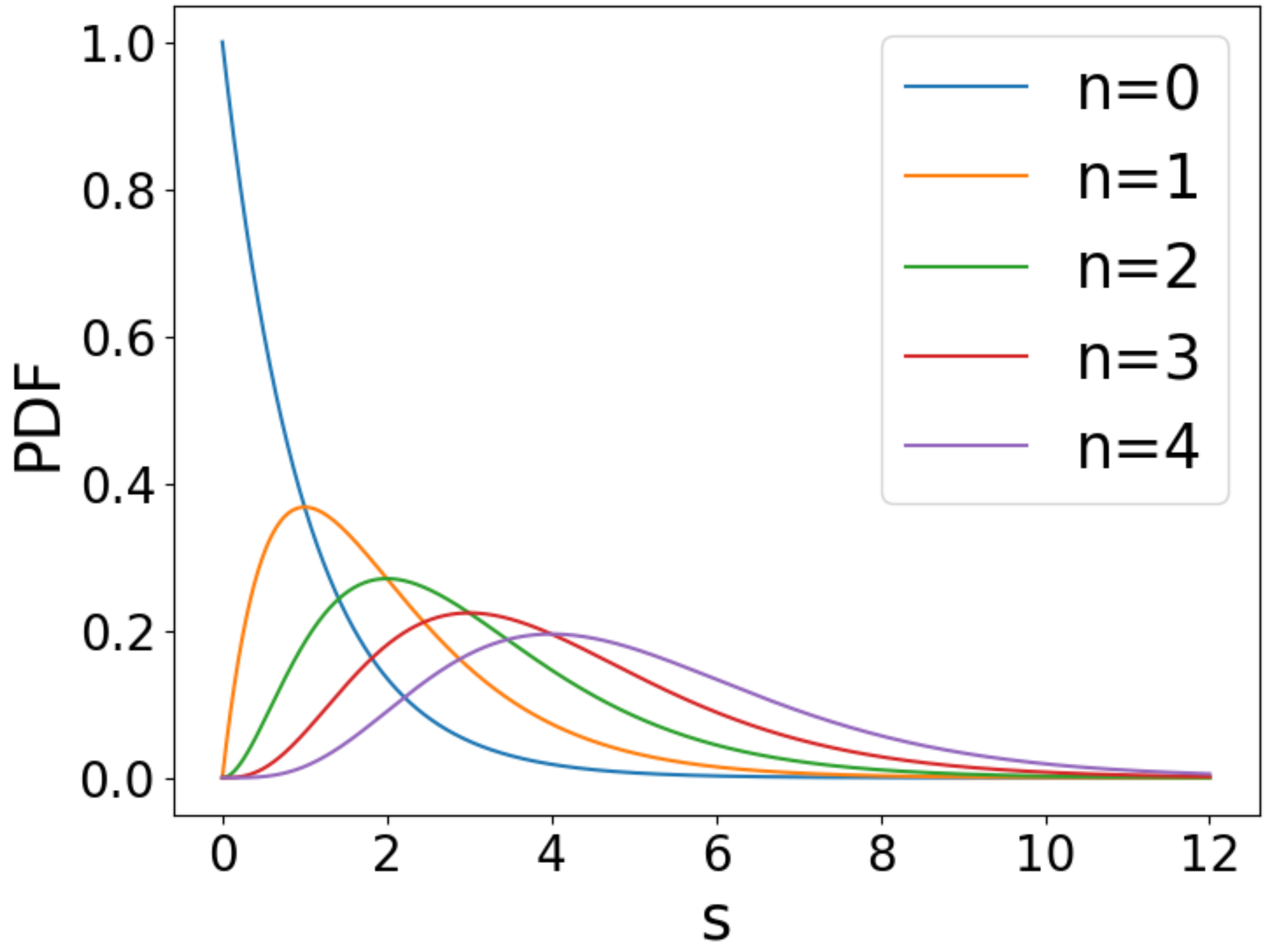
Example 3: Counting experiment

$$p(s\,|\,n) = \frac{p(n\,|\,s)p(s)}{p(n)} = \frac{p(n\,|\,s)p(s)}{\int ds\, p(n\,|\,s)p(s)} = \frac{s^n e^{-s} p(s)}{\int ds\, s^n e^{-s} p(s)}$$

Conjugate prior

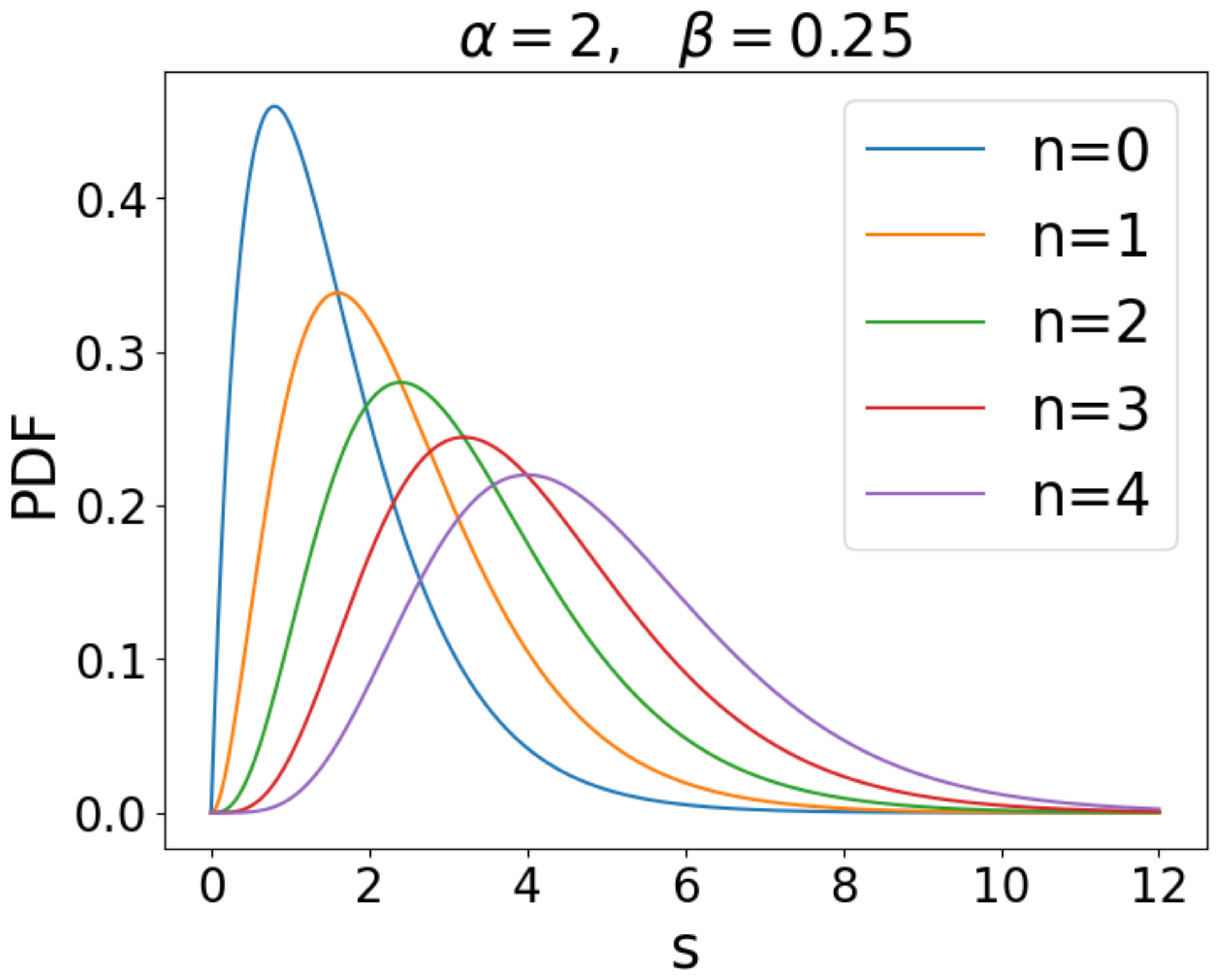
Prior: $p(s) = C$

Posterior: $p(s\,|\,n) = \frac{s^n}{n!} e^{-s}$



Prior: $p(s) = \text{Gamma}(\alpha, \beta) \propto \text{s}^{\alpha-1}\text{e}^{-\beta \text{s}}$

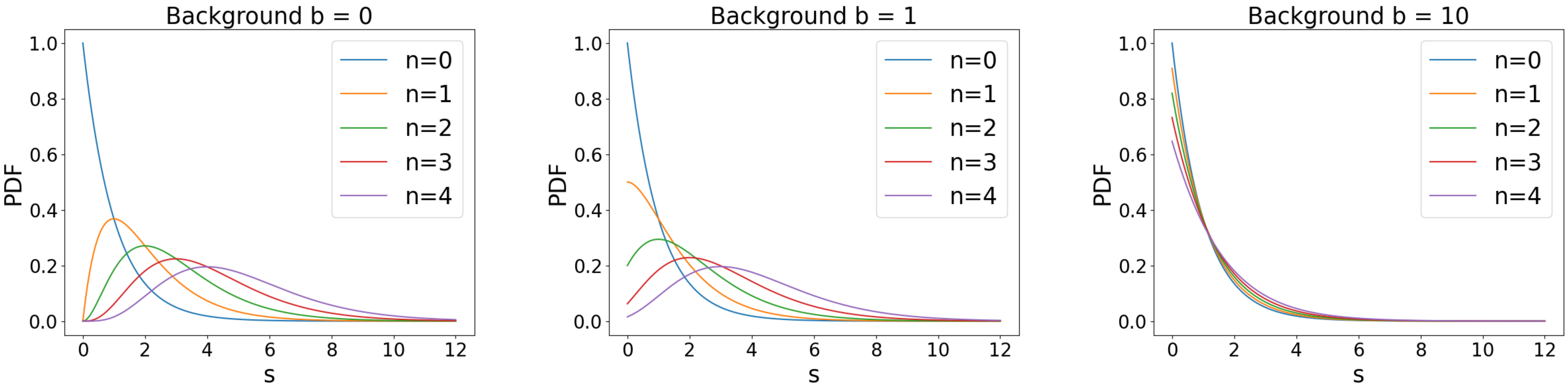
Posterior: $p(s\,|\,n) = \text{Gamma}(\alpha + \text{n}, \beta + 1)$



Example 3: Counting experiment

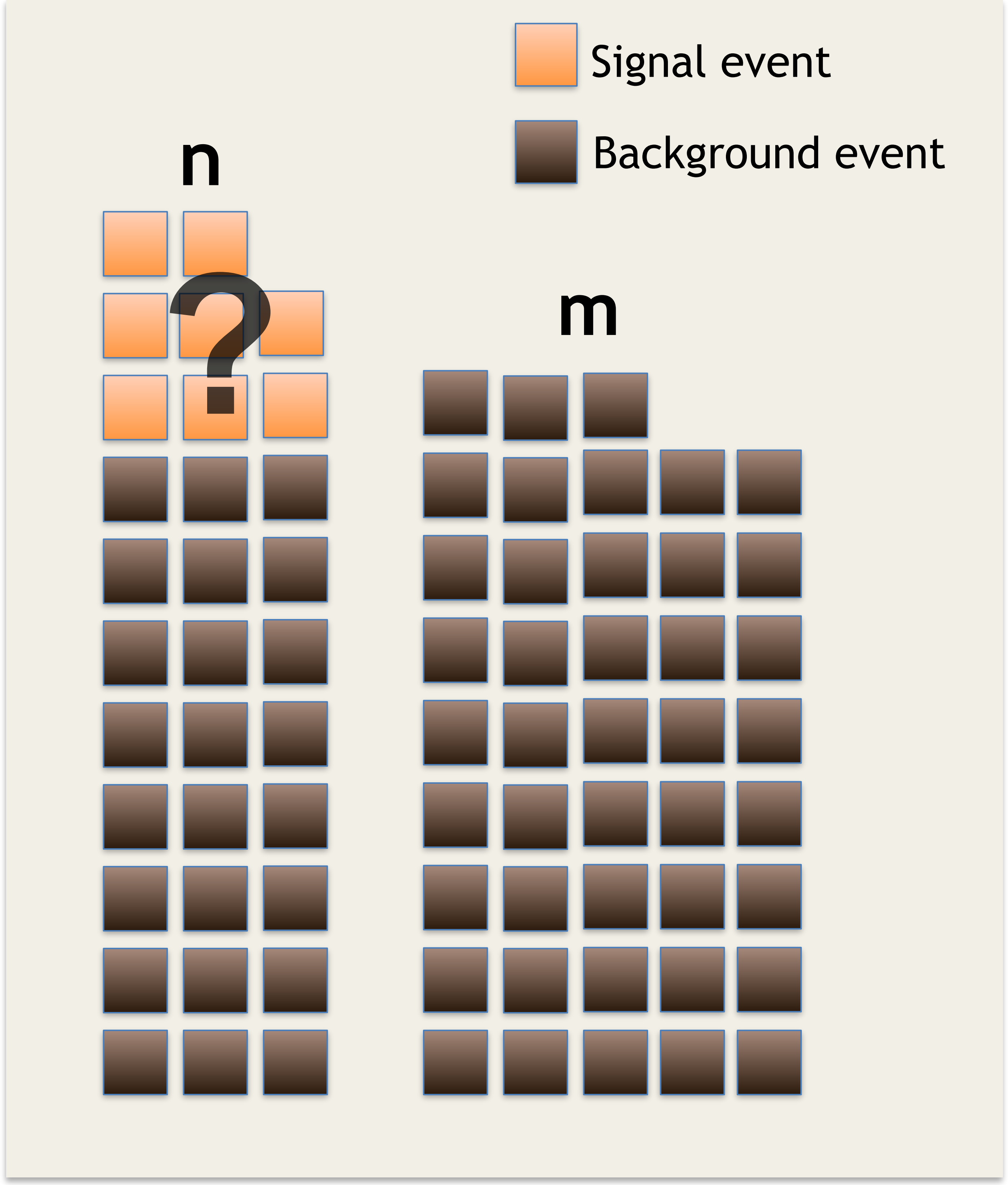
Adding the background $b = r_b \cdot T_{obs}$

Posterior: $p(s+b \,|\, n) \propto \dfrac{(s+b)^n}{n!} e^{-s-b}$



Example 3: Counting experiment

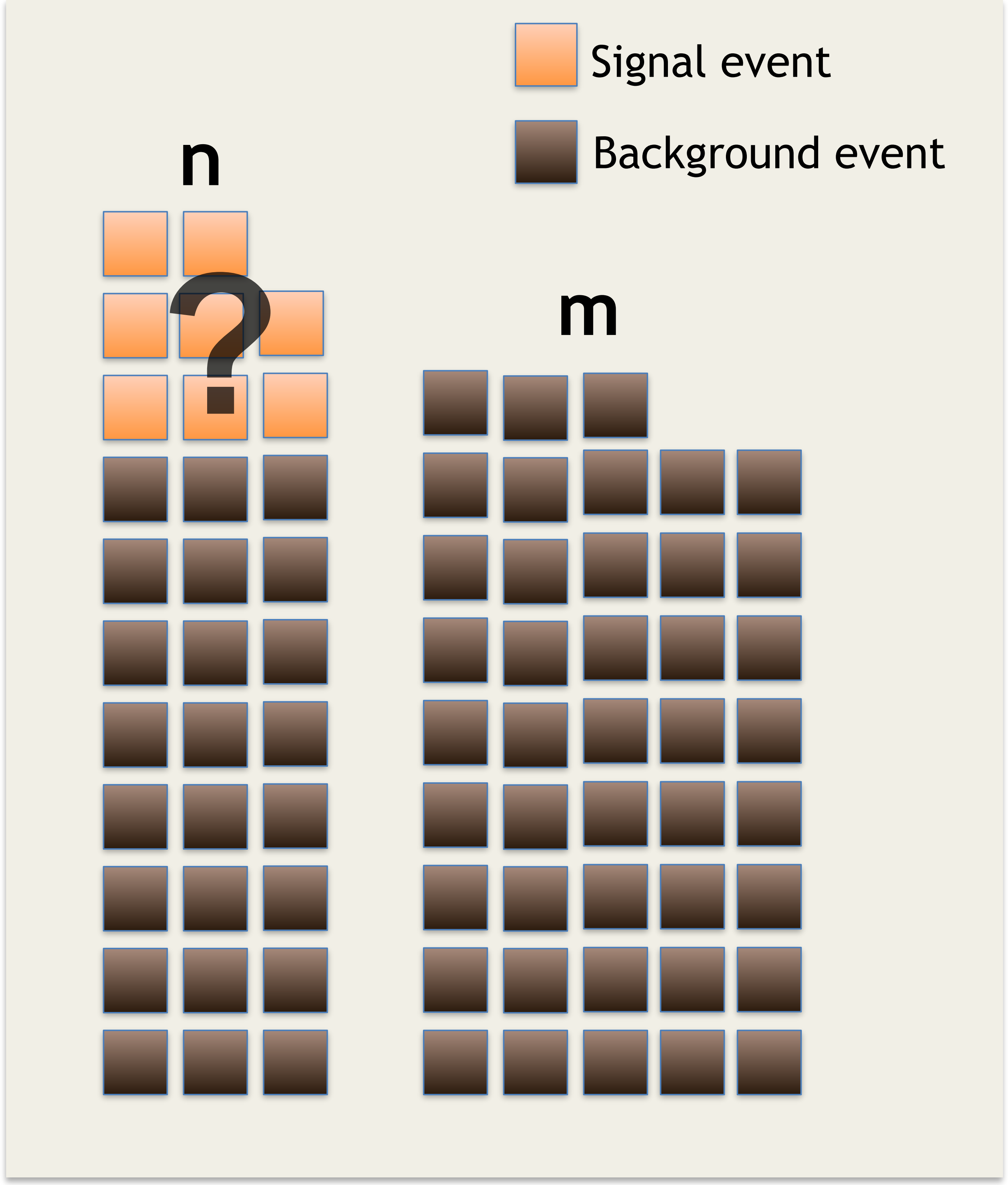
Adding the background estimated from an “**OFF region**”



$$\alpha = \frac{\text{size of ON region}}{\text{size of OFF region}}$$

Example 3: Counting experiment

Adding the background estimated from an "**OFF region**"



$$\alpha = \frac{\text{size of ON region}}{\text{size of OFF region}}$$

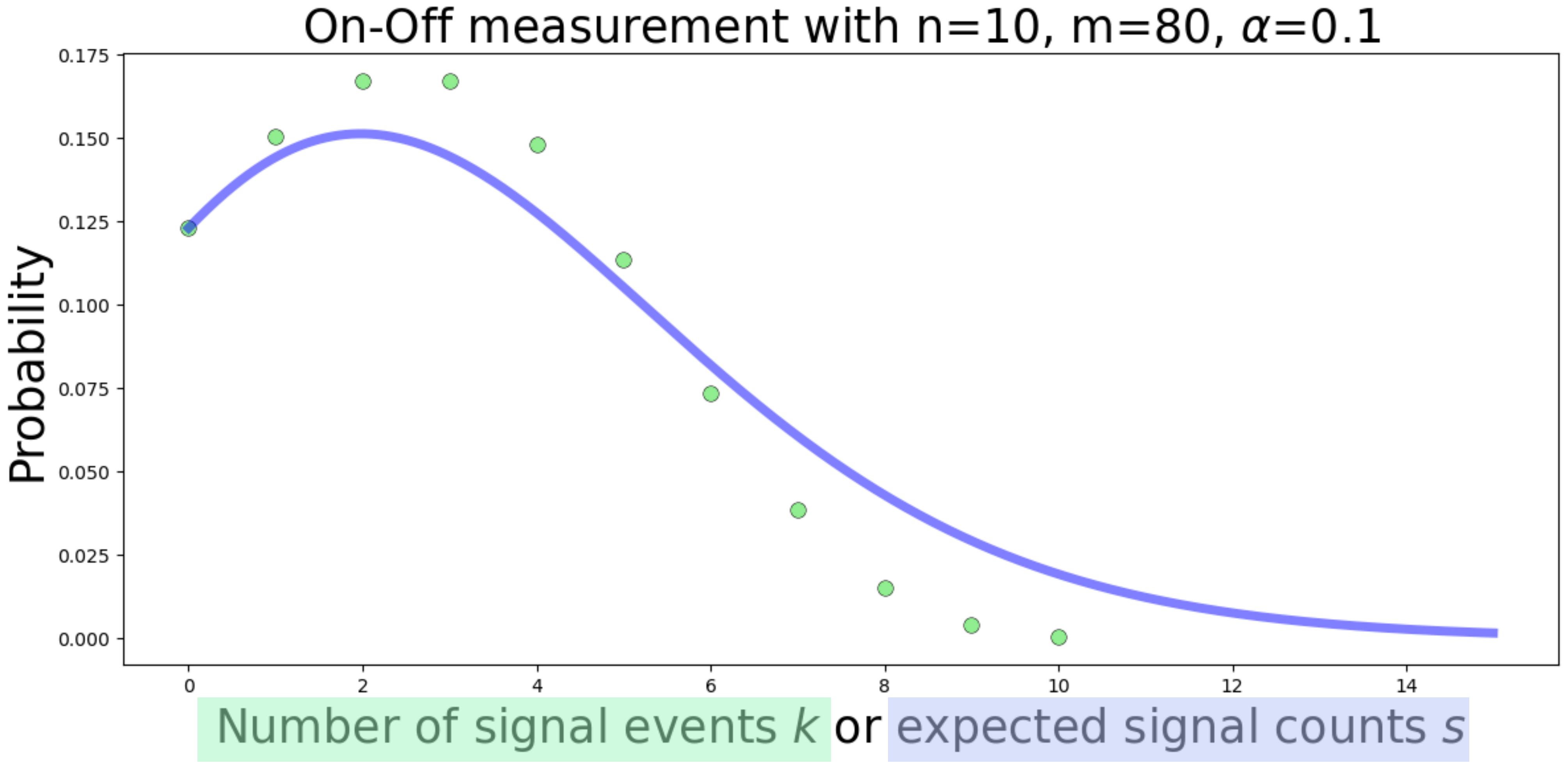
Likelihood: $$p(n, m \,|\, s, b) = \frac{(s + \alpha b)^n}{n!} e^{-(s+\alpha b)} \cdot \frac{b^m}{m!} e^{-b}$$

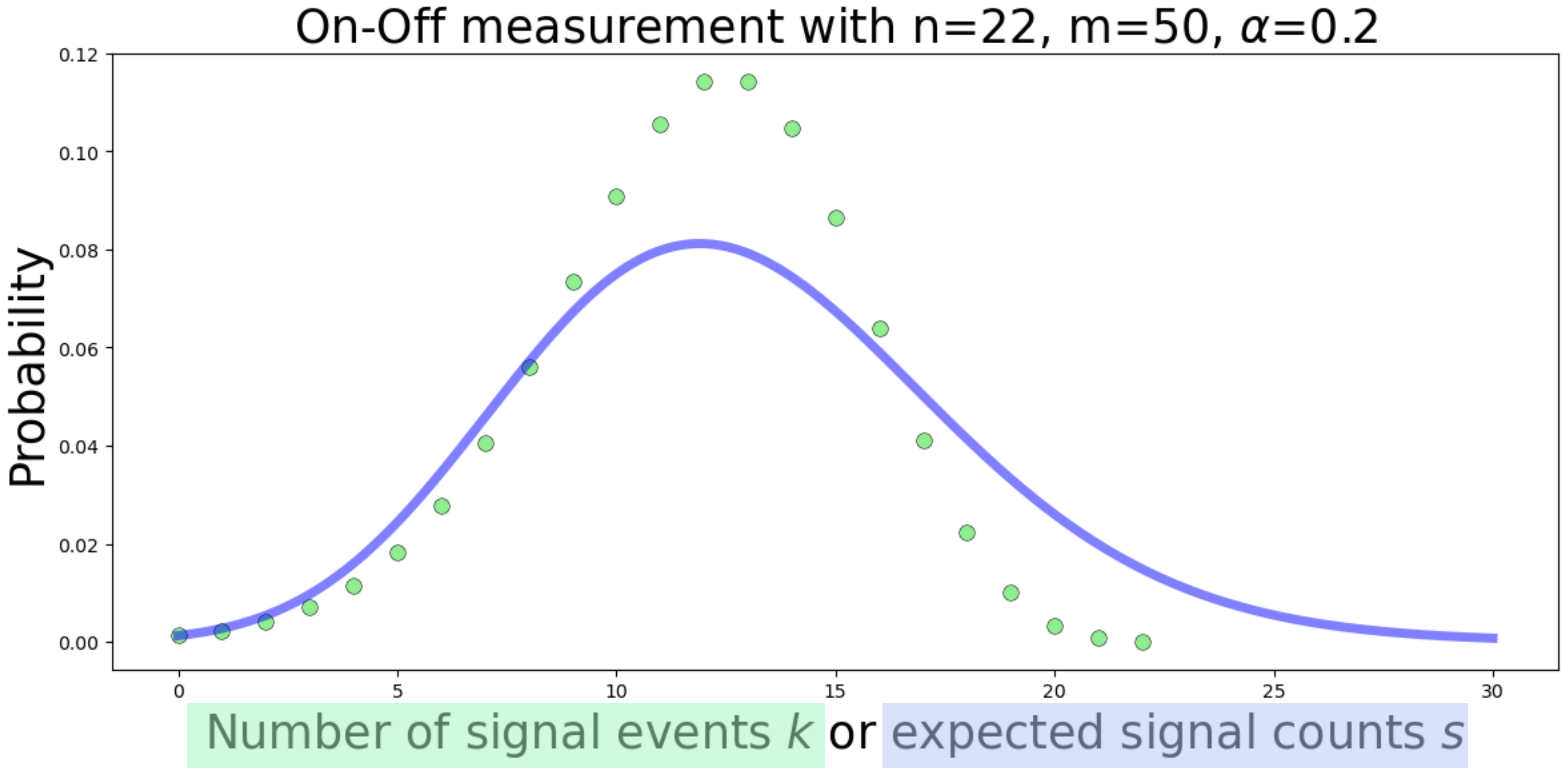
Posterior: $$p(s \,|\, n, m) = \frac{\int db\; p(n, m \mid s, b)\; p(b)\, p(s)}{\int ds\; db\; p(n, m \mid s, b)\; p(b)\, p(s)}$$

Example 3: Counting experiment

Assuming uniform priors $p(s) = p(b) = C$, it can be shown:

$$p(s \mid n, m) \propto \sum_{k=0}^{n} \frac{(n+m-k)!}{(1+1/\alpha)^{-k}(n-k)!} \cdot \frac{s^k}{k!}e^{-s} = \sum_{k=0}^{n} p(k \mid n, m) \cdot p(s \mid k)$$





When the integral becomes too complicated

Let’s assume we don’t know how to solve analytically

$$p(s\,|\,n,m) = \frac{\int db\; p(n,m \mid s,b)\; p(b)\, p(s)}{\int ds\; db\; p(n,m \mid s,b)\; p(b)\, p(s)}$$

Markov Chain Monte Carlo (MCMC)

Replaces **intractable** integrals with **sampling**

Most common algorithms:

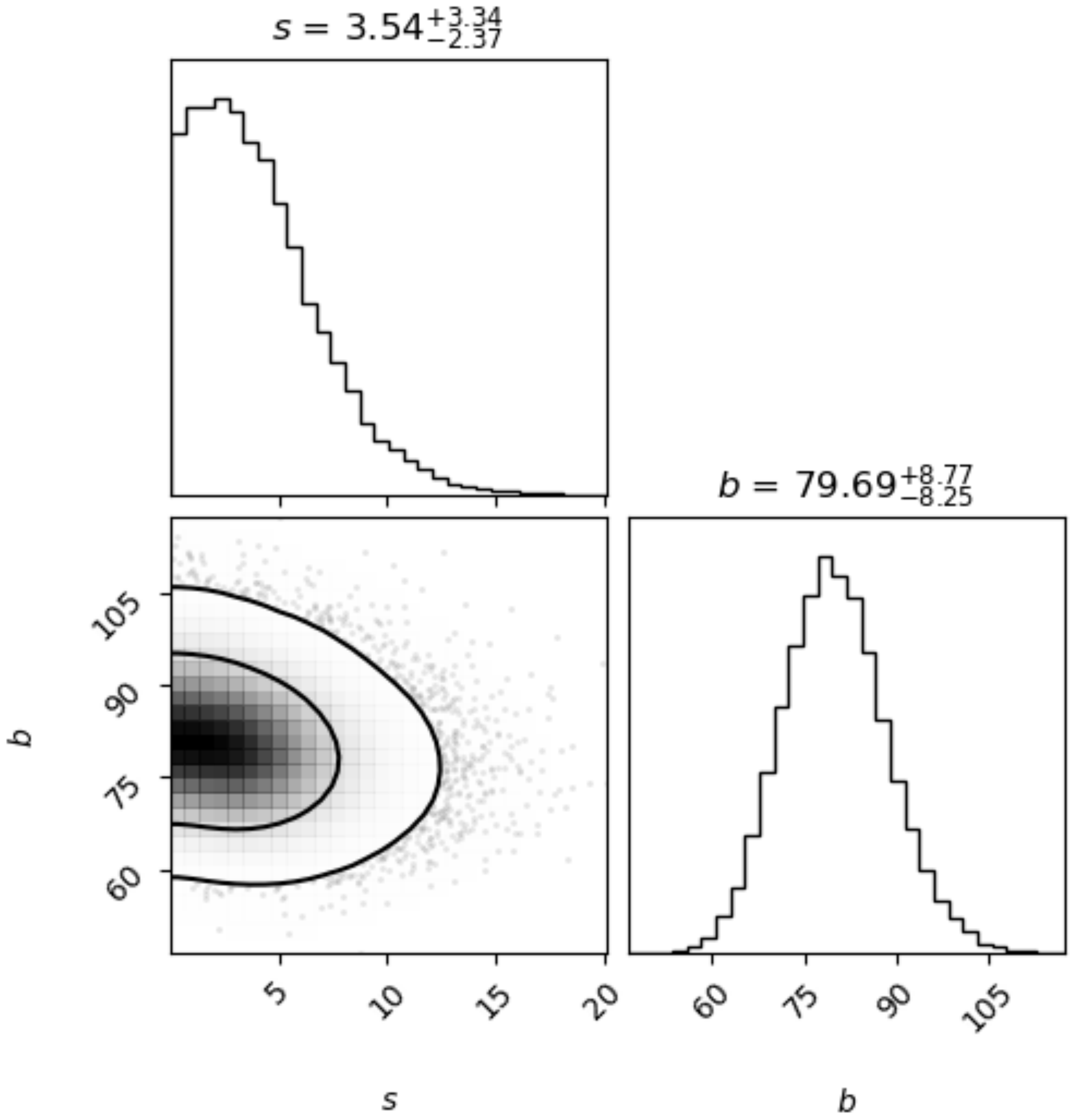
- Metropolis-Hasting
- Gibbs sampling
- …

When the integral becomes too complicated

Let’s assume we don’t know how to solve analytically

$$p(s\,|\,n,m) = \frac{\int db\; p(n,m \mid s,b)\; p(b)\, p(s)}{\int ds\; db\; p(n,m \mid s,b)\; p(b)\, p(s)}$$

Example with python package emcee



Markov Chain Monte Carlo (MCMC)

Replaces **intractable** integrals with **sampling**

Most common algorithms:

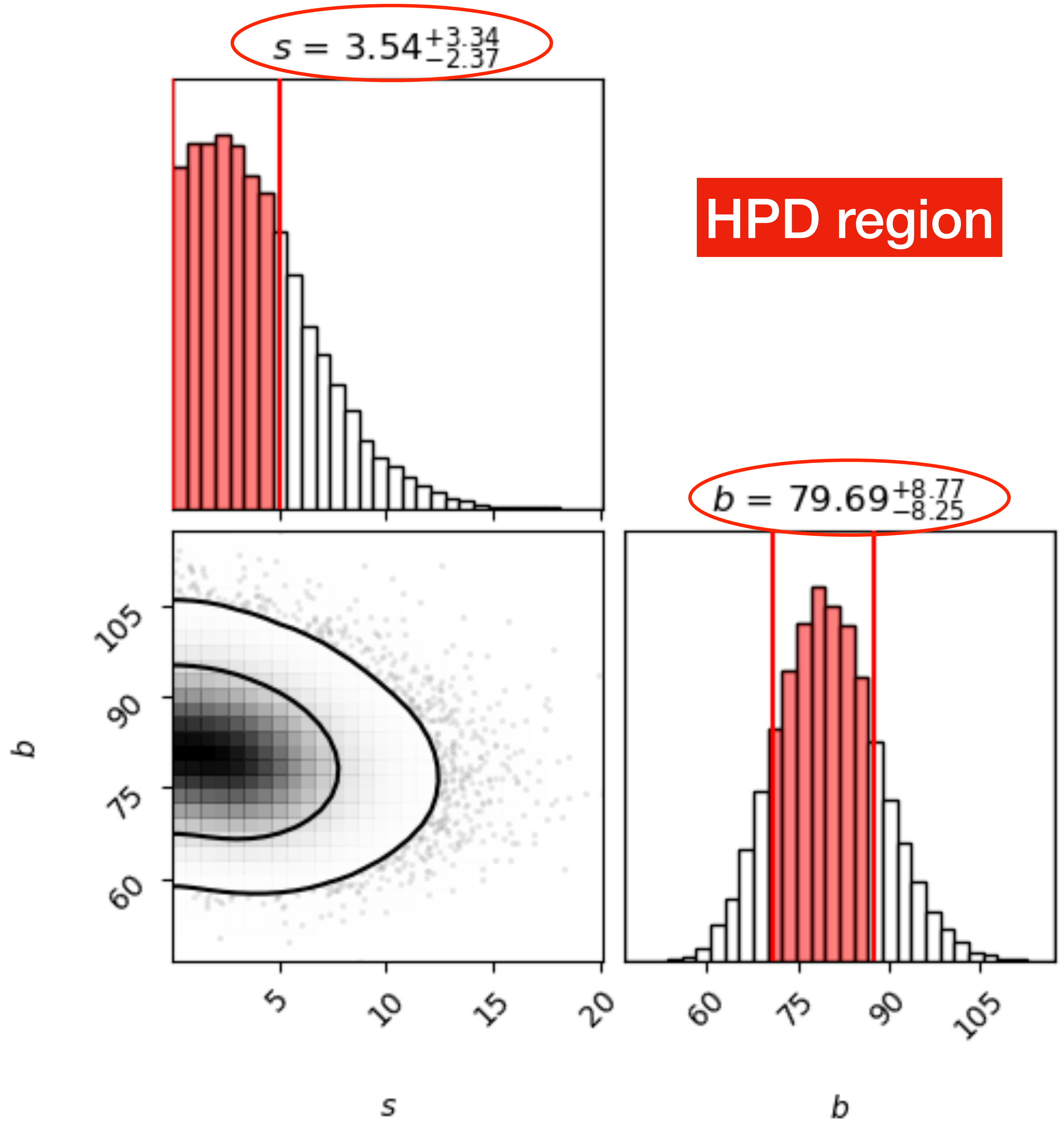
- Metropolis-Hasting
- Gibbs sampling
- …

When the integral becomes too complicated

Let’s assume we don’t know how to solve analytically

$$p(s\,|\,n,m) = \frac{\int db\; p(n,m\mid s,b)\; p(b)\, p(s)}{\int ds\; db\; p(n,m\mid s,b)\; p(b)\, p(s)}$$

Example with python package emcee



HPD region

Markov Chain Monte Carlo (MCMC)

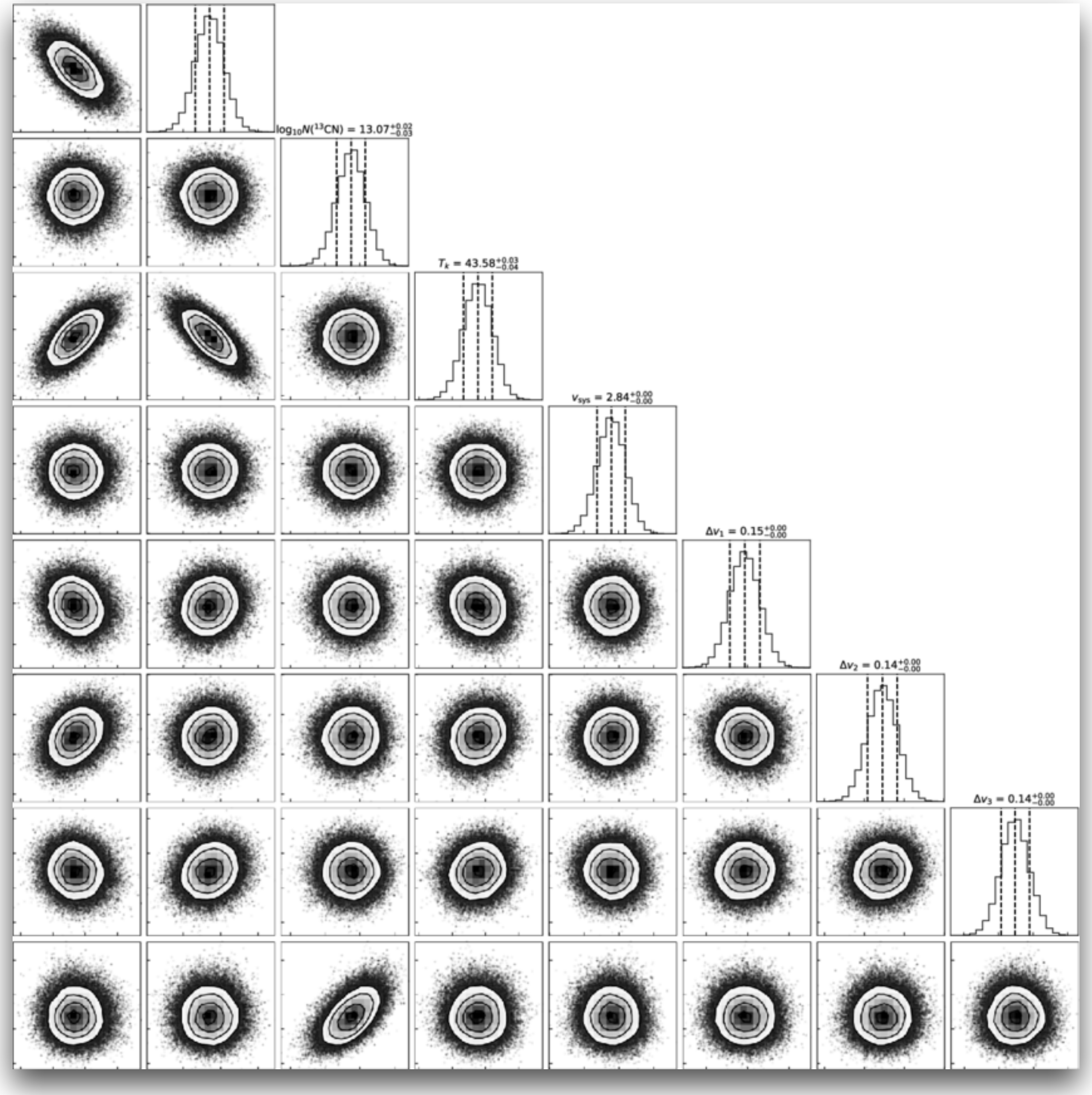
Replaces **intractable** integrals with **sampling**

Most common algorithms:

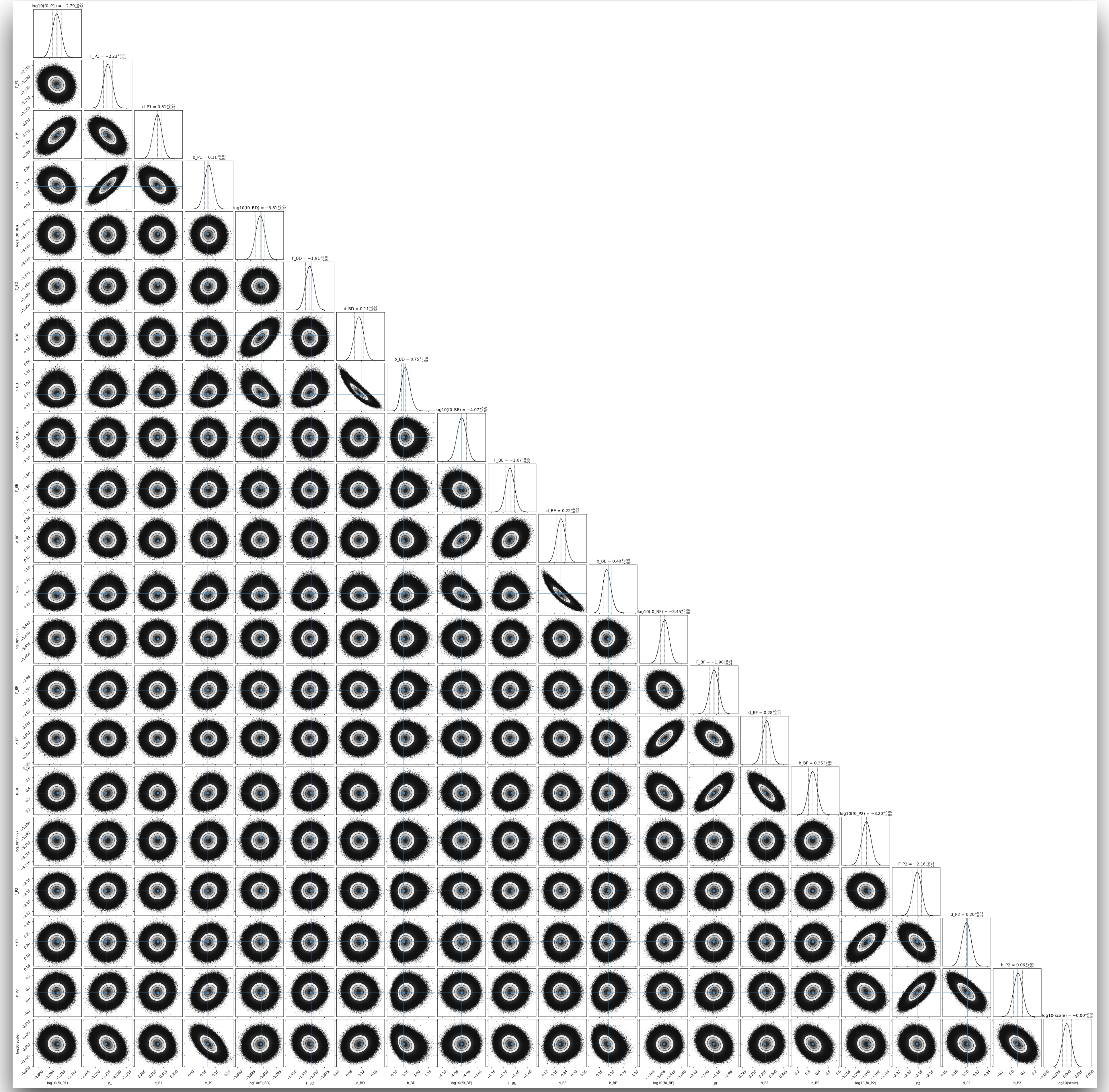
- Metropolis-Hasting
- Gibbs sampling
- …

When the integral becomes too complicated

… and even more complicated!



From DOI:10.3847/1538-4357/ad2fb4



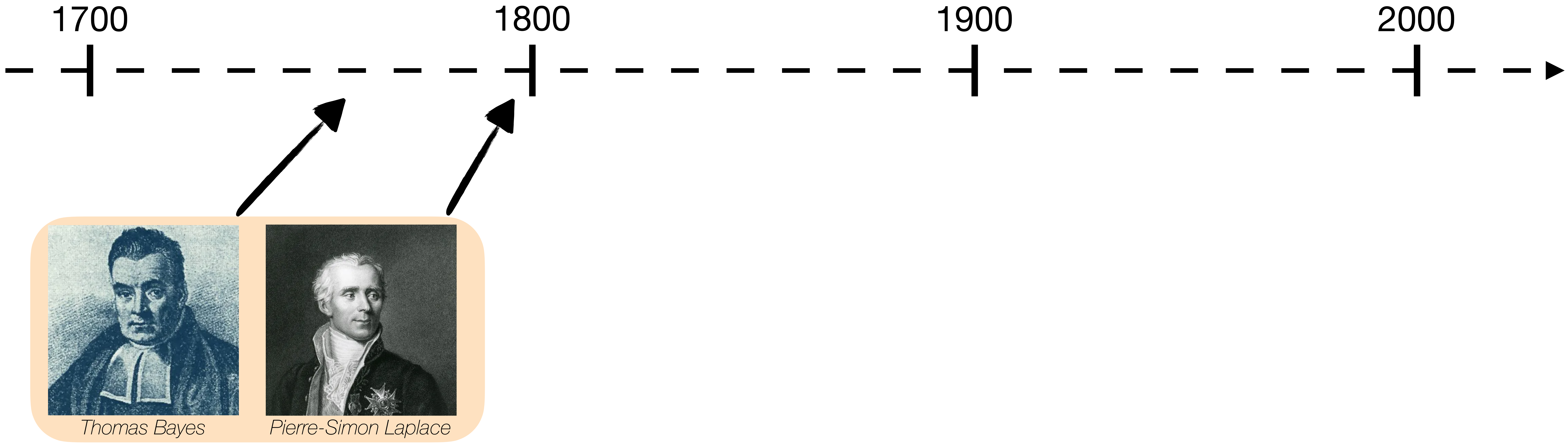
Credit: G. Ceribella

When the integral becomes too complicated

- Markov Chain Monte Carlo (MCMC)
  - Draws correlated samples from the posterior via a Markov chain whose stationary distribution is the target posterior
- Hamiltonian Monte Carlo (HMC)
  - Uses gradients of the log-posterior to simulate Hamiltonian dynamics and make long, efficient moves in parameter space
- Nested Sampling
  - Explores the posterior by iteratively restricting sampling to higher-likelihood regions using a set of “live” points
- Simulation-Based Inference
  - Performs Bayesian inference by repeatedly simulating data from a model and learning the relationship between parameters and observations, avoiding explicit likelihood evaluation
- …

The Frequentist approach

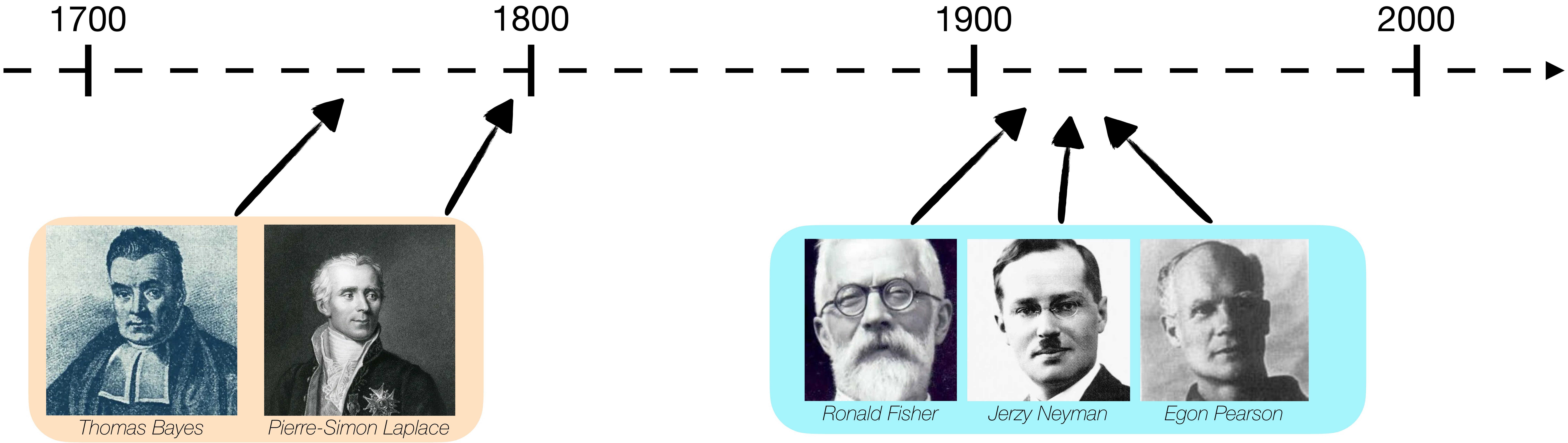
The Frequentist approach



Practical problem
Back then, there were no computers

Philosophical problem
The arbitrariness of the prior

The Frequentist approach



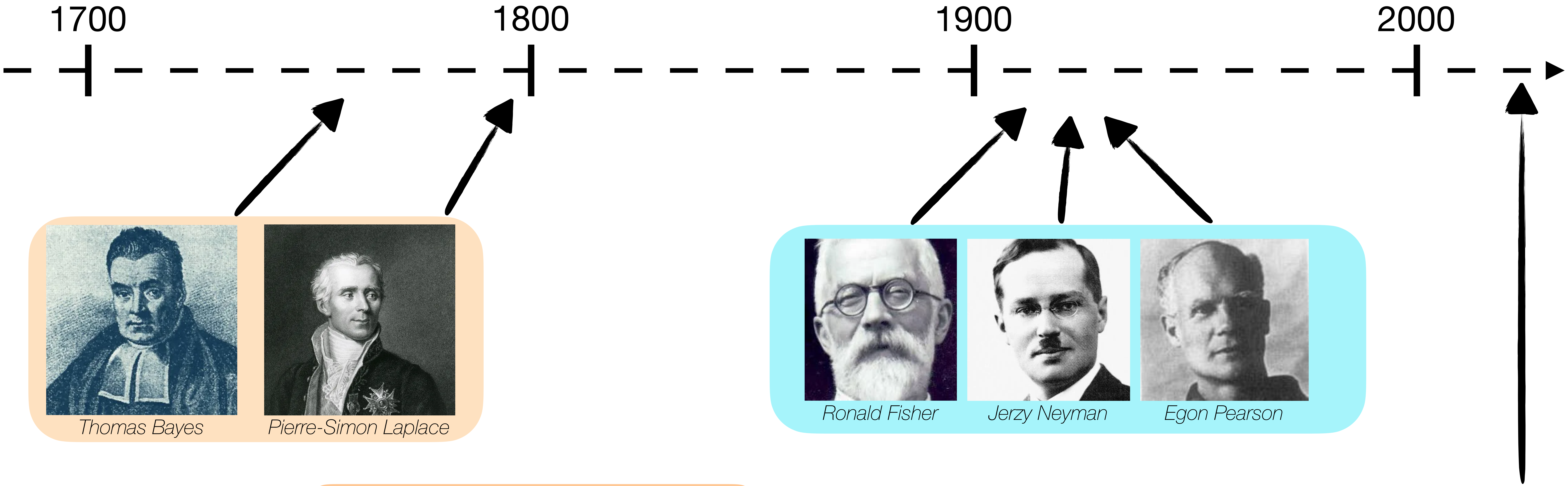
Practical problem

Back then, there were no computers

Philosophical problem

The arbitrariness of the prior

The Frequentist approach



Practical problem

Back then, there were no computers

Philosophical problem

The arbitrariness of the prior

Nowadays… a mix of both

The Frequentist approach

~~- The **Bayesian approach** tries to answer the question:~~

~~*Given our **prior** knowledge and the observed **data**, what is the **probability** that the model is true?*~~

- In the **Frequentist approach,** an inference analysis is performed by trying to answer the following question:

If I repeat the experiment an infinite number of times, assuming the model is true, with which frequency would I observe a value more extreme than the one actually observed?

The Frequentist approach

- ~~The **Bayesian approach** tries to answer the question:~~

~~*Given our **prior** knowledge and the observed **data**, what is the **probability** that the model is true?*~~

- In the **Frequentist approach,** an inference analysis is performed by trying to answer the following question:

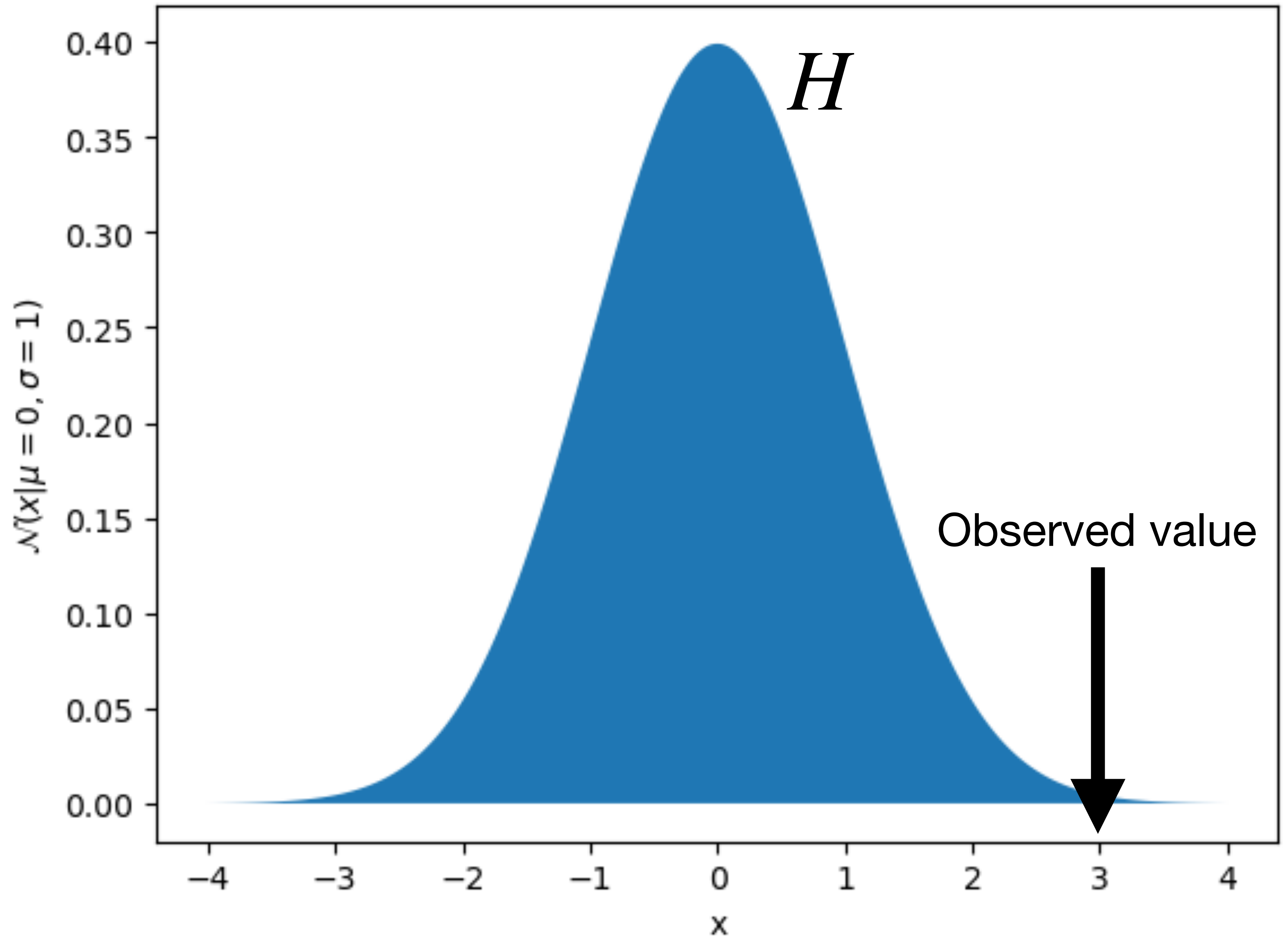
If I repeat the experiment an infinite number of times, assuming the model is **true**, with which **frequency** would I observe a **value** more **extreme** than the one actually observed?

The Frequentist approach

A trivial example:

I have performed an observation and got the experimental data D = 3.
According to the hypotheses H, D is a random variable that follows a normal distribution centered in zero and variance = 1

$$p(D|H) = \mathcal{N}(x=3|\mu=0,\sigma=1)$$

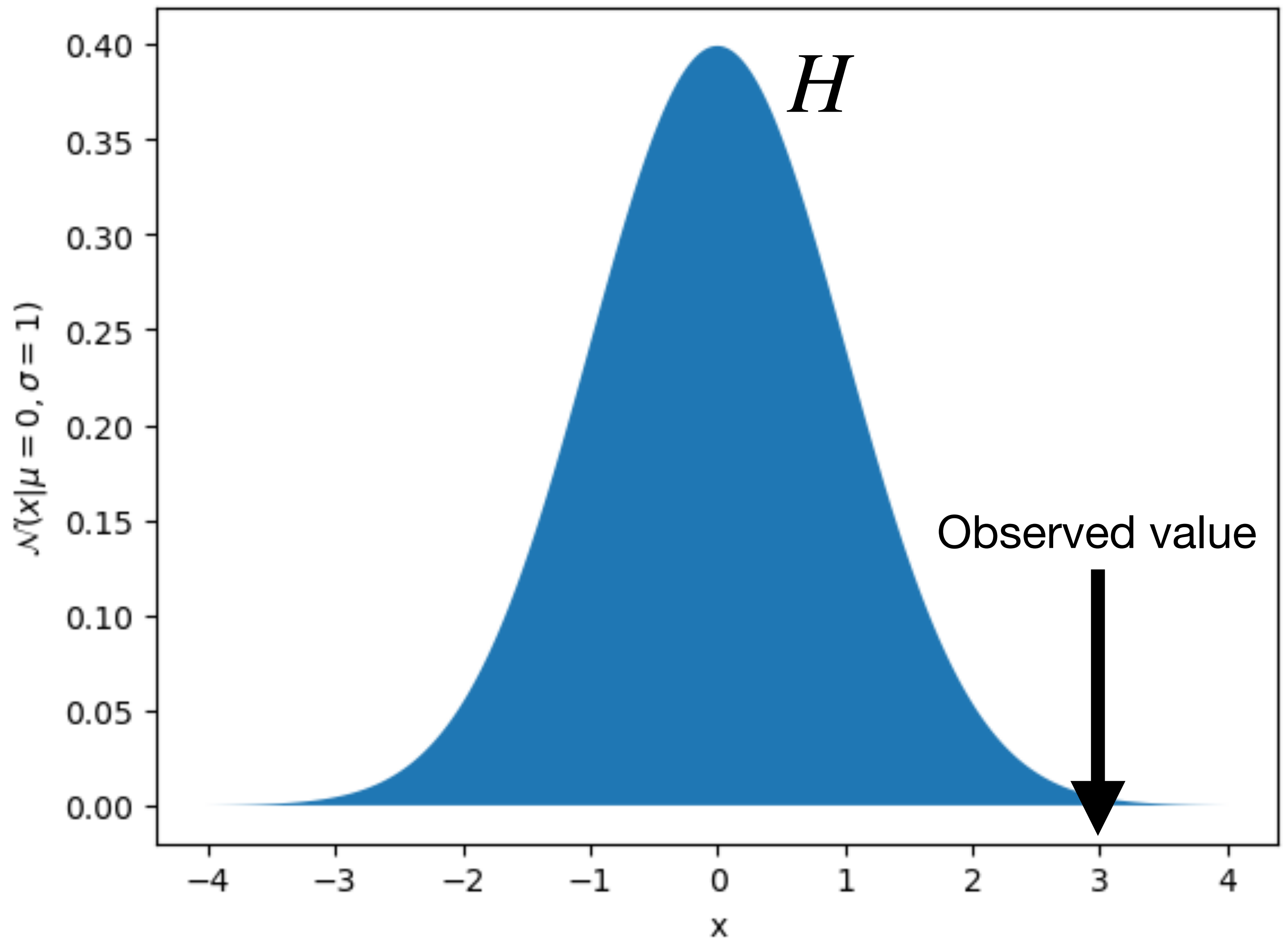


The Frequentist approach

A trivial example:

I have performed an observation and got the experimental data D = 3.
According to the hypotheses H, D is a random variable that follows a normal distribution centered in zero and variance = 1

$$p(D\,|\,H) = \mathcal{N}(x = 3\,|\,\mu = 0, \sigma = 1)$$



Conclusion of the inference analysis performed with the frequentist approach:

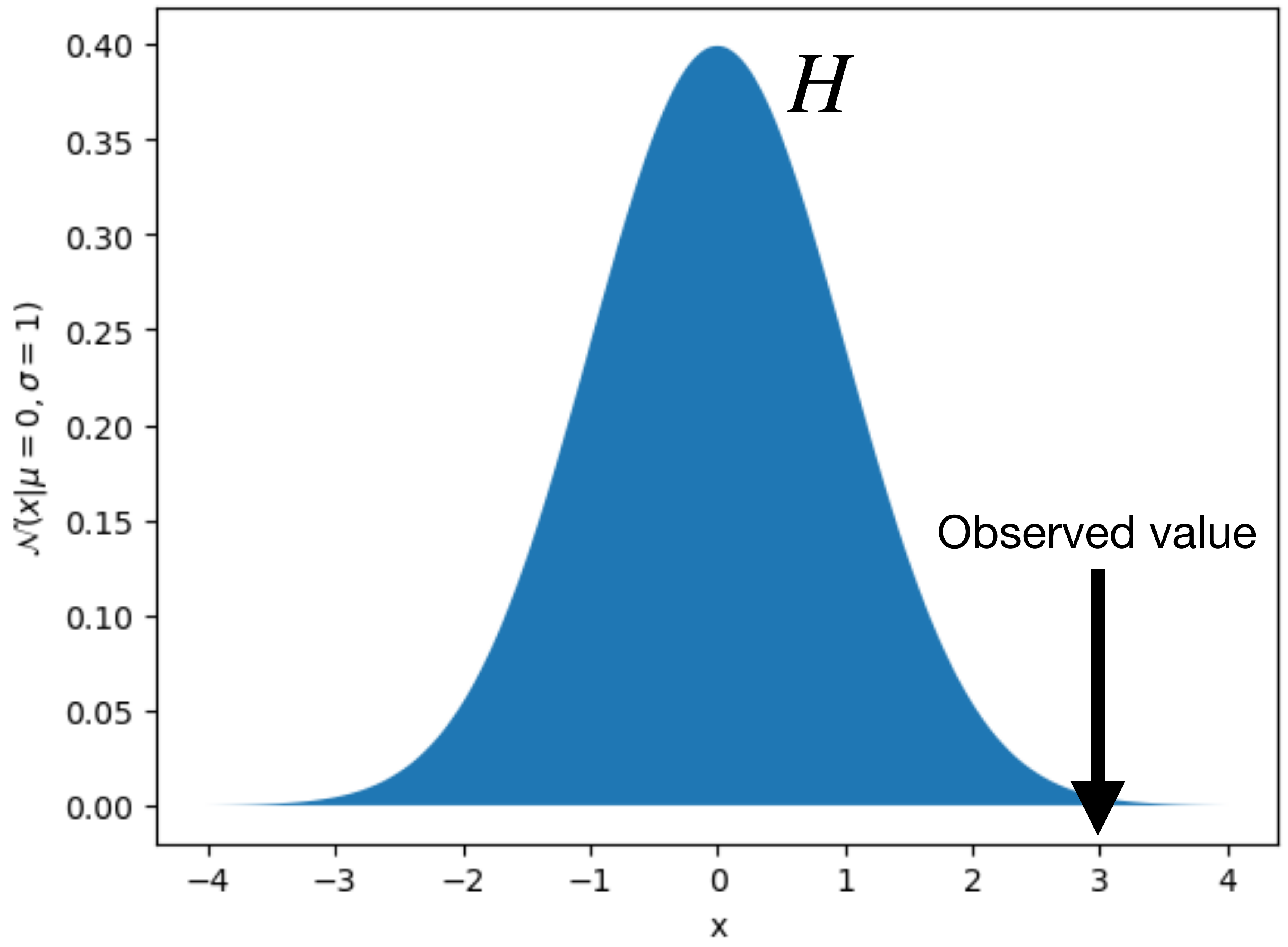
If I repeat the experiment an **infinitely** times, assuming H to be **true**, I would have observed ***D > 3*** only **0.27%** of the times

The Frequentist approach

A trivial example:

I have performed an observation and got the experimental data D = 3.
According to the hypotheses H, D is a random variable that follows a normal distribution centered in zero and variance = 1

$$p(D\,|\,H) = \mathcal{N}(x = 3\,|\,\mu = 0, \sigma = 1)$$



Conclusion of the inference analysis performed with the frequentist approach:

If I repeat the experiment an **infinitely** times, assuming H to be **true**, I would have observed ***D > 3*** only **0.27%** of the times

or

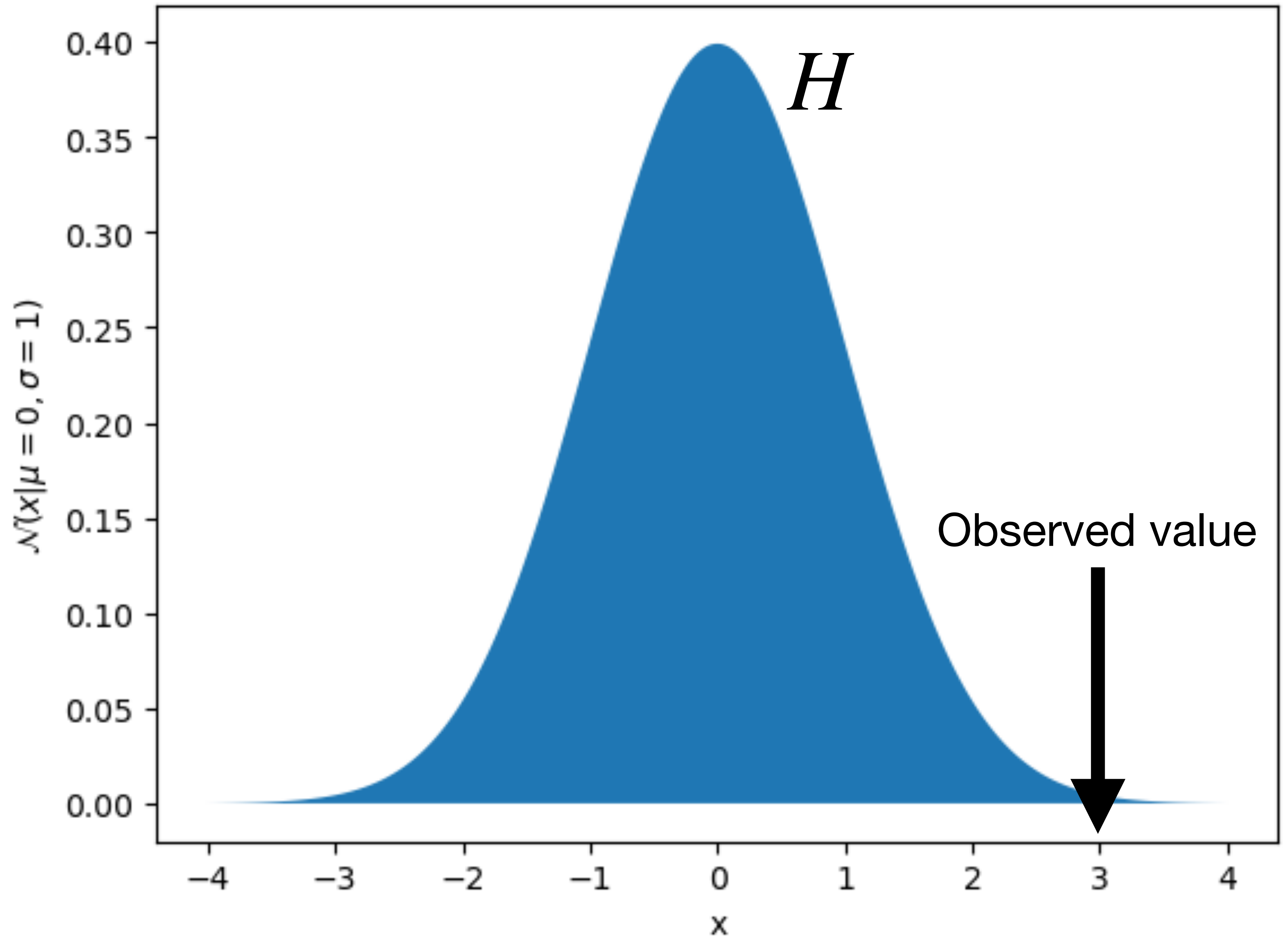
Hypothesis H is **rejected** at the 99.73% **Confidence Level (CL)**

The Frequentist approach

A trivial example:

I have performed an observation and got the experimental data D = 3.
According to the hypotheses H, D is a random variable that follows a normal distribution centered in zero and variance = 1

$$p(D \mid H) = \mathcal{N}(x = 3 \mid \mu = 0, \sigma = 1)$$



Conclusion of the inference analysis performed with the frequentist approach:

If I repeat the experiment an **infinitely** times, assuming H to be **true**, I would have observed ***D > 3*** only **0.27%** of the times

or

Hypothesis H is **rejected** at the 99.73% **Confidence Level (CL)**

or

Hypothesis H is **rejected** with a **significance** of 3 "***sigma***"

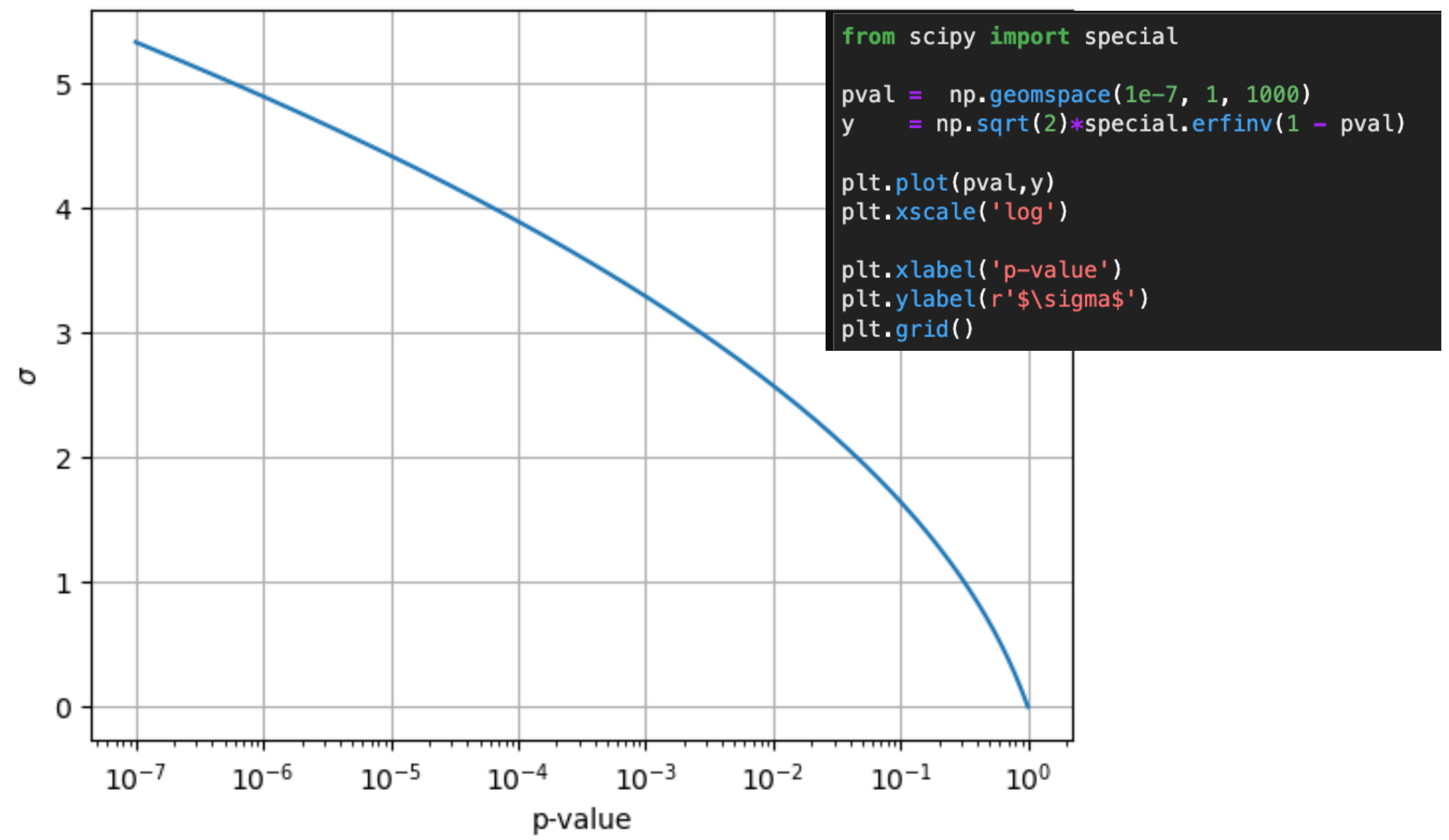
The Frequentist approach

A trivial example:

I have performed an observation and got the experimental data D = 3.
According to the hypotheses H, D is a random variable that follows a normal distribution centered in zero and variance = 1

A bit of terminology

$$\sigma = \sqrt{2} \cdot \mathrm{erf}^{-1}(\mathrm{CL}) = \sqrt{2} \cdot \mathrm{erf}^{-1}(1 - \text{p-value})$$



```
from scipy import special

pval =  np.geomspace(1e-7, 1, 1000)
y    = np.sqrt(2)*special.erfinv(1 - pval)

plt.plot(pval,y)
plt.xscale('log')

plt.xlabel('p-value')
plt.ylabel(r'$\sigma$')
plt.grid()
```

Conclusion of the inference analysis performed with the frequentist approach:

If I repeat the experiment an **infinitely** times, assuming H to be **true**, I would have observed ***D > 3*** only **0.27%** of the times

or

p-value

CL = 1- p-value

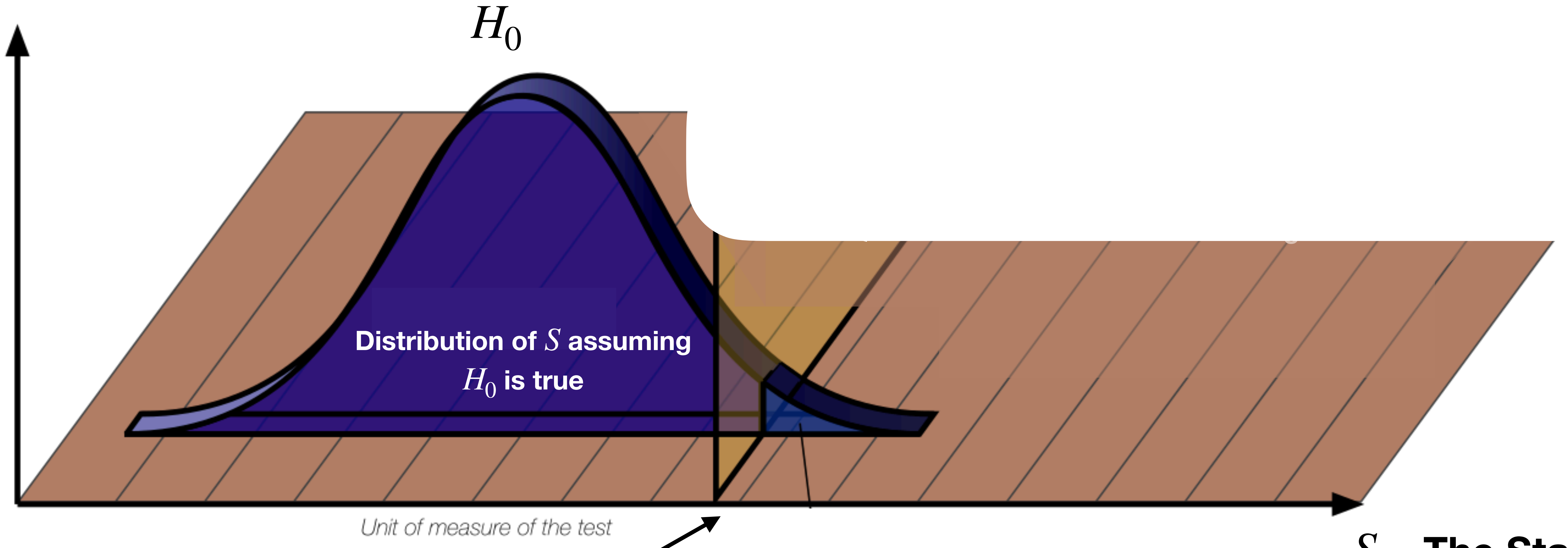
Hypothesis H is **rejected** at the 99.73% **Confidence Level (CL)**

or

$\sigma = \sqrt{2} \cdot \mathrm{erf}^{-1}(\mathrm{CL})$

Hypothesis H is **rejected** with a **significance** of 3 "***sigma***"

The Frequentist approach: more terminology…



Cutoff

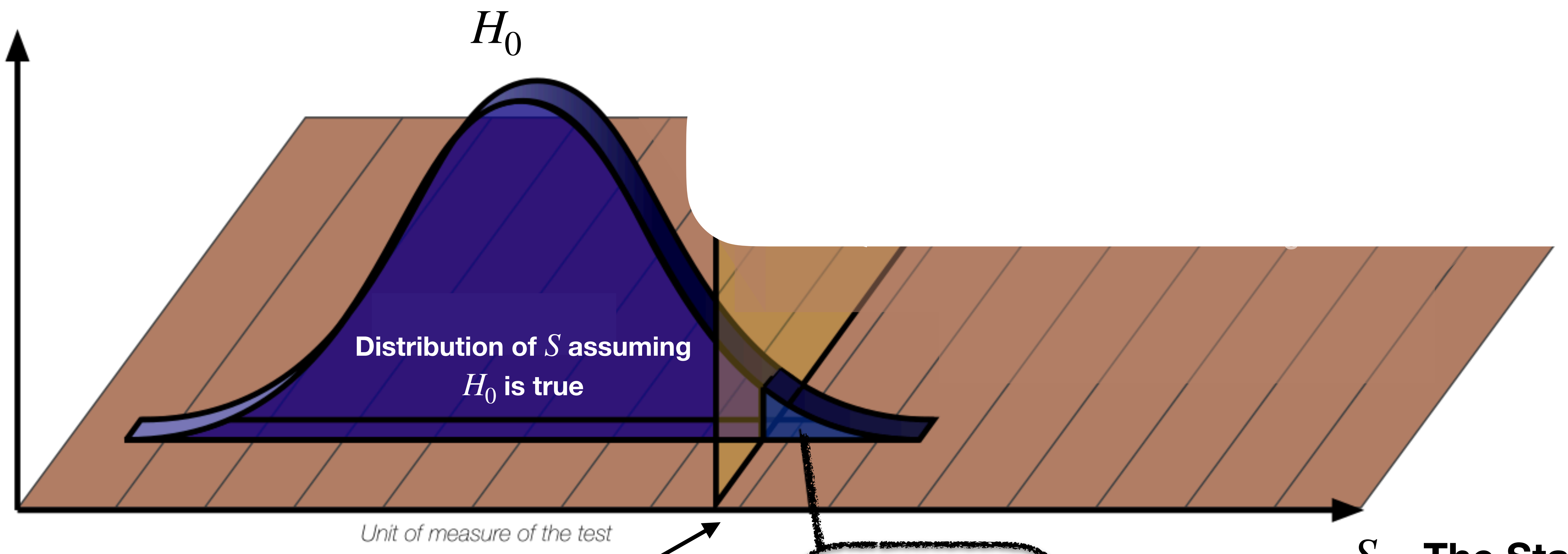
If the statistic is above the cutoff *I reject* H_0

S = The Statistic

The statistic is a single number obtained from the observed data and the assumed model

$S = f(D, H)$

The Frequentist approach: more terminology…



Cutoff

If the statistic is above the cutoff *I reject* H_0

α

Type-I error

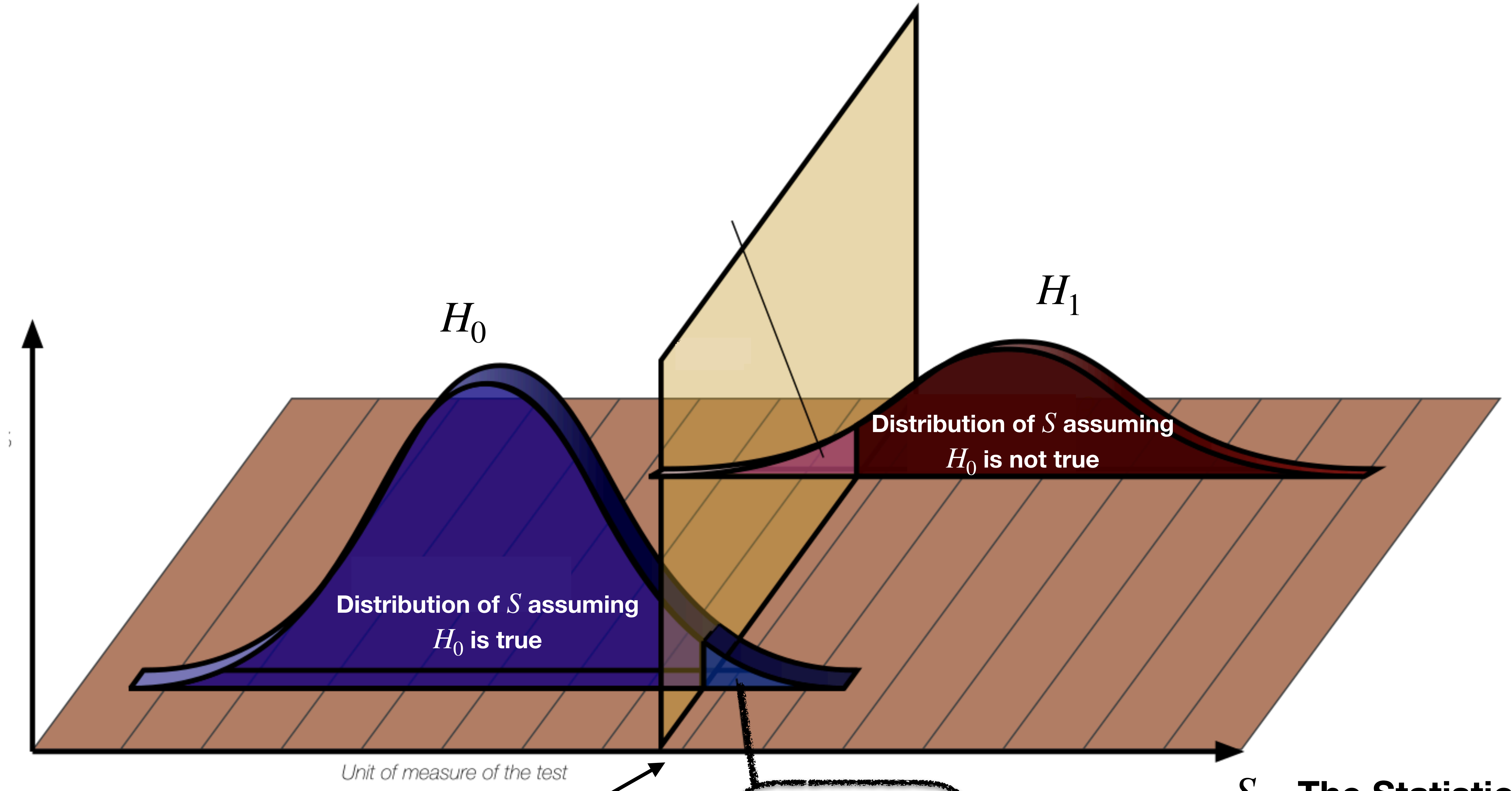
False positive:
Probability of **rejecting** H_0 when H_0 is true

S = **The Statistic**

The statistic is a single number obtained from the observed data and the assumed model

$S = f(D, H)$

The Frequentist approach: more terminology…



Cutoff

If the statistic is above the cutoff *I reject* H_0

α

Type-I error

False positive:
Probability of **rejecting** H_0
when H_0 is true

S = The Statistic

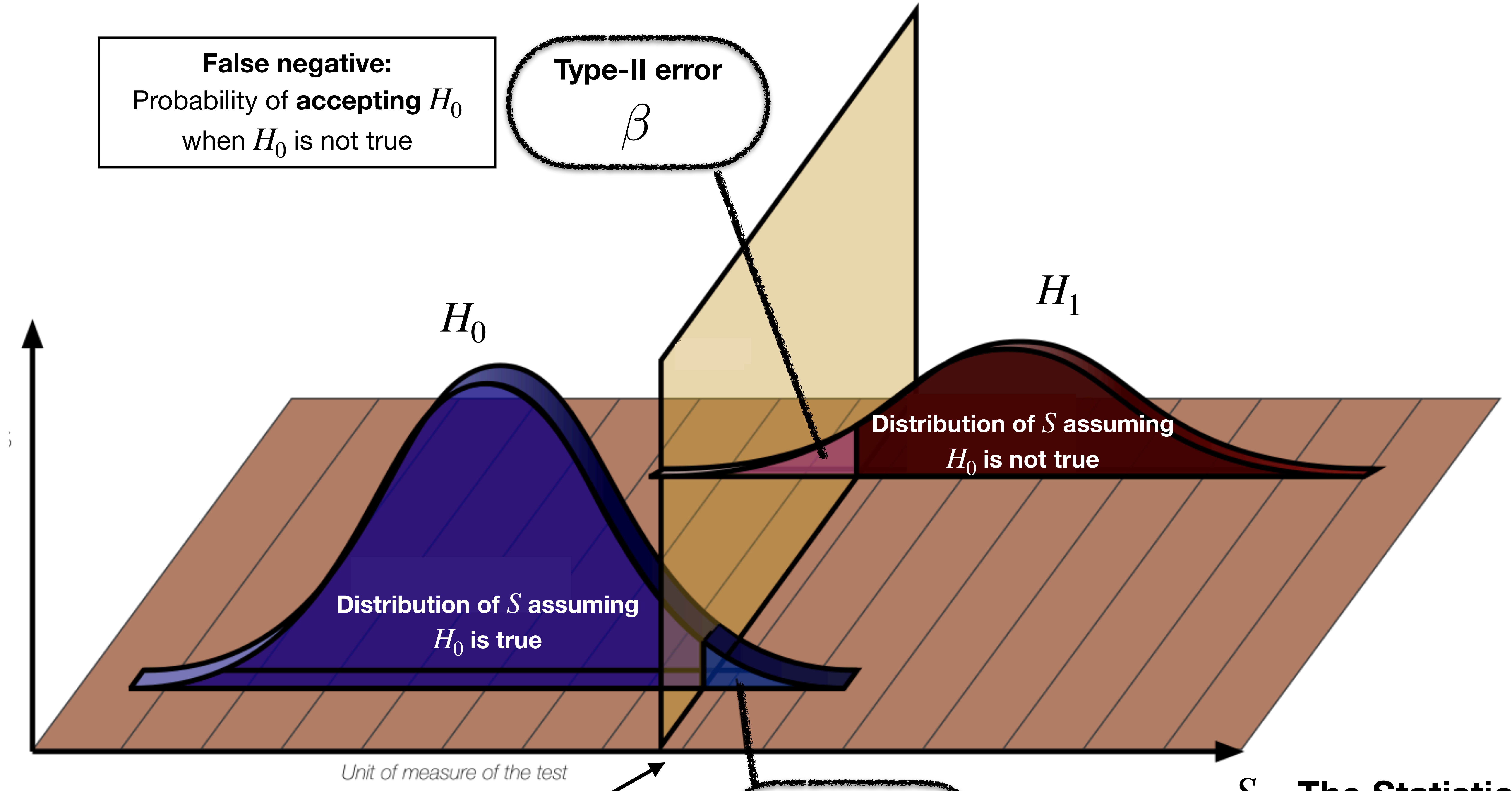
The statistic is a single number obtained from the observed data and the assumed model

$S = f(D, H)$

The Frequentist approach: more terminology...

False negative:
Probability of **accepting** H_0
when H_0 is not true

Type-II error
β



α
Type-I error

False positive:
Probability of **rejecting** H_0
when H_0 is true

Cutoff

If the statistic is above the cutoff *I reject* H_0

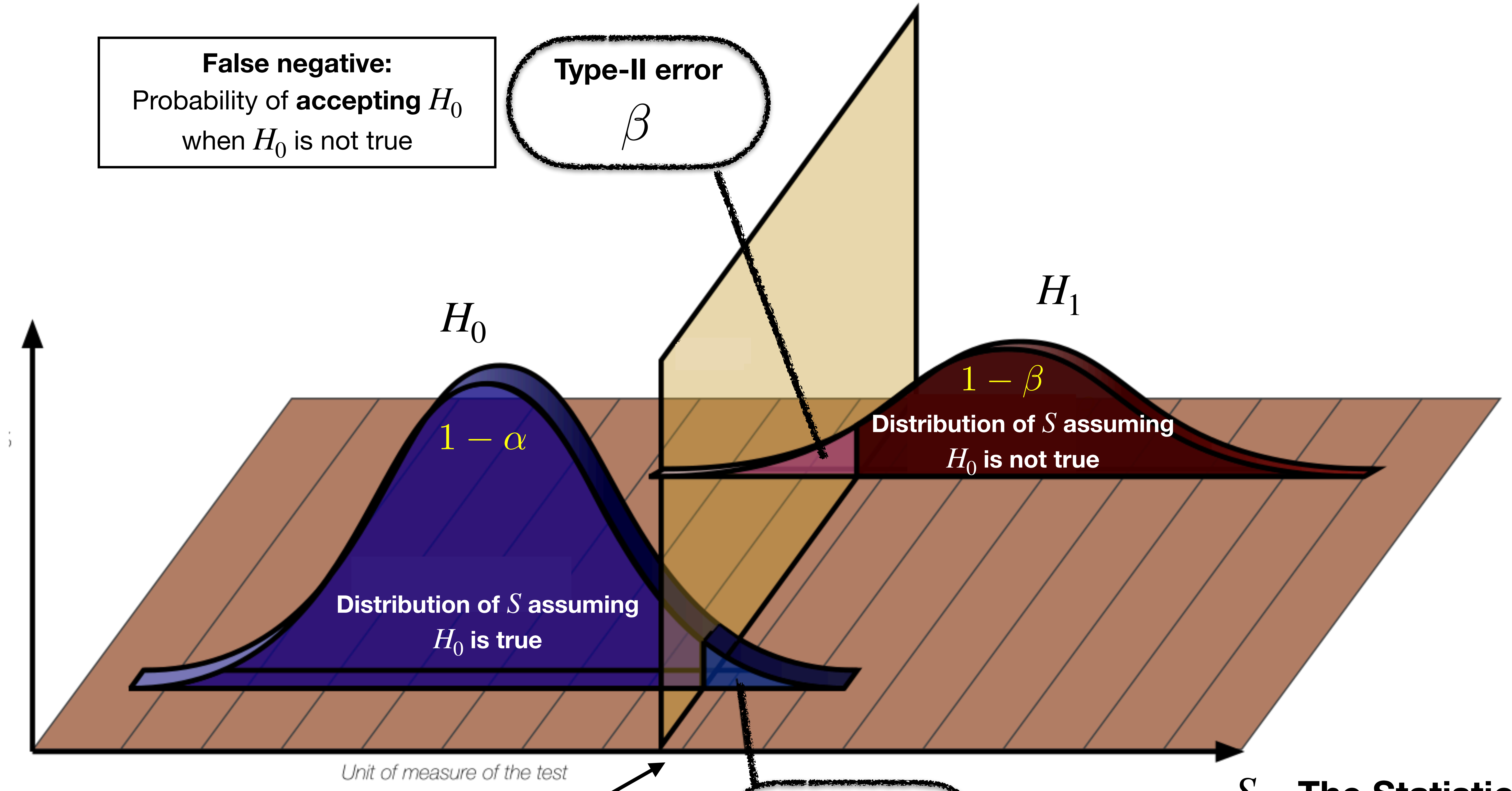
S = The Statistic

The statistic is a single number obtained from the observed data and the assumed model

$S = f(D, H)$

The Frequentist approach: more terminology…

False negative:
Probability of **accepting** H_0
when H_0 is not true



Cutoff

If the statistic is above the cutoff I reject H_0

α
Type-I error

False positive:
Probability of **rejecting** H_0
when H_0 is true

Power of the test

$$1-\beta$$

Probability of **rejecting** H_0
when H_0 is not true

Confidence level

$$1-\alpha$$

Probability of **accepting** H_0
when H_0 is true

S = The Statistic

The statistic is a single number obtained from the observed data and the assumed model

$S = f(D, H)$

Time for a quick recap

1. The **Bayesian** approach allows us to quantify our “opinion” on a given model from the observed data using the rules of **probability theory**
    - **Pros:** Alternative hypotheses are taken into account. No need to define a statistic and to know its distribution.
    - **Cons:** One needs a prior distribution.

Time for a quick recap

1. The **Bayesian** approach allows us to quantify our “opinion” on a given model from the observed data using the rules of **probability theory**
    - **Pros:** Alternative hypotheses are taken into account. No need to define a statistic and to know its distribution.
    - **Cons:** One needs a prior distribution.

2. The **frequentist** approach makes us exclude a model with given confidence by looking at infinity repetitions of the experiments in which the model is assumed to be true
    - **Pros:** No need for priors
    - **Cons:** Choice of the statistic is arbitrary. Alternative hypothesis not taken into account. Type I and II errors.

Time for a quick recap

A quote from Prof. Luis Lyons

the probability of a model to be true

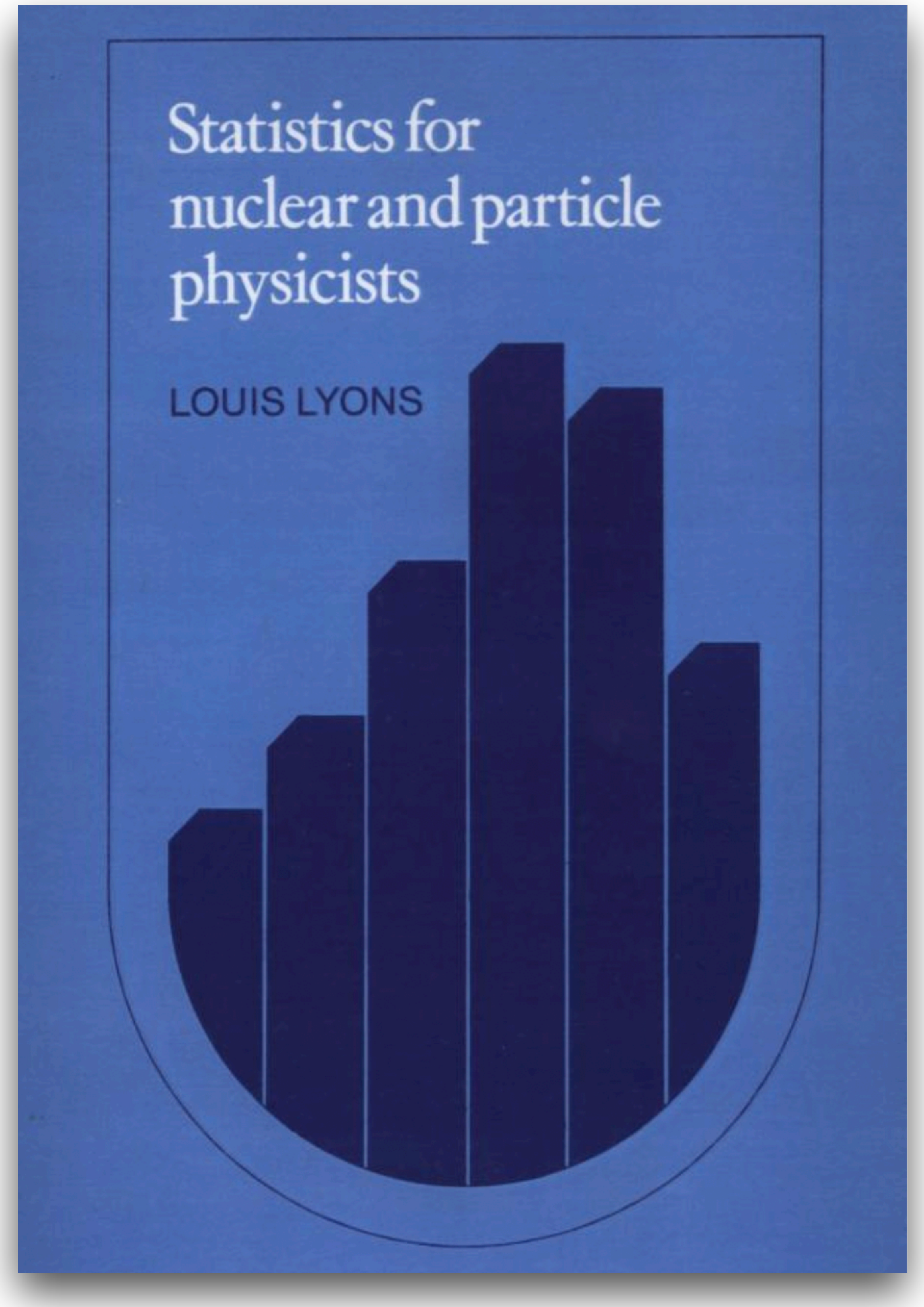
"Bayesians address the question everyone is interested in, by using assumptions no one believes"

the priors

test statistic

"Frequentists use impeccable logic to deal with an issue of no interest to anyone"

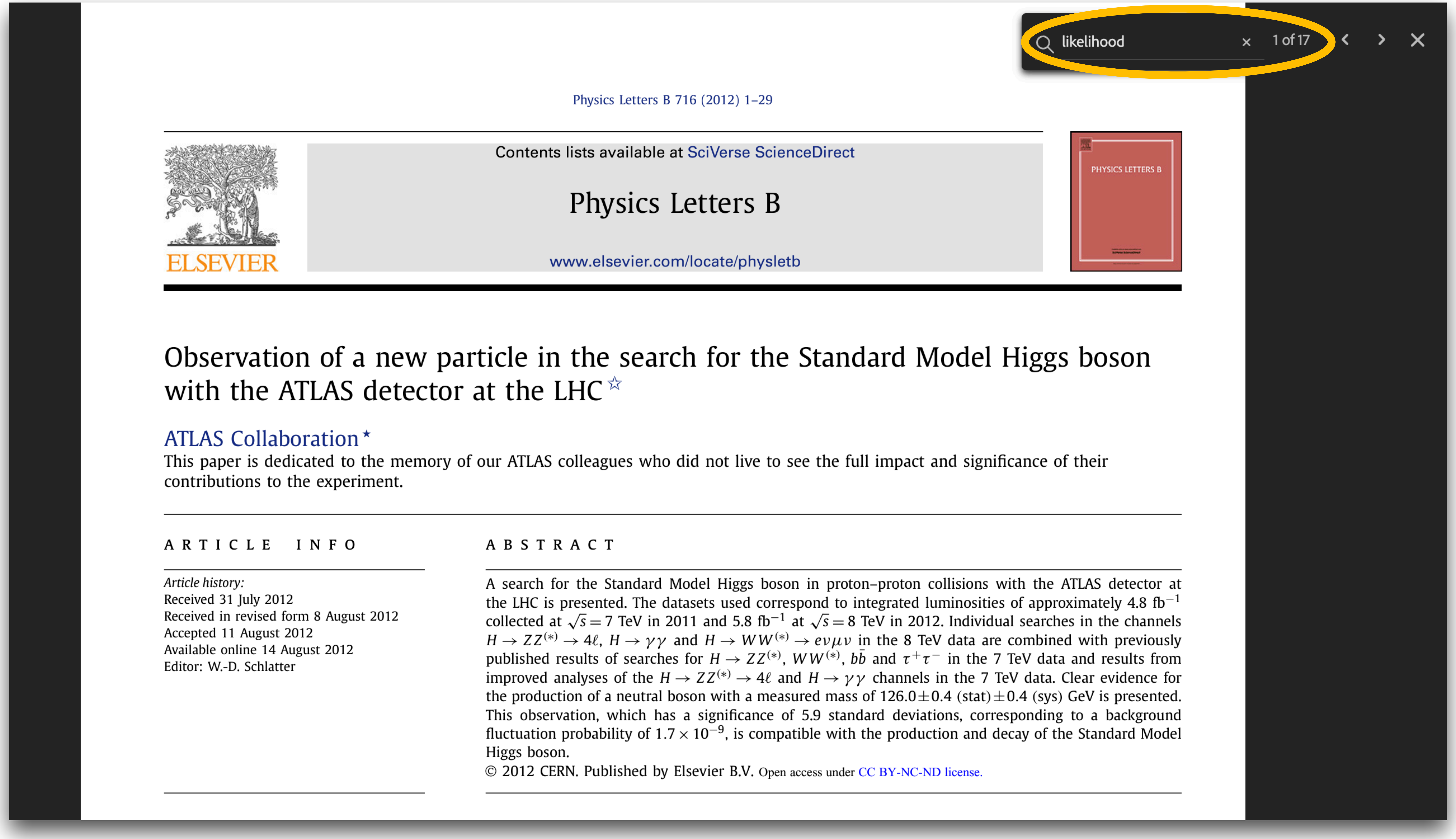
what happen if I repeat the experiment an infinite times assuming a model to be always true



The Likelihood

The Likelihood

Let’s take a look at some recent papers



Physics Letters B 716 (2012) 1–29

Contents lists available at SciVerse ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC ☆

ATLAS Collaboration *

This paper is dedicated to the memory of our ATLAS colleagues who did not live to see the full impact and significance of their contributions to the experiment.

ARTICLE INFO

Article history:
Received 31 July 2012
Received in revised form 8 August 2012
Accepted 11 August 2012
Available online 14 August 2012
Editor: W.-D. Schlatter

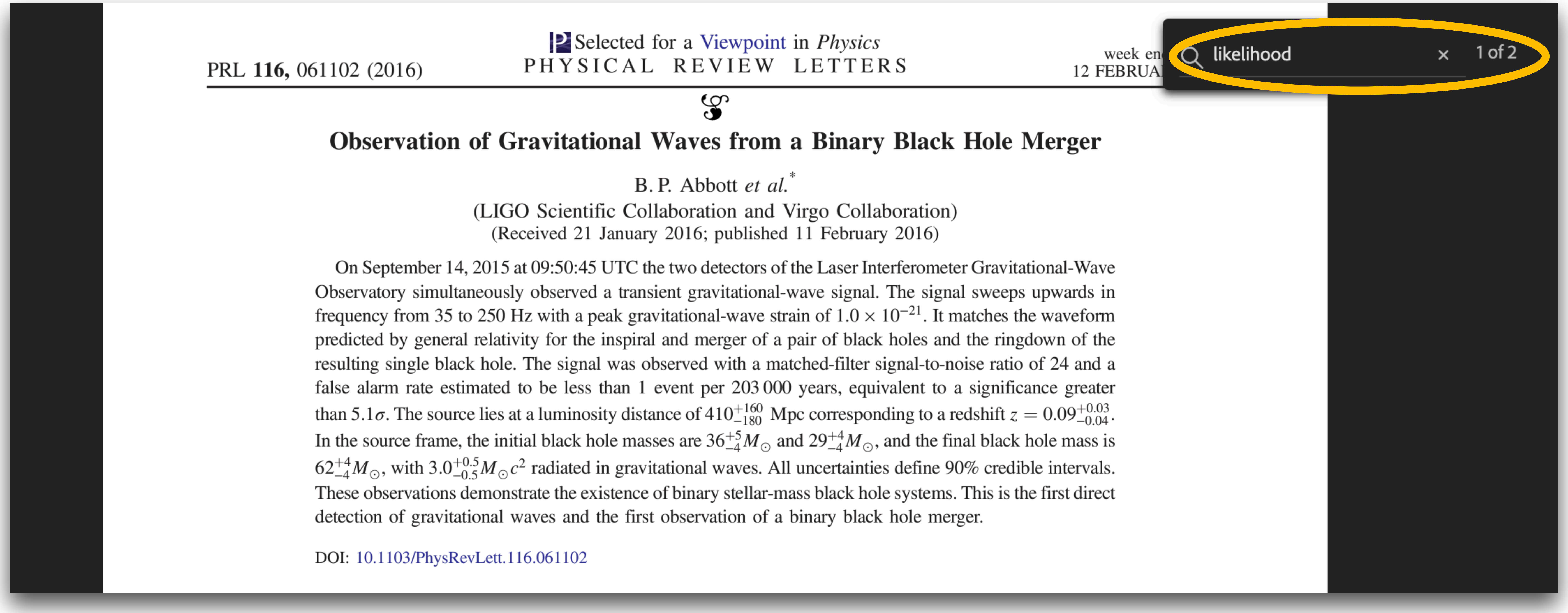
ABSTRACT

A search for the Standard Model Higgs boson in proton–proton collisions with the ATLAS detector at the LHC is presented. The datasets used correspond to integrated luminosities of approximately 4.8 fb^{-1} collected at $\sqrt{s} = 7$ TeV in 2011 and 5.8 fb^{-1} at $\sqrt{s} = 8$ TeV in 2012. Individual searches in the channels $H \to ZZ^{(*)} \to 4\ell$, $H \to \gamma\gamma$ and $H \to WW^{(*)} \to e\nu\mu\nu$ in the 8 TeV data are combined with previously published results of searches for $H \to ZZ^{(*)}$, $WW^{(*)}$, $b\bar{b}$ and $\tau^+\tau^-$ in the 7 TeV data and results from improved analyses of the $H \to ZZ^{(*)} \to 4\ell$ and $H \to \gamma\gamma$ channels in the 7 TeV data. Clear evidence for the production of a neutral boson with a measured mass of 126.0 ± 0.4 (stat) ± 0.4 (sys) GeV is presented. This observation, which has a significance of 5.9 standard deviations, corresponding to a background fluctuation probability of 1.7×10^{-9}, is compatible with the production and decay of the Standard Model Higgs boson.

© 2012 CERN. Published by Elsevier B.V. Open access under CC BY-NC-ND license.

The Likelihood

Let's take a look at some recent papers



Selected for a Viewpoint in *Physics*

PRL **116,** 061102 (2016) PHYSICAL REVIEW LETTERS week en
12 FEBRUA

Observation of Gravitational Waves from a Binary Black Hole Merger

B. P. Abbott *et al.**

(LIGO Scientific Collaboration and Virgo Collaboration)

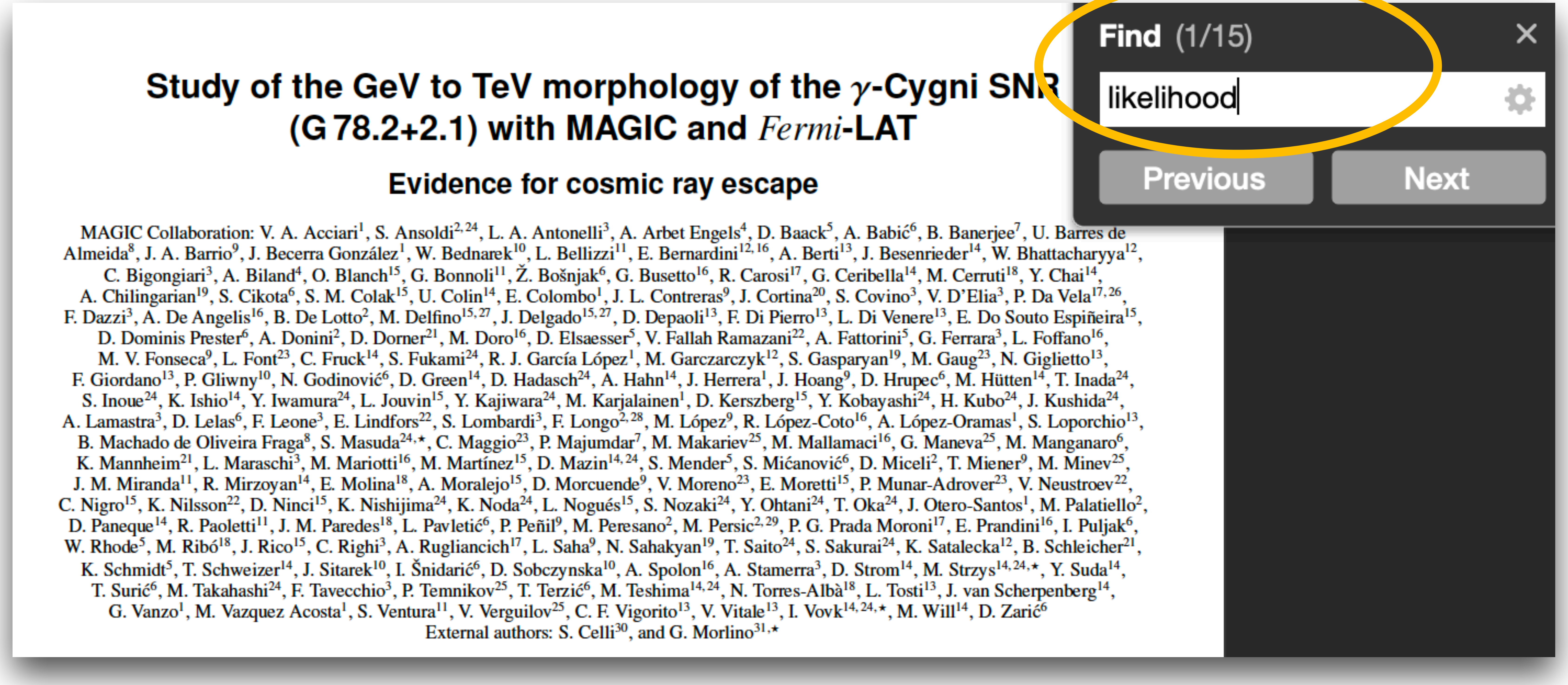
(Received 21 January 2016; published 11 February 2016)

On September 14, 2015 at 09:50:45 UTC the two detectors of the Laser Interferometer Gravitational-Wave Observatory simultaneously observed a transient gravitational-wave signal. The signal sweeps upwards in frequency from 35 to 250 Hz with a peak gravitational-wave strain of 1.0×10^{-21}. It matches the waveform predicted by general relativity for the inspiral and merger of a pair of black holes and the ringdown of the resulting single black hole. The signal was observed with a matched-filter signal-to-noise ratio of 24 and a false alarm rate estimated to be less than 1 event per 203 000 years, equivalent to a significance greater than 5.1σ. The source lies at a luminosity distance of 410^{+160}_{-180} Mpc corresponding to a redshift $z = 0.09^{+0.03}_{-0.04}$. In the source frame, the initial black hole masses are $36^{+5}_{-4} M_\odot$ and $29^{+4}_{-4} M_\odot$, and the final black hole mass is $62^{+4}_{-4} M_\odot$, with $3.0^{+0.5}_{-0.5} M_\odot c^2$ radiated in gravitational waves. All uncertainties define 90% credible intervals. These observations demonstrate the existence of binary stellar-mass black hole systems. This is the first direct detection of gravitational waves and the first observation of a binary black hole merger.

DOI: 10.1103/PhysRevLett.116.061102

The Likelihood

Let’s take a look at some recent papers



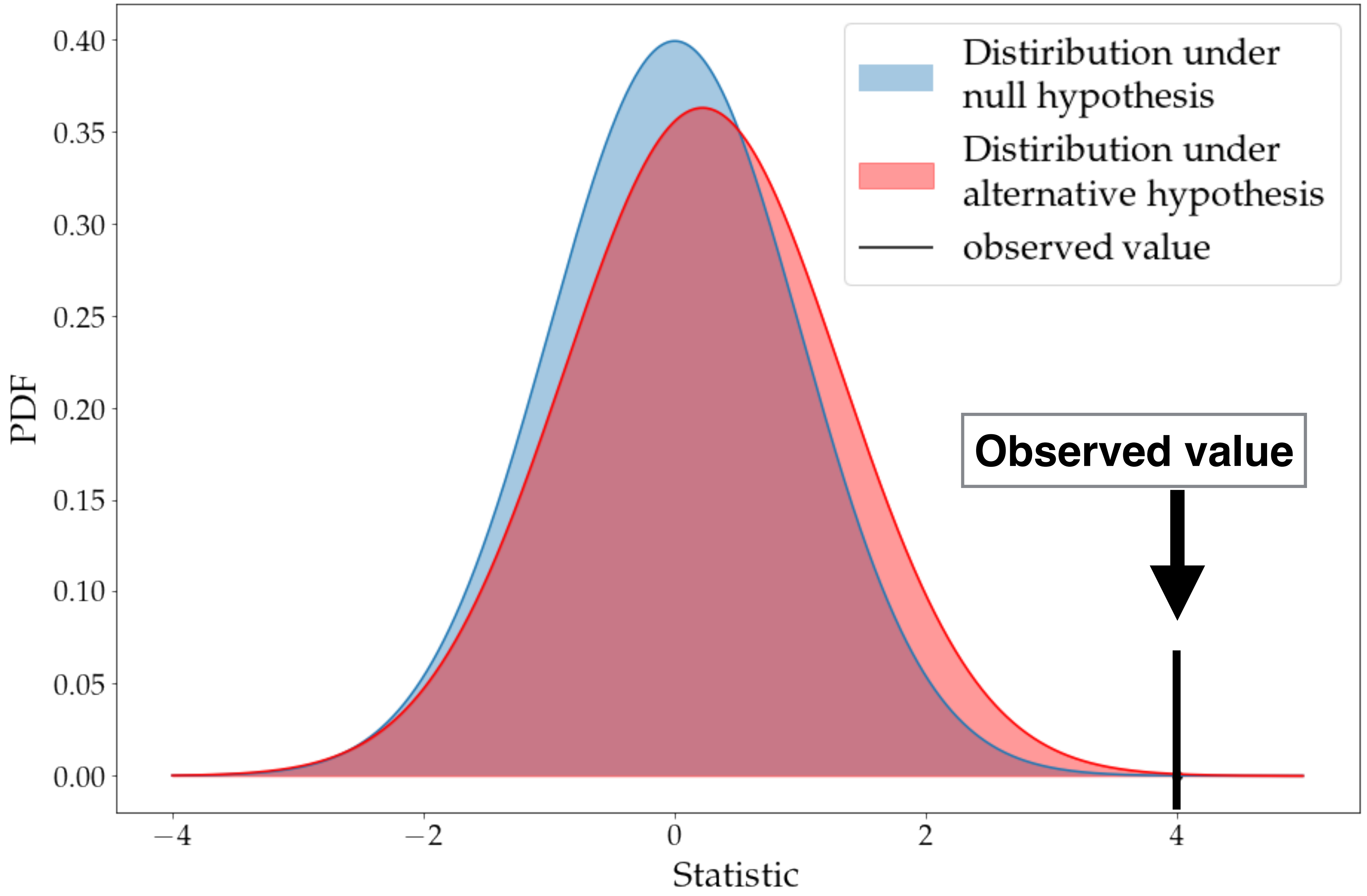
Study of the GeV to TeV morphology of the γ-Cygni SNR (G 78.2+2.1) with MAGIC and *Fermi*-LAT

Evidence for cosmic ray escape

MAGIC Collaboration: V. A. Acciari[1], S. Ansoldi[2,24], L. A. Antonelli[3], A. Arbet Engels[4], D. Baack[5], A. Babić[6], B. Banerjee[7], U. Barres de Almeida[8], J. A. Barrio[9], J. Becerra González[1], W. Bednarek[10], L. Bellizzi[11], E. Bernardini[12,16], A. Berti[13], J. Besenrieder[14], W. Bhattacharyya[12], C. Bigongiari[3], A. Biland[4], O. Blanch[15], G. Bonnoli[11], Ž. Bošnjak[6], G. Busetto[16], R. Carosi[17], G. Ceribella[14], M. Cerruti[18], Y. Chai[14], A. Chilingarian[19], S. Cikota[6], S. M. Colak[15], U. Colin[14], E. Colombo[1], J. L. Contreras[9], J. Cortina[20], S. Covino[3], V. D’Elia[3], P. Da Vela[17,26], F. Dazzi[3], A. De Angelis[16], B. De Lotto[2], M. Delfino[15,27], J. Delgado[15,27], D. Depaoli[13], F. Di Pierro[13], L. Di Venere[13], E. Do Souto Espiñeira[15], D. Dominis Prester[6], A. Donini[2], D. Dorner[21], M. Doro[16], D. Elsaesser[5], V. Fallah Ramazani[22], A. Fattorini[5], G. Ferrara[3], L. Foffano[16], M. V. Fonseca[9], L. Font[23], C. Fruck[14], S. Fukami[24], R. J. García López[1], M. Garczarczyk[12], S. Gasparyan[19], M. Gaug[23], N. Giglietto[13], F. Giordano[13], P. Gliwny[10], N. Godinović[6], D. Green[14], D. Hadasch[24], A. Hahn[14], J. Herrera[1], J. Hoang[9], D. Hrupec[6], M. Hütten[14], T. Inada[24], S. Inoue[24], K. Ishio[14], Y. Iwamura[24], L. Jouvin[15], Y. Kajiwara[24], M. Karjalainen[1], D. Kerszberg[15], Y. Kobayashi[24], H. Kubo[24], J. Kushida[24], A. Lamastra[3], D. Lelas[6], F. Leone[3], E. Lindfors[22], S. Lombardi[3], F. Longo[2,28], M. López[9], R. López-Coto[16], A. López-Oramas[1], S. Loporchio[13], B. Machado de Oliveira Fraga[8], S. Masuda[24,⋆], C. Maggio[23], P. Majumdar[7], M. Makariev[25], M. Mallamaci[16], G. Maneva[25], M. Manganaro[6], K. Mannheim[21], L. Maraschi[3], M. Mariotti[16], M. Martínez[15], D. Mazin[14,24], S. Mender[5], S. Mićanović[6], D. Miceli[2], T. Miener[9], M. Minev[25], J. M. Miranda[11], R. Mirzoyan[14], E. Molina[18], A. Moralejo[15], D. Morcuende[9], V. Moreno[23], E. Moretti[15], P. Munar-Adrover[23], V. Neustroev[22], C. Nigro[15], K. Nilsson[22], D. Ninci[15], K. Nishijima[24], K. Noda[24], L. Nogués[15], S. Nozaki[24], Y. Ohtani[24], T. Oka[24], J. Otero-Santos[1], M. Palatiello[2], D. Paneque[14], R. Paoletti[11], J. M. Paredes[18], L. Pavletić[6], P. Peñil[9], M. Peresano[2], M. Persic[2,29], P. G. Prada Moroni[17], E. Prandini[16], I. Puljak[6], W. Rhode[5], M. Ribó[18], J. Rico[15], C. Righi[3], A. Rugliancich[17], L. Saha[9], N. Sahakyan[19], T. Saito[24], S. Sakurai[24], K. Satalecka[12], B. Schleicher[21], K. Schmidt[5], T. Schweizer[14], J. Sitarek[10], I. Šnidarić[6], D. Sobczynska[10], A. Spolon[16], A. Stamerra[3], D. Strom[14], M. Strzys[14,24,⋆], Y. Suda[14], T. Surić[6], M. Takahashi[24], F. Tavecchio[3], P. Temnikov[25], T. Terzić[6], M. Teshima[14,24], N. Torres-Albà[18], L. Tosti[13], J. van Scherpenberg[14], G. Vanzo[1], M. Vazquez Acosta[1], S. Ventura[11], V. Verguilov[25], C. F. Vigorito[13], V. Vitale[13], I. Vovk[14,24,⋆], M. Will[14], D. Zarić[6]
External authors: S. Celli[30], and G. Morlino[31,⋆]

The Likelihood: why is it so important?

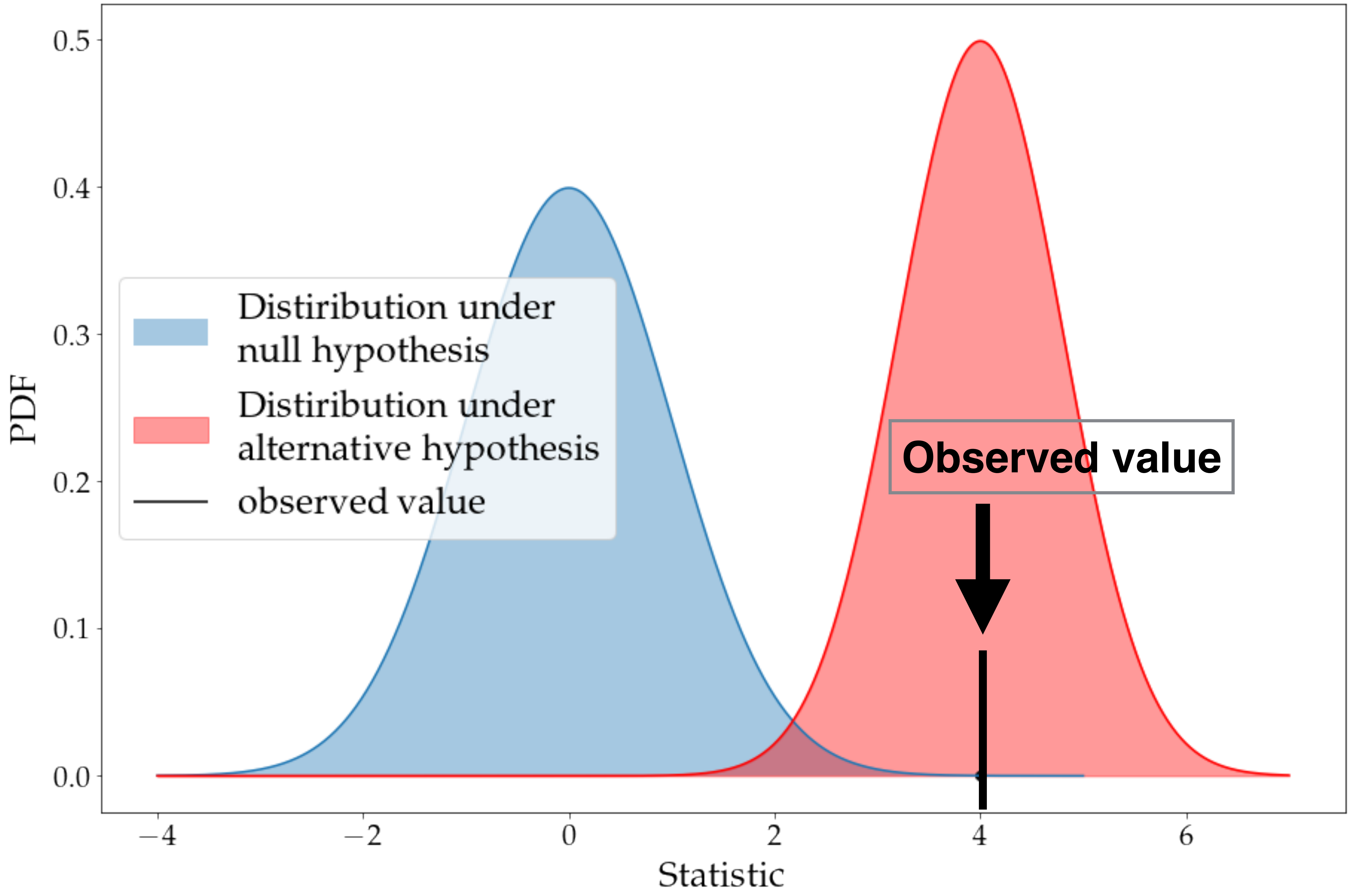
SCENARIO 1



Frequentist conclusion:

The null hypothesis is rejected at 4 sigma level

SCENARIO 2

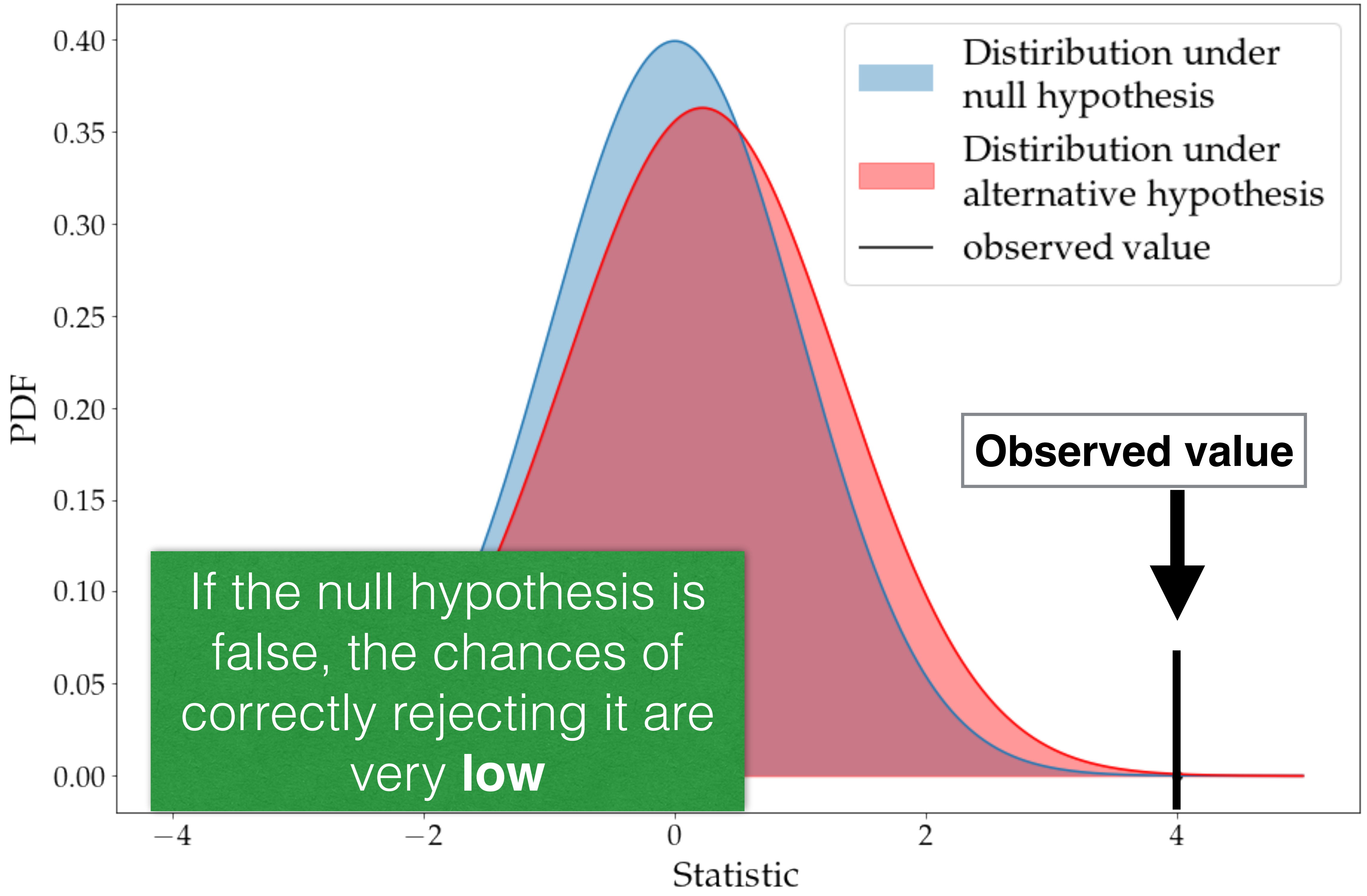


Frequentist conclusion:

The null hypothesis is rejected at 4 sigma level

The Likelihood: why is it so important?

SCENARIO 1

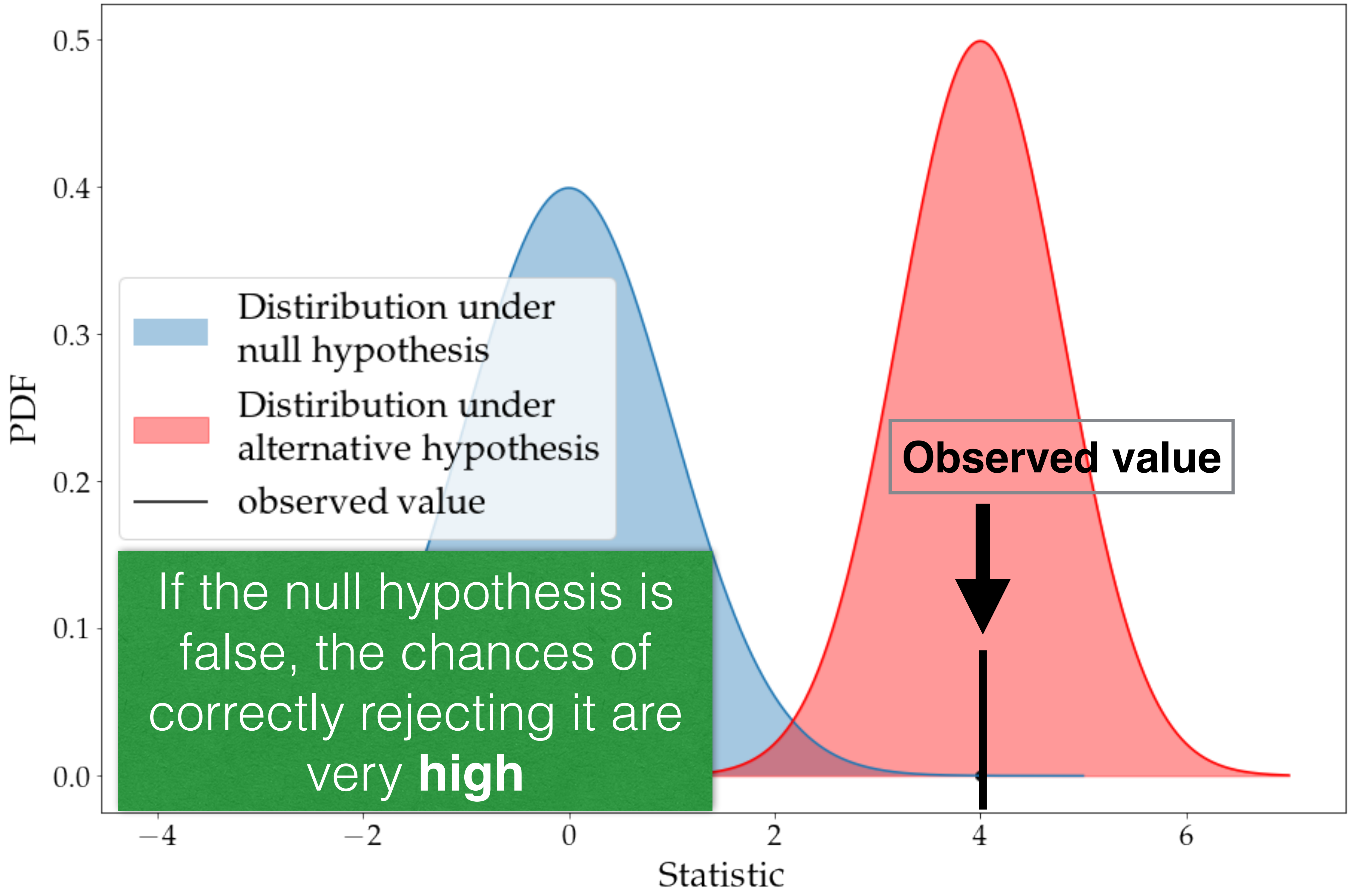


If the null hypothesis is false, the chances of correctly rejecting it are very **low**

Frequentist conclusion:

The null hypothesis is rejected at 4 sigma level

SCENARIO 2



If the null hypothesis is false, the chances of correctly rejecting it are very **high**

Frequentist conclusion:

The null hypothesis is rejected at 4 sigma level

The Likelihood: why is it so important?

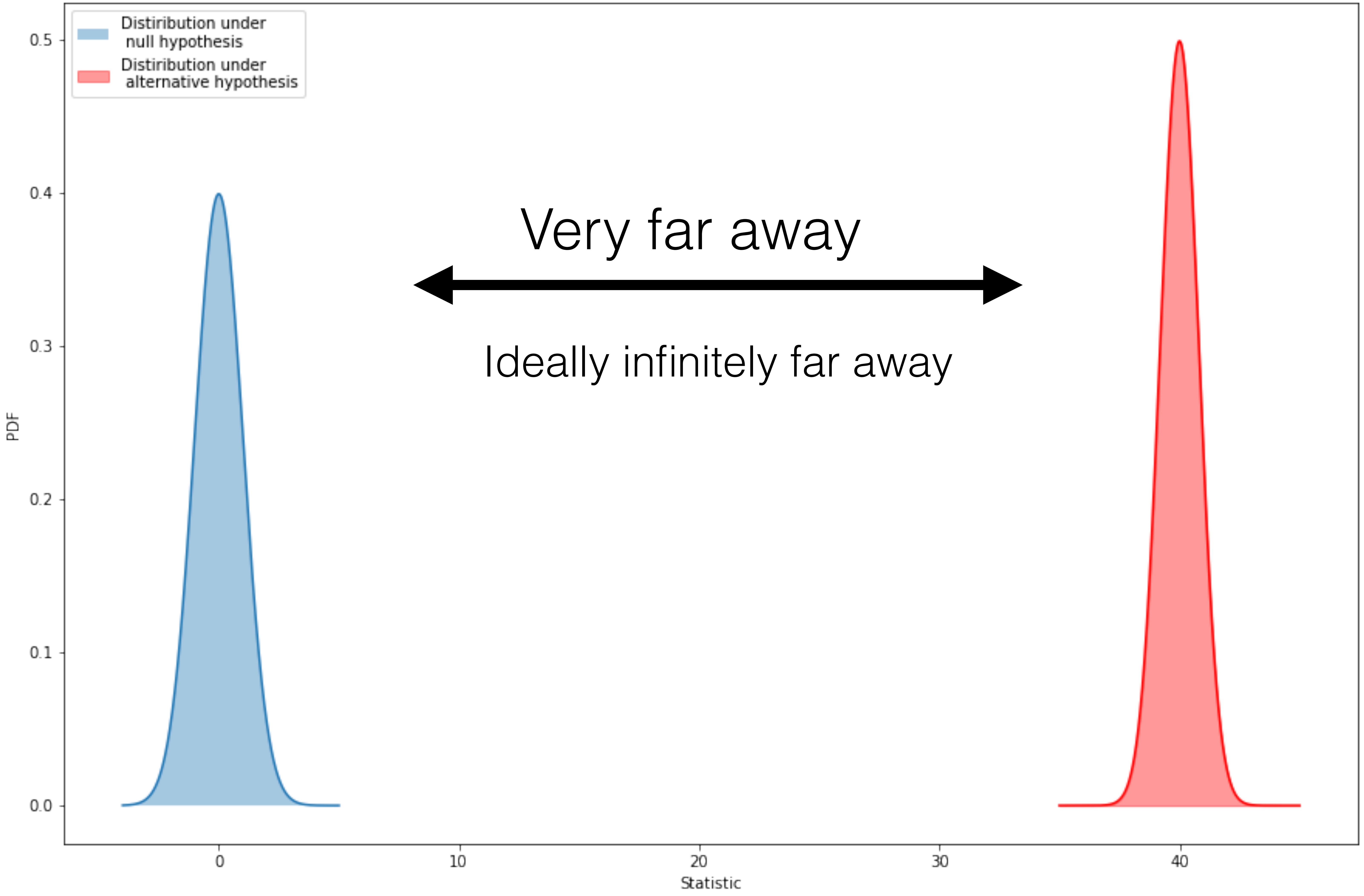
You want the statistic to give you a high chance of rejecting a hypothesis that is false

You want your statistic to be

POWERFUL!

The Likelihood: why is it so important?

Ideal case →

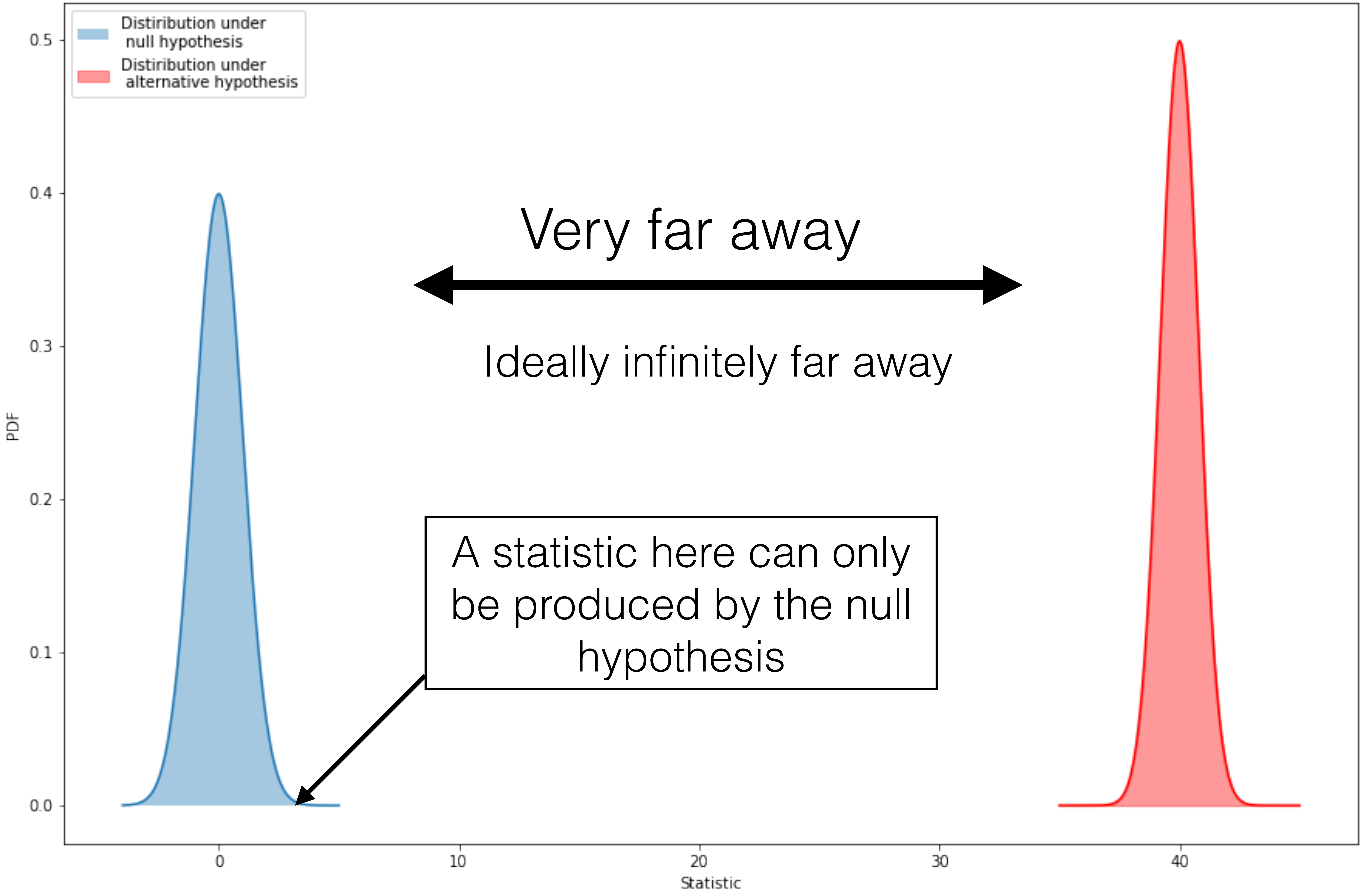


Very far away

Ideally infinitely far away

The Likelihood: why is it so important?

Ideal case →



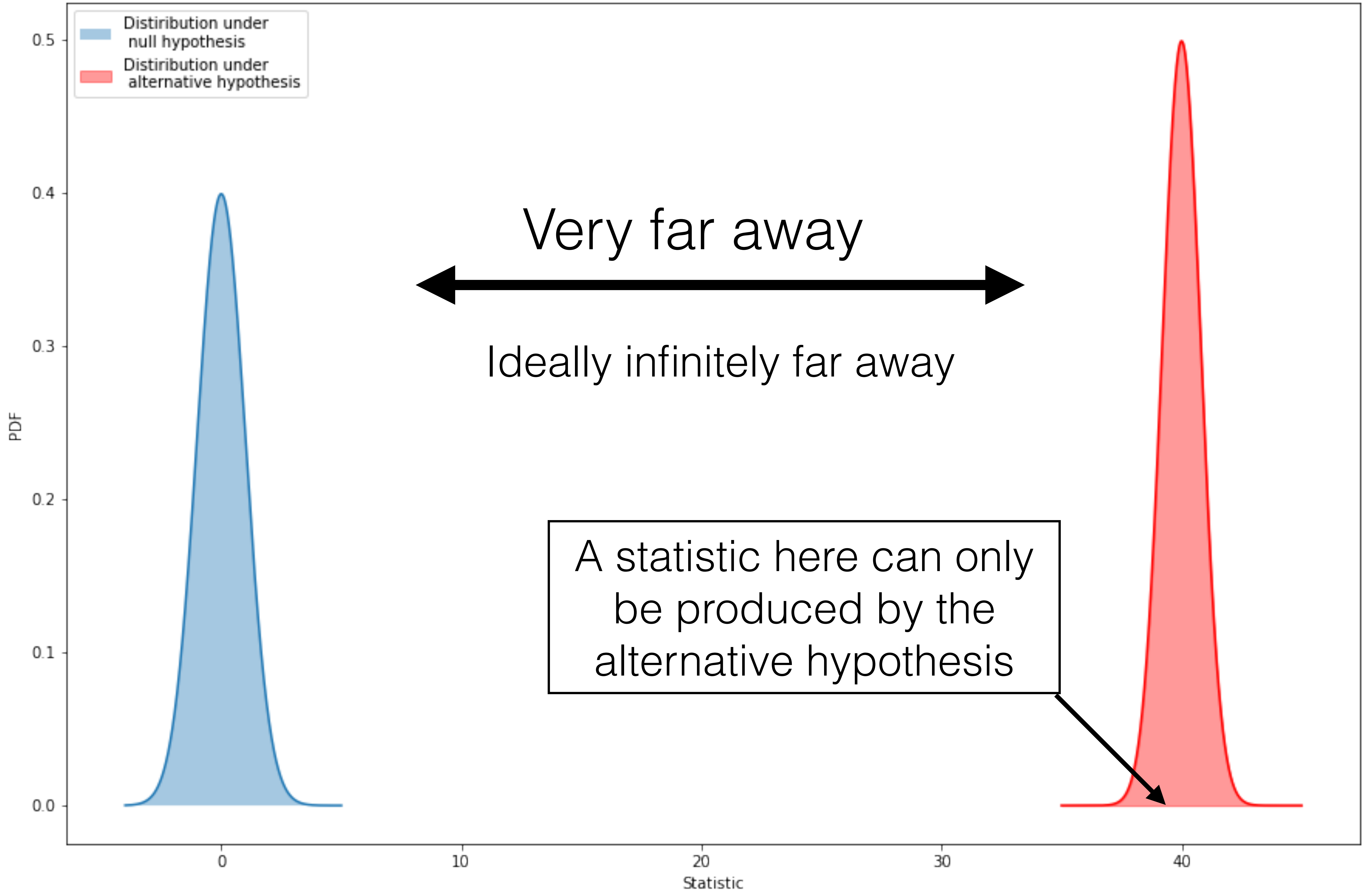
Very far away

Ideally infinitely far away

A statistic here can only be produced by the null hypothesis

The Likelihood: why is it so important?

Ideal case



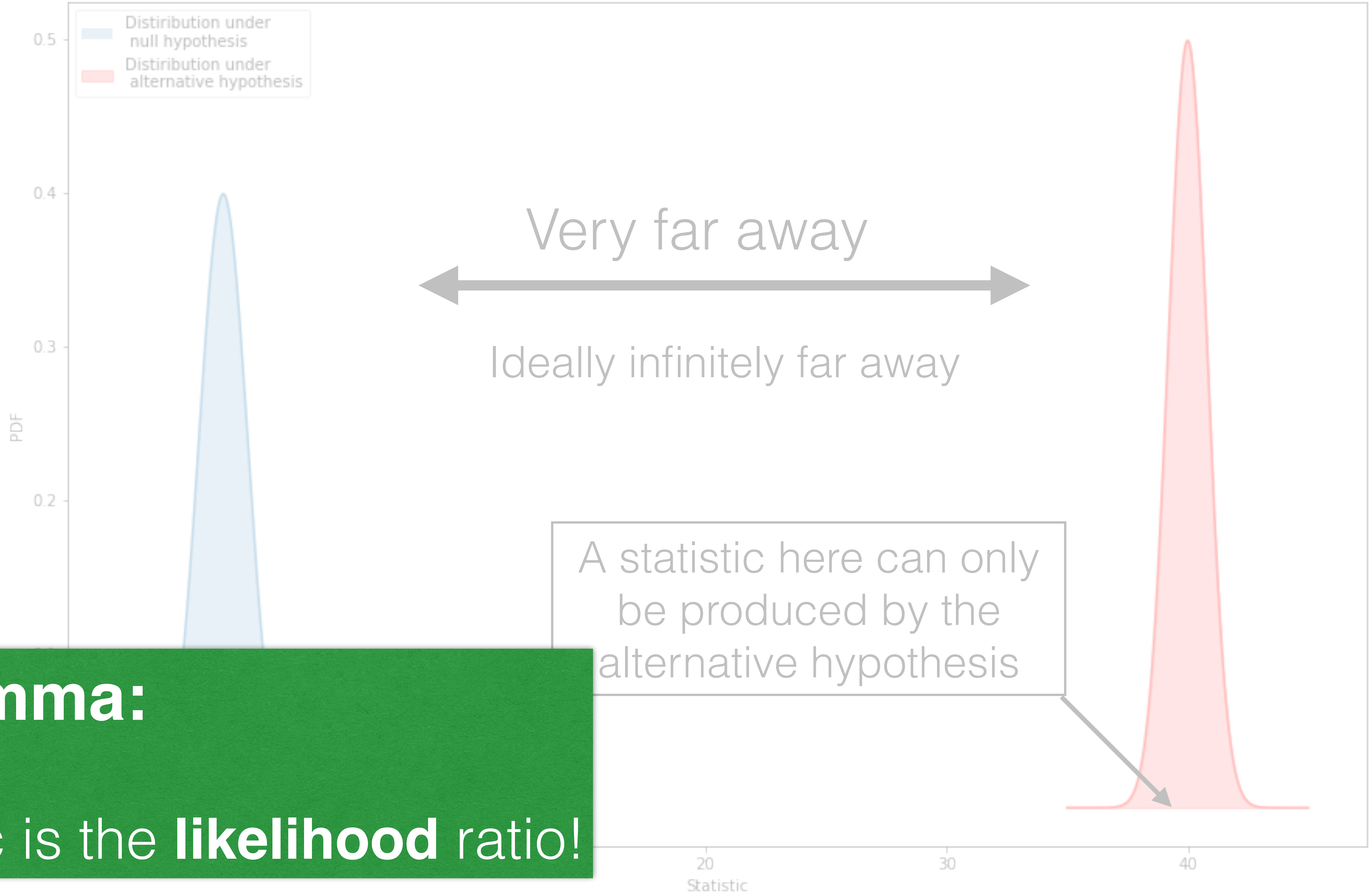
Very far away

Ideally infinitely far away

A statistic here can only be produced by the alternative hypothesis

The Likelihood: why is it so important?

Ideal case →



Very far away

Ideally infinitely far away

A statistic here can only be produced by the alternative hypothesis

In reality

↓

Neyman–Pearson lemma:

the most powerful statistic is the **likelihood** ratio!

The Likelihood ratio

First, let’s recall that:

The likelihood is a **function** of the model **parameters,** defined as the **probability** of observing the **data** assuming the model to be **true**

$$\mathcal{L}(\theta|D_{obs}) = p(D_{obs}|\theta)$$

Parameter of the hypothesis you want to test

The data we observed

The Likelihood ratio

Parameter of the hypothesis
you want to test

$$\frac{\mathcal{L}(\theta|D_{obs})}{\mathcal{L}(\hat{\theta}|D_{obs})}$$

Best fit or value
that maximises
the likelihood

Observed data

The Likelihood ratio

Parameter of the hypothesis you want to test

$$-2\log \frac{\mathcal{L}(\theta|D_{obs})}{\mathcal{L}(\hat{\theta}|D_{obs})}$$

⚠ Best fit or value that maximises the likelihood

Observed data

The Likelihood ratio

Parameter of the hypothesis you want to test

THE WILKS' THEOREM

$$-2\log\frac{\mathcal{L}(\theta|D_{obs})}{\mathcal{L}(\hat{\theta}|D_{obs})} \sim \chi^2$$

number of degrees
=
number of dimensions of θ

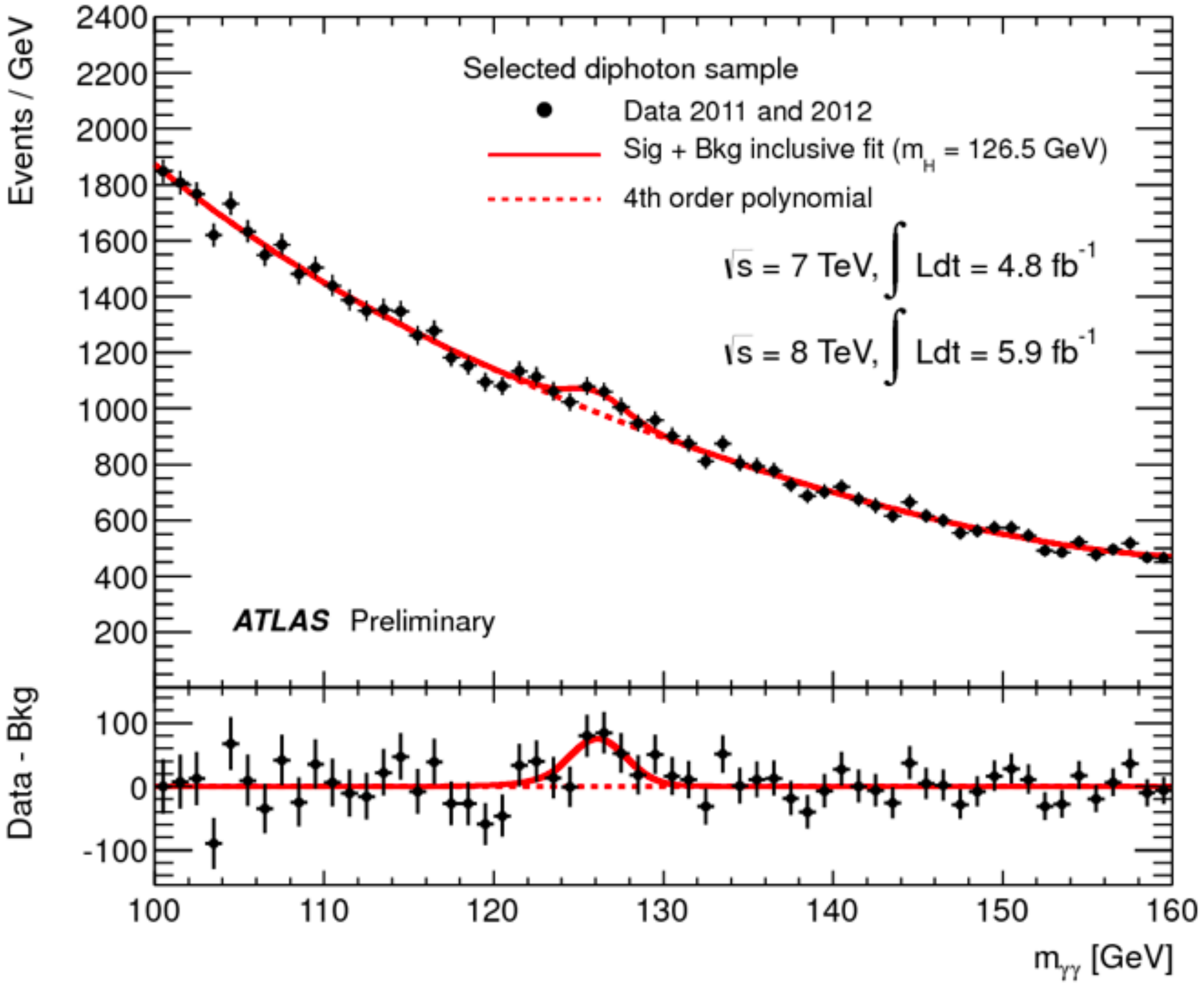
⚠ Best fit or value that maximises the likelihood

Observed data

The Likelihood: Higgs example

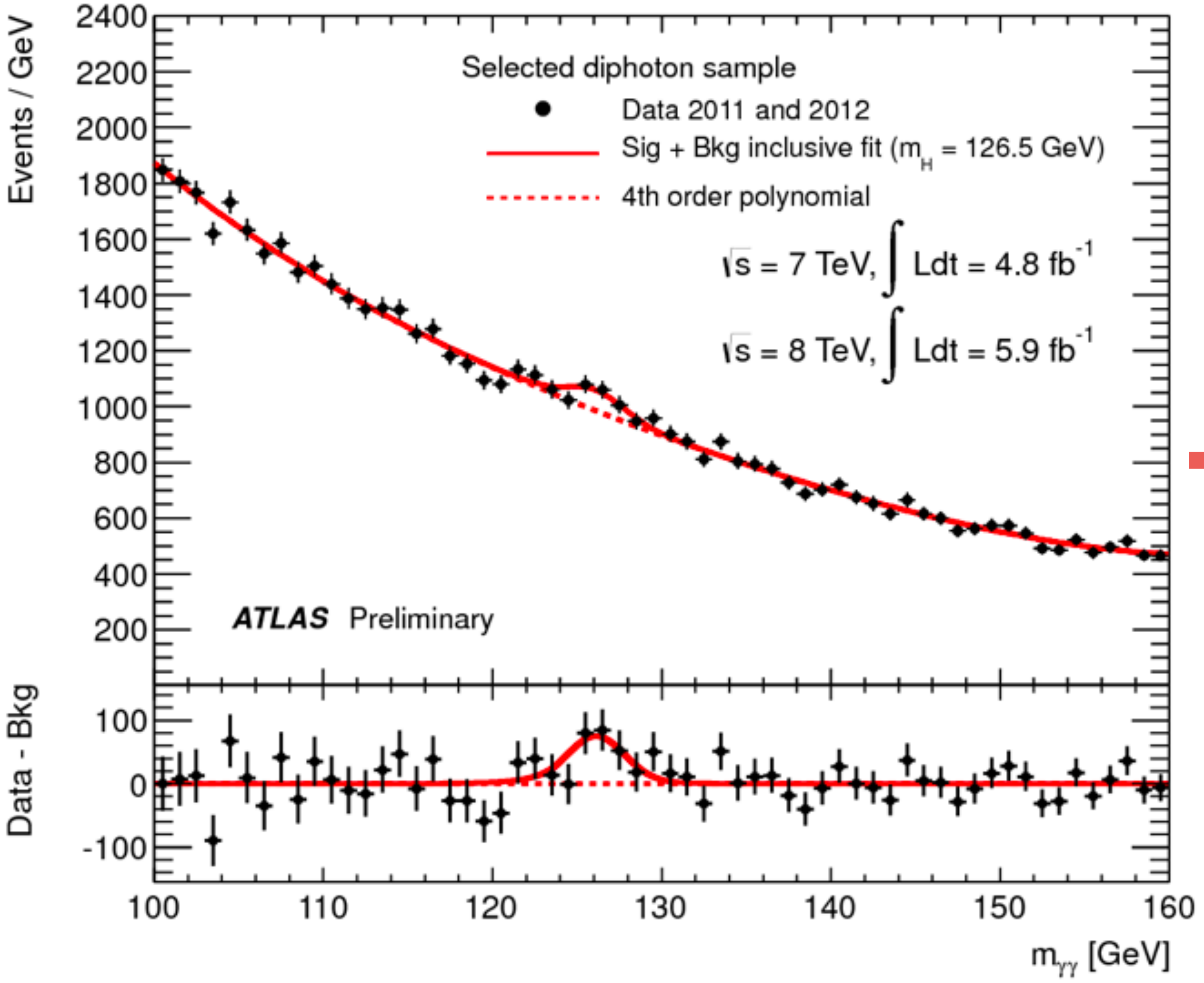
Example:

Below is the plot that led ATLAS to claim the **discovery** of the HIGGS.



Let's figure out how they were able to make such a claim with a **Toy Model** and with the **theory** we have learned so far!

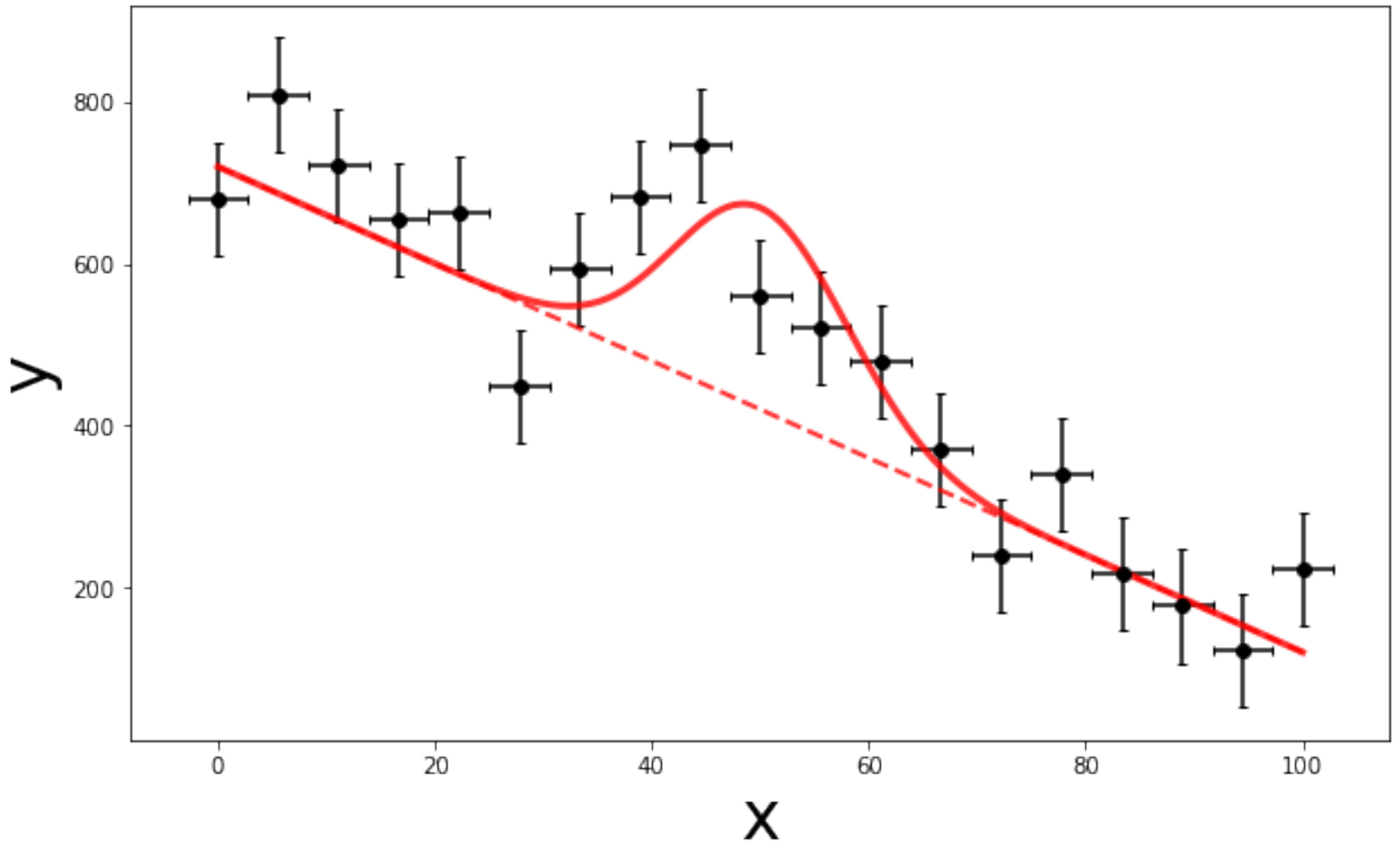
The Likelihood: Higgs example



Toy model:

$$y' = \boxed{mx + q} + \boxed{a \cdot \mathcal{G}(x; \mu = 50, \sigma = 8)}$$

noise: $y \sim \mathcal{N}(\mu = y', \sigma = 70)$



Null hypothesis H_0 - dashed line

a = 0

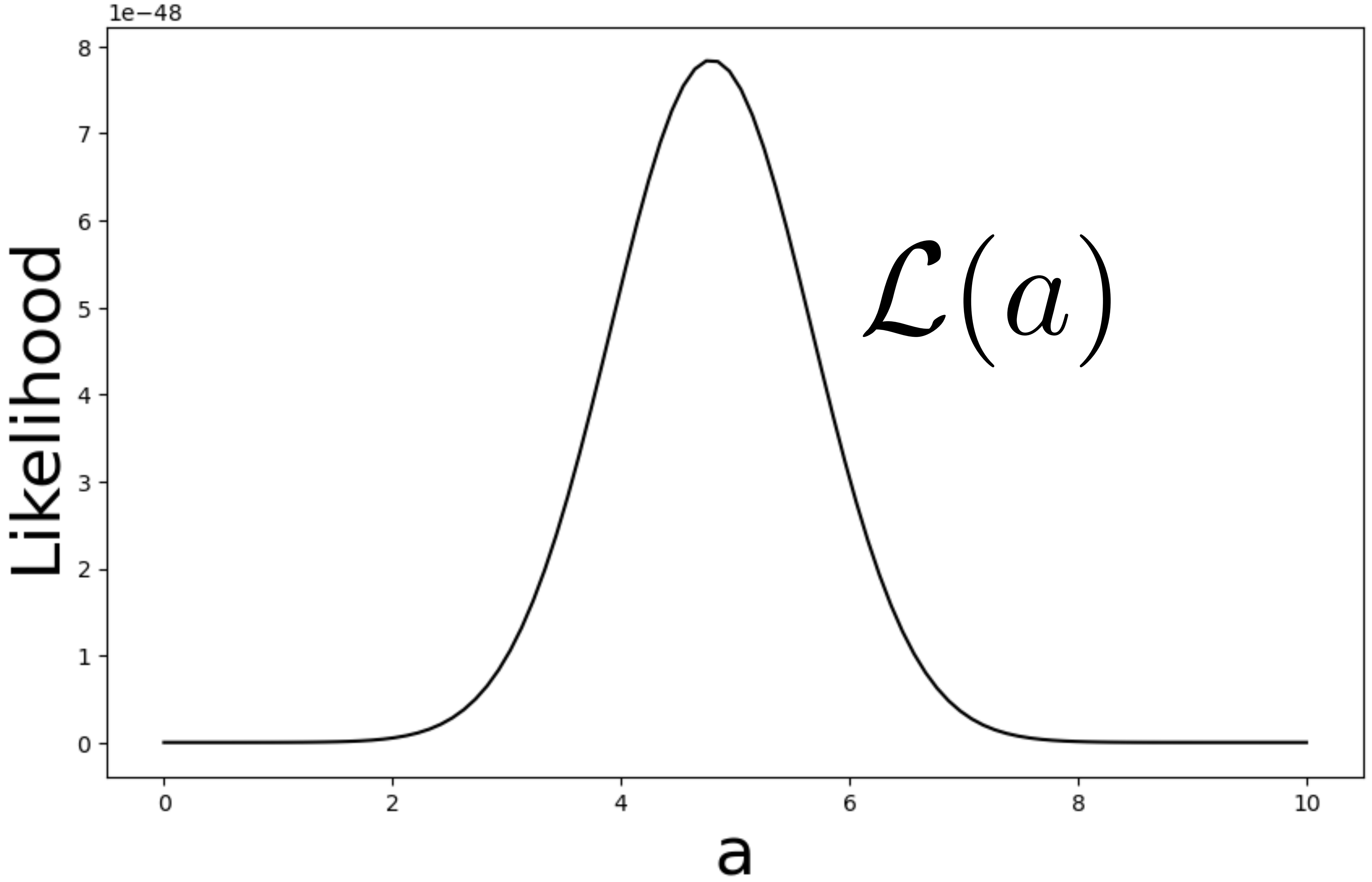
Alternative hypothesis H_1 - solid line

a $\neq$ 0

The Likelihood: Higgs example

Likelihood

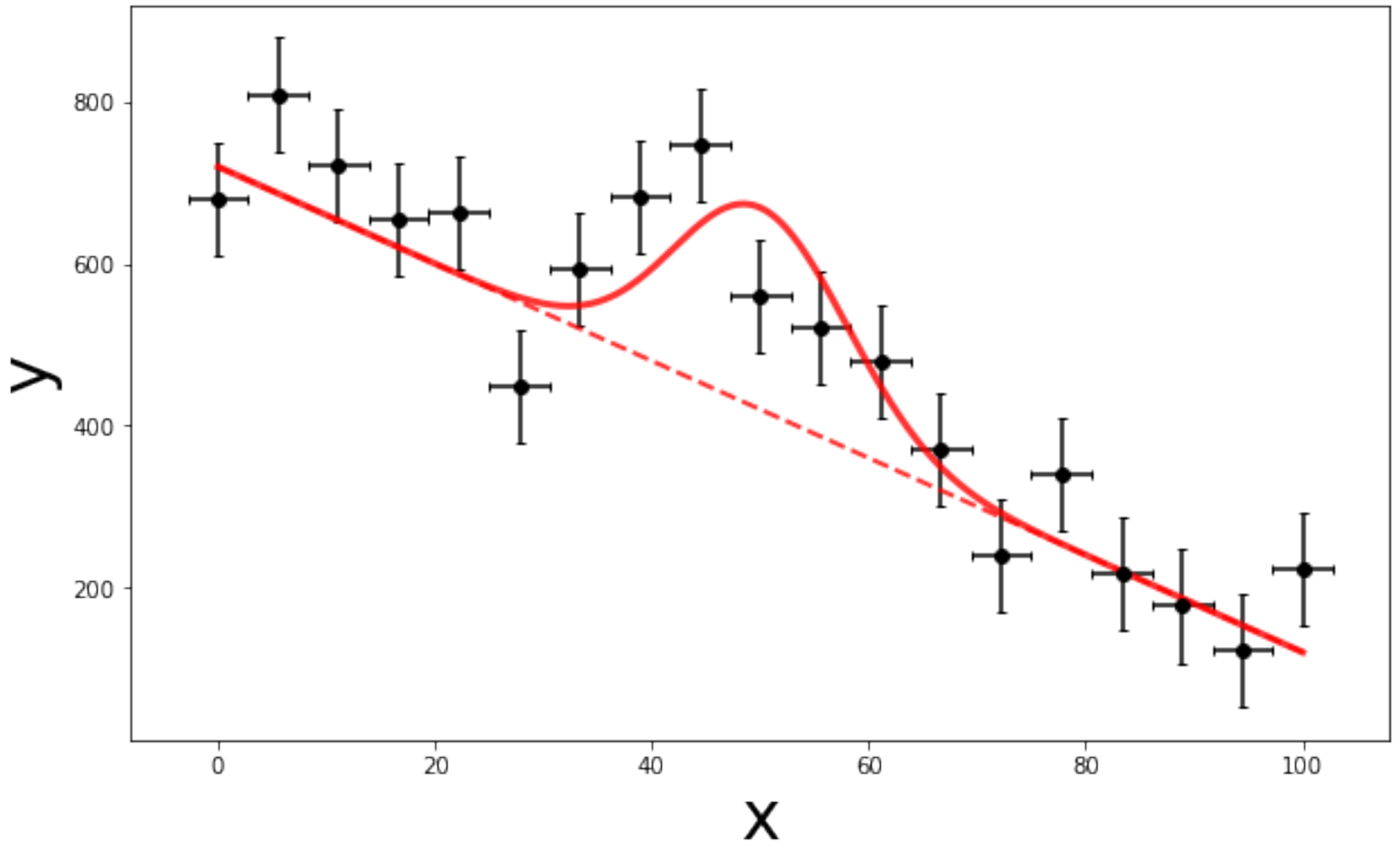
$$\mathscr{L}(a) = \prod p(x_i, y_i \mid a) = \prod \mathscr{N}(y_i \mid y_i'(a), \sigma)$$



Toy model:

$$y' = \boxed{mx + q} + \boxed{a \cdot \mathcal{G}(x; \mu = 50, \sigma = 8)}$$

noise: $y \sim \mathcal{N}(\mu = y', \sigma = 70)$



Null hypothesis H_0 **-** dashed line

a = 0

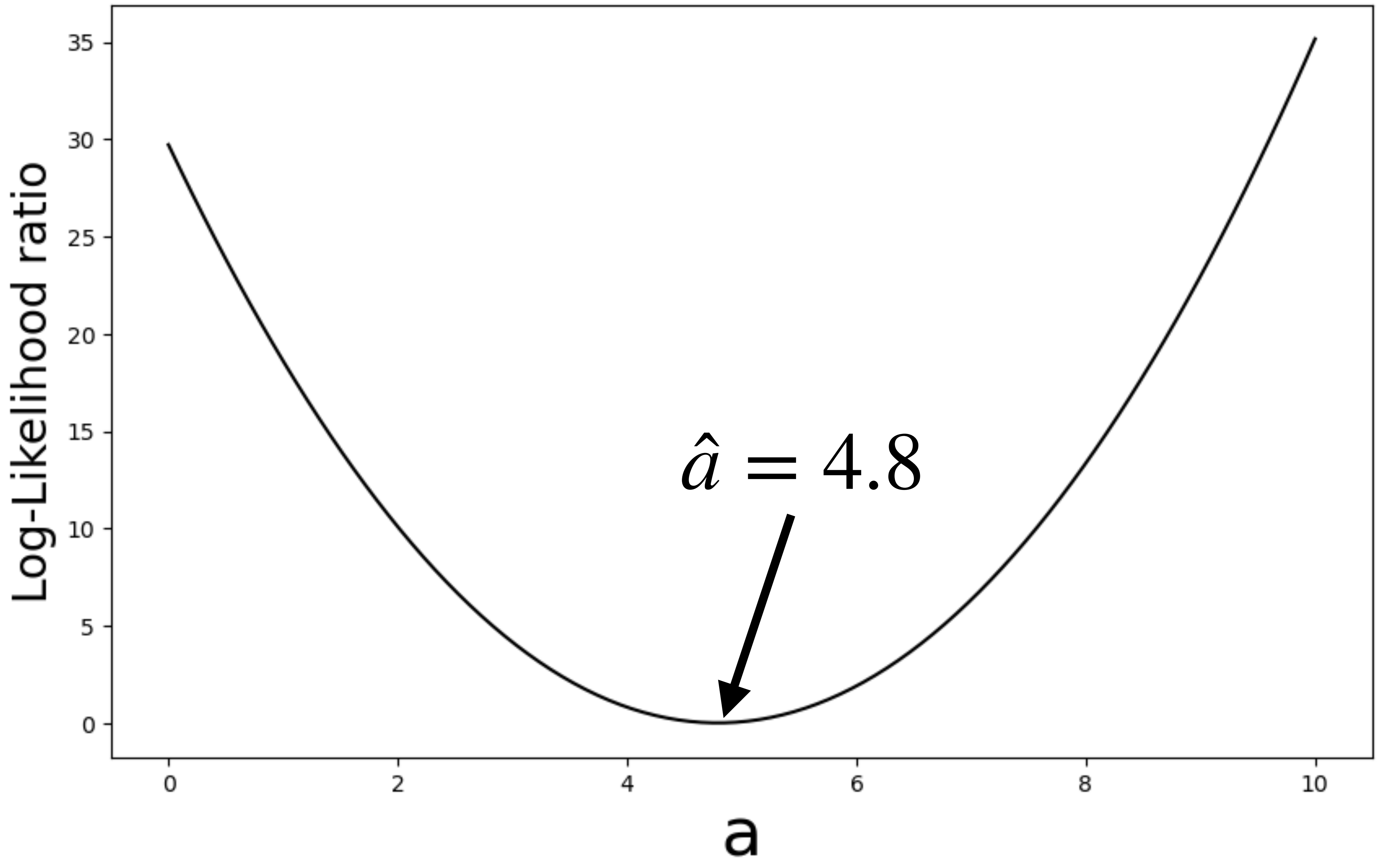
Alternative hypothesis H_1 **-** solid line

a $\neq$ 0

The Likelihood: Higgs example

Log-Likelihood ratio

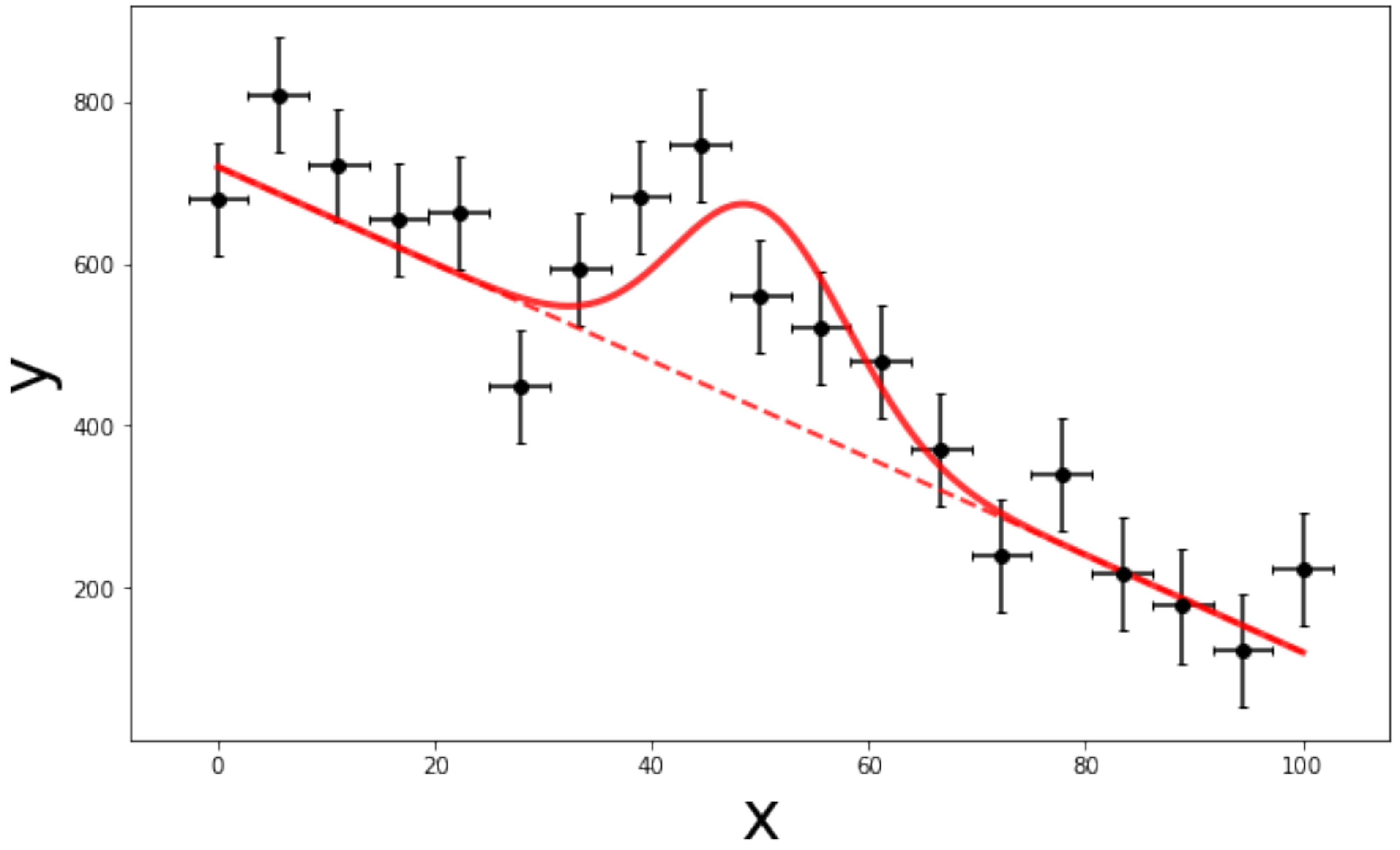
$$-2\log\frac{\mathscr{L}(a)}{\mathscr{L}(\hat{a})} = -2\log\frac{\mathscr{L}(a)}{\mathscr{L}(4.8)}$$



Toy model:

$$y' = \boxed{mx + q} + \boxed{a \cdot \mathcal{G}(x; \mu = 50, \sigma = 8)}$$

noise: $y \sim \mathcal{N}(\mu = y', \sigma = 70)$



Null hypothesis H_0 **-** dashed line

a = 0

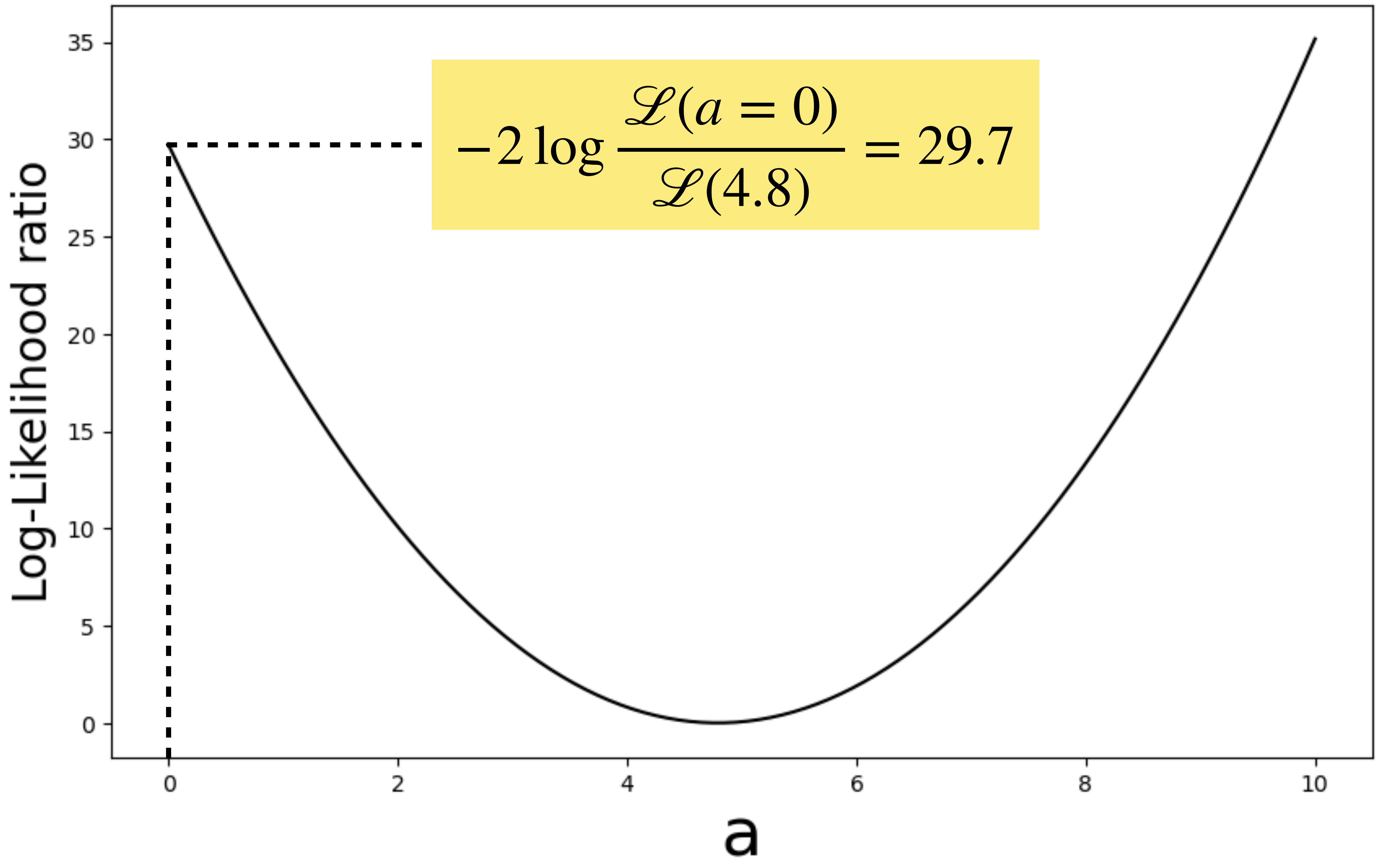
Alternative hypothesis H_1 **-** solid line

a $\neq$ 0

The Likelihood: Higgs example

Log-Likelihood ratio

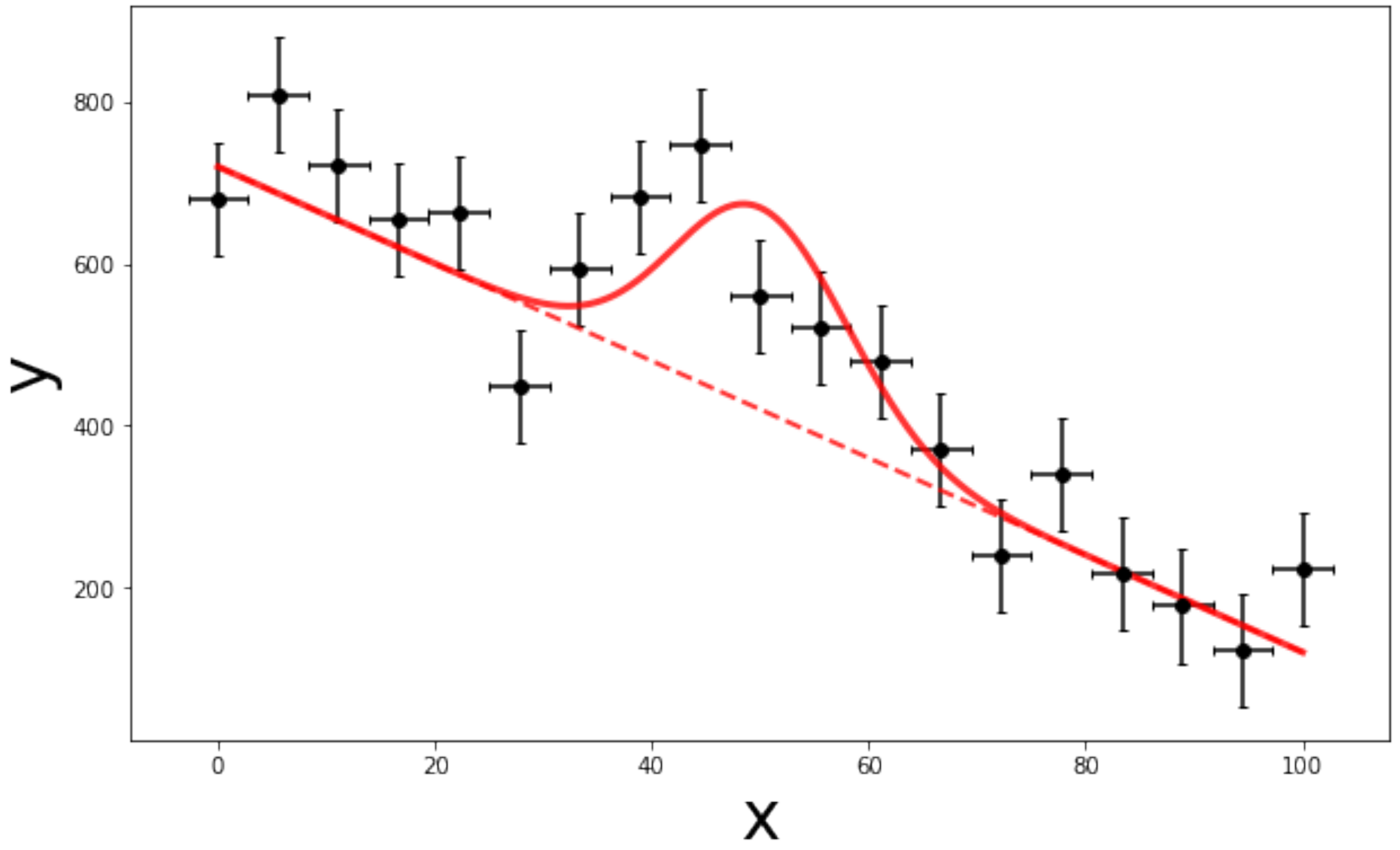
$$-2\log\frac{\mathscr{L}(a)}{\mathscr{L}(\hat{a})} = -2\log\frac{\mathscr{L}(a)}{\mathscr{L}(4.8)}$$



Toy model:

$$y' = \boxed{mx+q} + \boxed{a\cdot\mathcal{G}(x;\mu=50,\sigma=8)}$$

noise: $y \sim \mathcal{N}(\mu = y', \sigma = 70)$



Null hypothesis H_0 - dashed line

a = 0

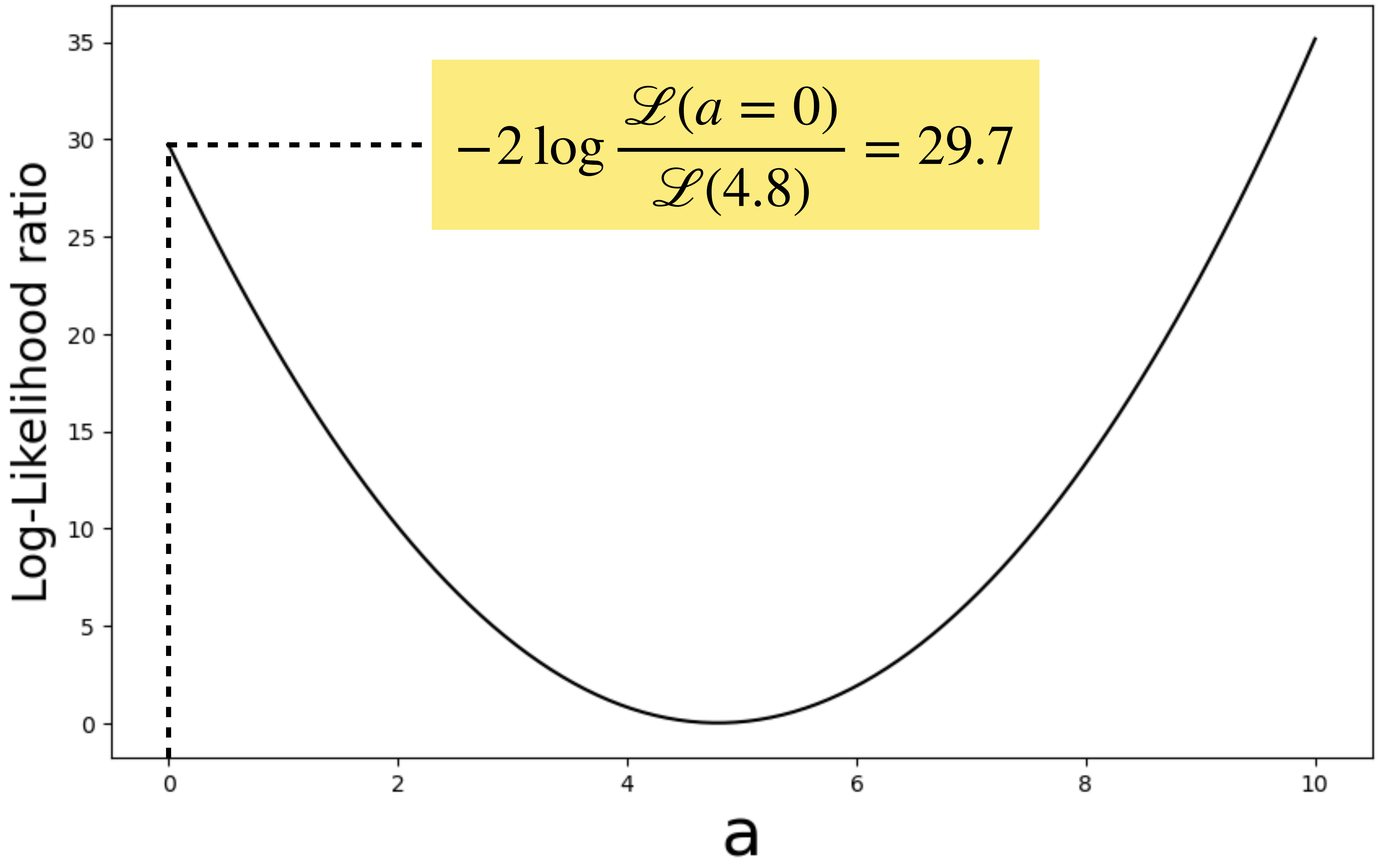
Alternative hypothesis H_1 - solid line

a $\neq$ 0

The Likelihood: Higgs example

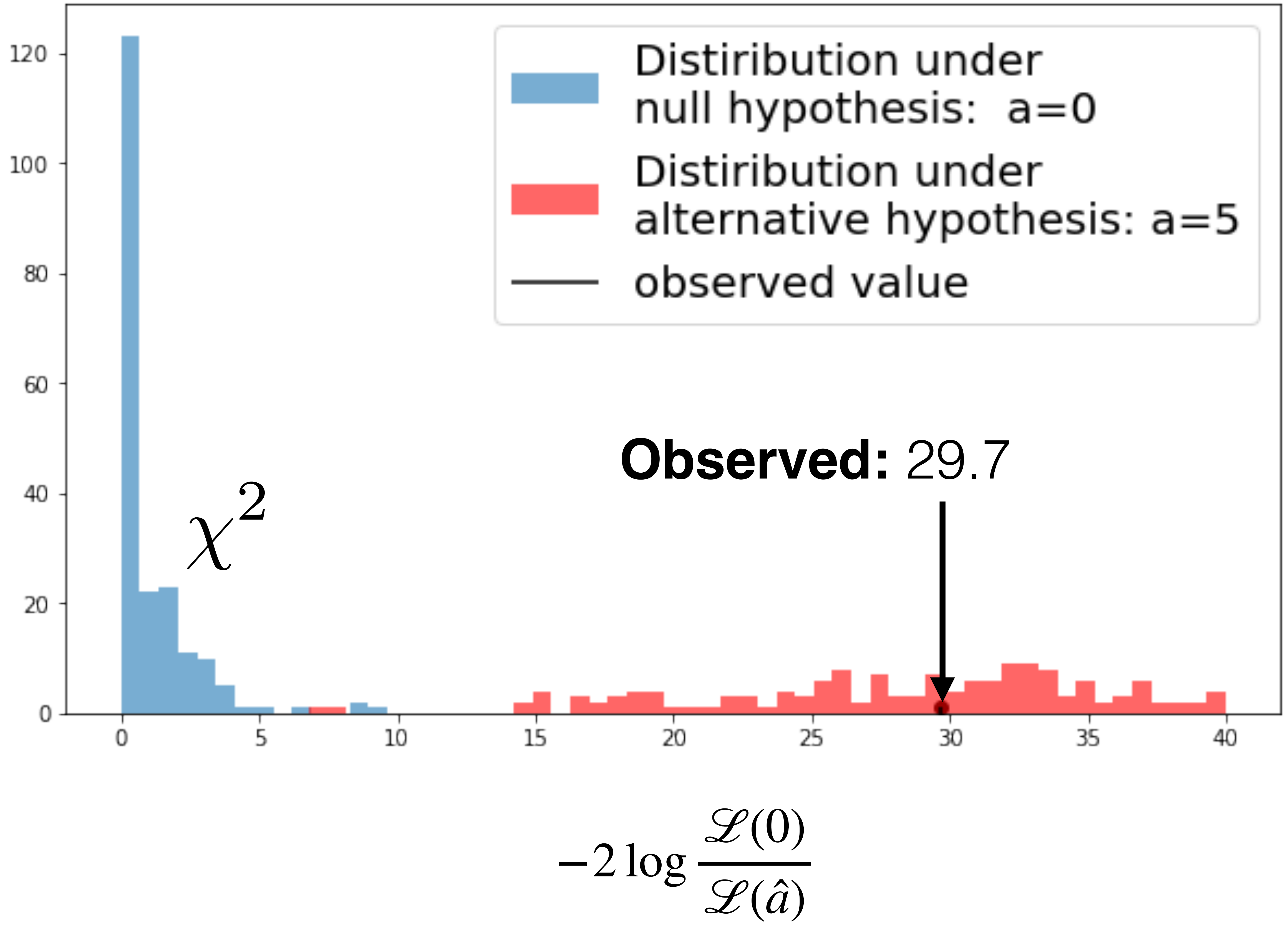
Log-Likelihood ratio

$$-2\log\frac{\mathscr{L}(a)}{\mathscr{L}(\hat{a})} = -2\log\frac{\mathscr{L}(a)}{\mathscr{L}(4.8)}$$



$$-2\log\frac{\mathscr{L}(a=0)}{\mathscr{L}(4.8)} = 29.7$$

From MC simulations of $-2\log\frac{\mathscr{L}(a=0)}{\mathscr{L}(\hat{a})}$

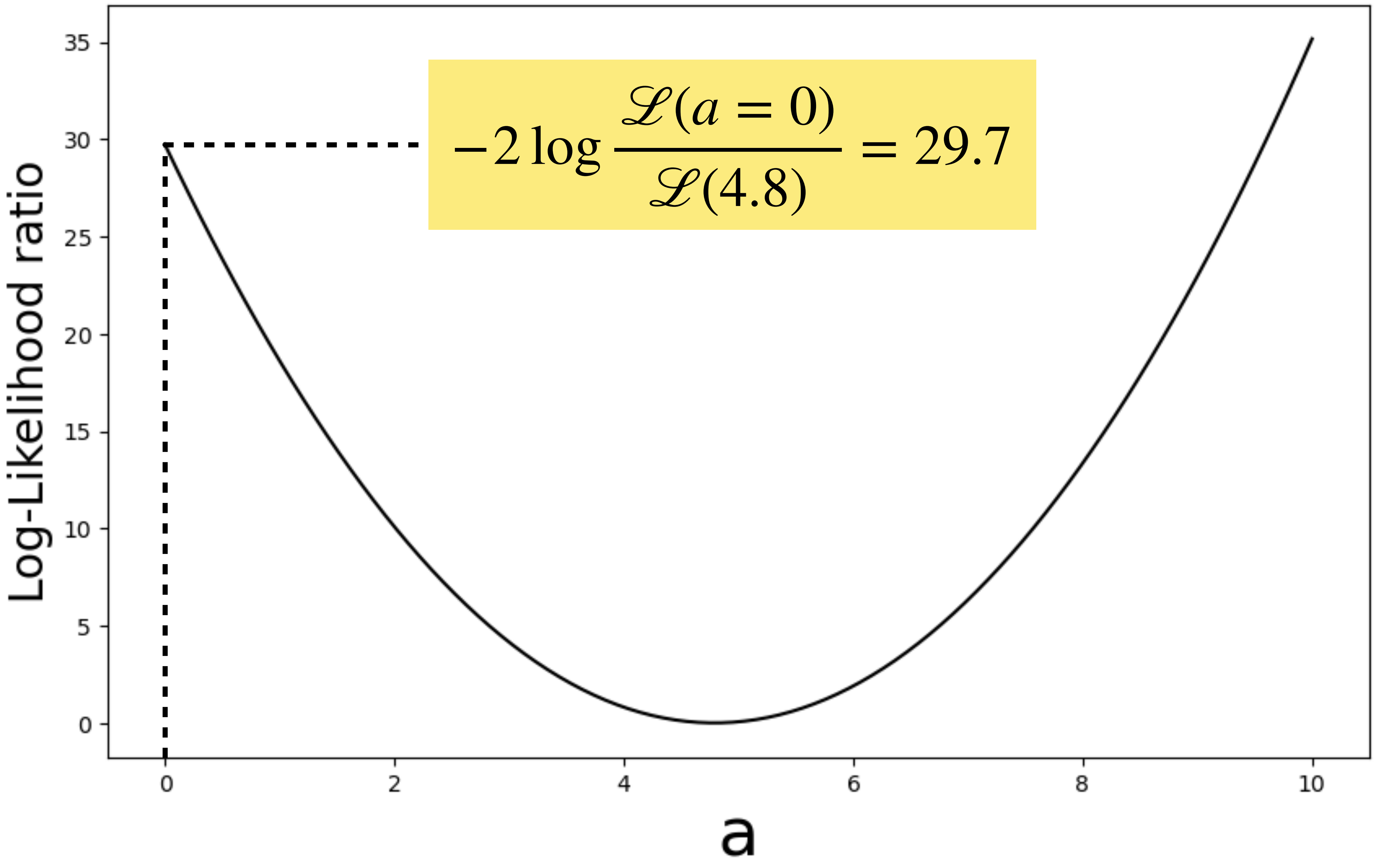


The Likelihood: Higgs example

Log-Likelihood ratio

$$-2\log\frac{\mathscr{L}(a)}{\mathscr{L}(\hat{a})} = -2\log\frac{\mathscr{L}(a)}{\mathscr{L}(4.8)}$$

$$\text{p-value} = \int_{29.7}^{\infty} dx\, \chi^2(x) \simeq 5\cdot 10^{-8}$$



$$-2\log\frac{\mathscr{L}(a=0)}{\mathscr{L}(4.8)} = 29.7$$

Converting the p-value to a “sigma”

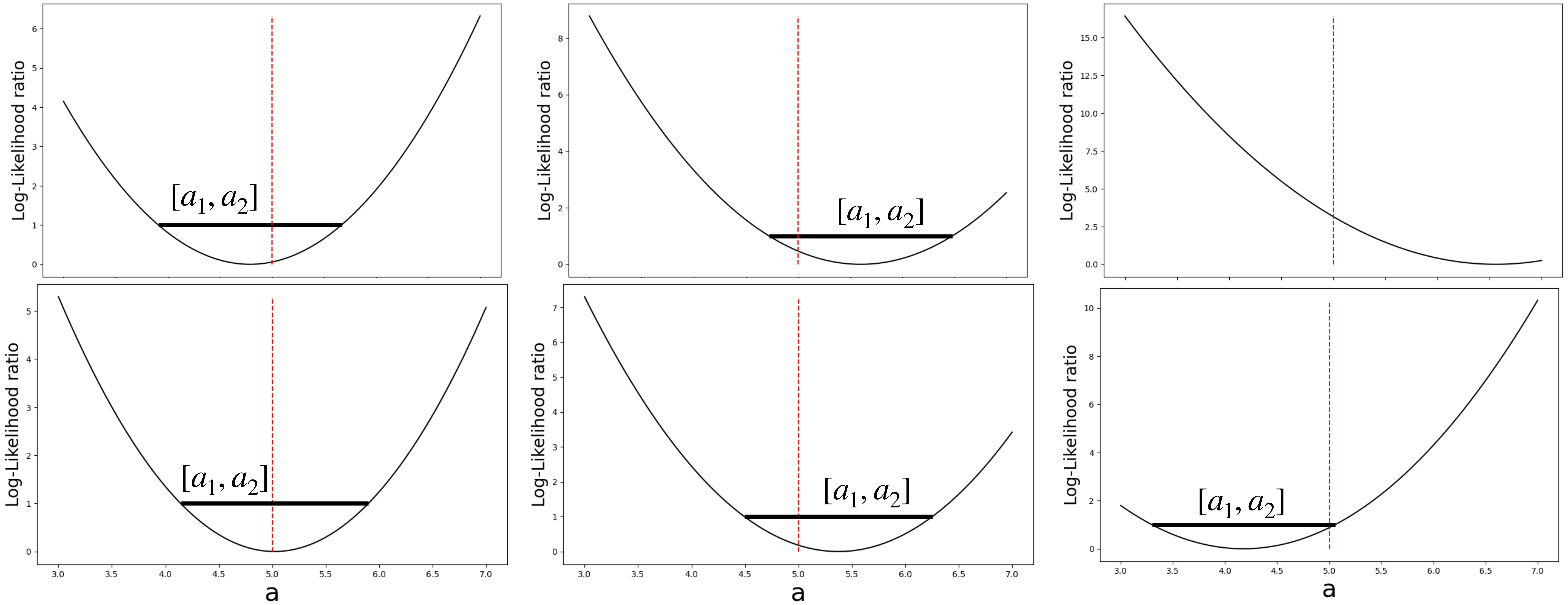
$$\sqrt{2}\cdot \mathrm{erf}^{-1}(1-5\cdot 10^{-8}) \simeq 5.45$$

Notice that $\sqrt{29.7} \simeq 5.45$

Why?

Coverage and confidence intervals

By construction, the **interval** $[a_1, a_2]$ such that $-2\log\mathscr{L}(a_{1,2})/\mathscr{L}(\hat{a}) = 1$ will **include** the true value $a_{true} = 5$ **68%** of the times



Coverage and confidence intervals

By construction, the **interval** $[a_1, a_2]$ such that $-2\log \mathscr{L}(a_{1,2})/\mathscr{L}(\hat{a}) = 1$ will **include** the true value $a_{true} = 5$ **68%** of the times

The **interval** $[a_1, a_2]$ is called a 68% **confidence interval**

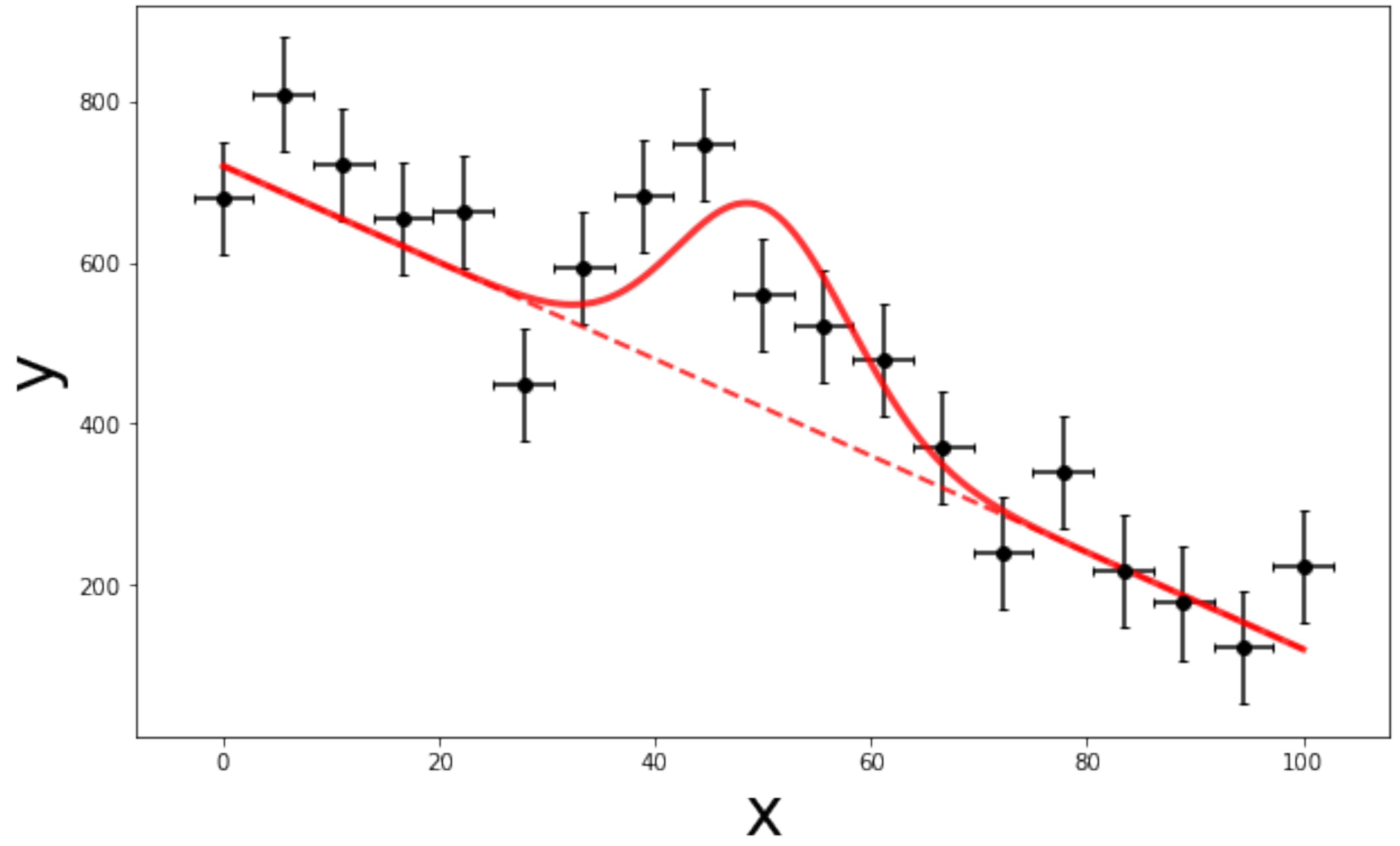
and is said to “**cover**” the true value 68% of the times

Coverage and confidence intervals

By construction, the **interval** $[a_1, a_2]$ such that $-2\log\mathscr{L}(a_{1,2})/\mathscr{L}(\hat{a}) = 1$ will **include** the true value $a_{true} = 5$ **68%** of the times

The **interval** $[a_1, a_2]$ is called a 68% **confidence interval**

and is said to "**cover**" the true value 68% of the times



→

$a = 5.33$

68% CL $= [4.46\,,\,6.22]$

or

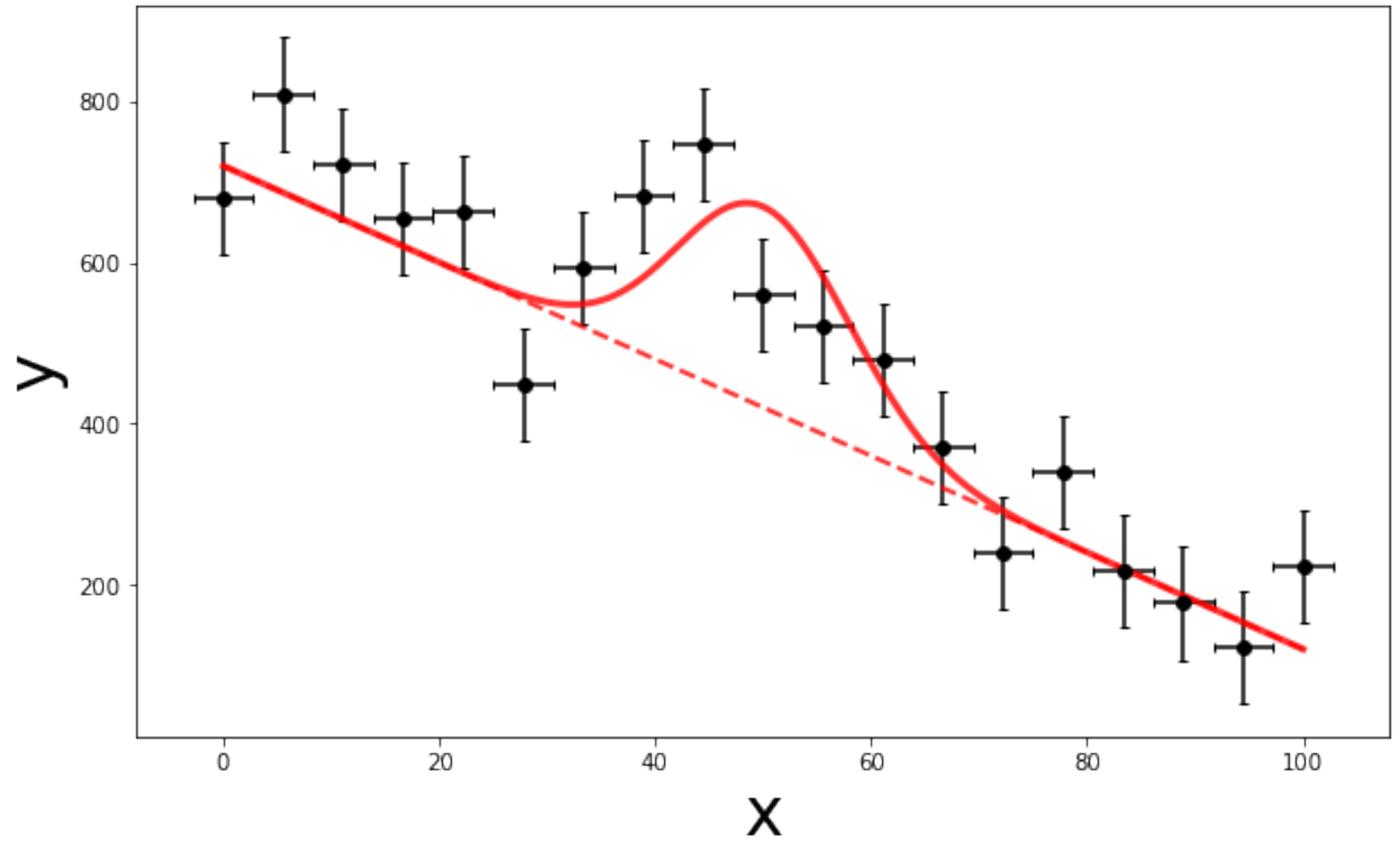
$a = 5.3^{+0.88}_{-0.87}$

Coverage and confidence intervals

By construction, the **interval** $[a_1, a_2]$ such that $-2\log \mathscr{L}(a_{1,2})/\mathscr{L}(\hat{a}) = 1$ will **include** the true value $a_{true} = 5$ **68%** of the times

The **interval** $[a_1, a_2]$ is called a 68% **confidence interval**

and is said to “**cover**” the true value 68% of the times



→

$a = 5.33$

68% CL $= [4.46\,, 6.22]$

or

$a = 5.3^{+0.88}_{-0.87}$

⚠

The range $[4.46\,, 6.22]$ DOES NOT cover the true value 68% of the times

It is the procedure that gives $[a_1, a_2]$ to be “right” 68% of the times

The likelihood profile

The **likelihood profile** (or **profile likelihood**) is a way to study one parameter of **interest** while accounting for the best possible values of the **nuisance parameters**

$$\mathscr{L}(\underset{\substack{\nearrow \\ \text{parameter of interest}}}{\theta}, \underset{\substack{\nwarrow \\ \text{nuisance parameters}}}{\nu})$$

$$\boxed{\mathscr{L}_p(\theta) = \max_{\nu} \mathscr{L}(\theta, \nu) \equiv \mathscr{L}(\theta, \hat{\nu}(\theta))}$$

Later, we will use it for ON/OFF measurements…

Conclusions Part I

- In this part, we have introduced some **basic** (and **advanced**) concepts of statistics
    - Bayesian approach
    - Frequentist approach
    - Likelihood analysis
- for more details:
    - My review paper on statistics for IACTs: arXiv:2202.04590
    - Bayesian for astrophysics: astro.cornell.edu/~loredo/bayes/promise.pdf
    - …
- Things we did not cover:
    - Bayes Factors, Model Selection, Trial Factors, Threshold Calibration, Time Series Analysis, and more …

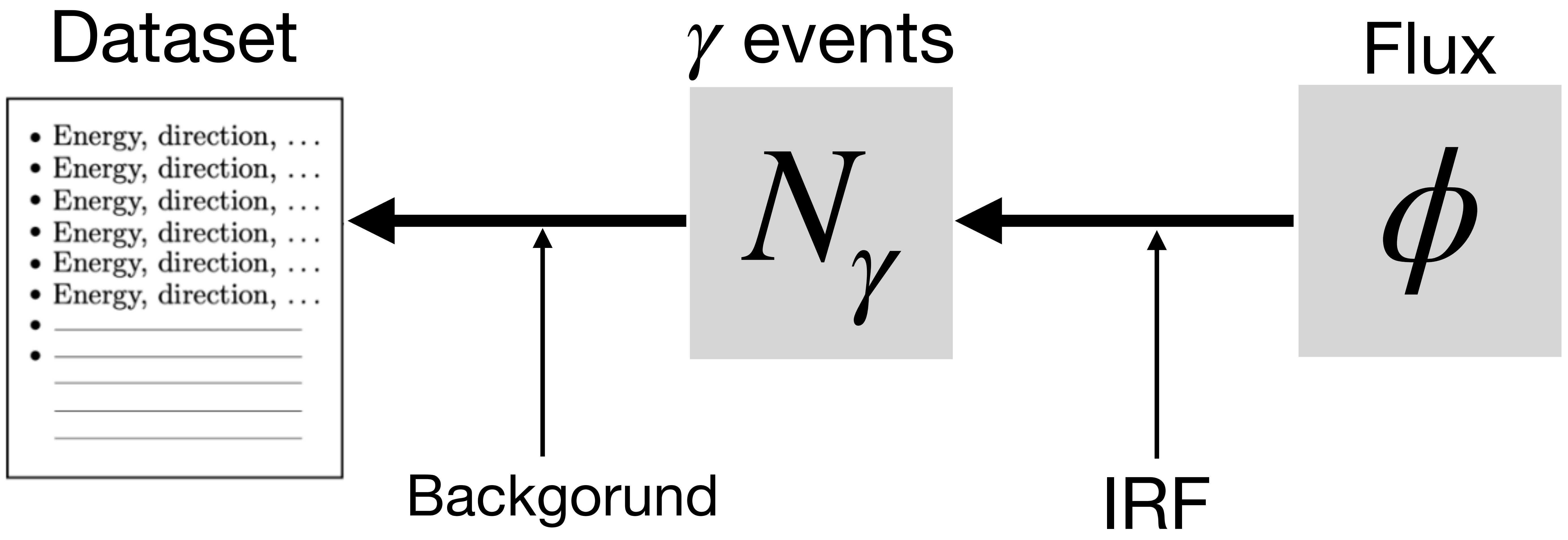
PART II

Outline - Part II

- Estimating the excess
    - On/Off measurement
    - Confidence levels and upper limits
    - Li&Ma significance

- From the excess to the γ flux
    - A first approximate estimation
    - Unfolding
    - Forward folding
    - The light curve

A typical analysis in γ-ray astronomy



Dataset

- Energy, direction, ...
- Energy, direction, ...
- Energy, direction, ...
- Energy, direction, ...
- Energy, direction, ...
- Energy, direction, ...

γ events

N_γ

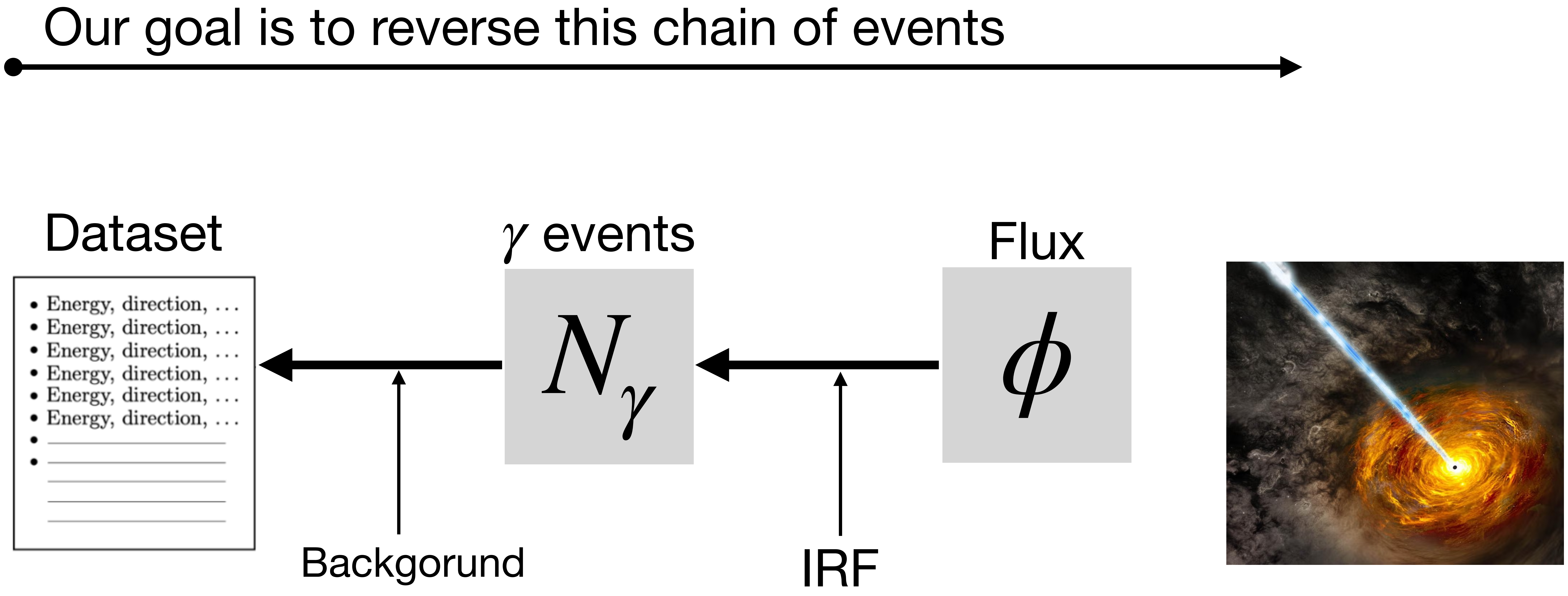
Flux

ϕ

Backgorund

IRF

A typical analysis in γ-ray astronomy



Our goal is to reverse this chain of events

Dataset

- Energy, direction, ...
- Energy, direction, ...
- Energy, direction, ...
- Energy, direction, ...
- Energy, direction, ...
- Energy, direction, ...

γ events

N_γ

Flux

ϕ

Backgorund

IRF

First a bit of terminology

① γ rate: $\dfrac{dN_\gamma}{dt}$ *unit:* $\dfrac{1}{\mathrm{s}}$

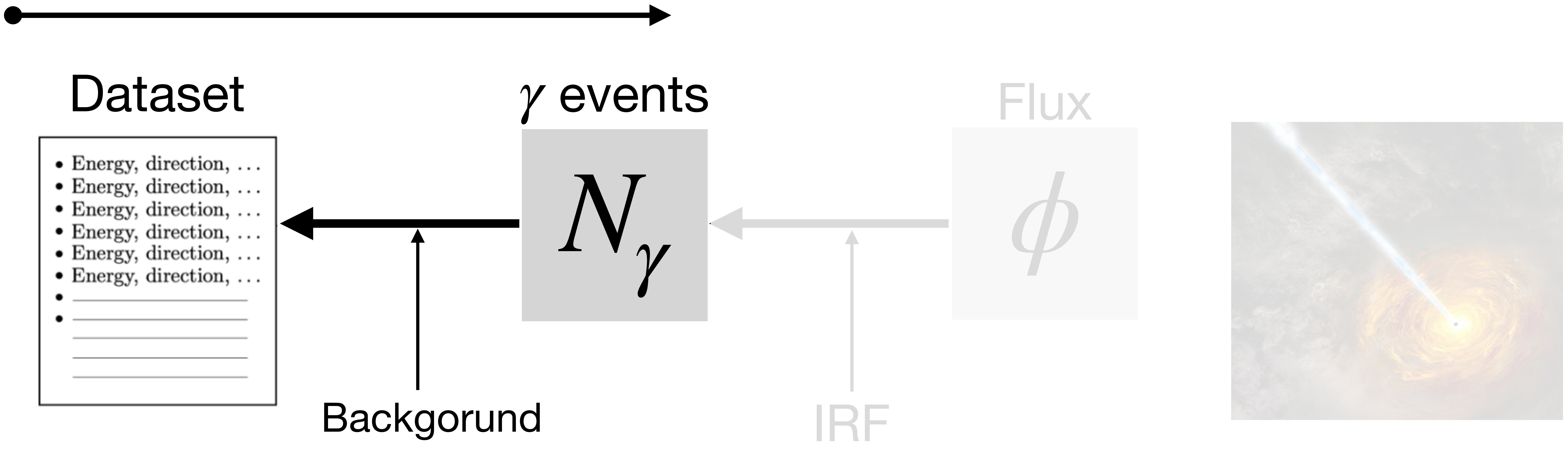
② Flux: $\phi = \dfrac{d^2N_\gamma}{dt\,dA}$ *unit:* $\dfrac{1}{\mathrm{s\ cm^2}}$

③ Differential energy spectrum: $\dfrac{d\phi}{dE} = \dfrac{d^3N_\gamma}{dt\,dA\,dE}$ *unit:* $\dfrac{1}{\mathrm{s\ cm^2\ TeV}}$

④ Light Curve: $\displaystyle\int_{E_{th}} dE\,\frac{d\phi(t_i)}{dE}$ with $t_0, \ldots, t_i, t_{i+1}, \ldots$ *unit:* $\dfrac{1}{\mathrm{s\ cm^2}}$

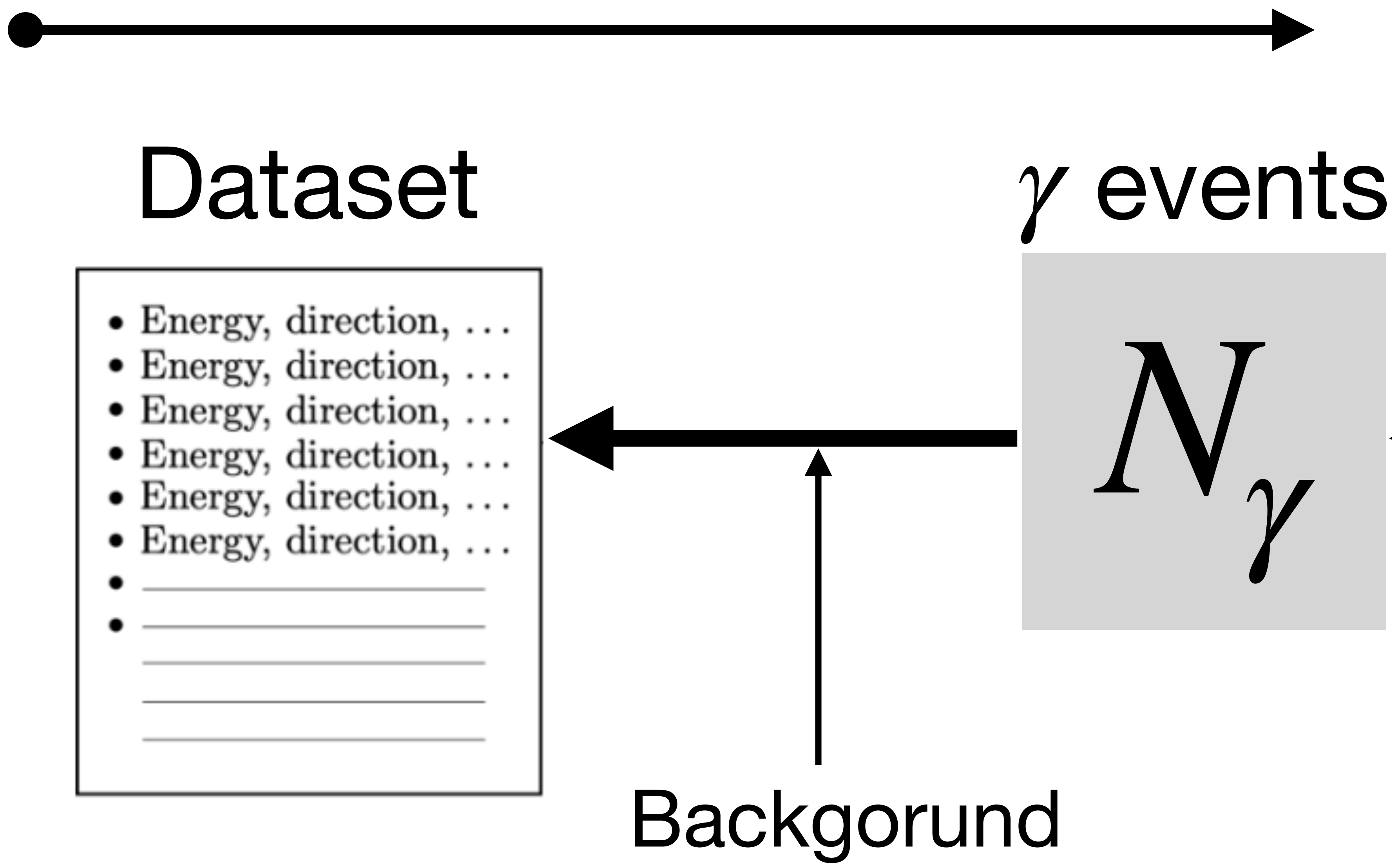
Estimating the excess

In this part, we will estimate the excess N_γ



Estimating the excess

In this part, we will estimate the excess N_γ



Dataset

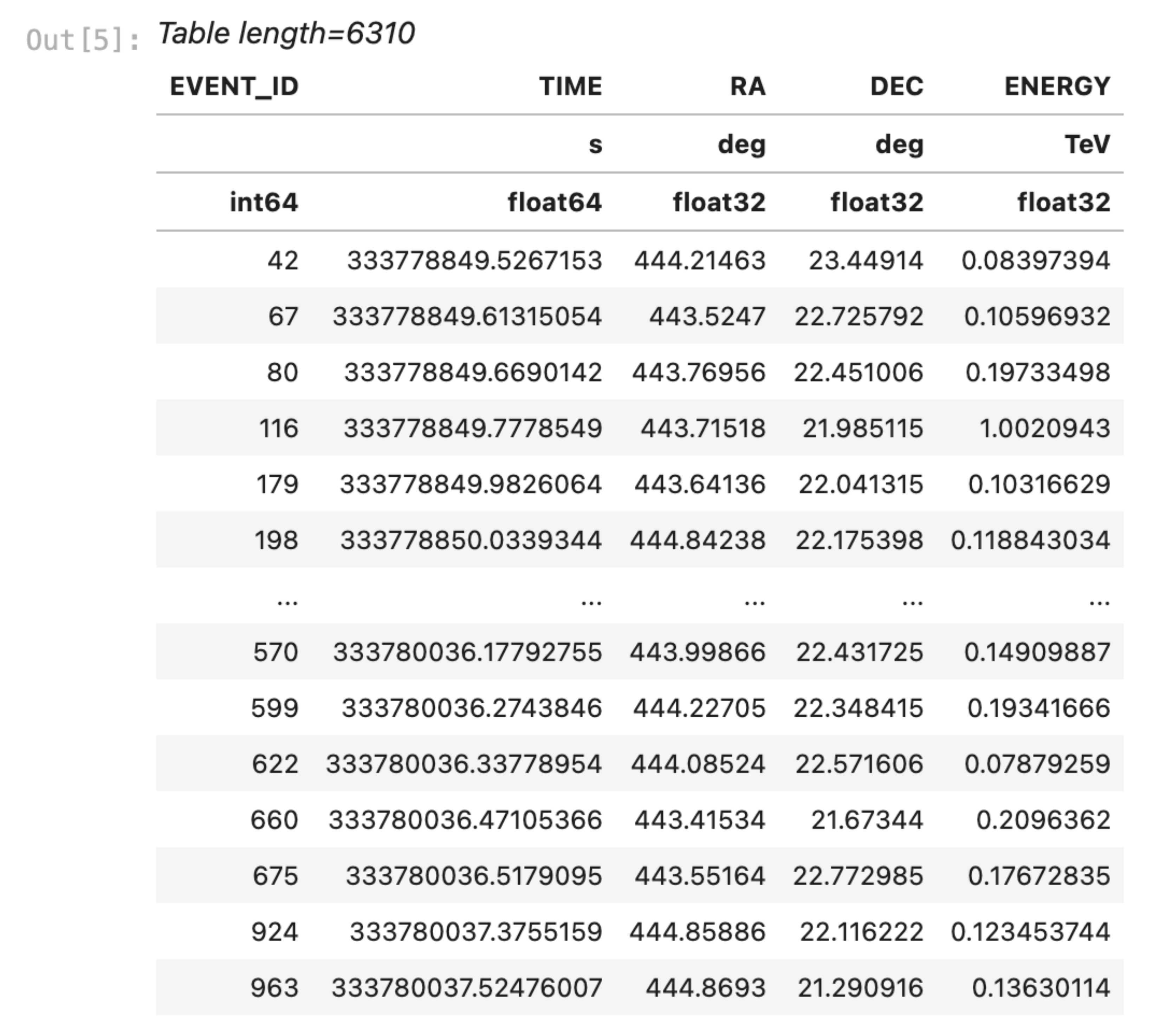
- Energy, direction, . . .
- Energy, direction, . . .
- Energy, direction, . . .
- Energy, direction, . . .
- Energy, direction, . . .
- Energy, direction, . . .

γ events

N_γ

Backgorund

Given your event list what's the expected number of γ ray?



Out[5]: *Table length=6310*

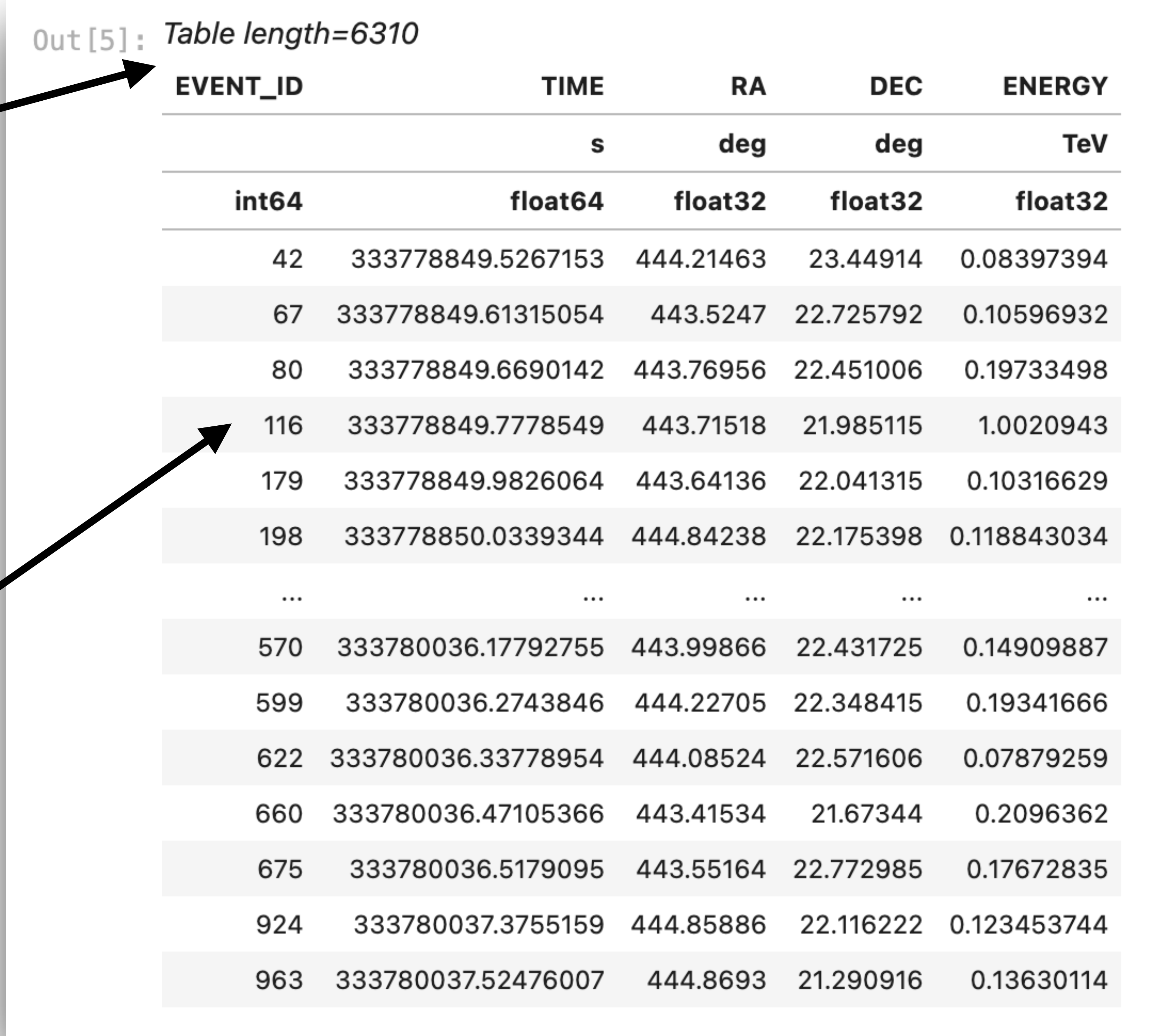
EVENT_ID	TIME	RA	DEC	ENERGY
	s	deg	deg	TeV
int64	float64	float32	float32	float32
42	333778849.5267153	444.21463	23.44914	0.08397394
67	333778849.61315054	443.5247	22.725792	0.10596932
80	333778849.6690142	443.76956	22.451006	0.19733498
116	333778849.7778549	443.71518	21.985115	1.0020943
179	333778849.9826064	443.64136	22.041315	0.10316629
198	333778850.0339344	444.84238	22.175398	0.118843034
...	...	...	...	...
570	333780036.17792755	443.99866	22.431725	0.14909887
599	333780036.2743846	444.22705	22.348415	0.19341666
622	333780036.33778954	444.08524	22.571606	0.07879259
660	333780036.47105366	443.41534	21.67344	0.2096362
675	333780036.5179095	443.55164	22.772985	0.17672835
924	333780037.3755159	444.85886	22.116222	0.123453744
963	333780037.52476007	444.8693	21.290916	0.13630114

Estimating the excess

We have 6310 events (in a given temporal, energetic, and spatial window). Does that mean that the number of gamma-ray events is 6310?

Consider this event at 1 TeV. Is it a **signal** event (a gamma-ray) or a **background** event (a muon, proton, etc…)?

Given your event list what's the expected number of γ ray?

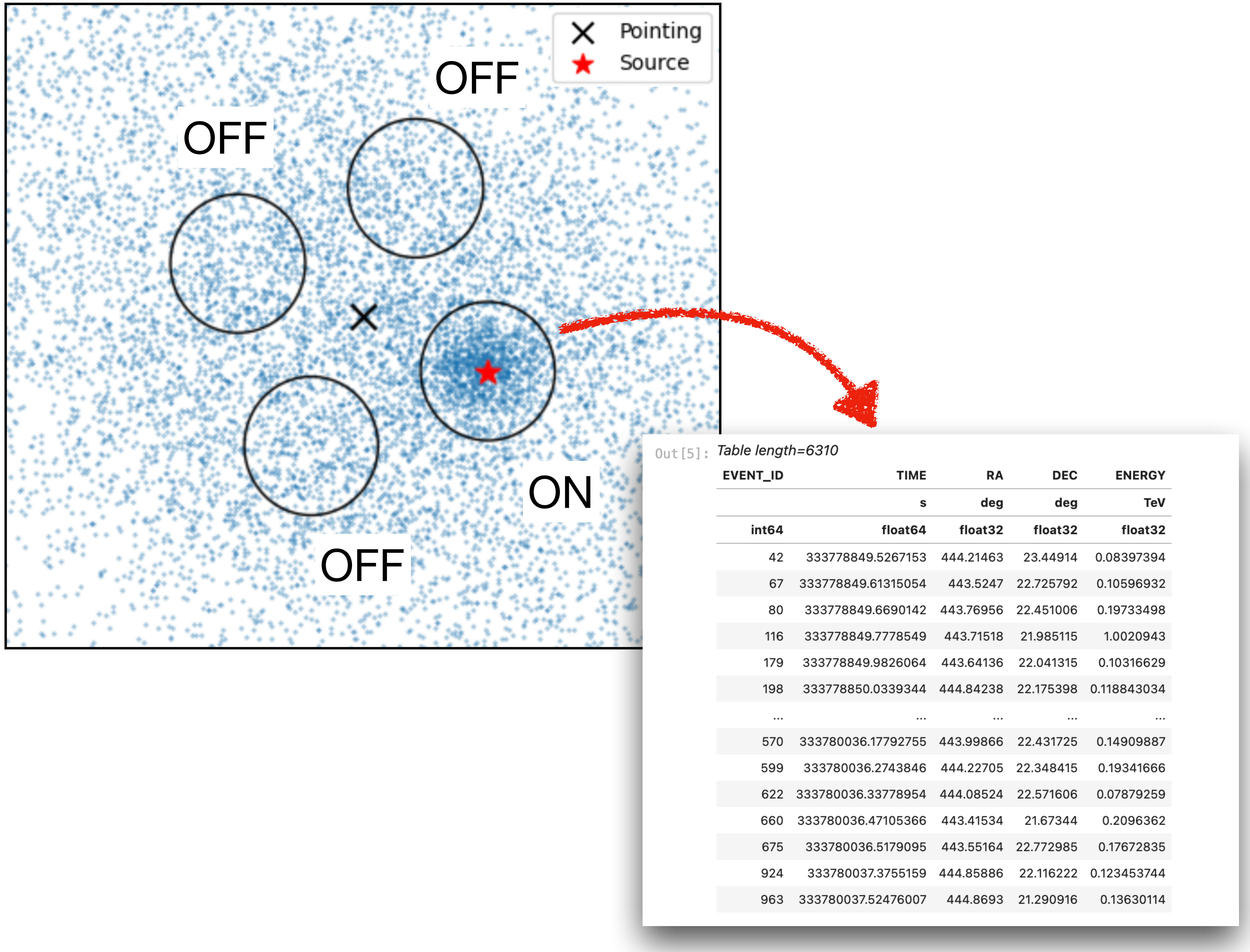


Out[5]: *Table length=6310*

EVENT_ID	TIME	RA	DEC	ENERGY
	s	deg	deg	TeV
int64	float64	float32	float32	float32
42	333778849.5267153	444.21463	23.44914	0.08397394
67	333778849.61315054	443.5247	22.725792	0.10596932
80	333778849.6690142	443.76956	22.451006	0.19733498
116	333778849.7778549	443.71518	21.985115	1.0020943
179	333778849.9826064	443.64136	22.041315	0.10316629
198	333778850.0339344	444.84238	22.175398	0.118843034
...	...	...	...	...
570	333780036.17792755	443.99866	22.431725	0.14909887
599	333780036.2743846	444.22705	22.348415	0.19341666
622	333780036.33778954	444.08524	22.571606	0.07879259
660	333780036.47105366	443.41534	21.67344	0.2096362
675	333780036.5179095	443.55164	22.772985	0.17672835
924	333780037.3755159	444.85886	22.116222	0.123453744
963	333780037.52476007	444.8693	21.290916	0.13630114

On/Off measurement

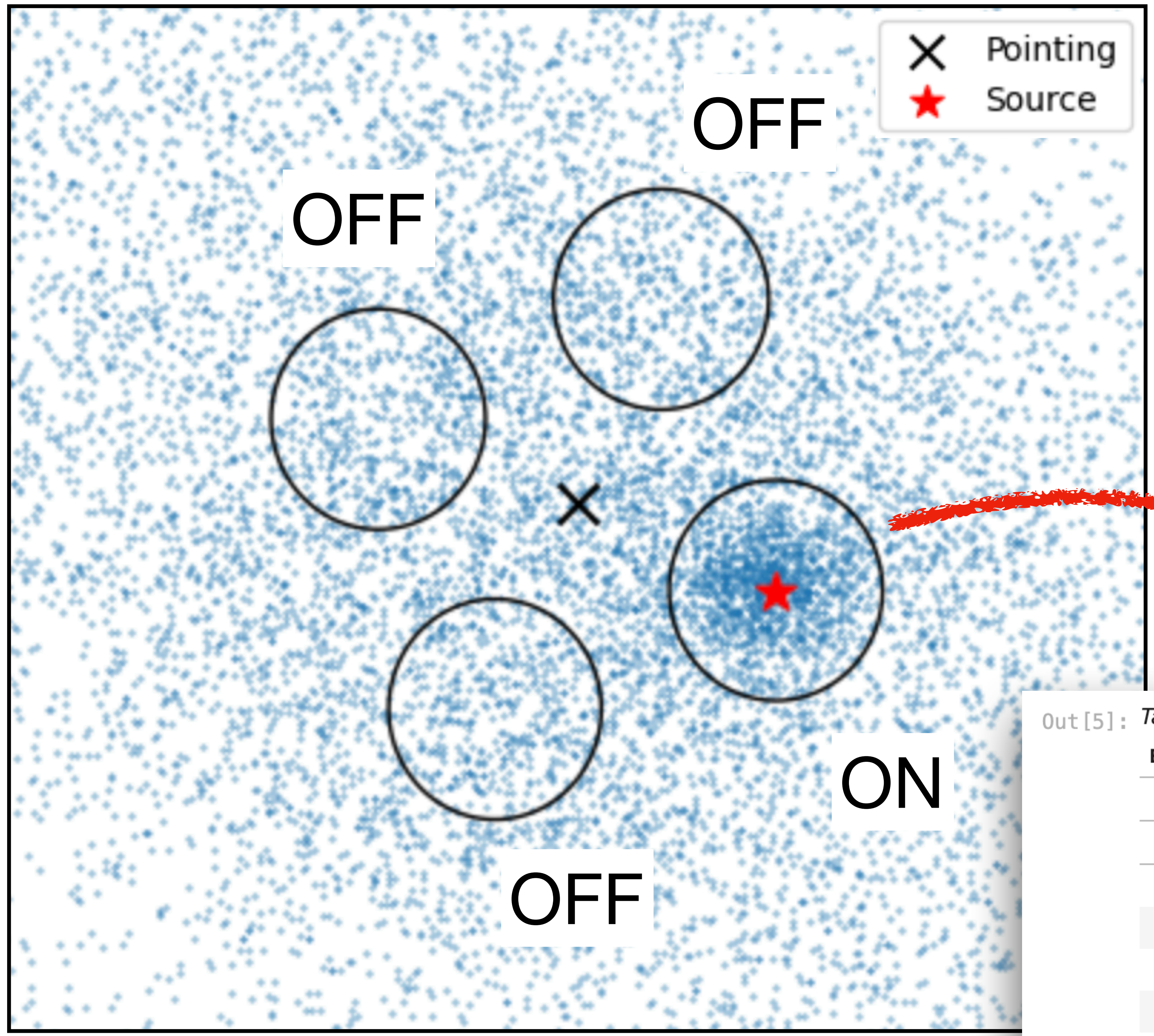
Observed events in the sky

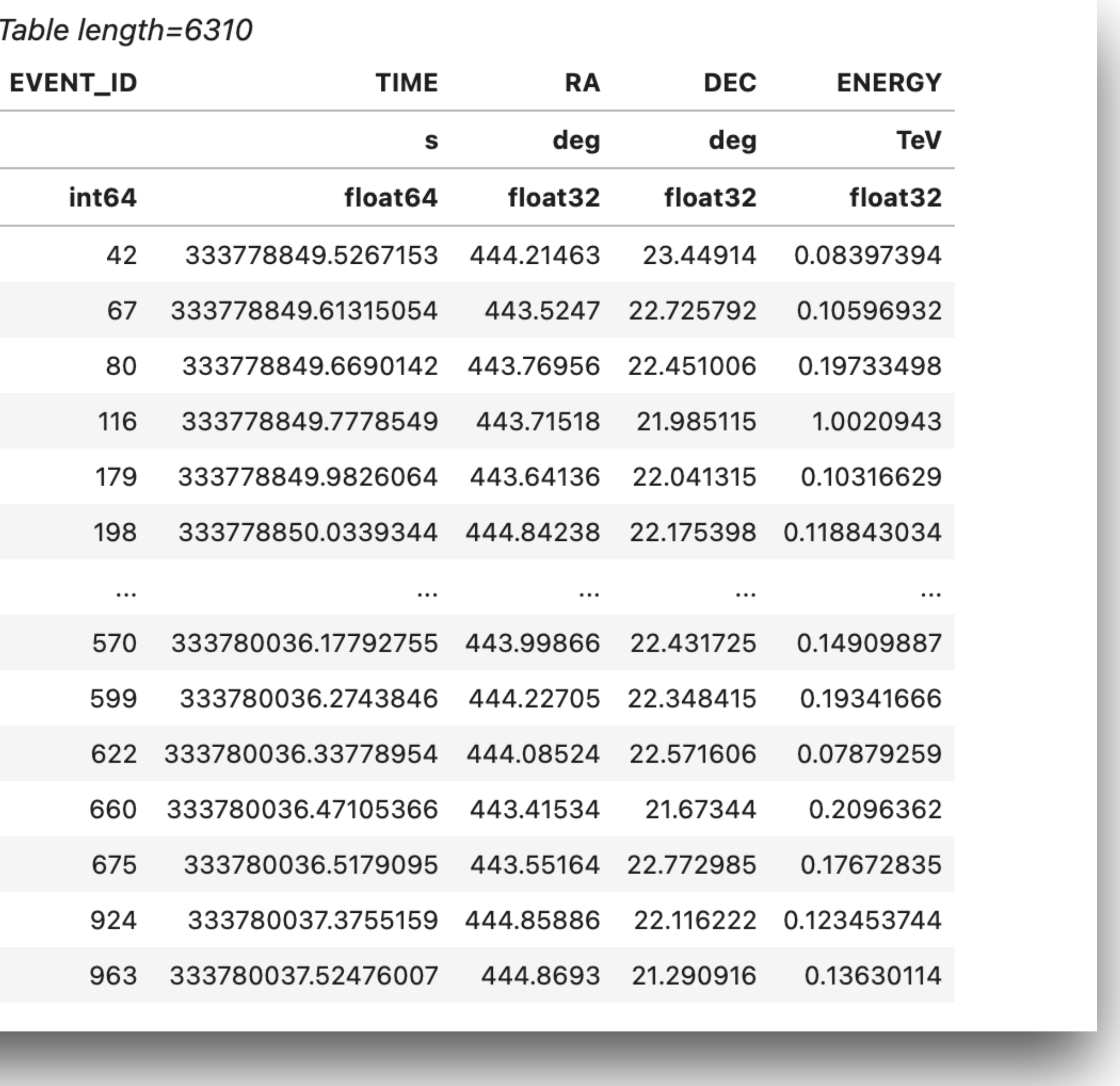


```
Out[5]: Table length=6310
EVENT_ID        TIME          RA        DEC       ENERGY
                   s         deg        deg          TeV
   int64     float64     float32    float32      float32
      42  333778849.5267153  444.21463  23.44914  0.08397394
      67  333778849.61315054  443.5247  22.725792  0.10596932
      80  333778849.6690142  443.76956  22.451006  0.19733498
     116  333778849.7778549  443.71518  21.985115  1.0020943
     179  333778849.9826064  443.64136  22.041315  0.10316629
     198  333778850.0339344  444.84238  22.175398  0.118843034
     ...                ...        ...        ...         ...
     570  333780036.17792755  443.99866  22.431725  0.14909887
     599  333780036.2743846  444.22705  22.348415  0.19341666
     622  333780036.33778954  444.08524  22.571606  0.07879259
     660  333780036.47105366  443.41534  21.67344  0.2096362
     675  333780036.5179095  443.55164  22.772985  0.17672835
     924  333780037.3755159  444.85886  22.116222  0.123453744
     963  333780037.52476007  444.8693  21.290916  0.13630114
```

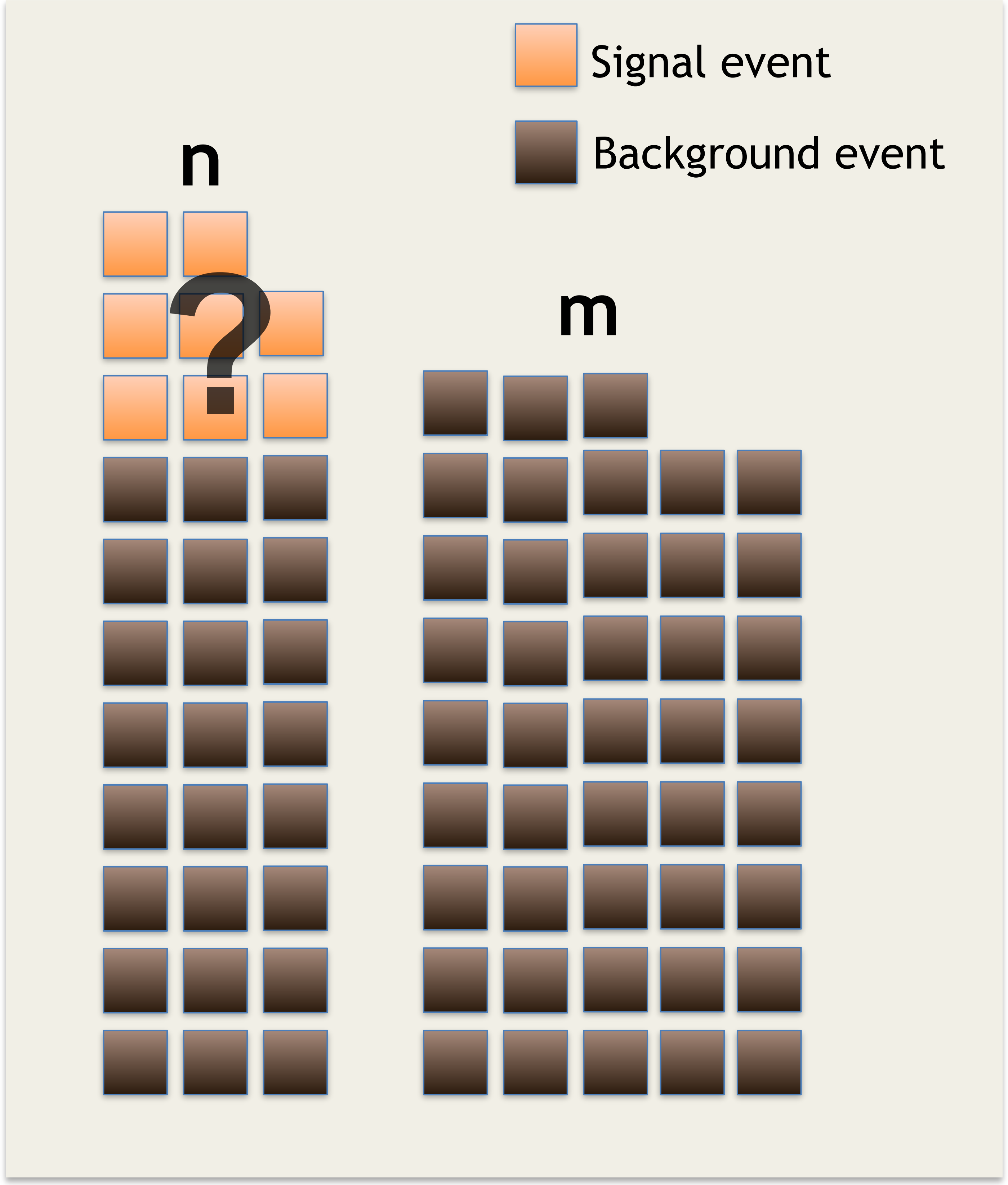
On/Off measurement

Observed events in the sky

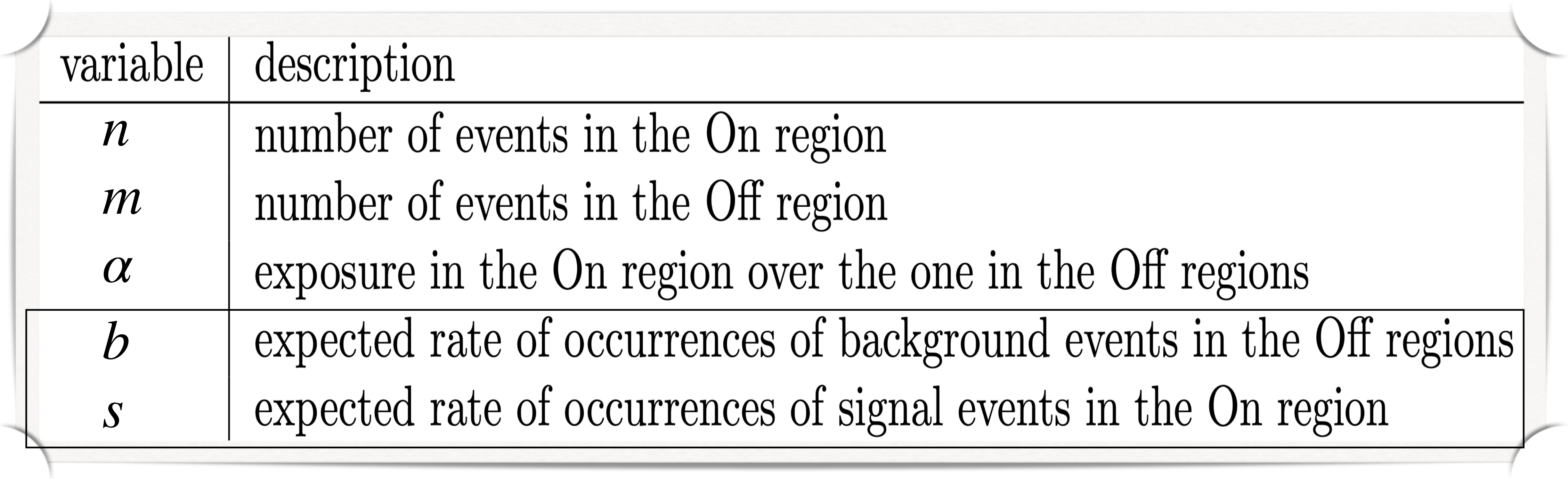




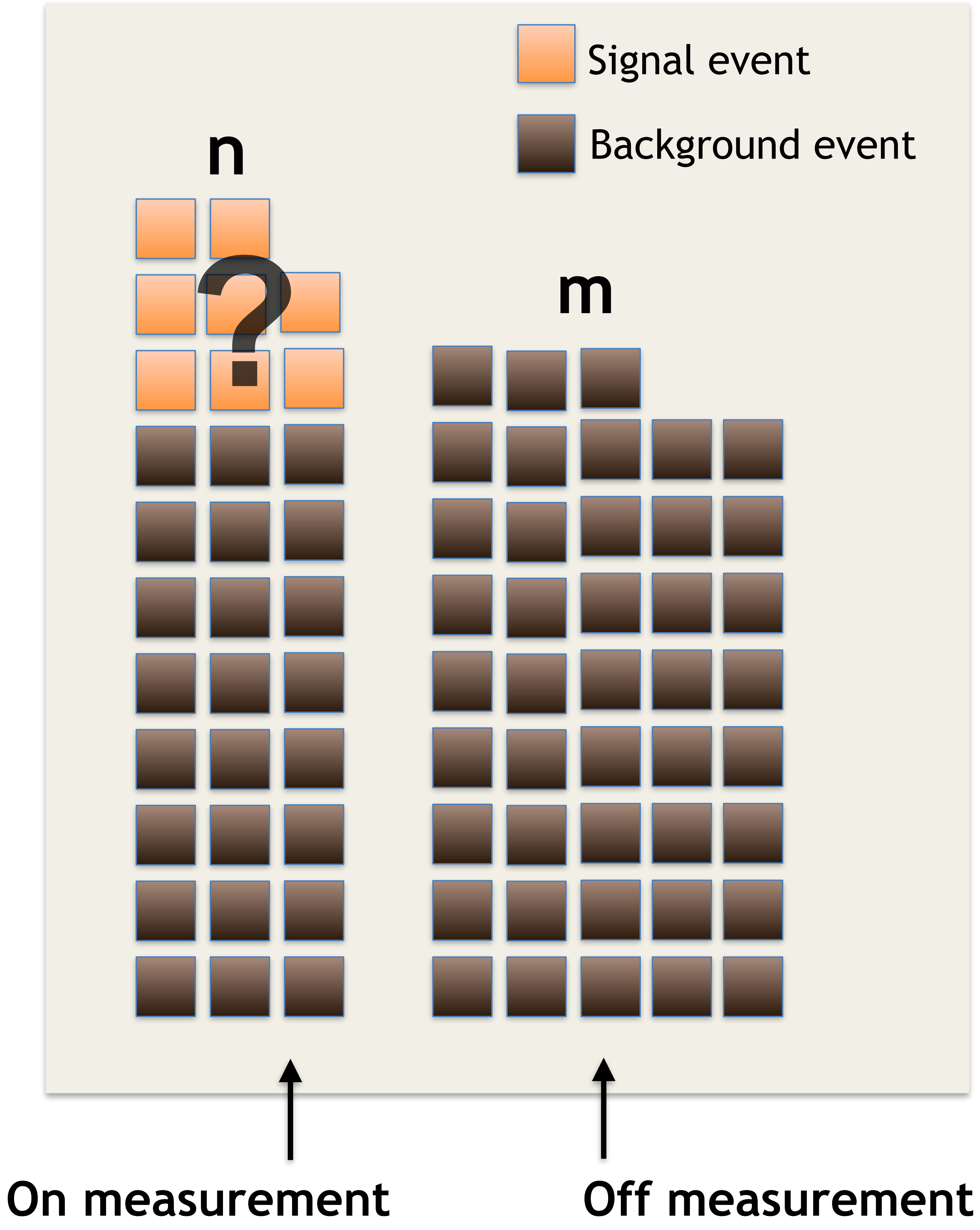
```
Out[5]: Table length=6310
EVENT_ID              TIME        RA        DEC      ENERGY
                         s       deg        deg         TeV
   int64           float64   float32    float32     float32
      42  333778849.5267153  444.21463  23.44914  0.08397394
      67  333778849.61315054  443.5247  22.725792  0.10596932
      80  333778849.6690142  443.76956  22.451006  0.19733498
     116  333778849.7778549  443.71518  21.985115  1.0020943
     179  333778849.9826064  443.64136  22.041315  0.10316629
     198  333778850.0339344  444.84238  22.175398  0.118843034
     ...                ...        ...        ...         ...
     570  333780036.17792755  443.99866  22.431725  0.14909887
     599  333780036.2743846  444.22705  22.348415  0.19341666
     622  333780036.33778954  444.08524  22.571606  0.07879259
     660  333780036.47105366  443.41534  21.67344  0.2096362
     675  333780036.5179095  443.55164  22.772985  0.17672835
     924  333780037.3755159  444.85886  22.116222  0.123453744
     963  333780037.52476007  444.8693  21.290916  0.13630114
```



On/Off measurement



variable	description
n	number of events in the On region
m	number of events in the Off region
α	exposure in the On region over the one in the Off regions
b	expected rate of occurrences of background events in the Off regions
s	expected rate of occurrences of signal events in the On region



Likelihood:

$$\mathscr{L}(s,b) \equiv p(n,m\,|\,s,b) = \frac{(s+\alpha b)^n}{n\,!}e^{-(s+\alpha b)} \cdot \frac{b^m}{m\,!}e^{-b}$$

Likelihood analysis

Get the likelihood profile:

$$\frac{\partial \mathscr{L}(s,b)}{\partial b} = 0 \longrightarrow \frac{b^{-1+m} e^{-s-b(1+\alpha)} (s+b\alpha)^{-1+n}}{m!\,n!} \left[\alpha(1+\alpha)\, b^2 + \left(s(1+\alpha) - (m+n)\alpha\right)\, b - ms\right] = 0$$

Likelihood analysis

Get the likelihood profile:

$$\frac{\partial \mathscr{L}(s,b)}{\partial b} = 0 \quad \longrightarrow \quad \frac{b^{-1+m} e^{-s-b(1+\alpha)} (s+b\alpha)^{-1+n}}{m!\,n!} \left[\alpha(1+\alpha)\, b^2 + \big(s(1+\alpha) - (m+n)\alpha\big)\; b - ms\right] = 0$$

$$\hat{b} = 0$$

Zero expected counts in the OFF region

$$\hat{b}(s) = -\frac{s}{\alpha}$$

Zero expected counts in the ON region

$$\boxed{\hat{b}(s) = \frac{n+m-(1+1/\alpha)s \pm \sqrt{(n+m-(1+1/\alpha)s)^2 + 4(1+1/\alpha)s\,m}}{2\alpha(1+\alpha)}}$$

Likelihood analysis

Get the likelihood profile:

$$\frac{\partial \mathscr{L}(s,b)}{\partial b} = 0 \longrightarrow \frac{b^{-1+m} e^{-s-b(1+\alpha)} (s+b\alpha)^{-1+n}}{m!\,n!}\left[\alpha(1+\alpha)\,b^2 + \big(s(1+\alpha) - (m+n)\alpha\big)\;b - ms\right] = 0$$

$$\hat{b} = 0$$

Zero expected counts in the OFF region

$$\hat{b}(s) = -\frac{s}{\alpha}$$

Zero expected counts in the ON region

$$\boxed{\hat{b}(s) = \frac{n+m-(1+1/\alpha)s \pm \sqrt{(n+m-(1+1/\alpha)s)^2 + 4(1+1/\alpha)s\,m}}{2\alpha(1+\alpha)}}$$

Get the maximum value of the likelihood:

$$s + \alpha b = n$$

$$b = m$$

$$\hat{\mathscr{L}} = \frac{n^n}{n!}e^{-n} \cdot \frac{m^m}{m!}e^{-m}$$

Likelihood analysis

Putting all together

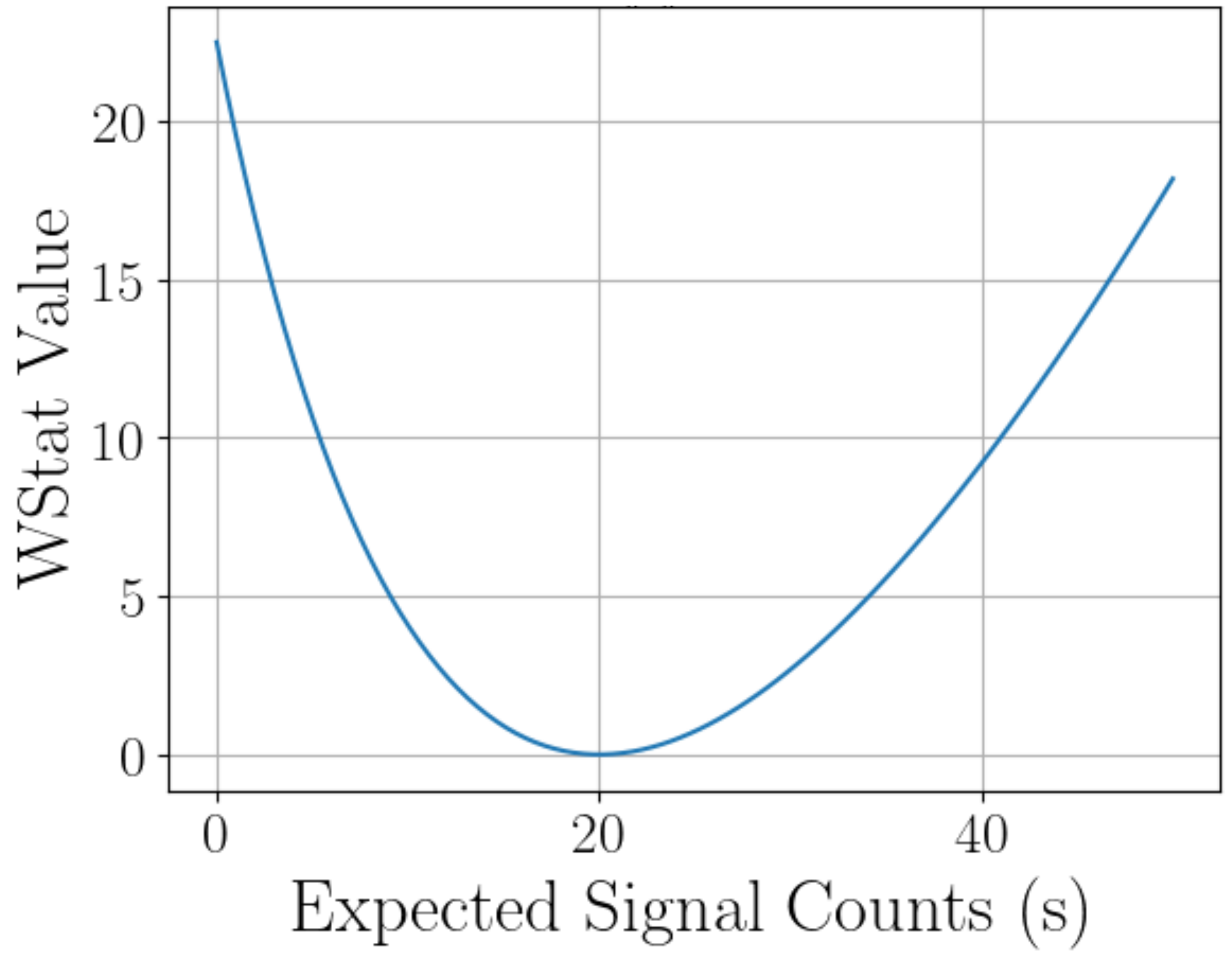
$$-2\Delta\log\mathscr{L}(s) \equiv -2\log\mathscr{L}(s,\hat{b}) + 2\log\hat{\mathscr{L}} = 2\left(s + (1+\alpha)\hat{b} \;- n\log(s+\alpha\hat{b}) - m\log(\hat{b}) - C\right)$$

$$\hat{b}(s) = \frac{n+m-(1+1/\alpha)s+\sqrt{(n+m-(1+1/\alpha)s)^2+4(1+1/\alpha)s\,m}}{2(1+\alpha)}$$

$$C = n + m - n\log(n) - m\log(m)$$

W-stat in gammapy

Example with $n = 30$, $m = 100$ and $\alpha = 0.1$



Likelihood analysis

$$W(s) \equiv -2\log \mathscr{L}(s,\hat{b}) + 2\log \hat{\mathscr{L}}_i = 2\left(s + (1+\alpha)\hat{b} - n\log(s+\alpha\hat{b}) - m\log(\hat{b}) - C\right)$$

Introducing a new nuisance parameter: “**systematics in the collection area**”

$s \to s(1+\varepsilon)$ with a Gaussian prior $\mathcal{N}(0,\sigma)$

$$\text{Signal Counts} \propto \mathrm{Flux} \times \text{Coll. Area} \times (1+\varepsilon)$$

Likelihood analysis

$$W(s) \equiv -2\log \mathscr{L}(s,\hat{b}) + 2\log \hat{\mathscr{L}}_i = 2\left(s + (1+\alpha)\hat{b} \; - n\log(s+\alpha\hat{b}) - m\log(\hat{b}) - C \right)$$

Introducing a new nuisance parameter: “**systematics in the collection area**”

$$s \rightarrow s(1+\varepsilon)$$

with a Gaussian prior $\mathscr{N}(0,\sigma)$

$$W_\varepsilon(s) \equiv -2\log \mathscr{L}(s,\hat{b},\hat{\varepsilon}) + 2\log \hat{\mathscr{L}}_i = 2\left(s(1+\hat{\varepsilon}) + (1+\alpha)\hat{b} \; - n\log(s(1+\hat{\varepsilon}) + \alpha\hat{b}) - m\log(\hat{b}) + \frac{\hat{\varepsilon}^2}{2\sigma^2} - C \right)$$

Likelihood analysis

$$W(s) \equiv -2\log\mathscr{L}(s,\hat{b}) + 2\log\hat{\mathscr{L}}_i = 2\left(\; s + (1+\alpha)\hat{b} \;\; - n\log(s+\alpha\hat{b}) - m\log(\hat{b}) - C\right)$$

Introducing a new nuisance parameter: “**systematics in the collection area**”

$s \rightarrow s(1+\varepsilon)$ with a Gaussian prior $\mathscr{N}(0,\sigma)$

$$W_\varepsilon(s) \equiv -2\log\mathscr{L}(s,\hat{b},\hat{\varepsilon}) + 2\log\hat{\mathscr{L}}_i = 2\left(\; s(1+\hat{\varepsilon}) + (1+\alpha)\hat{b} \;\; - n\log(s(1+\hat{\varepsilon}) + \alpha\hat{b}) - m\log(\hat{b}) + \frac{\hat{\varepsilon}^2}{2\sigma^2} - C\right)$$

$$\alpha\hat{\varepsilon}^3 + A\hat{\varepsilon}^2 + B\hat{\varepsilon} + C = 0$$

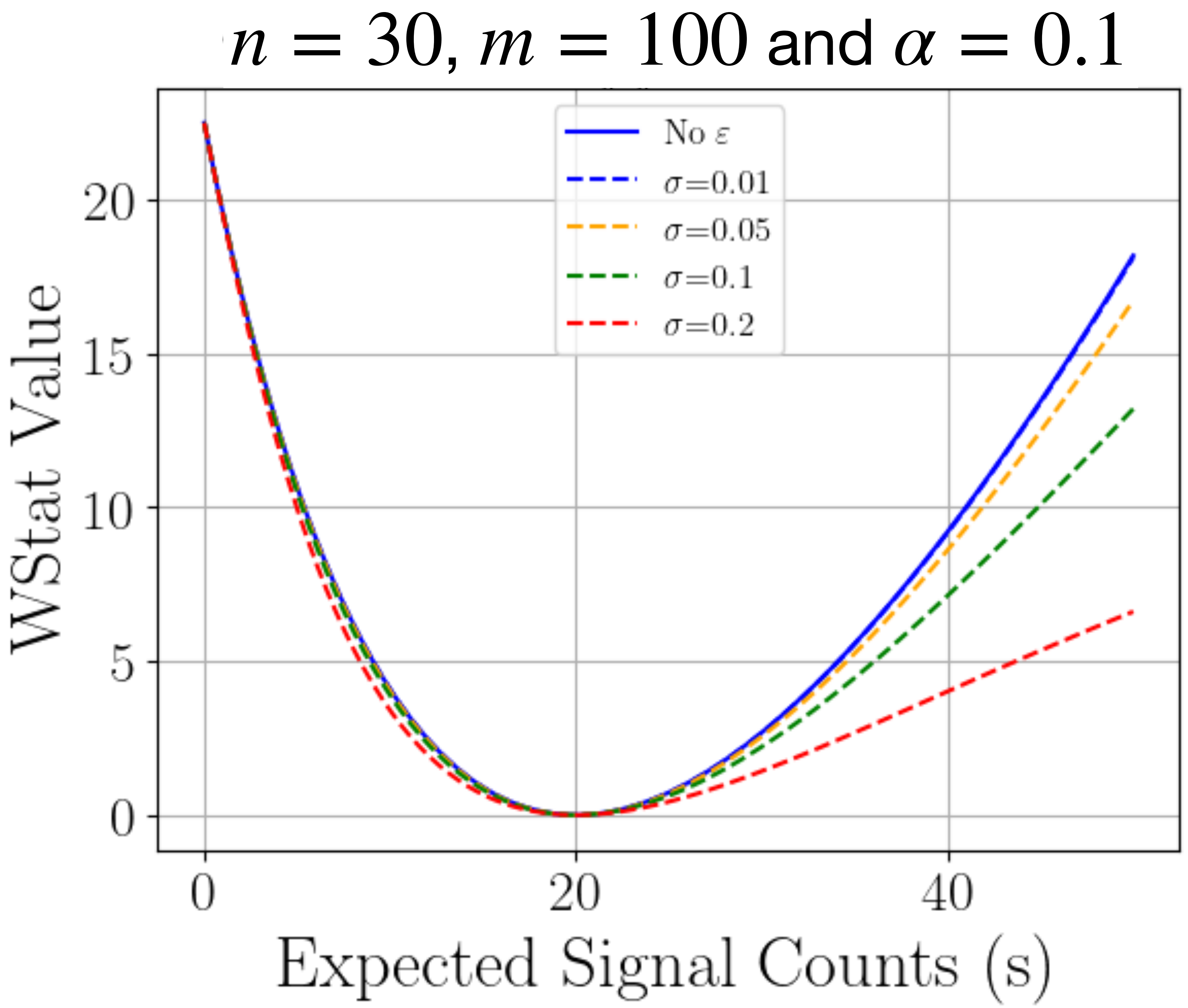
$$A = \; s\sigma^2(\alpha-1) + \alpha$$
$$B = \; \sigma^2\left(\alpha(s-m-n) - s - s^2\sigma^2\right]$$
$$C = \sigma^4 s\,(n - \alpha\,m - s)$$

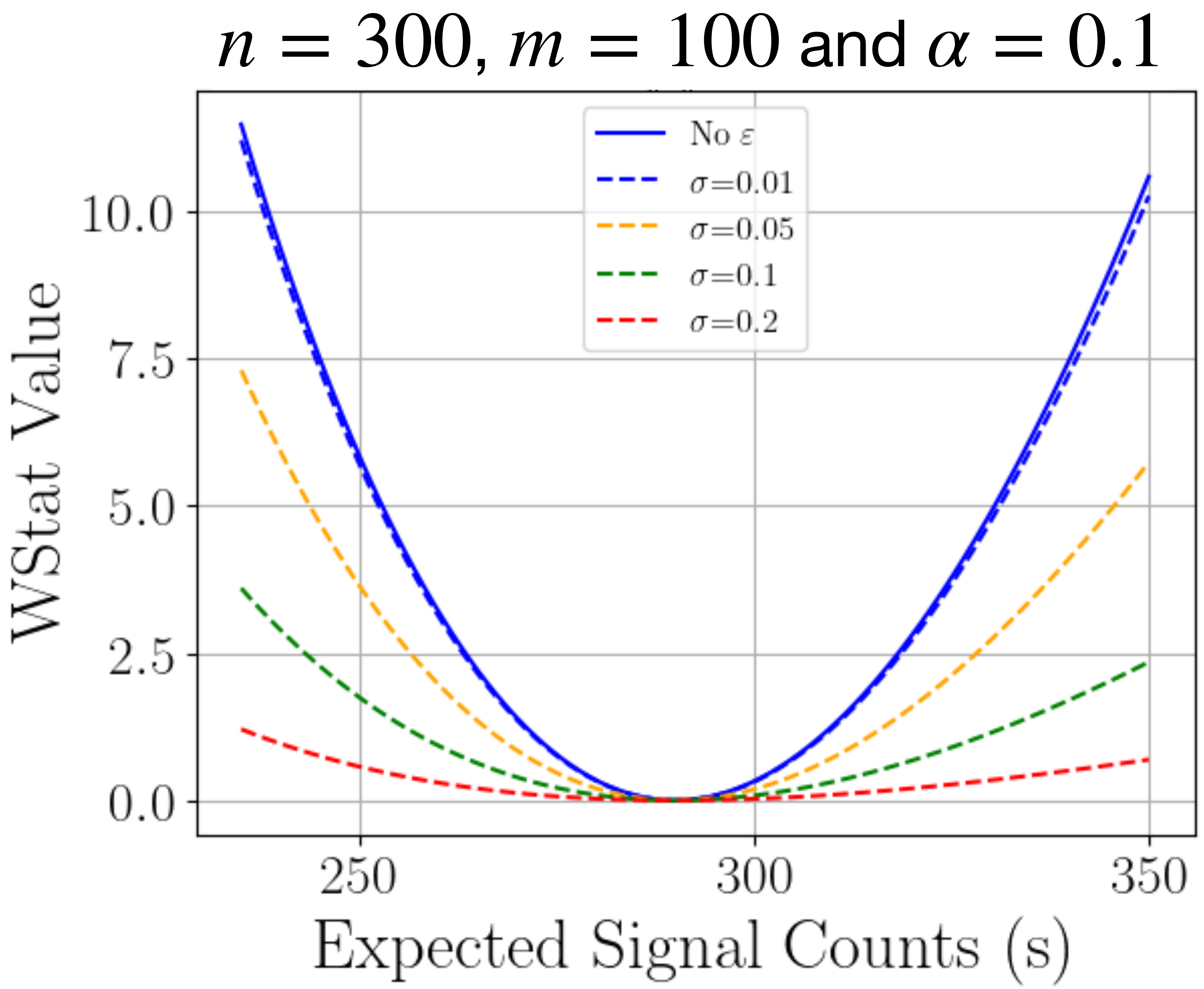
$$\hat{b}(s) = \frac{n + m - (1+1/\alpha)s(1+\varepsilon) + \sqrt{(n+m-(1+1/\alpha)s(1+\varepsilon))^2 + 4(1+1/\alpha)s(1+\varepsilon)\,m}}{2(1+\alpha)}$$

$$C = n + m - n\log(n) - m\log(m)$$

Likelihood analysis

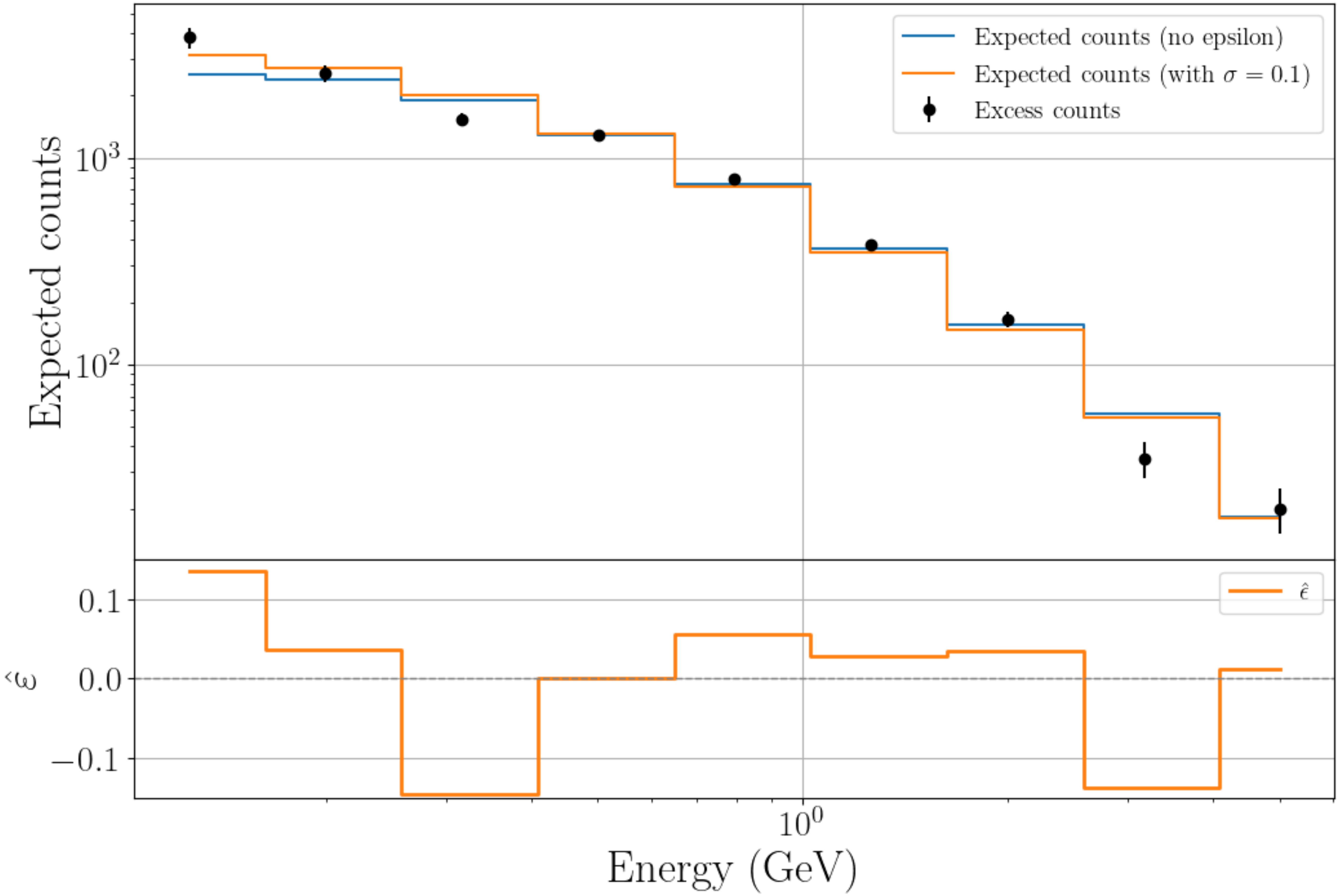
$$W_\varepsilon(s) \equiv -2\log \mathscr{L}(s,\hat{b},\hat{\varepsilon}) + 2\log \hat{\mathscr{L}}_i = 2\left(s(1+\hat{\varepsilon}) + (1+\alpha)\hat{b} - n\log(s(1+\hat{\varepsilon}) + \alpha\hat{b}) - m\log(\hat{b}) + \frac{\hat{\varepsilon}^2}{2\sigma^2} - C \right)$$





Likelihood analysis

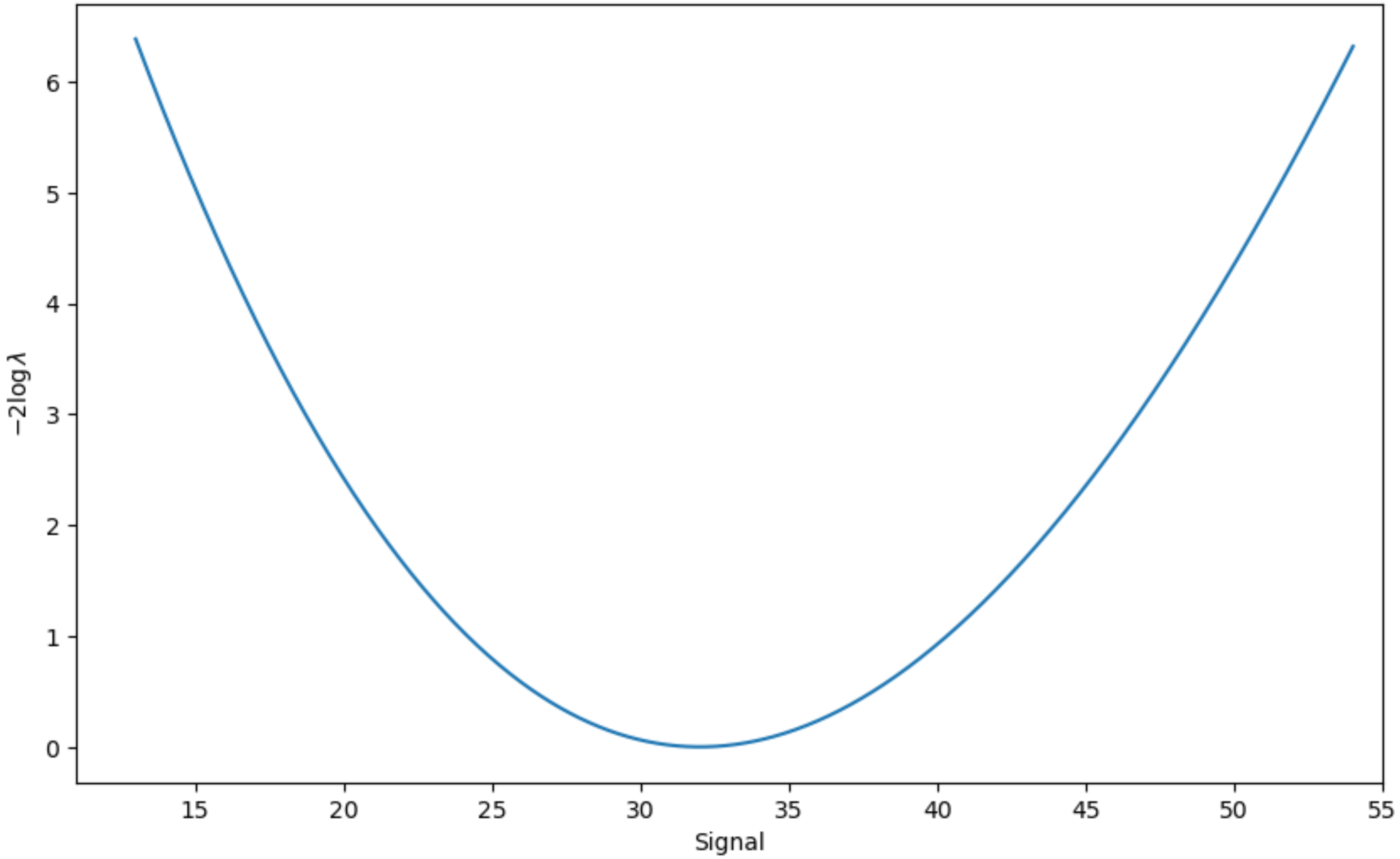
Example (modifying ***gammapy*** source code):



Likelihood analysis

$$-2\Delta\log\mathscr{L}(s) \equiv -2\log\mathscr{L}(s,\hat{b}) + 2\log\hat{\mathscr{L}} = 2\left(\ s + (1+\alpha)\hat{b}\ \ - n\log(s+\alpha\hat{b}) - m\log(\hat{b}) - C\right)$$

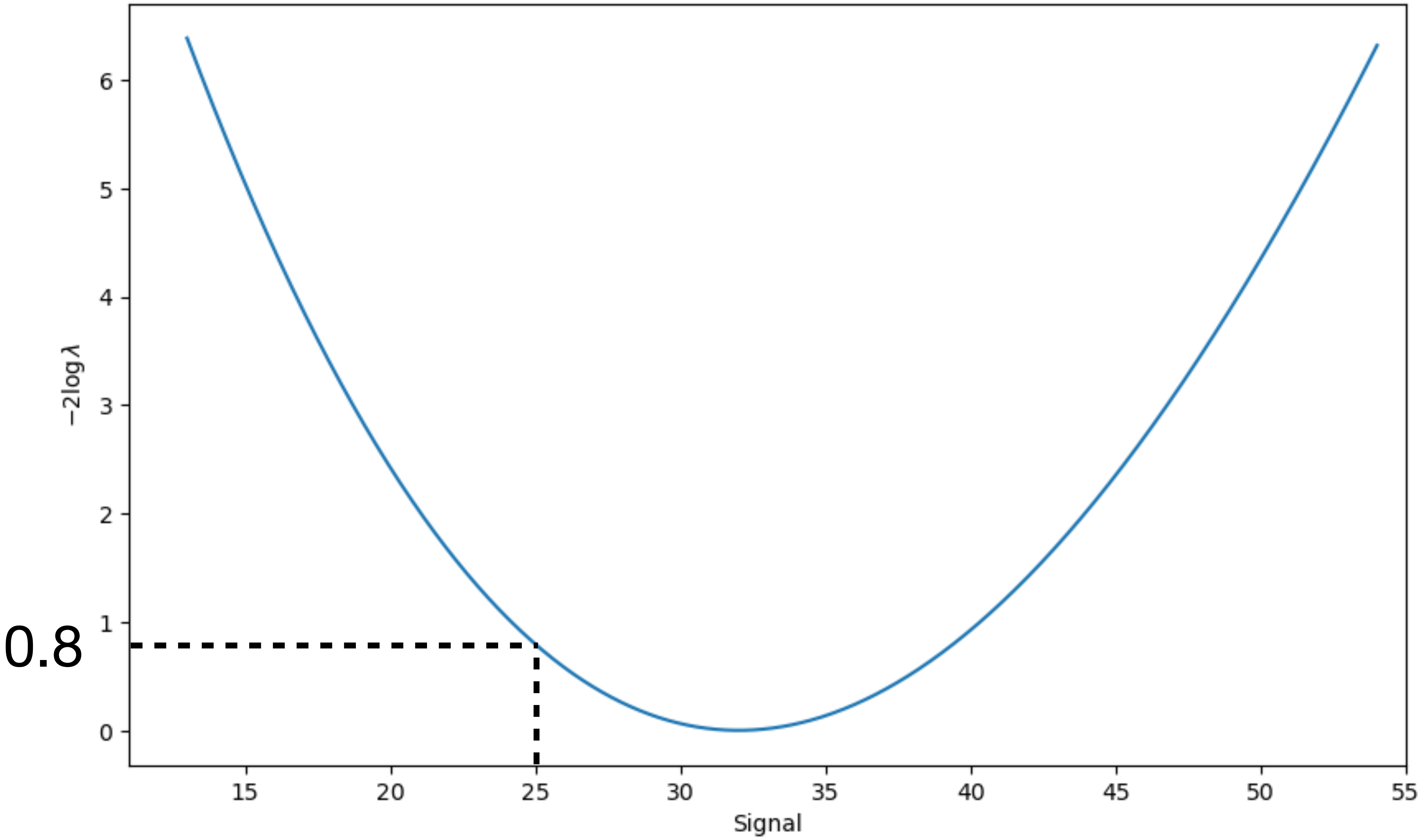
Example with $n = 54$, $m = 44$ and $\alpha = 0.5$



Likelihood analysis

$$-2\Delta\log\mathscr{L}(s) \equiv -2\log\mathscr{L}(s,\hat{b}) + 2\log\hat{\mathscr{L}} = 2\left(s + (1+\alpha)\hat{b} - n\log(s+\alpha\hat{b}) - m\log(\hat{b}) - C\right)$$

Example with $n = 54$, $m = 44$ and $\alpha = 0.5$



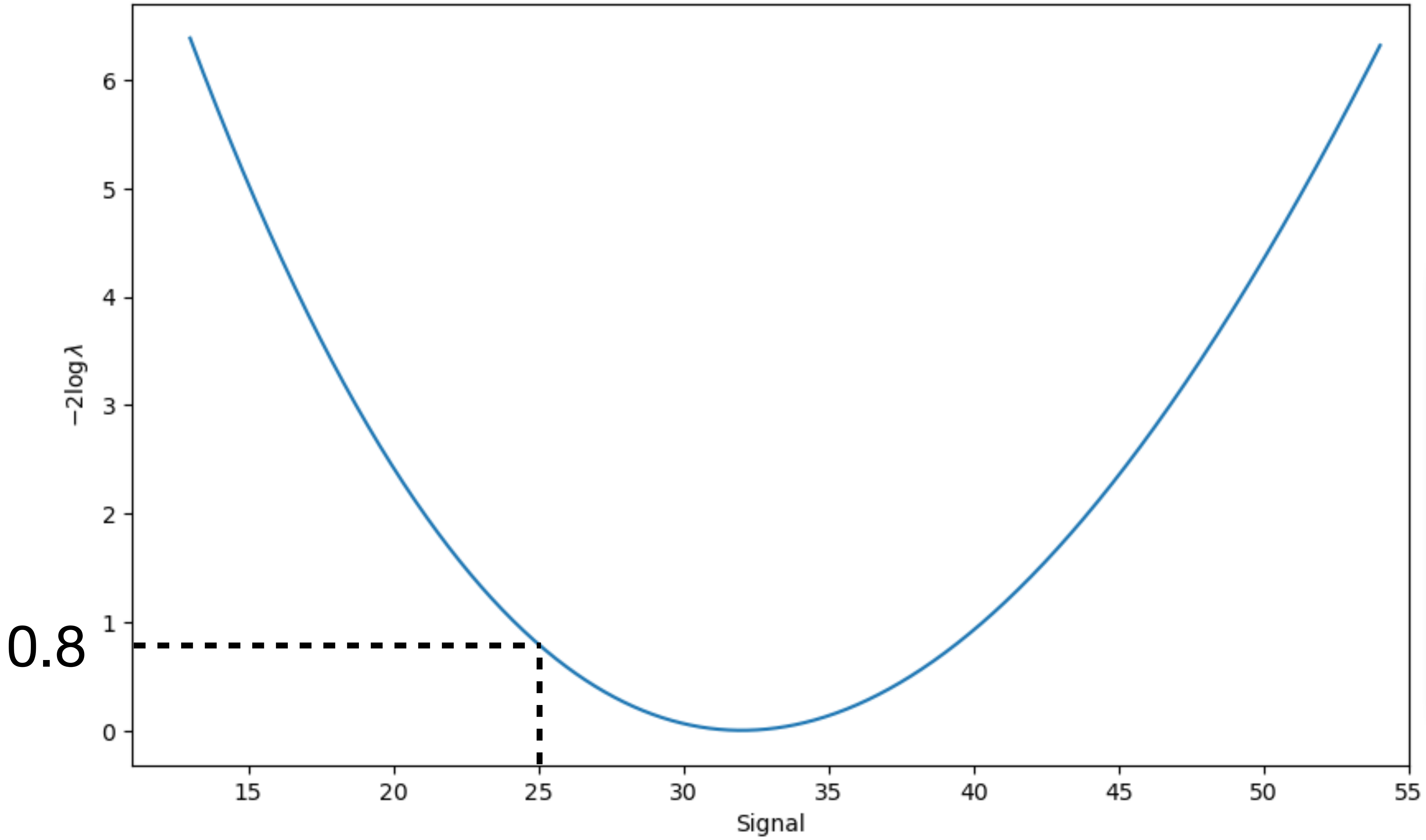
What does it mean that for s = 25, we have $-2\Delta\log\mathscr{L} = 0.8$?

Likelihood analysis

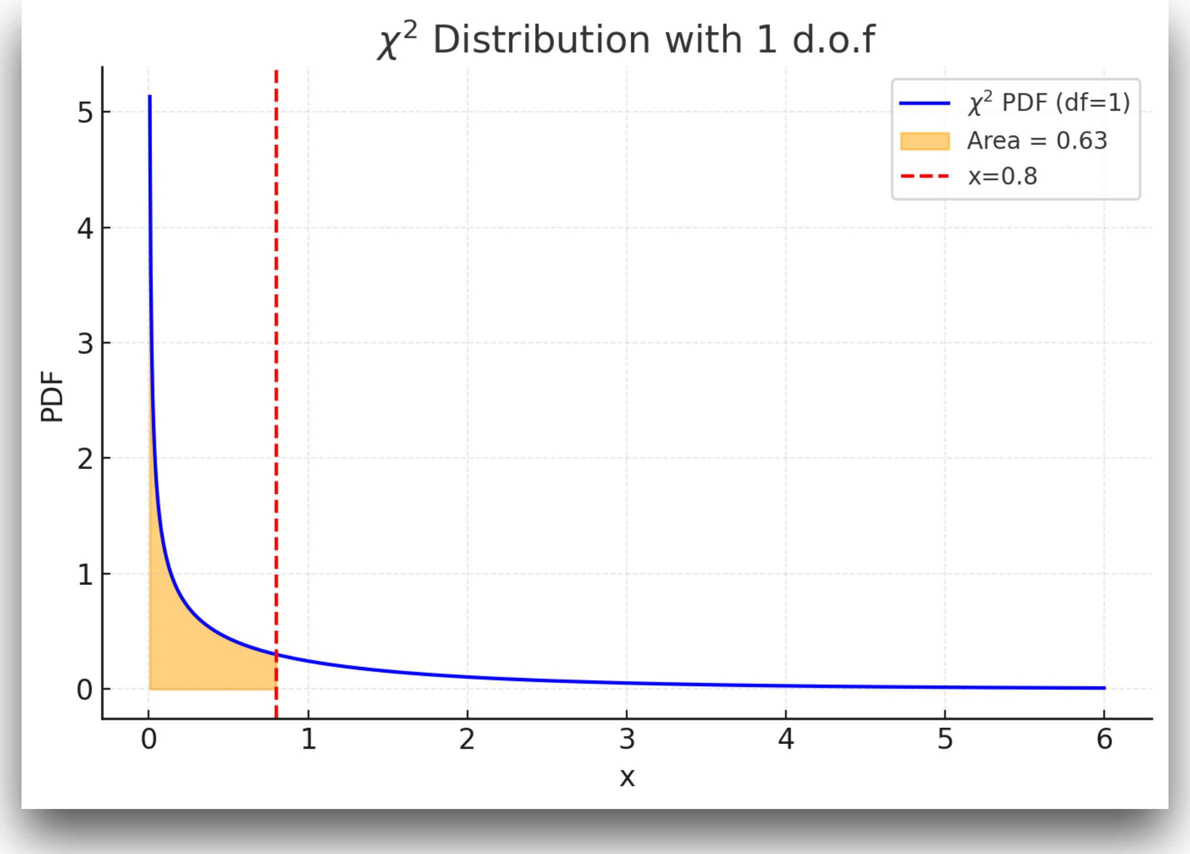
$$-2\Delta \log \mathscr{L}(s) \equiv -2\log \mathscr{L}(s,\hat{b}) + 2\log \hat{\mathscr{L}} = 2\left(s + (1+\alpha)\hat{b} \;\; - n\log(s+\alpha\hat{b}) - m\log(\hat{b}) - C \right)$$

What does it mean that for s = 25, we have $-2\Delta\log\mathscr{L} = 0.8$?

Example with $n = 54$, $m = 44$ and $\alpha = 0.5$



Since $-2\Delta\log\mathscr{L}$ is a χ^2 variable, we can exclude the hypothesis $s = 25$ with a 63% CL

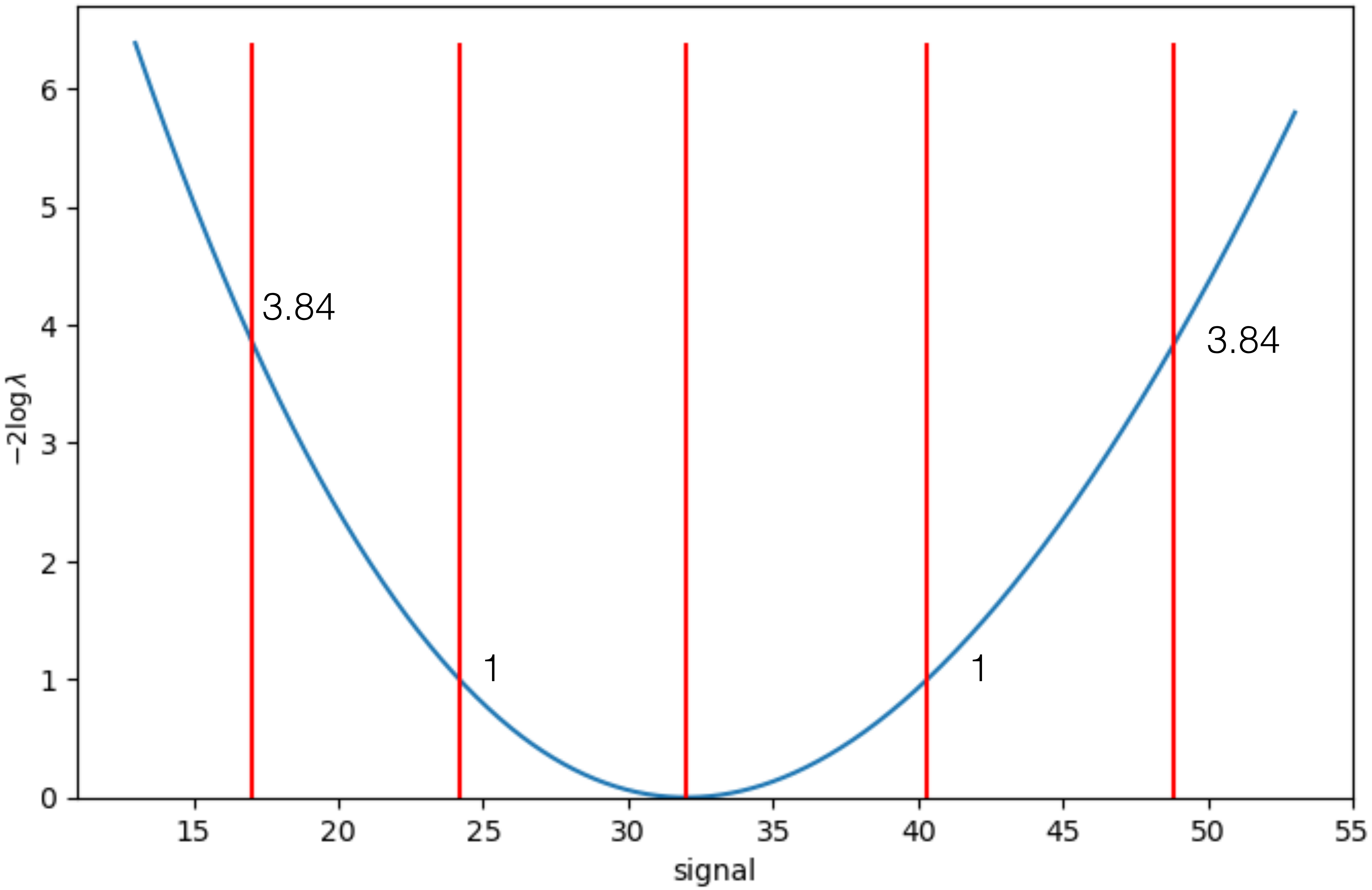


Likelihood analysis

Conventionally 3 confidence levels are reported:

- **0% CL** : which is by definition when the chi-squared is **zero**
- **68% CL** : which is when the chi-squared is **1**
- **95% CL** : which is when the chi-squared is **3.84**

Example with $n = 54$, $m = 44$ and $\alpha = 0.5$

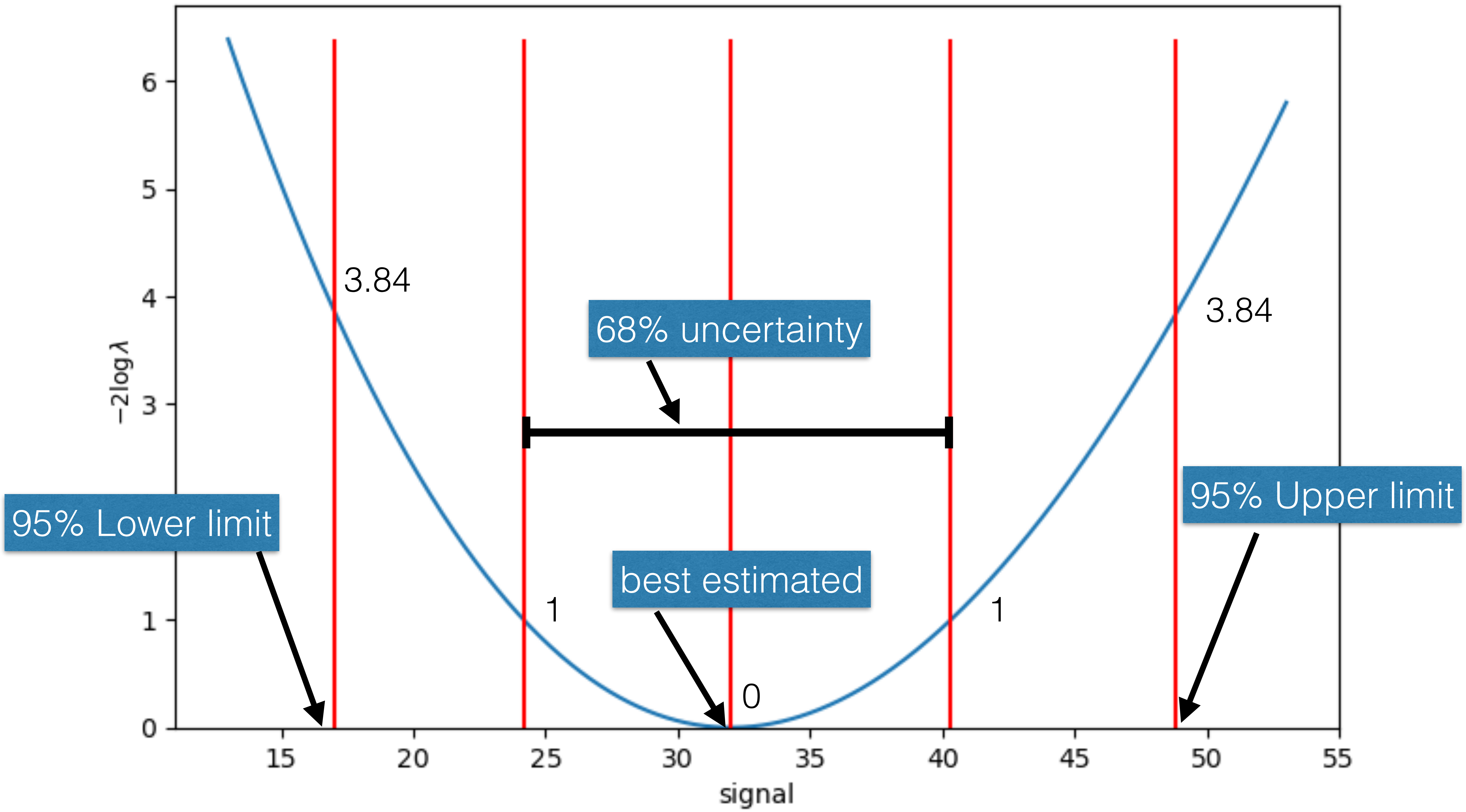


Likelihood analysis

Conventionally 3 confidence levels are reported:

- **0% CL** : which is by definition when the chi-squared is **zero**
- **68% CL** : which is when the chi-squared is **1**
- **95% CL** : which is when the chi-squared is **3.84**

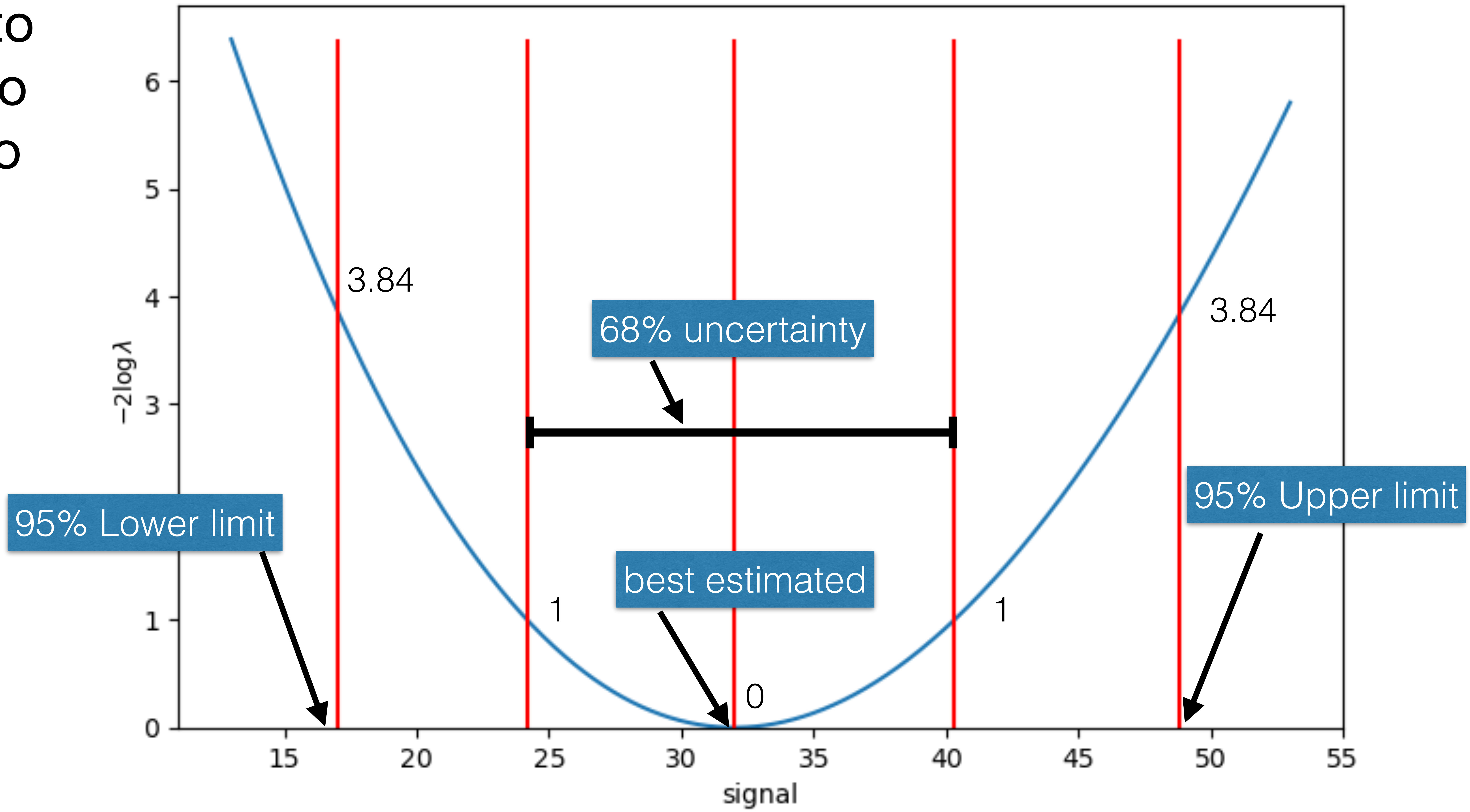
Example with $n = 54$, $m = 44$ and $\alpha = 0.5$



Likelihood analysis

So do we need each time to compute the likelihood ratio and find where it is equal to zero, one, and 3.84?

Example with $n = 54$, $m = 44$ and $\alpha = 0.5$

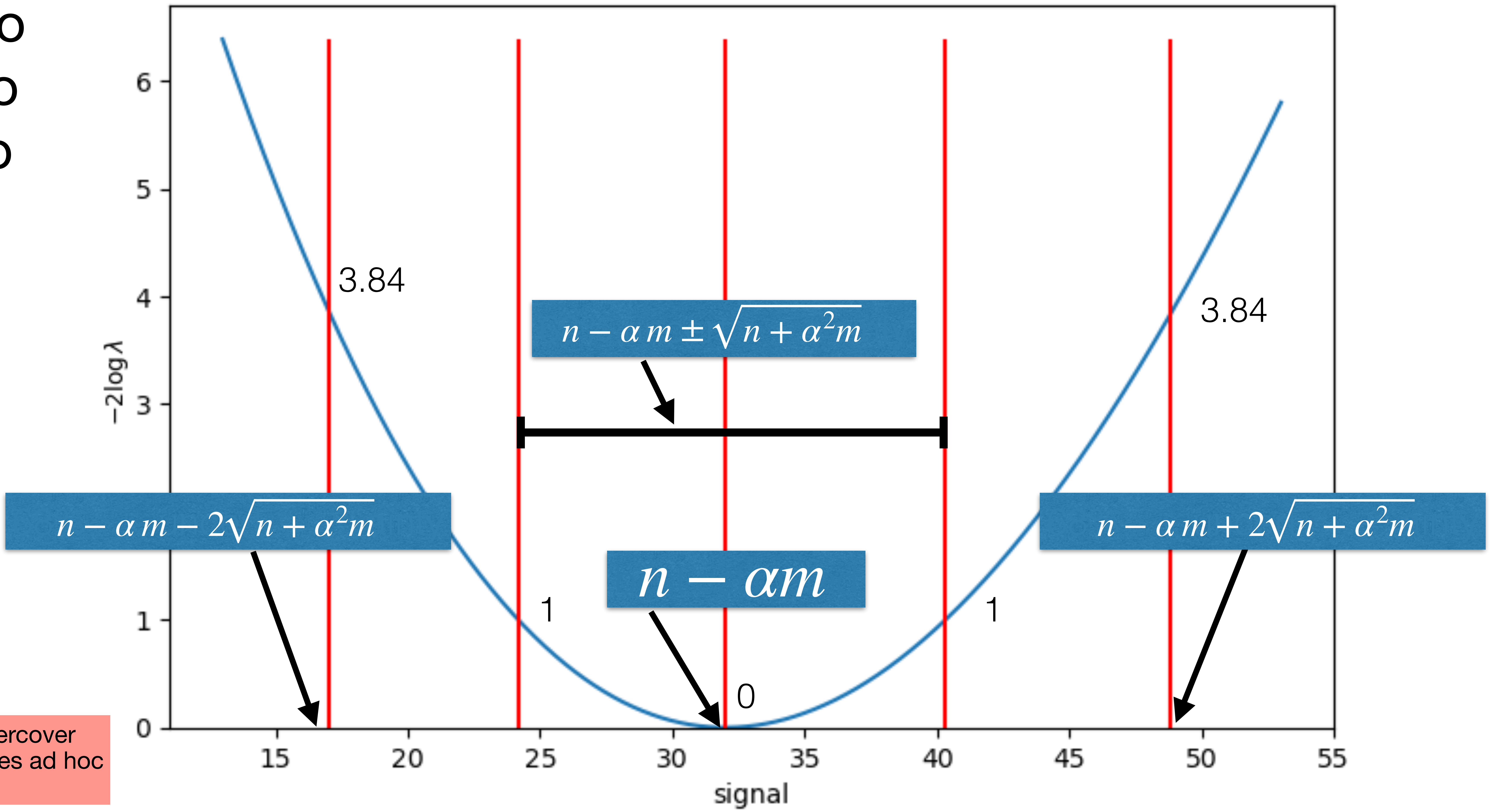


Likelihood analysis

So do we need each time to compute the likelihood ratio and find where it is equal to zero, one, and 3.84?

Thankfully in **most cases*** we can get a good approximation using the following expression

Example with $n = 54$, $m = 44$ and $\alpha = 0.5$



$n - \alpha\, m \pm \sqrt{n + \alpha^2 m}$

$n - \alpha\, m - 2\sqrt{n + \alpha^2 m}$

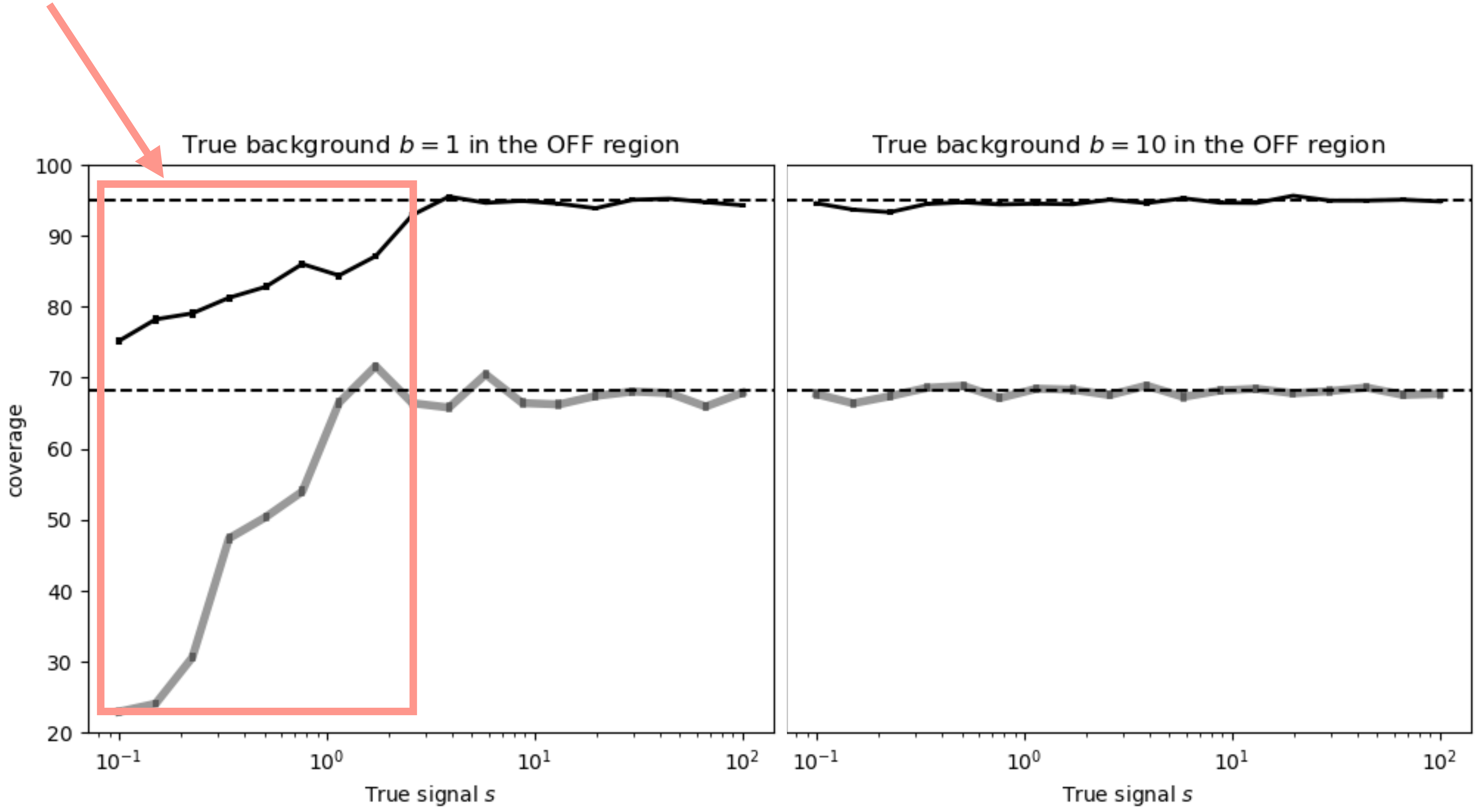
$n - \alpha m$

$n - \alpha\, m + 2\sqrt{n + \alpha^2 m}$

* When the counts are too small, these intervals tend to undercover the true value of s. Such a problem is well-known and requires ad hoc adjustments (see Rolke, W. A., & Lopez, A. M. (2001)).

Likelihood analysis

When the counts are too small, these intervals tend to undercover the true value of s. Such a problem is well-known and requires ad hoc adjustments (see Rolke, W. A., & Lopez, A. M. (2001))



Likelihood analysis

Final result:

$$n - \alpha\, m = 32$$

$$\sqrt{n + \alpha^2 m} = 8.06$$

- The signal estimation is:

32 ± 8

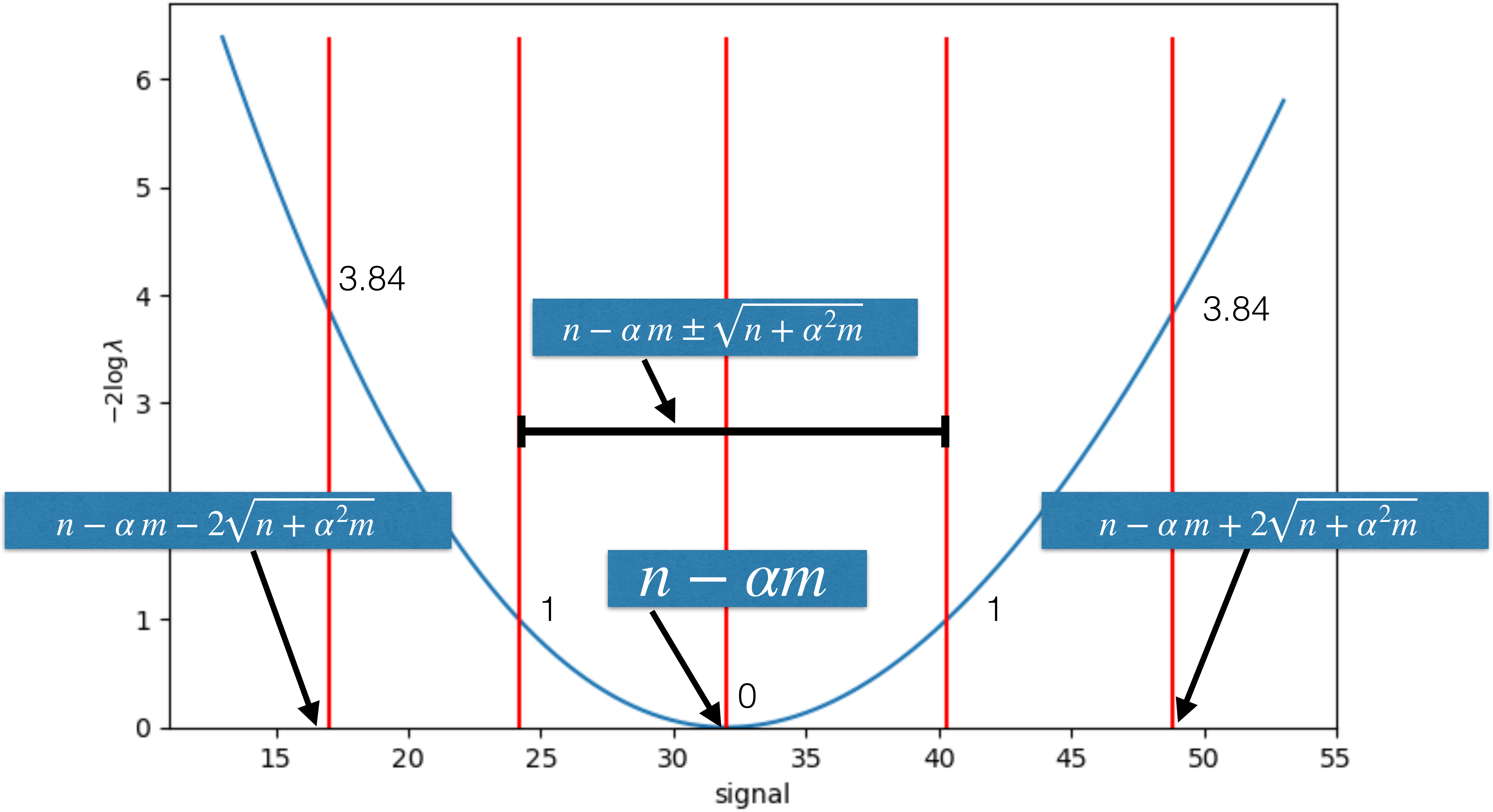
- with upper limit

48.1

- and lower limit

15.9

Example with $n = 54$, $m = 44$ and $\alpha = 0.5$



Li&Ma significance

Among all the possible hypotheses, there is a ‘special’ one we are interested in excluding…
… the one in which there is no signal, i.e. **s=0**

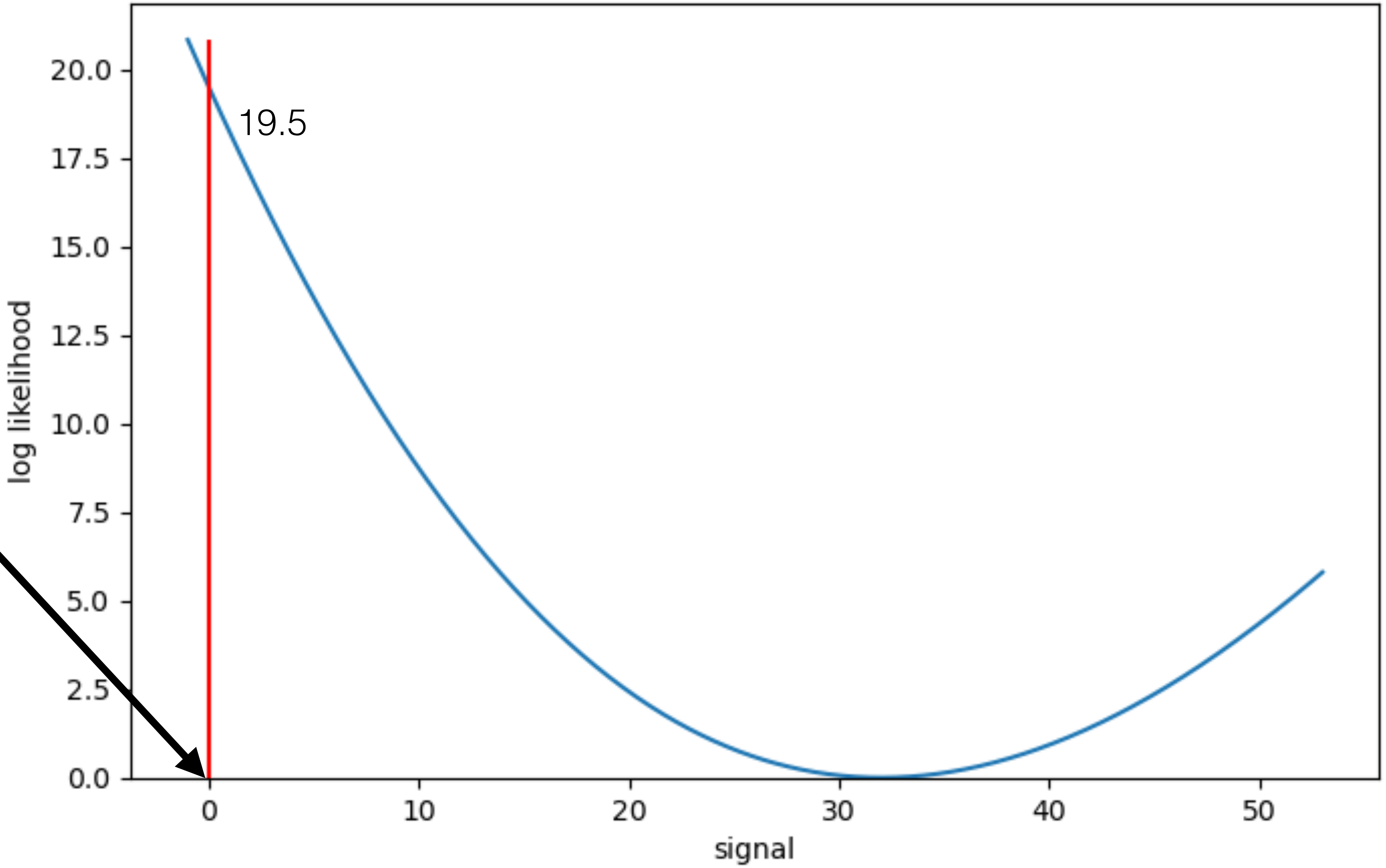
↓

$$-2\Delta \log \mathscr{L}(s=0)$$

↓ We then take the square root of it to get the significance

$$\sqrt{-2\Delta \log \mathscr{L}}$$

Example with $n = 54$, $m = 44$ and $\alpha = 0.5$



Li&Ma significance

Among all the possible hypotheses, there is a 'special' one we are interested in excluding…
… the one in which there is no signal, i.e. **s=0**

$$\downarrow$$

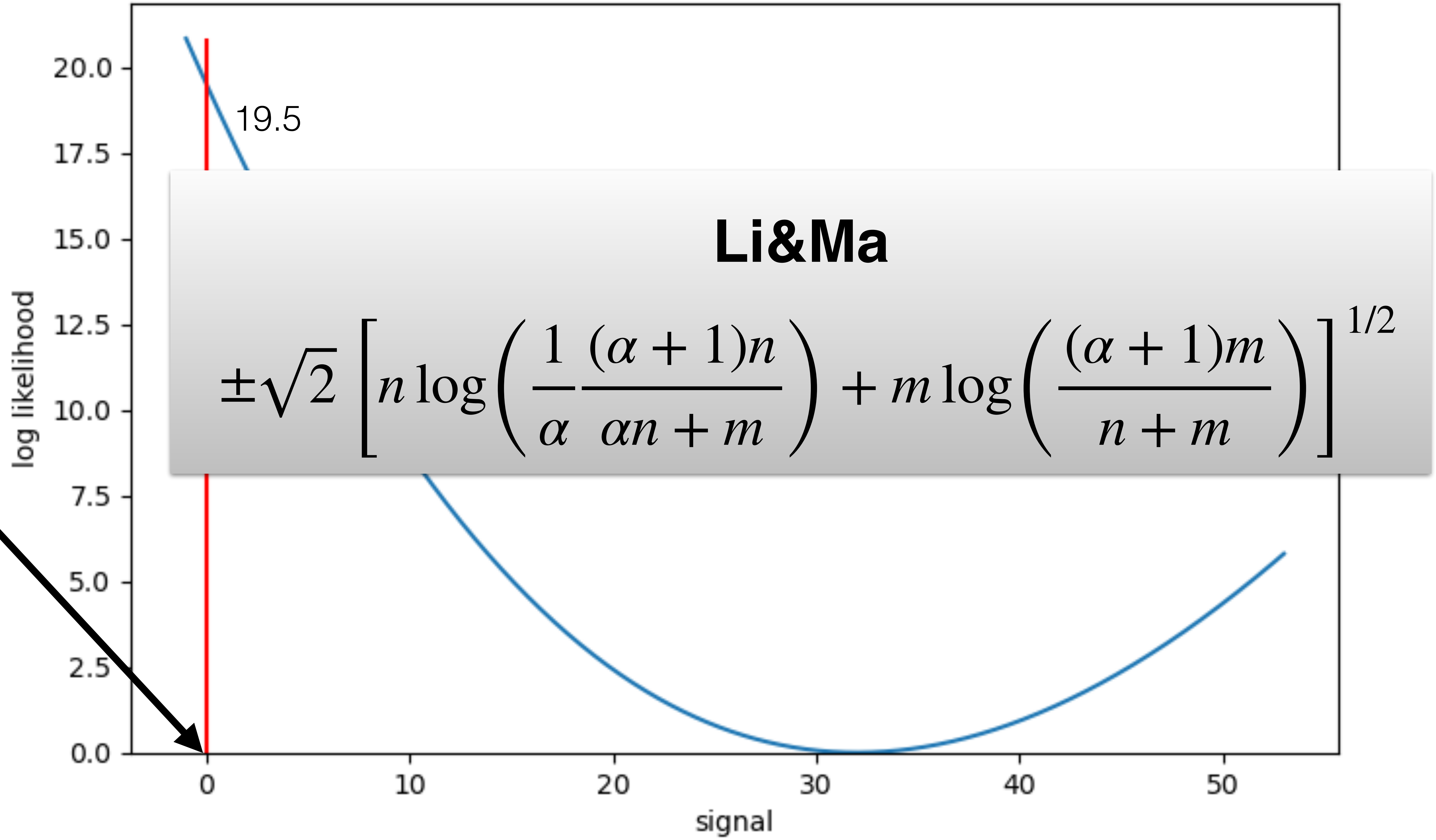
$$-2\Delta\log\mathscr{L}(s=0)$$

$$\downarrow$$

We then take the square root of it to get the significance

$$\sqrt{-2\Delta\log\mathscr{L}}$$

Example with $n = 54$, $m = 44$ and $\alpha = 0.5$

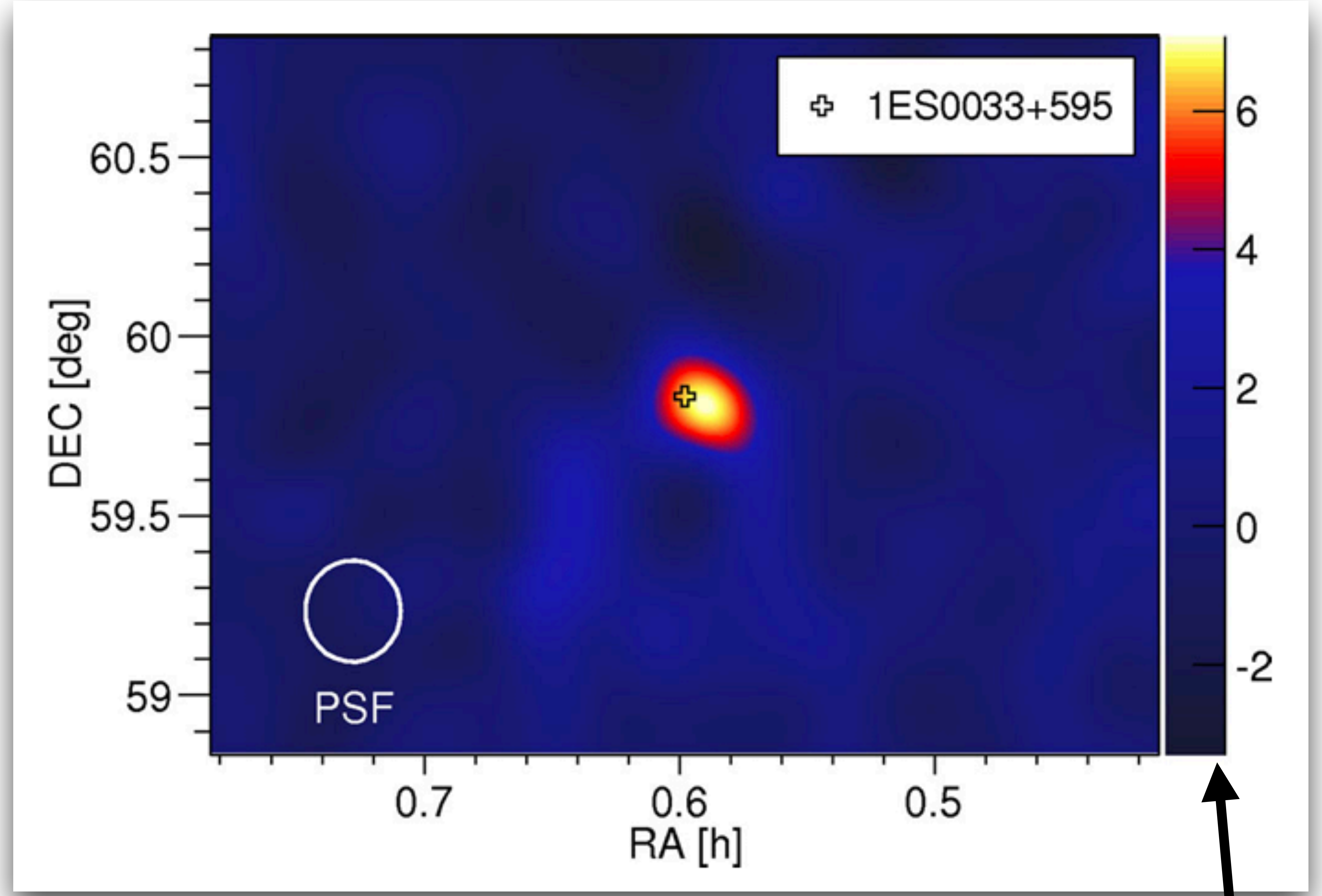


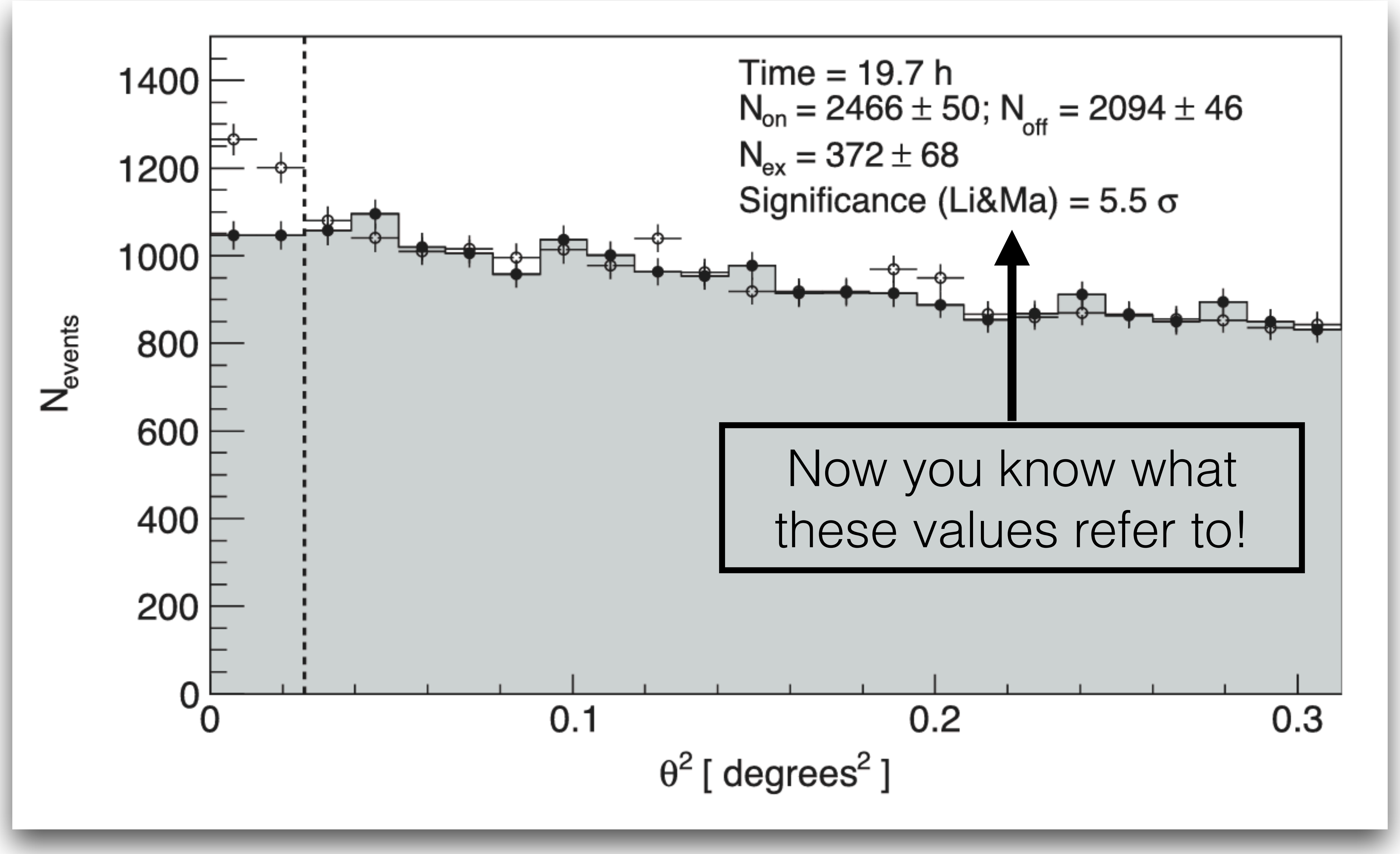
Li&Ma

$$\pm\sqrt{2}\left[n\log\left(\frac{1}{\alpha}\frac{(\alpha+1)n}{\alpha n+m}\right)+m\log\left(\frac{(\alpha+1)m}{n+m}\right)\right]^{1/2}$$

Li&Ma significance

Examples from published papers:





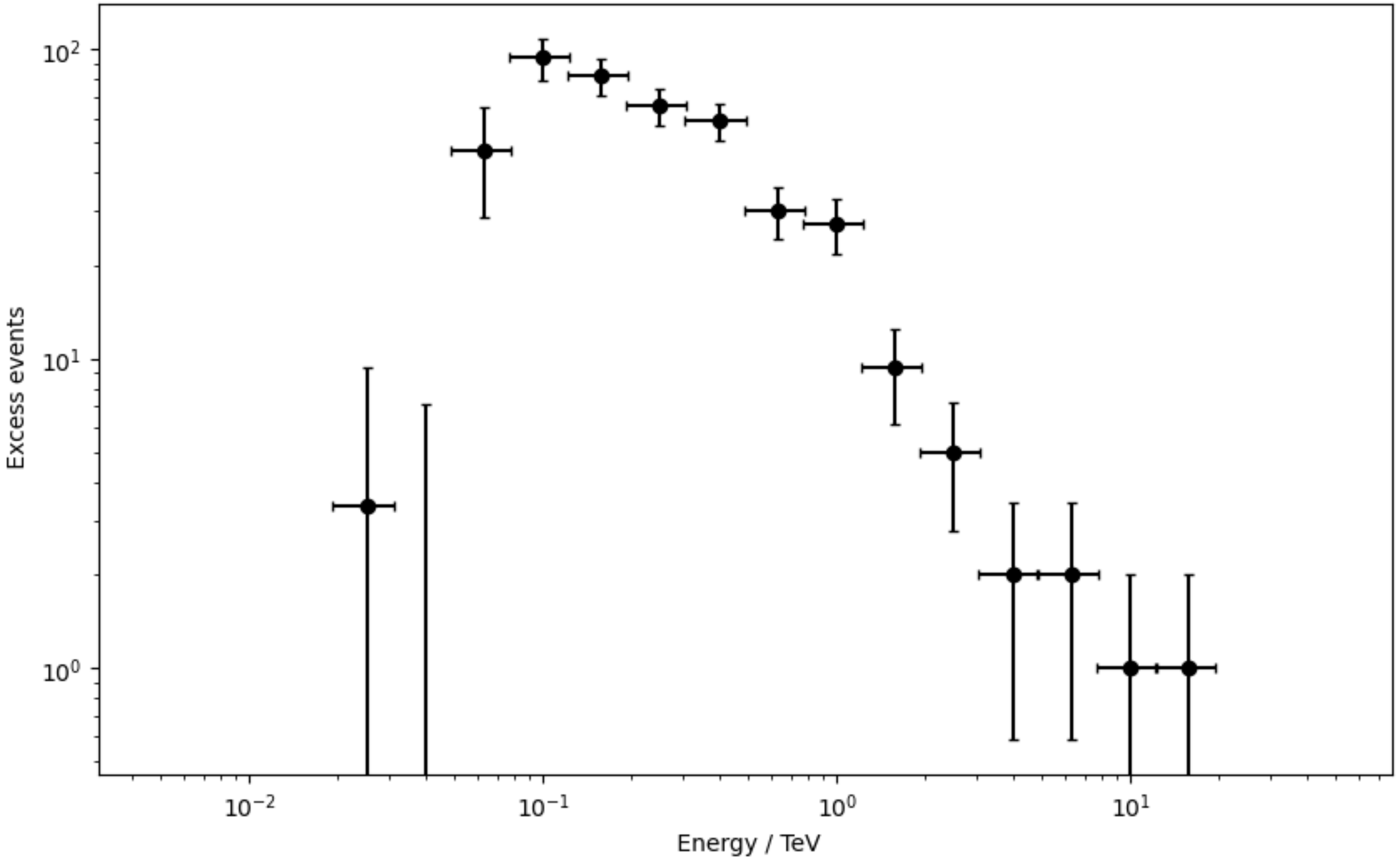
Now you know what these values refer to!

Here you are seeing the TS or the log-likelihood value obtained in each pixel for the null hypothesis

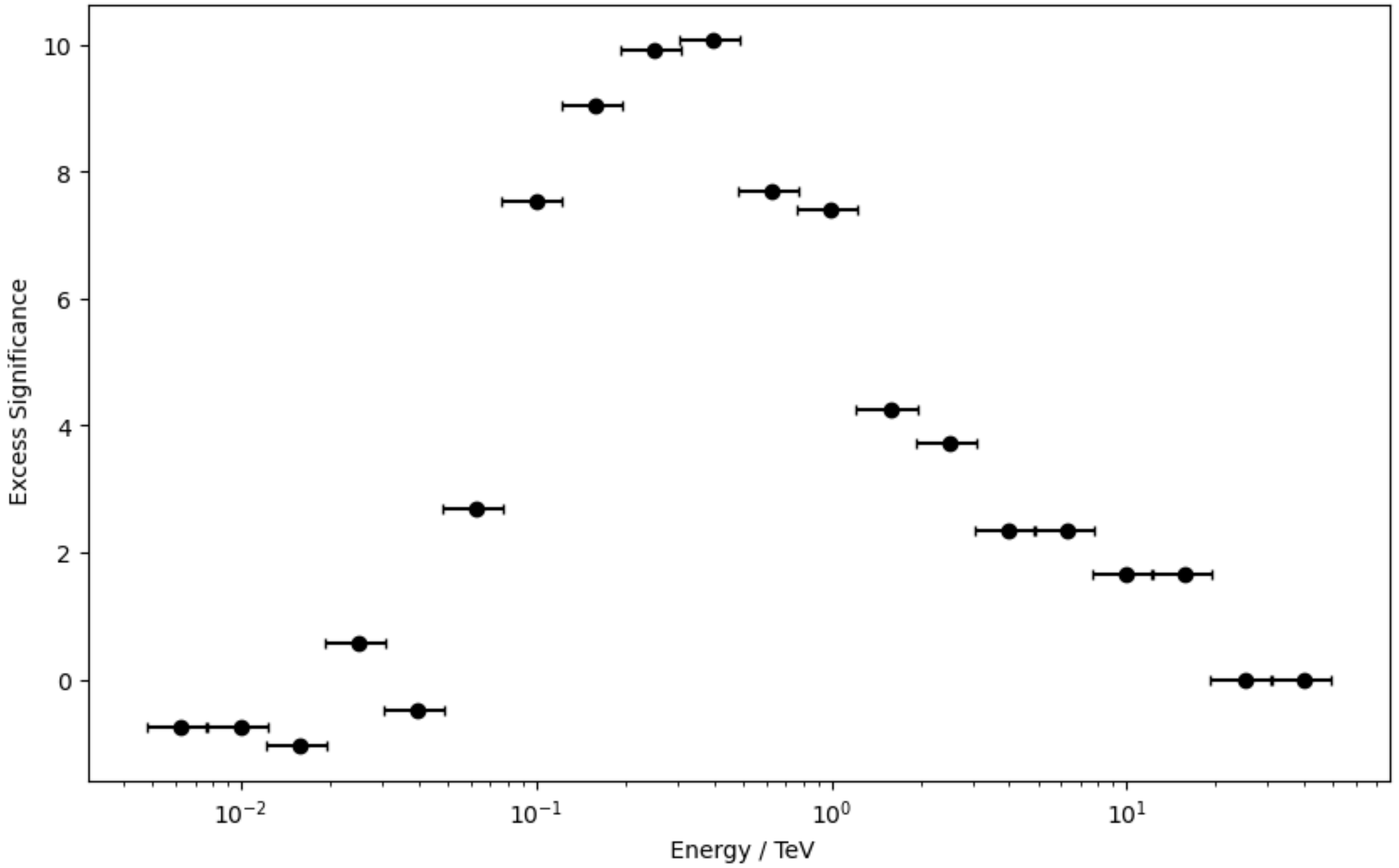
And finally the excess estimation

Taking the dataset I showed you a few slides back:

$$N_\gamma = n - \alpha\, m \pm \sqrt{n + \alpha^2 m}$$

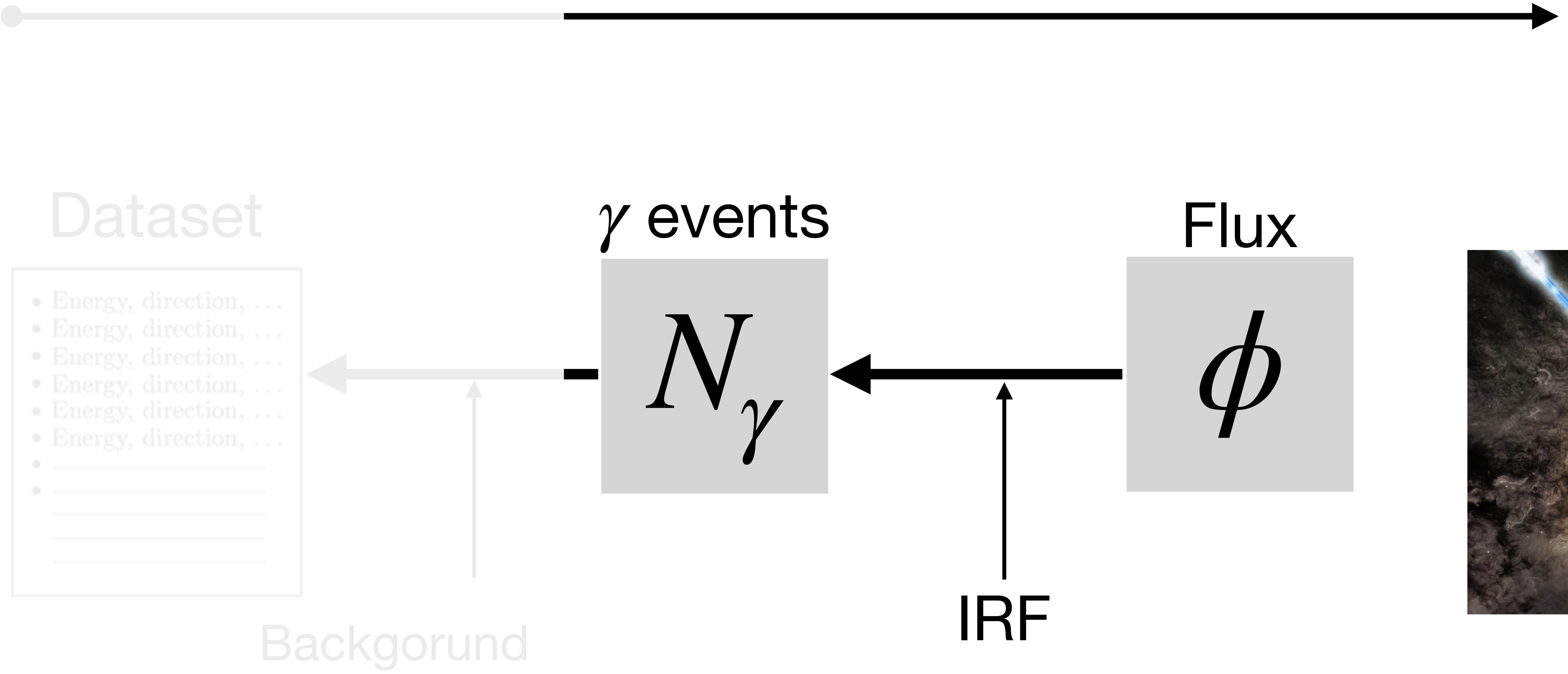


Li&Ma



Estimating the flux

In this part, we will estimate the flux from the excess N_γ



Dataset

- Energy, direction, …
- Energy, direction, …
- Energy, direction, …
- Energy, direction, …
- Energy, direction, …
- Energy, direction, …

Backgorund

γ events

N_γ

IRF

Flux

ϕ

Estimating the flux

Differential energy spectrum: $\dfrac{d\phi}{dE} = \dfrac{d^3N_\gamma}{dt\,dA\,dE}$

Estimating the flux

Differential energy spectrum: $\frac{d\phi}{dE} = \frac{d^3 N_\gamma}{dt\, dA\, dE} \approx \frac{N_\gamma}{t_{eff}\, A_{eff}\, \Delta E}$

Estimating the flux

Excess in a given energy and time window,
we now know how to compute it!

Differential energy spectrum: $$\frac{d\phi}{dE} = \frac{d^3N_\gamma}{dt\,dA\,dE} \approx \frac{\overset{\searrow}{N_\gamma}}{t_{eff}\,A_{eff}\,\Delta E}$$

Estimating the flux

Excess in a given energy and time window,
we now know how to compute it!

Differential energy spectrum: $\frac{d\phi}{dE} = \frac{d^3N_\gamma}{dt\,dA\,dE} \approx \frac{N_\gamma}{t_{eff}\,A_{eff}\,\Delta E}$

Energy width: $E_2 - E_1$

Estimating the flux

Excess in a given energy and time window,
we now know how to compute it!

Differential energy spectrum: $$\frac{d\phi}{dE} = \frac{d^3N_\gamma}{dt\,dA\,dE} \approx \frac{N_\gamma}{t_{eff}\,A_{eff}\,\Delta E}$$

Effective time: not the same as the elapsed time between the beginning and end of the observation, as gaps in the data taking and dead time after the recording of each event must be taken into account

Energy width: $E_2 - E_1$

Estimating the flux

Excess in a given energy and time window,
we now know how to compute it!

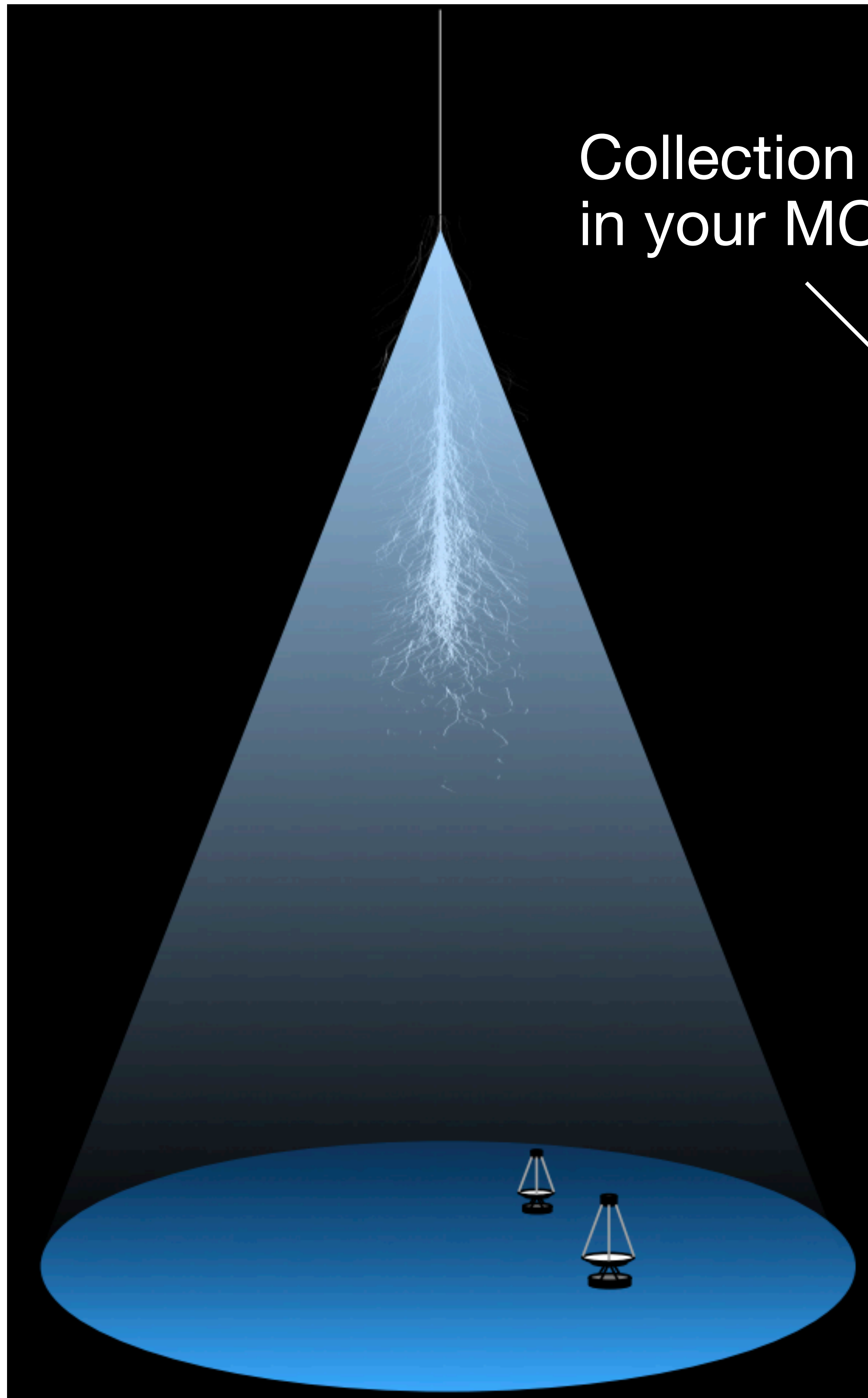
Differential energy spectrum: $$\frac{d\phi}{dE} = \frac{d^3N_\gamma}{dt\,dA\,dE} \approx \frac{N_\gamma}{t_{eff}\,A_{eff}\,\Delta E}$$

Effective time: not the same as the elapsed time between the beginning and end of the observation, as gaps in the data taking and dead time after the recording of each event must be taken into account

Energy width: $E_2 - E_1$

Effective collection area: area of an ideal detector that would collect the same γ rays as our real detector (see next slides)

Effective collection area



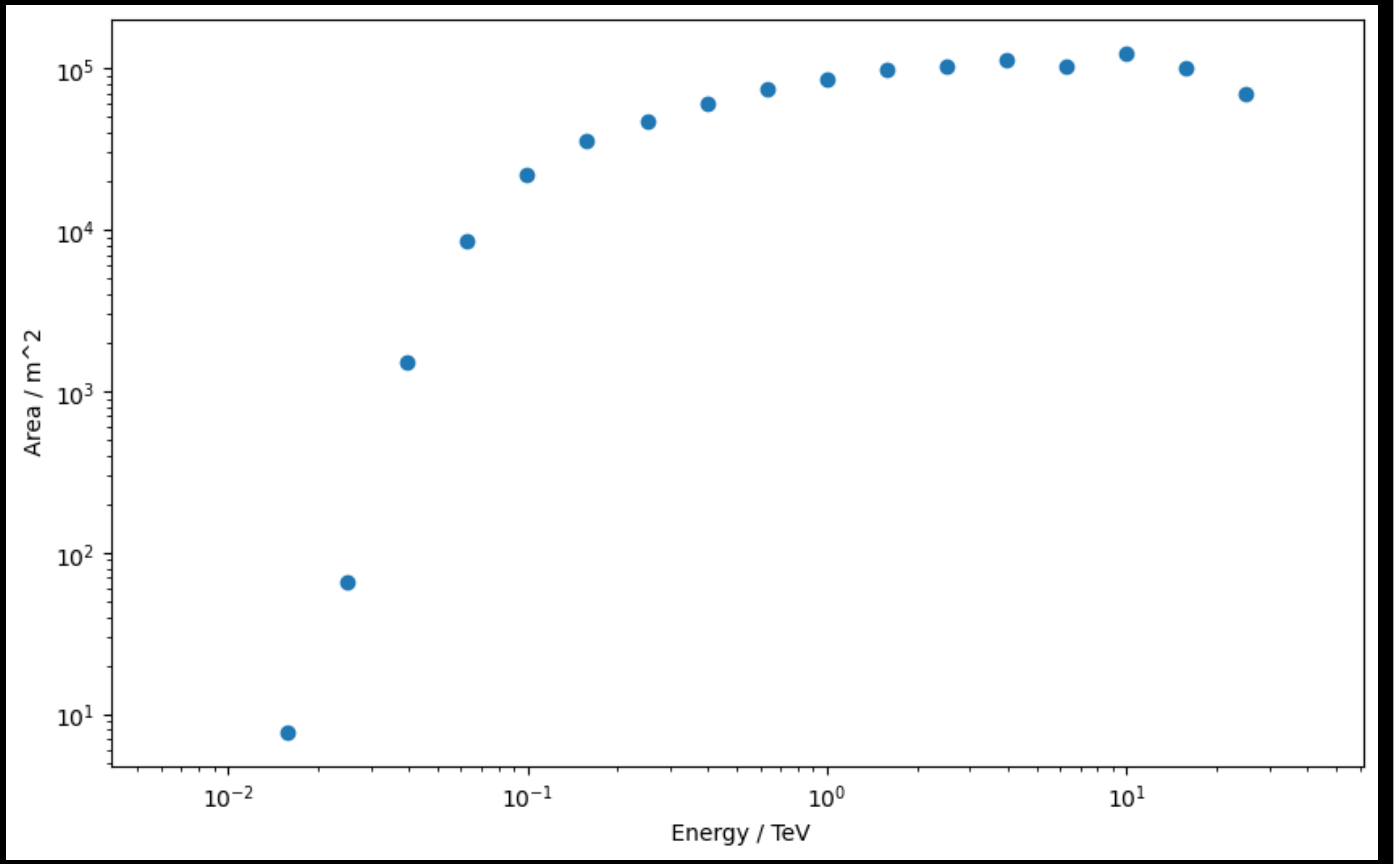
Collection area in your MC

Remaining γ after trigger, quality and selection cuts

$$\boxed{A_{\gamma,MC} \times \frac{N_{\gamma}^{\text{Survived}}}{N_{\gamma}^{\text{Total}}}}$$

Total γ simulated

Collection Area of the telescope (per each energy bin)

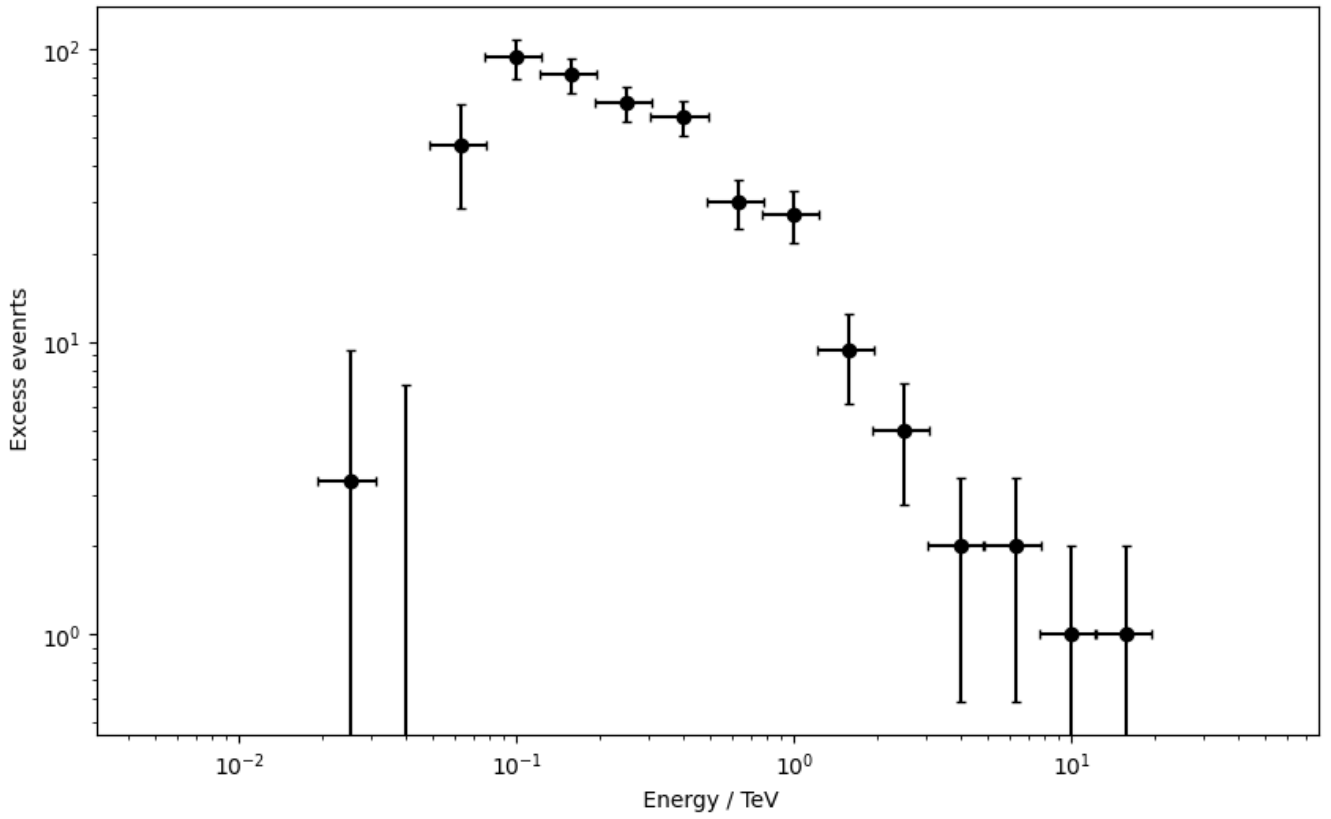


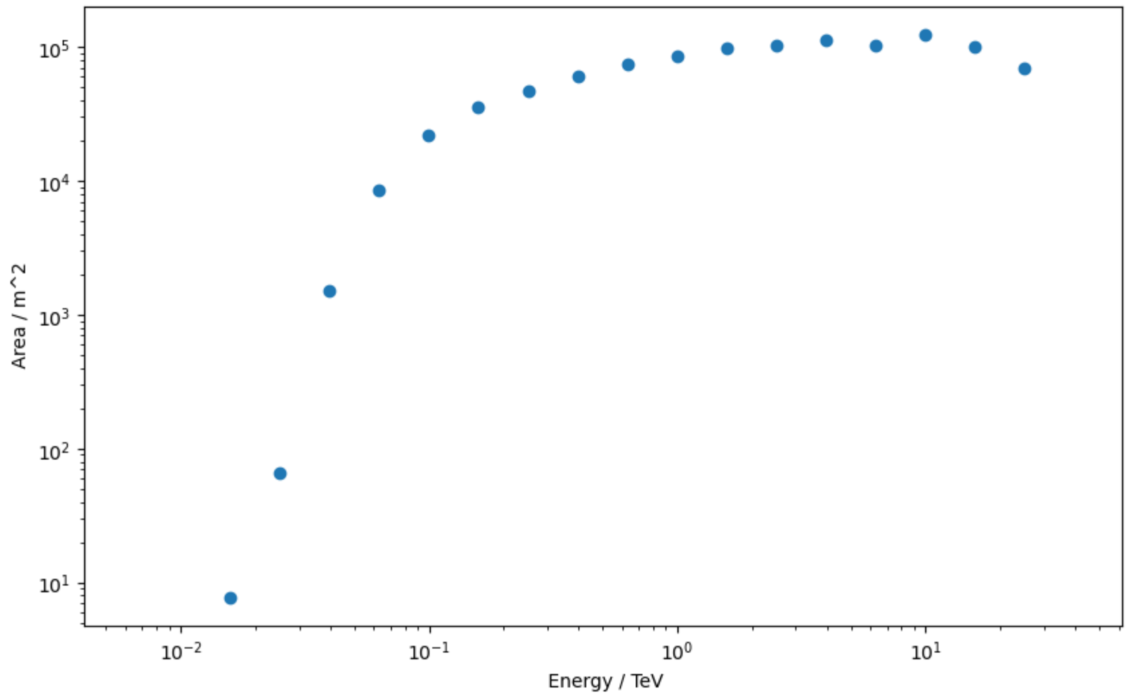
The effective area changes for different observations as it depends on many factors: Zenith, Azimuth, atm. conditions, the γ direction, etc…

Flux estimation

At a first approximation (there are some caveats that will be discussed later)

$$\frac{d\phi}{dE} \approx \frac{N_\gamma}{A_{eff}\, t_{eff}\, \Delta E} =$$



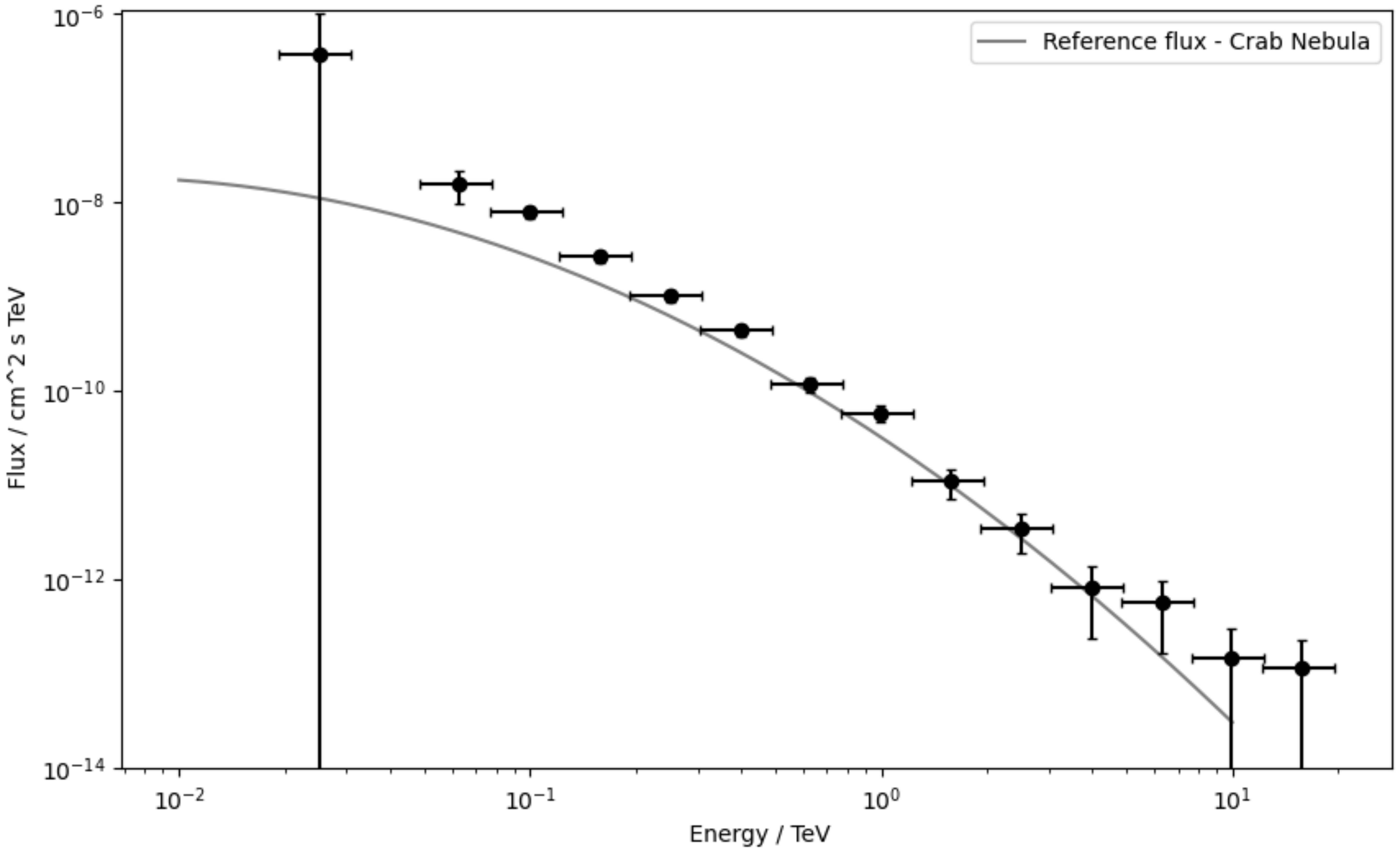


$$\times\, t_{eff}\, \Delta E$$

Flux estimation

At a first approximation (there are some caveats that will be discussed later)

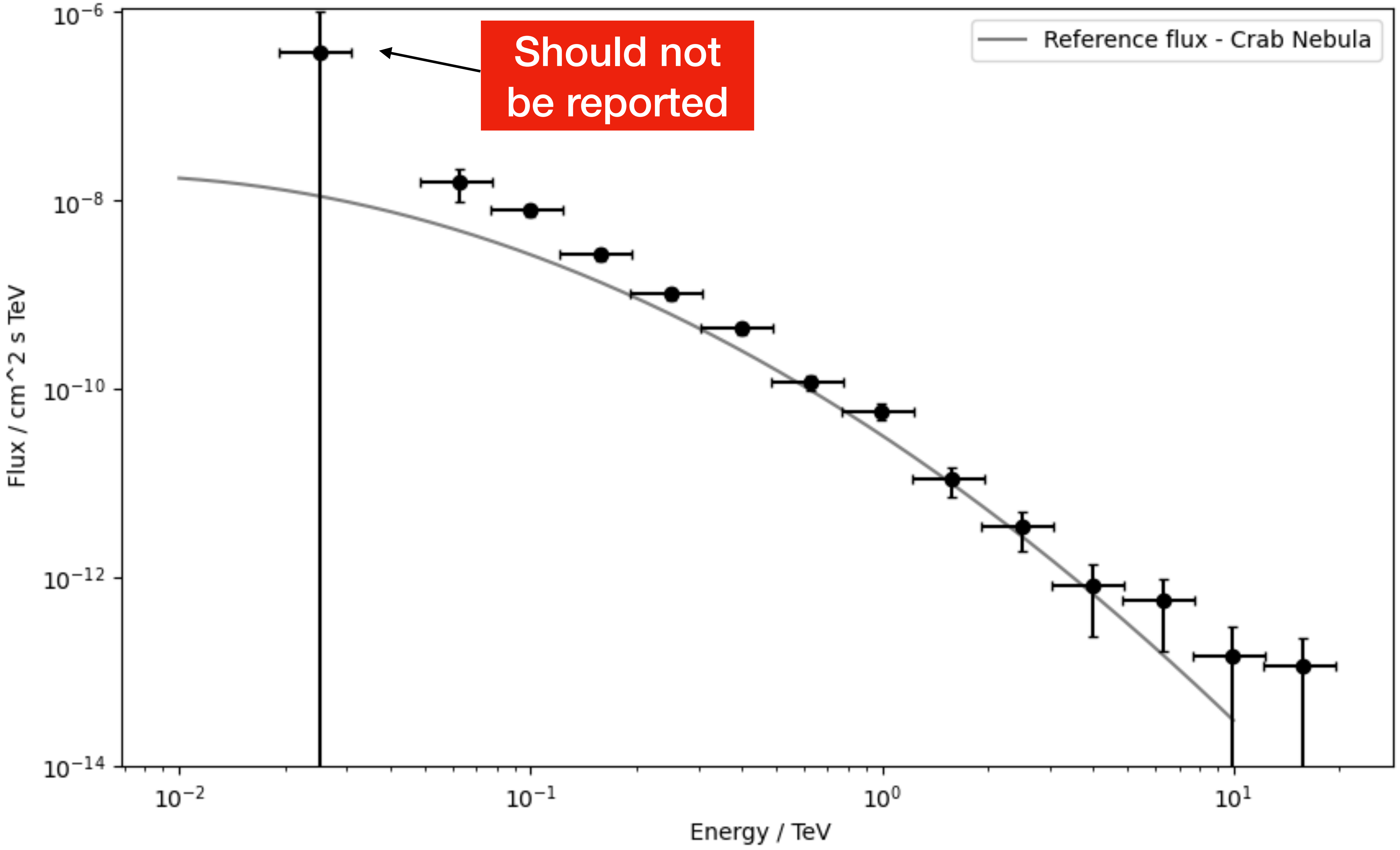
$$\frac{d\phi}{dE} \approx \frac{N_\gamma}{A_{eff}\, t_{eff}\, \Delta E} =$$



Flux estimation

At a first approximation (there are some caveats that will be discussed later)

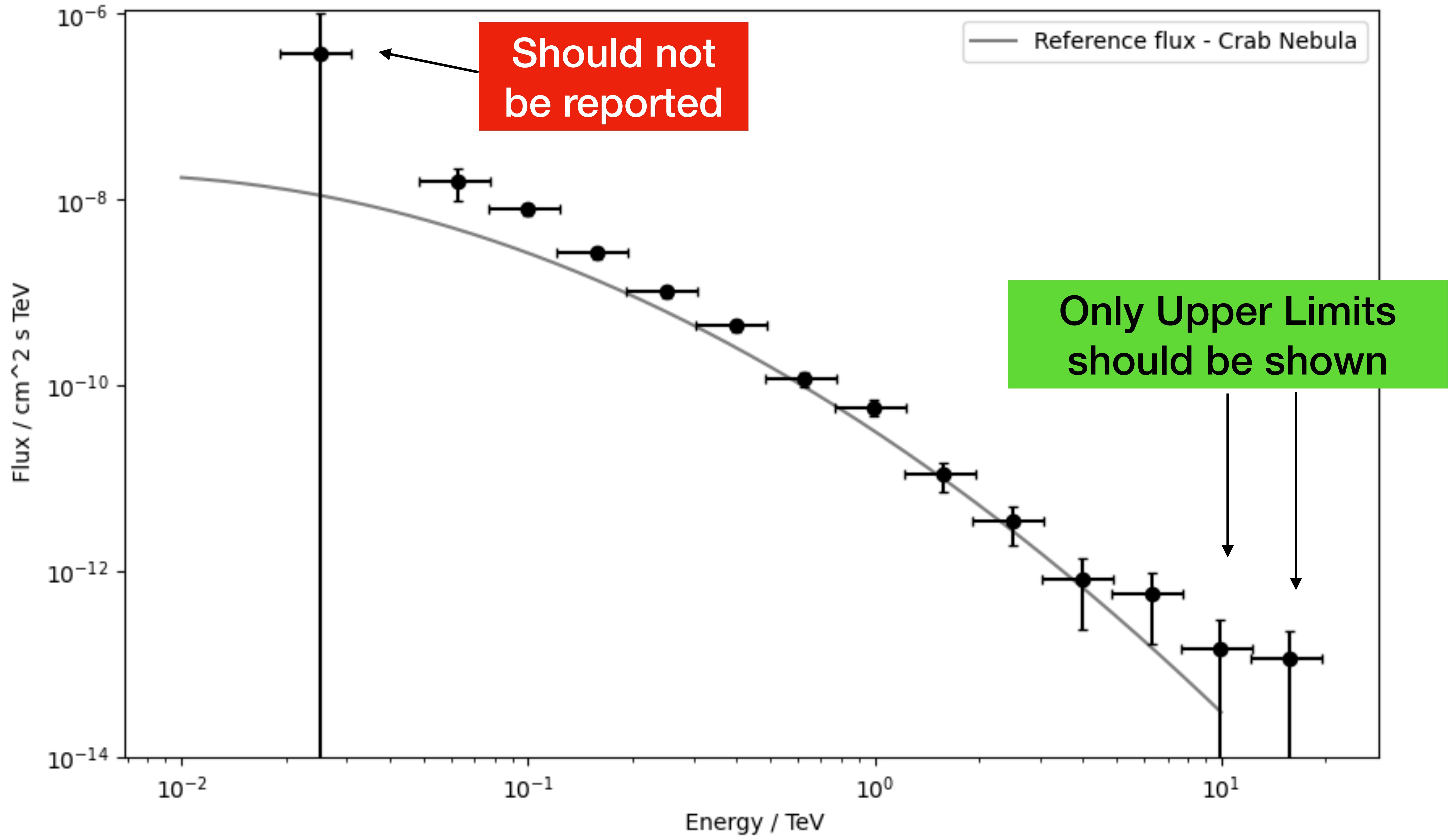
$$\frac{d\phi}{dE} \approx \frac{N_\gamma}{A_{eff}\, t_{eff}\, \Delta E} =$$



Flux estimation

At a first approximation (there are some caveats that will be discussed later)

$$\frac{d\phi}{dE} \approx \frac{N_\gamma}{A_{eff}\, t_{eff}\, \Delta E} =$$



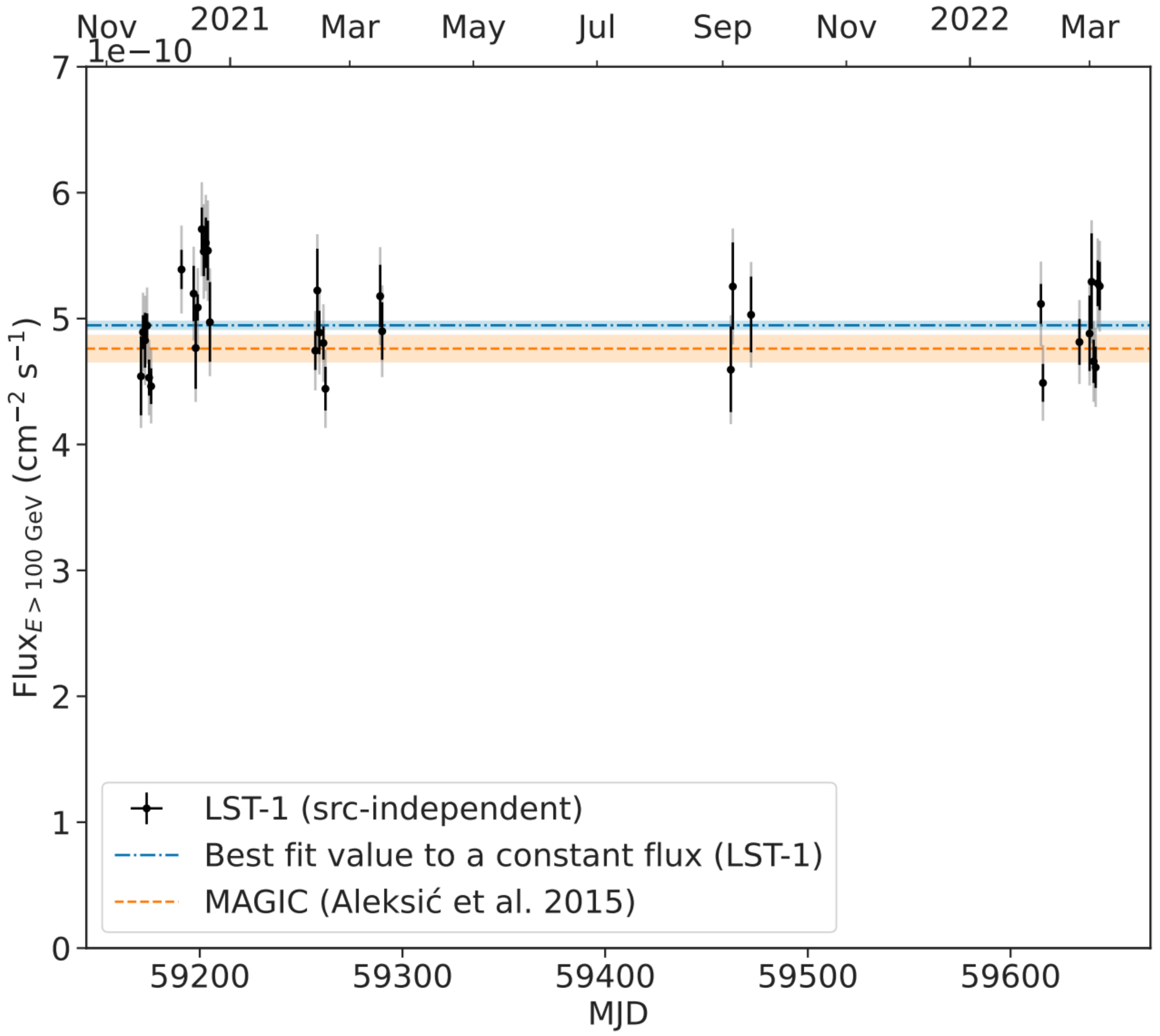
Should not be reported

Only Upper Limits should be shown

The Light Curve

The Light Curve (LC) is simply the **time evolution of the integral flux**

$$LC(t_i) = \int_{E_{th}} dE \, \frac{d\phi(t_i)}{dE} \quad \text{with } t_0, \ldots, t_i, t_{i+1}, \ldots$$



Example from arxiv 2306.12960:

- Source: Crab Nebula
- Instrument: LST-1
- $E_{th} = 100$ GeV

Flux estimation - Caveats

The energy we measure is not the true energy of the γ ray $E_{true} \neq E_{est}$

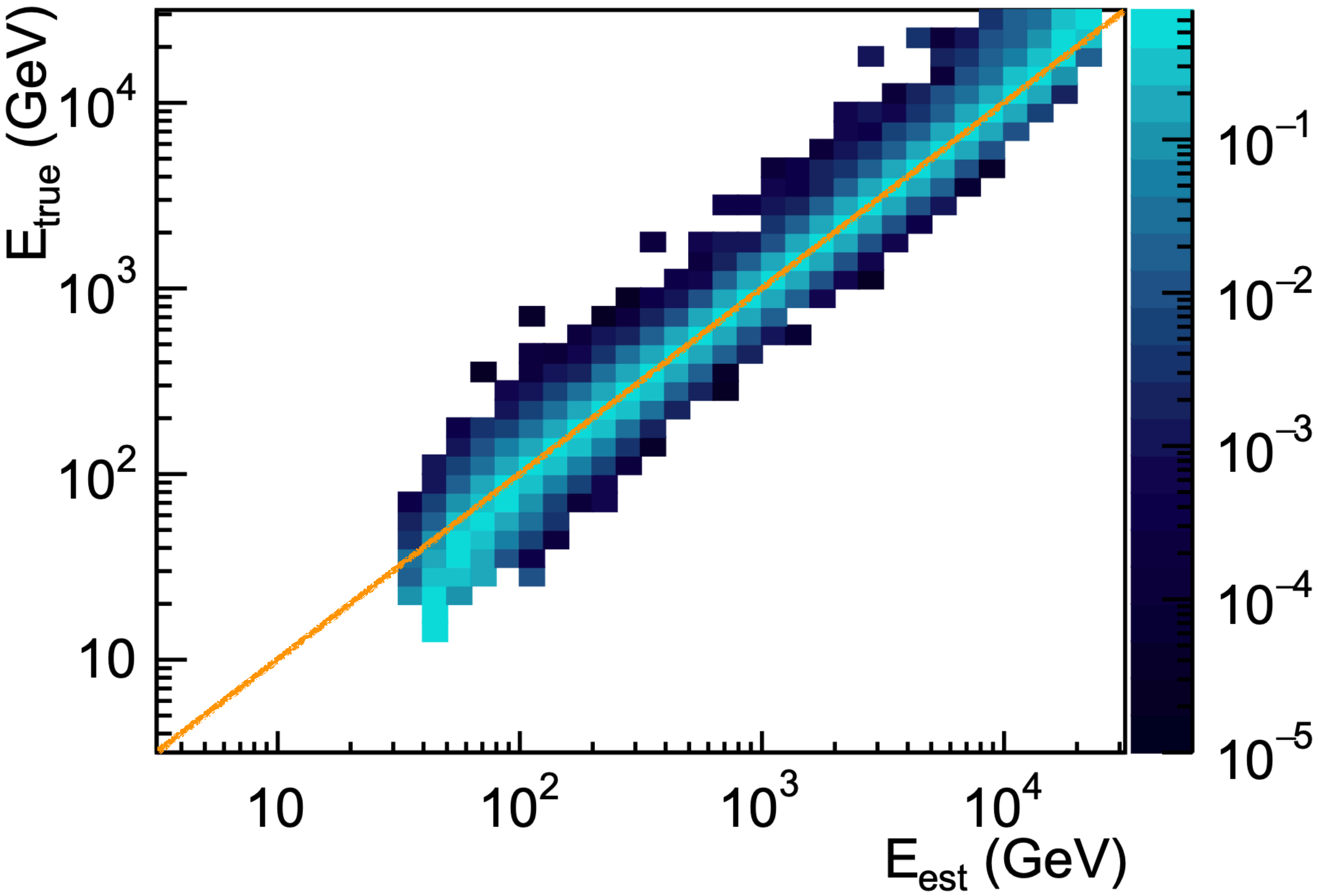


Figure adapted from Ishio, K., & Paneque, D. (2024). *Astroparticle Physics*, *158*, 102937.

Flux estimation - Caveats

The energy we measure is not the true energy of the γ ray

$$E_{true} \neq E_{est}$$

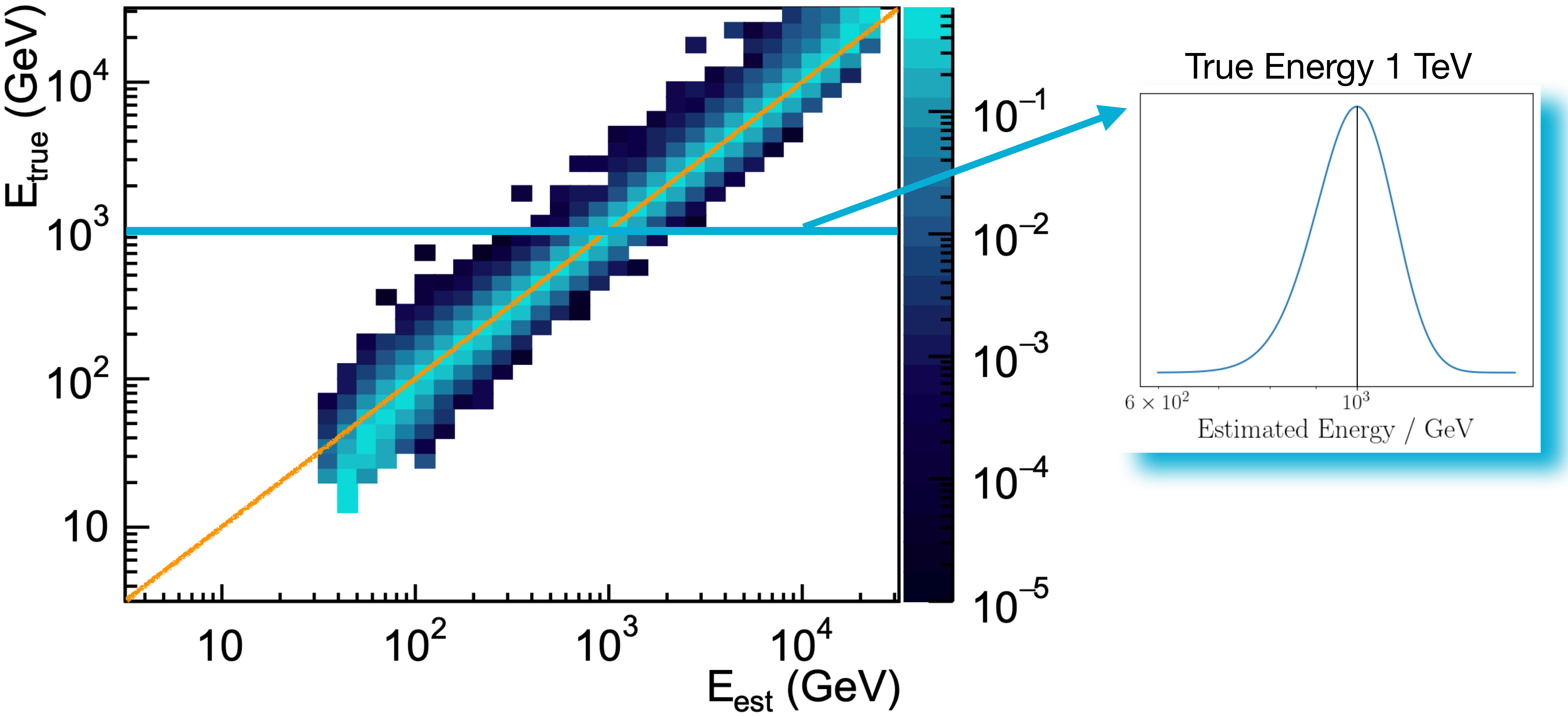
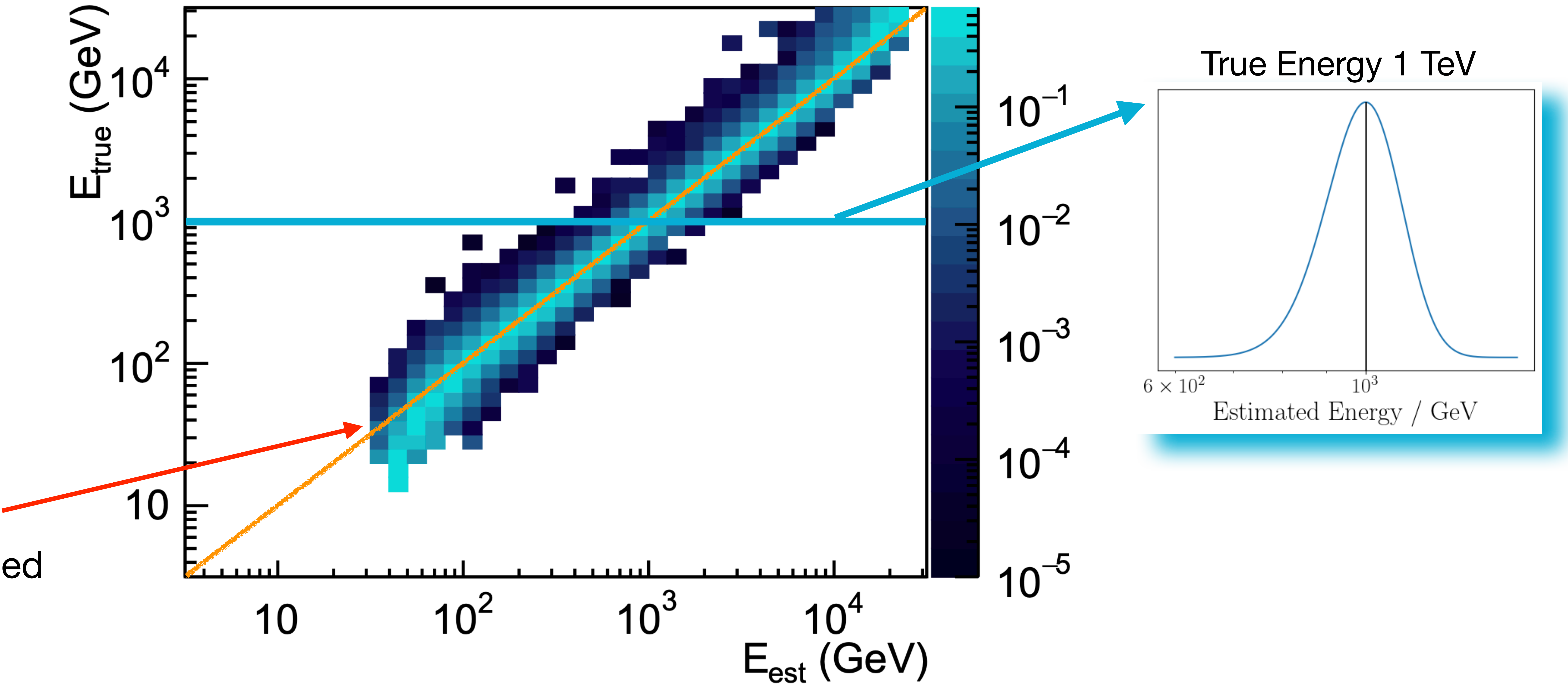


Figure adapted from Ishio, K., & Paneque, D. (2024). *Astroparticle Physics*, *158*, 102937.

Flux estimation - Caveats

The energy we measure is not the true energy of the γ ray $E_{true} \neq E_{est}$



Events at low energy tend to be reconstructed with higher energy

Figure adapted from Ishio, K., & Paneque, D. (2024). *Astroparticle Physics*, *158*, 102937.

Flux estimation - Unfolding

We can look at the problem as a ‘*geometrical*’ problem:

Estimated energy

Differeantial energy flux

Energy dispersion

$$s_i = \int_{\Delta E_i} dE_{est} \int dE_{true} \frac{d\phi(E_{true})}{dE_{true}} \; A_{eff}(E_{true}) \; D(E_{est} \,|\, E_{true}) \; t_{eff}$$

Expected γ counts in the i-th bin in **estimated** energy

True energy

Effective area

Effective time

Flux estimation - Unfolding

We can look at the problem as a '*geometrical*' problem:

$$\int_{a_0}^{a_N} = \int_{a_0}^{a_1} + \ldots + \int_{a_{N-1}}^{a_N}$$

$$s_i = \sum_j \int_{\Delta E_i} dE_{est} \int_{\Delta E_j} dE_{true} \frac{d\phi(E_{true})}{dE_{true}} \; A_{eff}(E_{true}) \; D(E_{est} \,|\, E_{true}) \; t_{eff}$$

Flux estimation - Unfolding

We can look at the problem as a '*geometrical*' problem:

$$s_i = \sum_j \int_{\Delta E_i} dE_{est} \int_{\Delta E_j} dE_{true} \frac{d\phi(E_{true})}{dE_{true}} \, A_{eff}(E_{true}) \, D(E_{est} \,|\, E_{true}) \, t_{eff}$$

$$= \sum_j \phi_j \, A_j \, D_{ij} \, t_{eff} \equiv \sum_j M_{ij} \, \phi_j$$

average flux and collection area in the j-th true energy bin

probability that a γ in the j-th bin is assigned the i-th bin

Flux estimation - Unfolding

We can look at the problem as a '*geometrical*' problem:

For more details on the computation of the average flux and collection area, and of the matrix M_{ij} see for instance, Albert, et al (2007). *583*(2-3), 494-506.

$$s_i = \sum_j \int_{\Delta E_i} dE_{est} \int_{\Delta E_j} dE_{true} \frac{d\phi(E_{true})}{dE_{true}} A_{eff}(E_{true})\, D(E_{est}\,|\,E_{true})\, t_{eff}$$

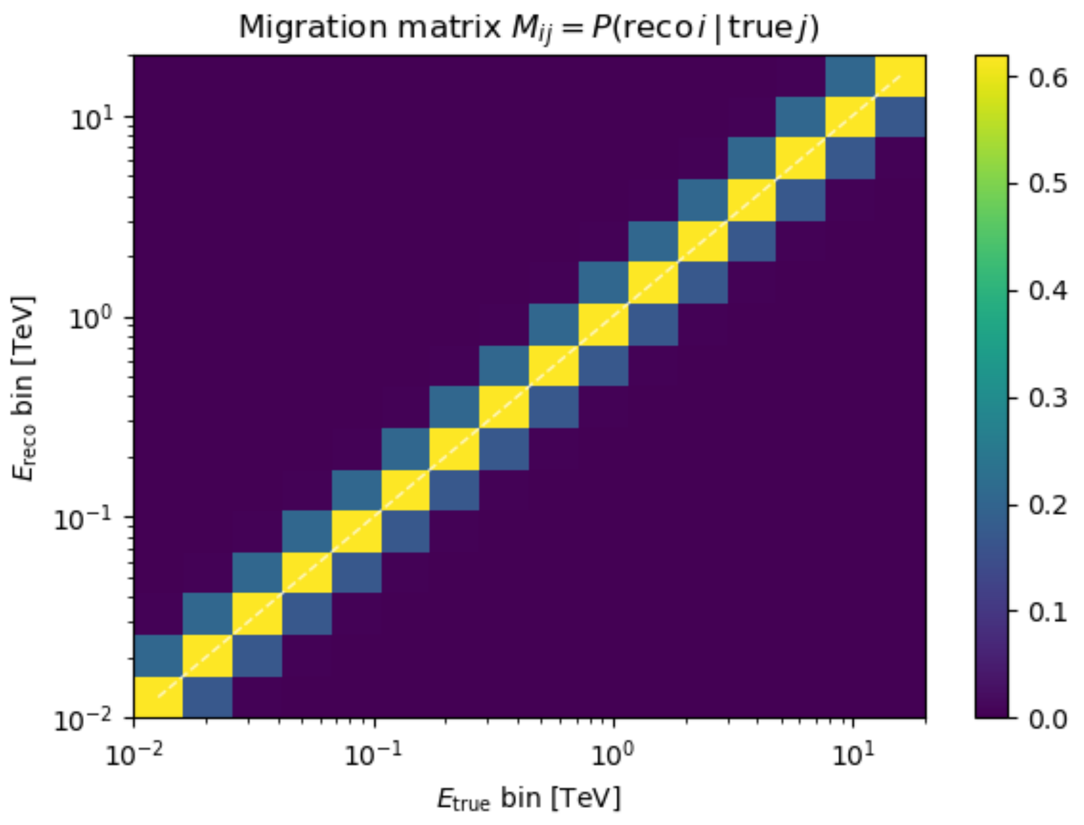
$$= \sum_j \phi_j\, A_j\, D_{ij}\, t_{eff} \equiv \sum_j M_{ij}\, \phi_j$$

average flux and collection area in the j-th true energy bin

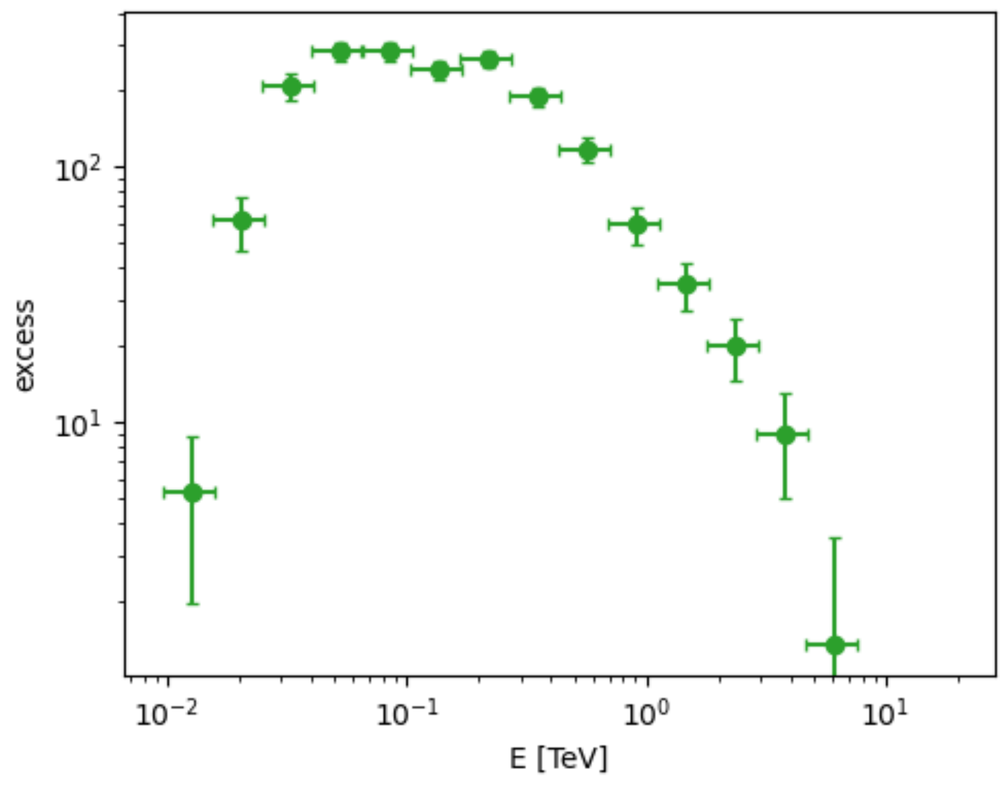
probability that a γ in the j-th bin is assigned the i-th bin

Flux estimation - Unfolding

We can look at the problem as a '*geometrical*' problem:



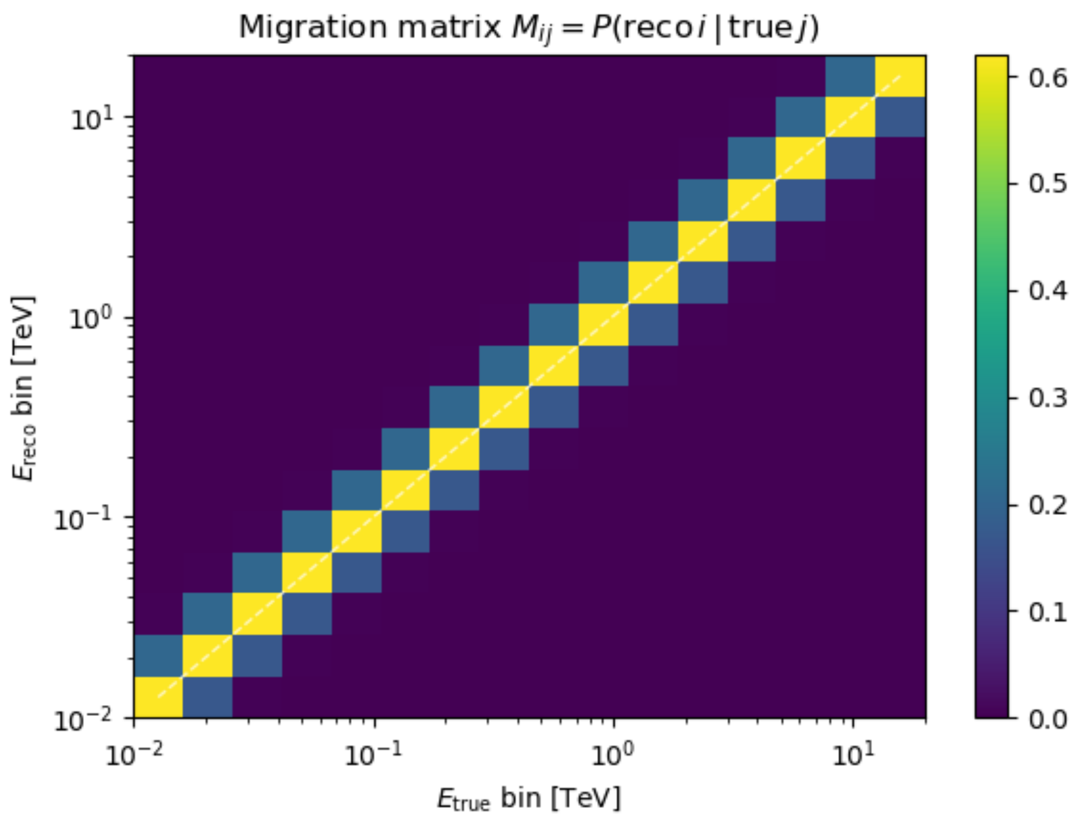
X ? =



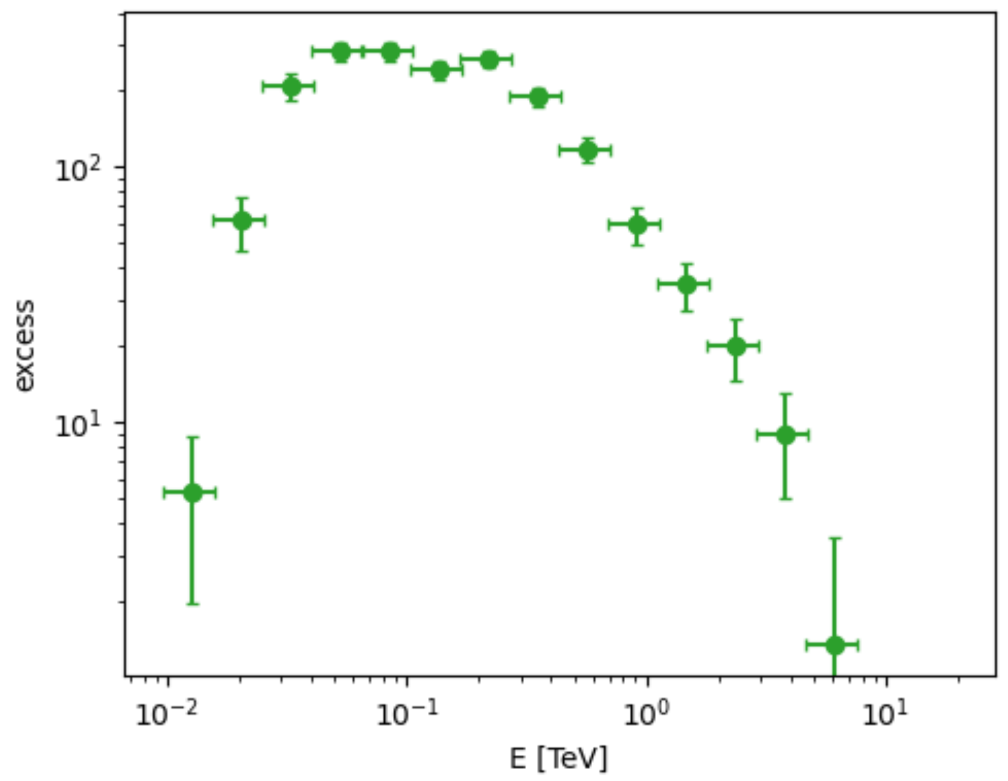
The goal of the unfolding procedure is to find a solution to this equation by inverting the matrix M_{ij}

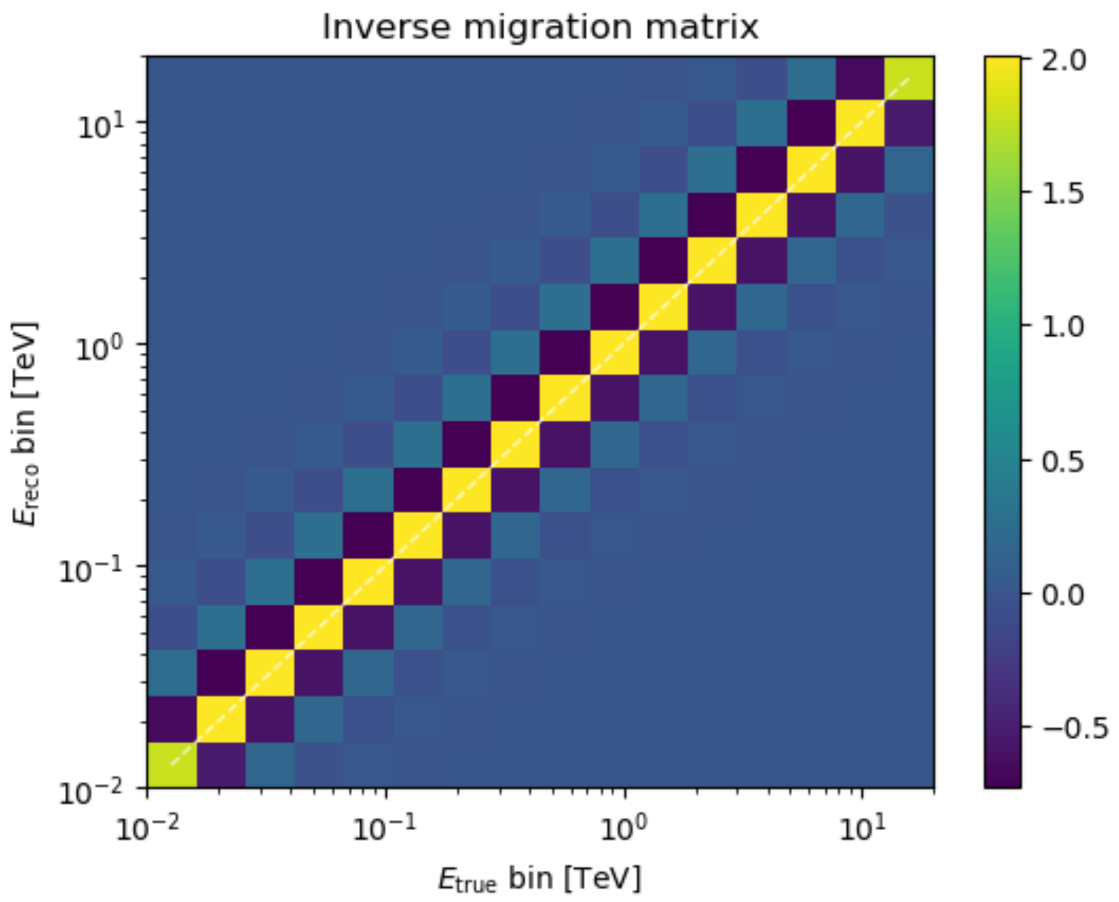
Flux estimation - Unfolding

We can look at the problem as a ‘*geometrical*’ problem:



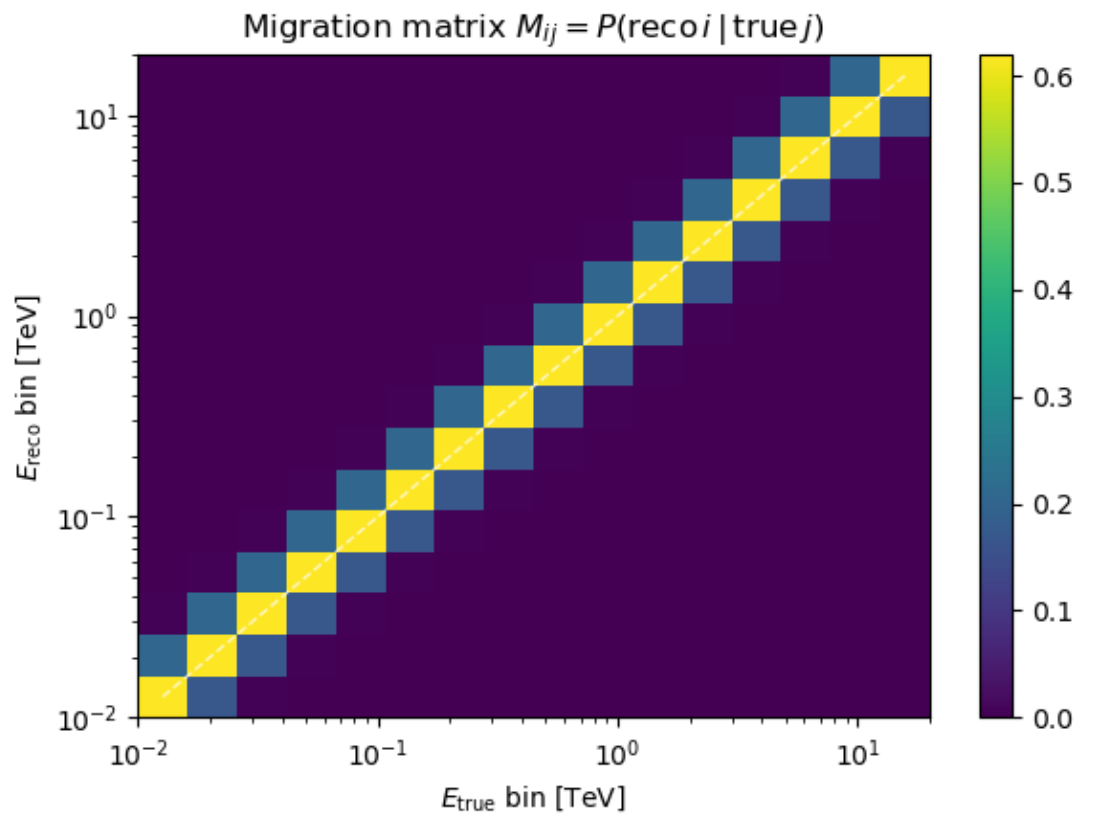
X ? =



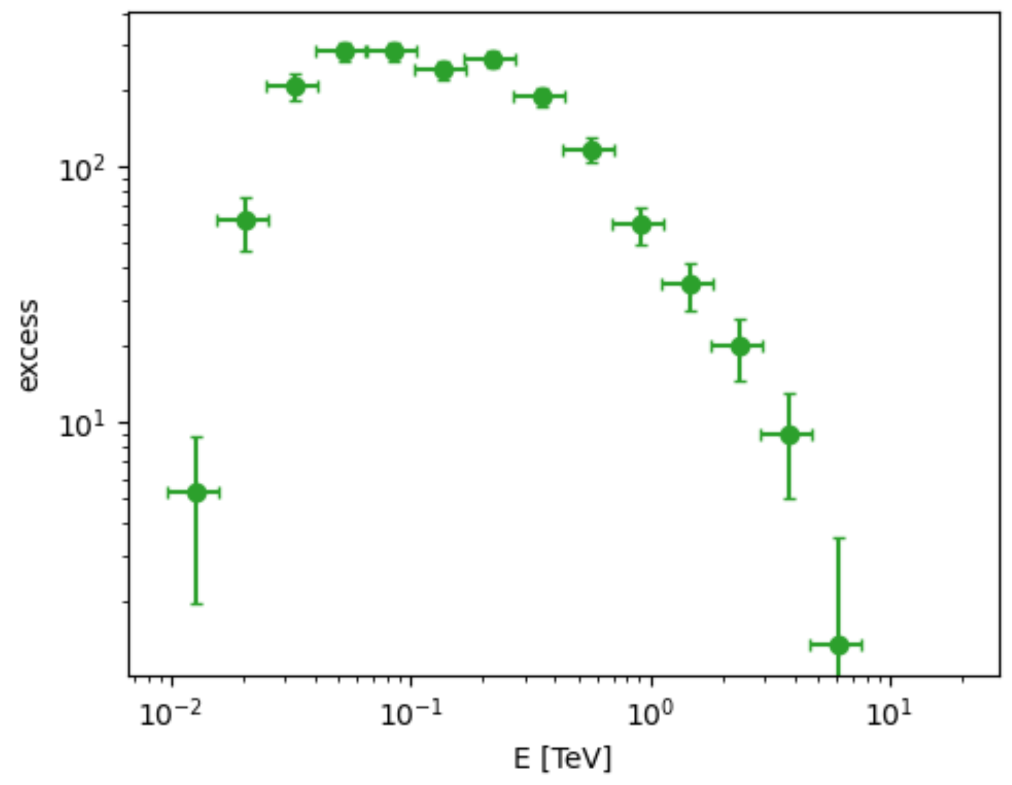


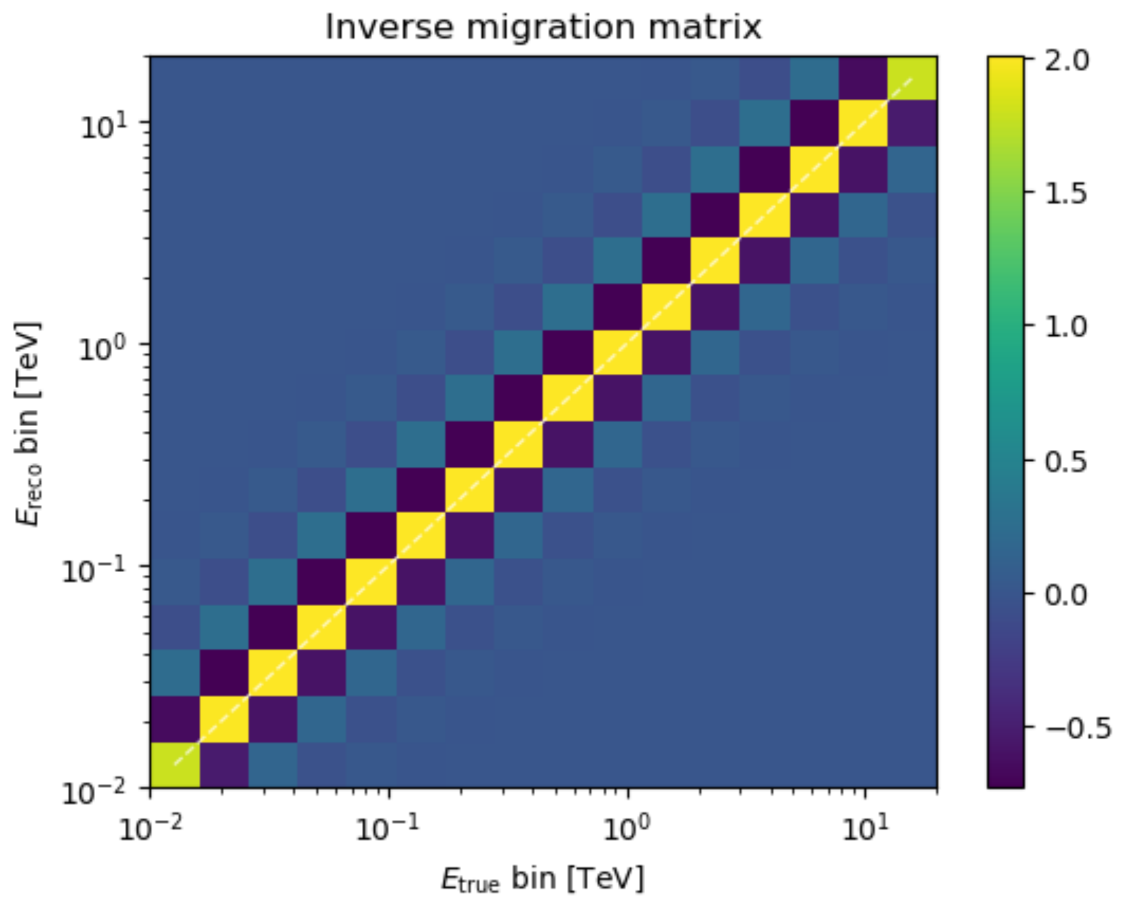
Flux estimation - Unfolding

We can look at the problem as a ‘*geometrical*’ problem:

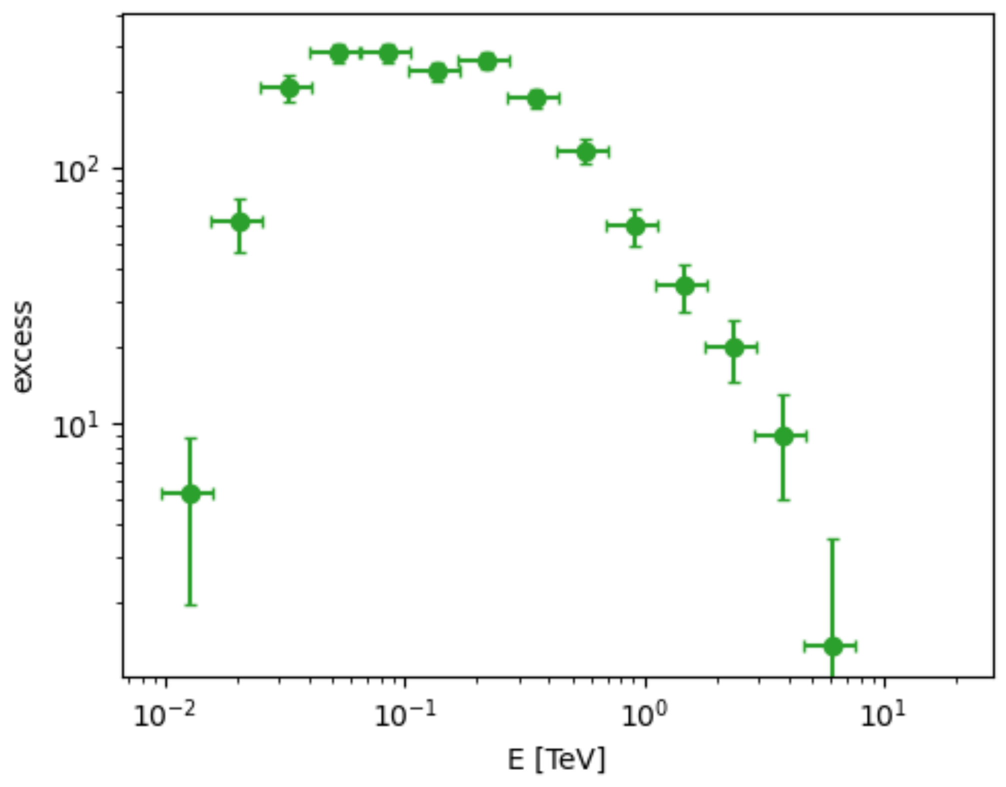


X ? =

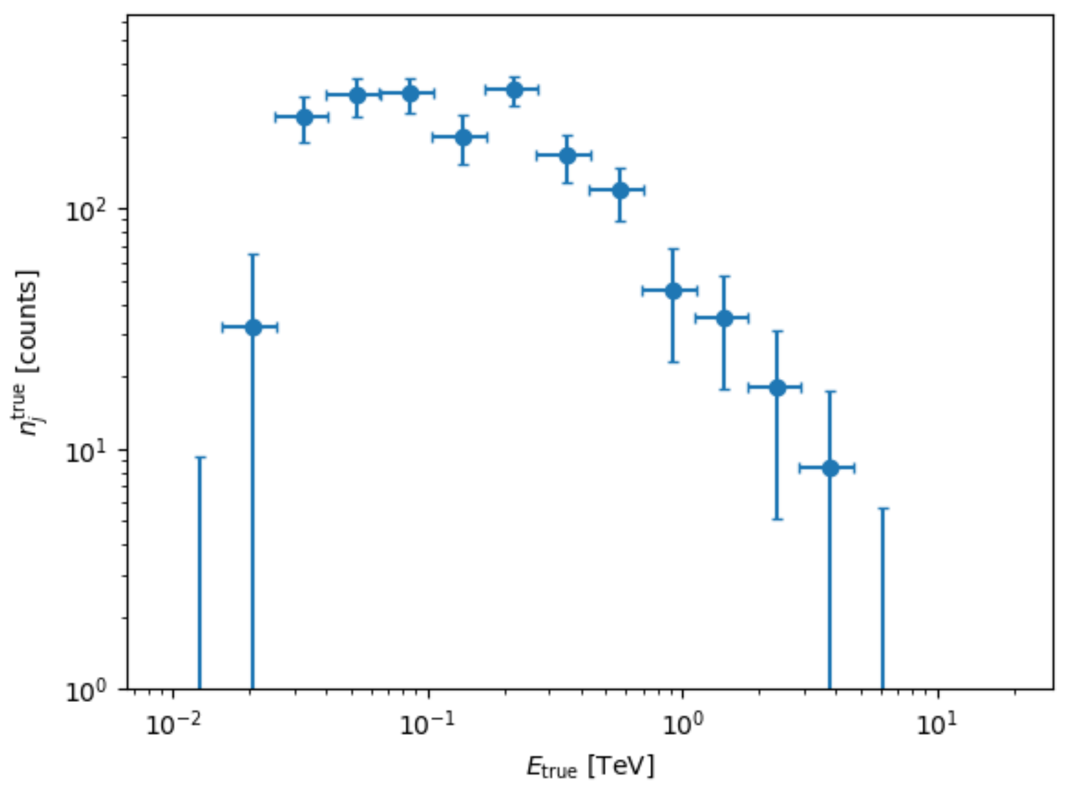




X



=



Flux estimation - Unfolding

We can look at the problem as a ‘*geometrical*’ problem:

$$s_i = \sum_j \phi_j \, A_j \, D_{ij} \, t_{eff} \equiv \sum_j M_{ij} \, \phi_j \qquad \longrightarrow \qquad \phi_j = \sum_i M^{-1}_{ji} \, s_i$$

It is an ***ill-posed*** problem

Well-posed problem:

1. Existence of a solution
2. Uniqueness
3. ~~Stability under small perturbations~~

Flux estimation - Unfolding

We can look at the problem as a ‘*geometrical*’ problem:

$$s_i = \sum_j \phi_j \, A_j \, D_{ij} \, t_{eff} \equiv \sum_j M_{ij} \, \phi_j \longrightarrow \phi_j = \sum_i M^{-1}_{ji} \, s_i$$

It is an ***ill-posed*** problem

Tikhonov

Tikhonov, A.N. 1963; Volume 151, pp. 501–504.

Bertero

Bertero, M. Adv. Electron. Electron Phys. 1989, 75, 1–120.

Schmelling

Schmelling, M. Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrometers Detect. Assoc. Equip. 1994, 340, 400–412

Bayesian unfolding

Choudalakis, G. Fully bayesian unfolding. arXiv 2012, arXiv:1201.4612.

and more …

Well-posed problem:

1. Existence of a solution
2. Uniqueness
3. ~~Stability under small perturbations~~

Forward folding

Unfolding:

Get the flux by making as few assumptions as possible on its shape

Forward folding:

Get the flux by assuming a (parametric) shape

$$\frac{d\phi(E,\theta)}{dE}$$

The goal of the forward folding is to estimate the **parameters** θ

Forward folding

$\mathscr{L}(\vec{\theta}) = P(D\,|\,\vec{\theta})$ probability of the data given a model with parameters $\vec{\theta} = (\theta_0, \ \dots \ , \theta_n)$

$$\mathscr{L}(\vec{\theta}) = \sum_i \left(\frac{\left(s_i + \alpha b_i\right)^{n_i}}{n_i!} e^{-(s_i+\alpha b_i)} \times \frac{b_i^{m_i}}{m_i!} e^{b_i} \right)$$

with

$$s_i = \int_{\Delta E_i} dE_{est} \int dE_{true} \frac{d\phi(E_{true}, \vec{\theta})}{dE_{true}} \ A_{eff}(E_{true}) \ D(E_{est}\,|\,E_{true}) \ t_{eff}$$

Forward folding

$\mathscr{L}(\vec{\theta}) = P(D\,|\,\vec{\theta})$ probability of the data given a model with parameters $\vec{\theta} = (\theta_0, \ldots, \theta_n)$

$$\mathscr{L}(\vec{\theta}) = \sum_i \left(\frac{\left(s_i + \alpha b_i\right)^{n_i}}{n_i!} e^{-(s_i + \alpha b_i)} \times \frac{b_i^{m_i}}{m_i!} e^{b_i} \right)$$

with

$$s_i = \int_{\Delta E_i} dE_{est} \int dE_{true} \frac{d\phi(E_{true}, \vec{\theta})}{dE_{true}} \, A_{eff}(E_{true}) \, D(E_{est}\,|\,E_{true}) \, t_{eff}$$

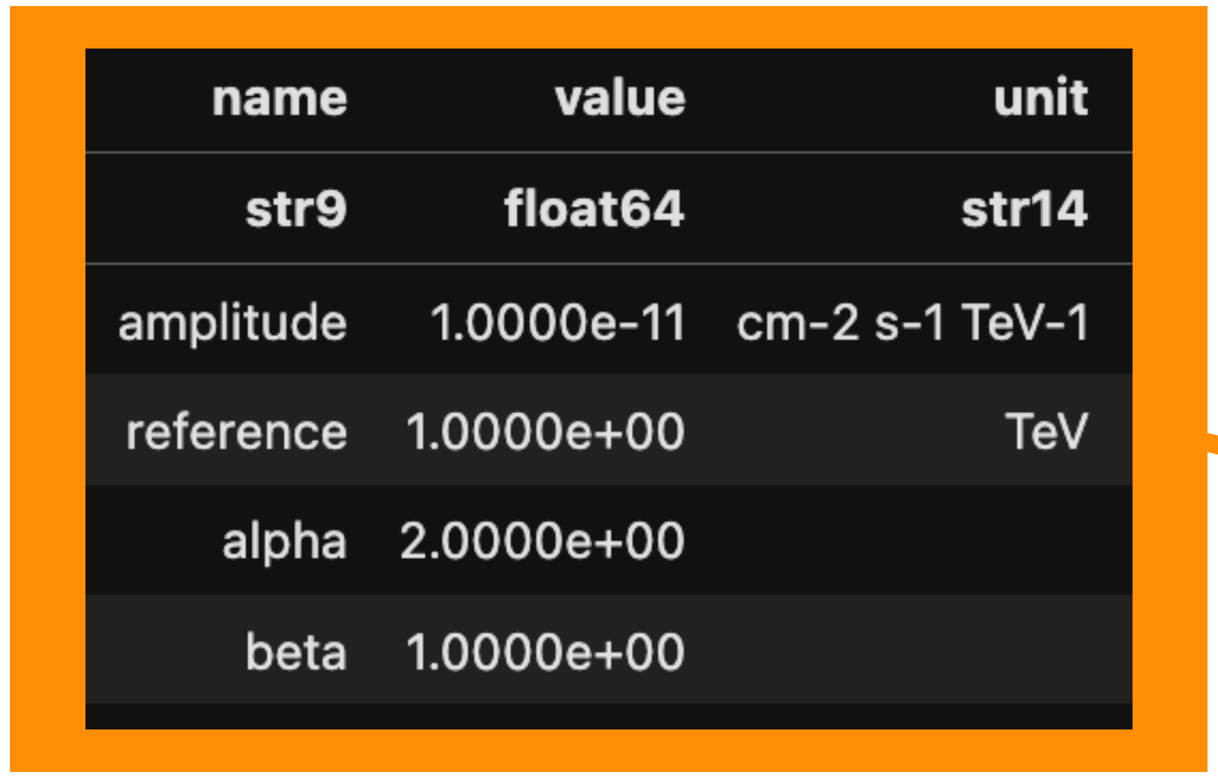
$$\Delta TS \equiv -2 \log \frac{\mathscr{L}(\vec{\theta})}{\mathscr{L}(\hat{\theta})} \sim \chi^2$$

Values of $\vec{\theta}$ that maximize the likelihood

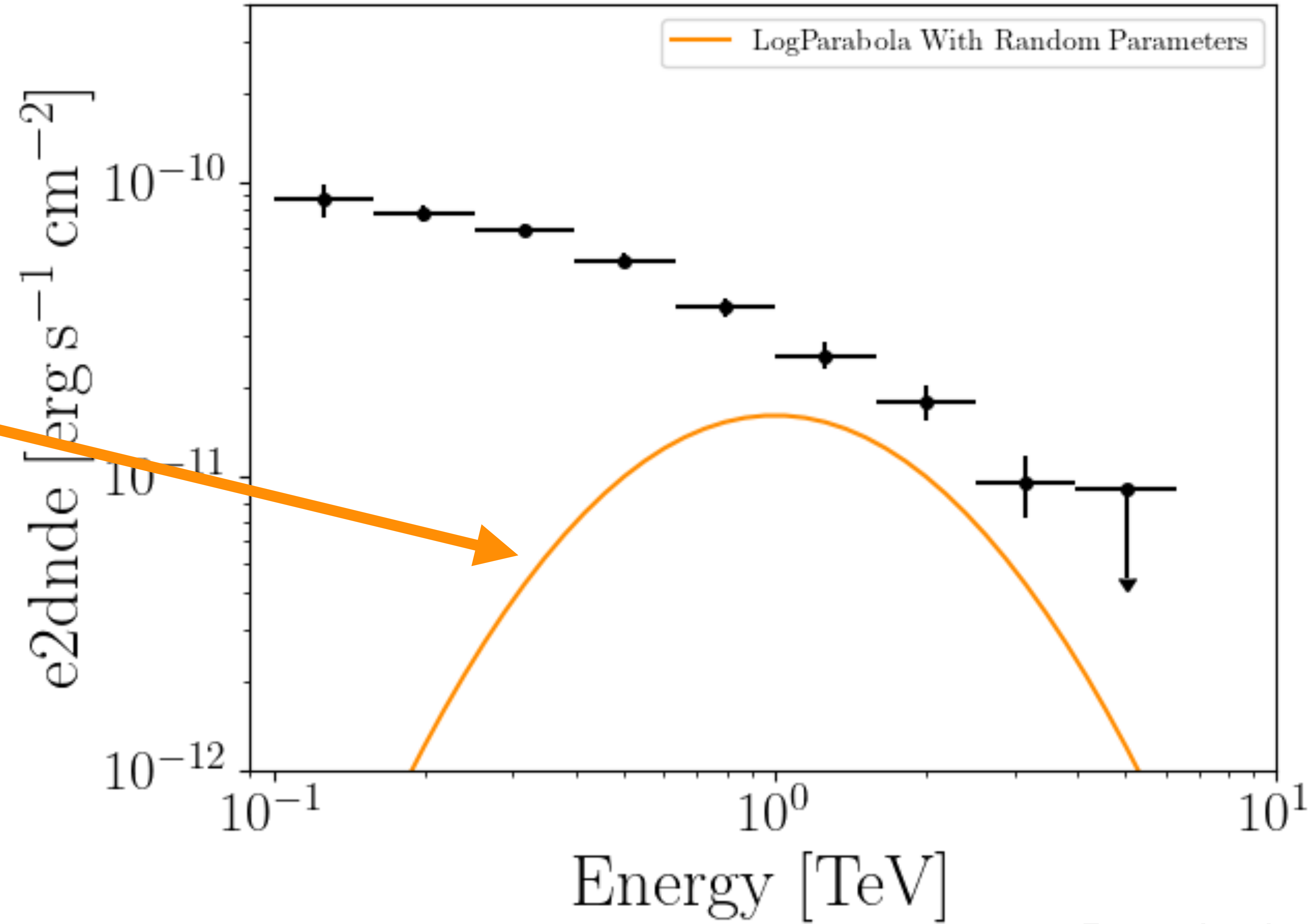
Forward folding

Example with Log Parabola model

$$\frac{d\phi(E)}{dE} = \phi_0 \left(\frac{E}{E_0}\right)^{-\alpha-\beta\log\left(\frac{E}{E_0}\right)}$$



name	value	unit
str9	float64	str14
amplitude	1.0000e-11	cm-2 s-1 TeV-1
reference	1.0000e+00	TeV
alpha	2.0000e+00	
beta	1.0000e+00	



Plot obtained with gammapy

Forward folding

Example with Log Parabola model

$$\frac{d\phi(E)}{dE} = \phi_0 \left(\frac{E}{E_0}\right)^{-\alpha-\beta\log\left(\frac{E}{E_0}\right)}$$

Values of α and ϕ_0 that maximize the likelihood for a fixed β

$$\Delta TS(\beta) \equiv -2\log\frac{\mathscr{L}(\beta,\hat{\alpha},\hat{\phi}_0)}{\hat{\mathscr{L}}}$$

Likelihood profile!

Maximum value of the likelihood over the entire parameter space

Forward folding

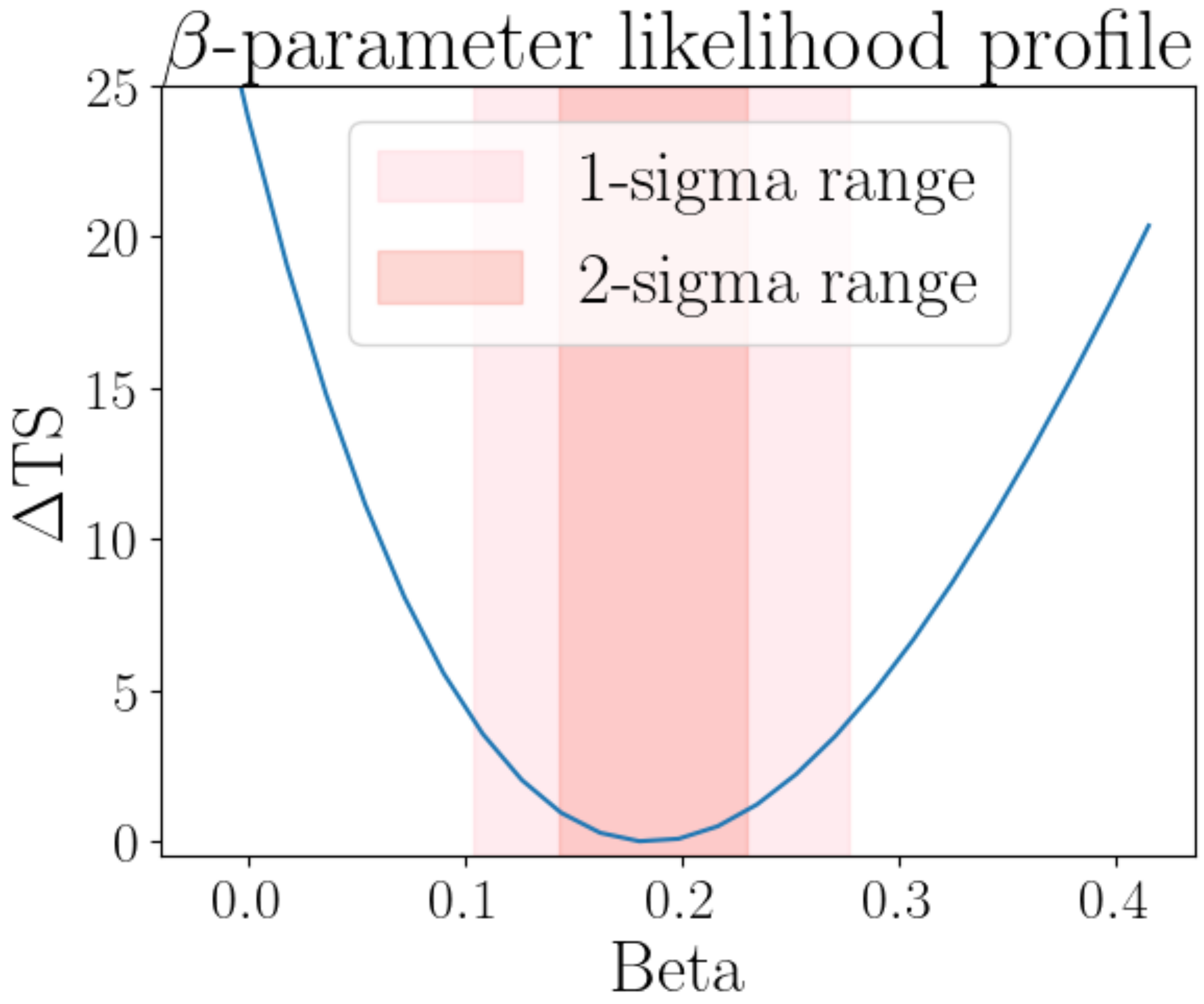
Example with Log Parabola model

$$\frac{d\phi(E)}{dE} = \phi_0 \left(\frac{E}{E_0} \right)^{-\alpha - \beta \log\left(\frac{E}{E_0}\right)}$$

Values of α and ϕ_0 that maximize the likelihood for a fixed β

$$\Delta TS(\beta) \equiv -2 \log \frac{\mathscr{L}(\beta, \hat{\alpha}, \hat{\phi}_0)}{\hat{\mathscr{L}}}$$

Maximum value of the likelihood over the entire parameter space



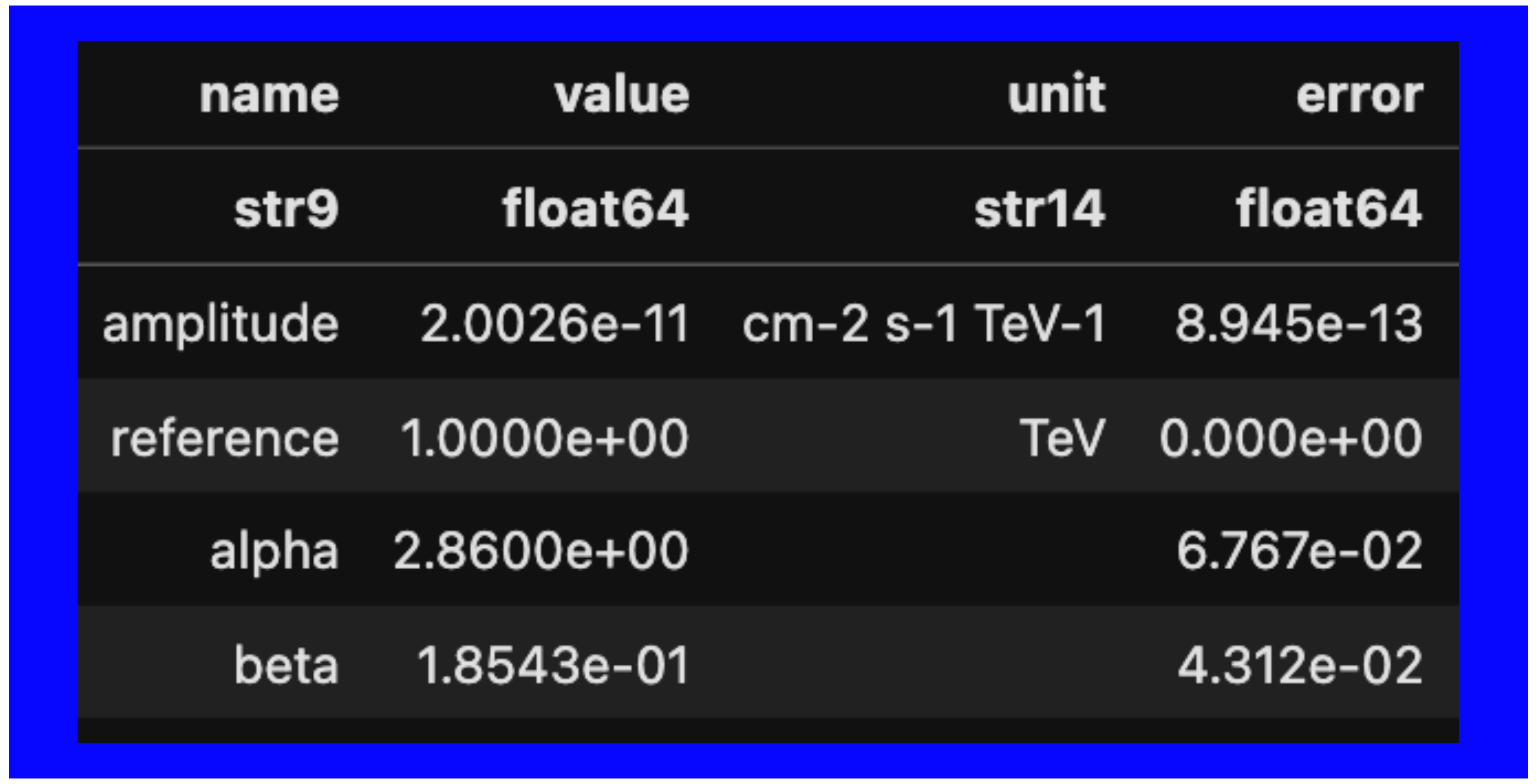
Plot obtained with gammapy

Forward folding

Example with Log Parabola model

$$\frac{d\phi(E)}{dE} = \phi_0 \left(\frac{E}{E_0}\right)^{-\alpha-\beta\log\left(\frac{E}{E_0}\right)}$$

Output of the optimization algorithm in gammapy



name	value	unit	error
str9	**float64**	**str14**	**float64**
amplitude	2.0026e-11	cm-2 s-1 TeV-1	8.945e-13
reference	1.0000e+00	TeV	0.000e+00
alpha	2.8600e+00		6.767e-02
beta	1.8543e-01		4.312e-02

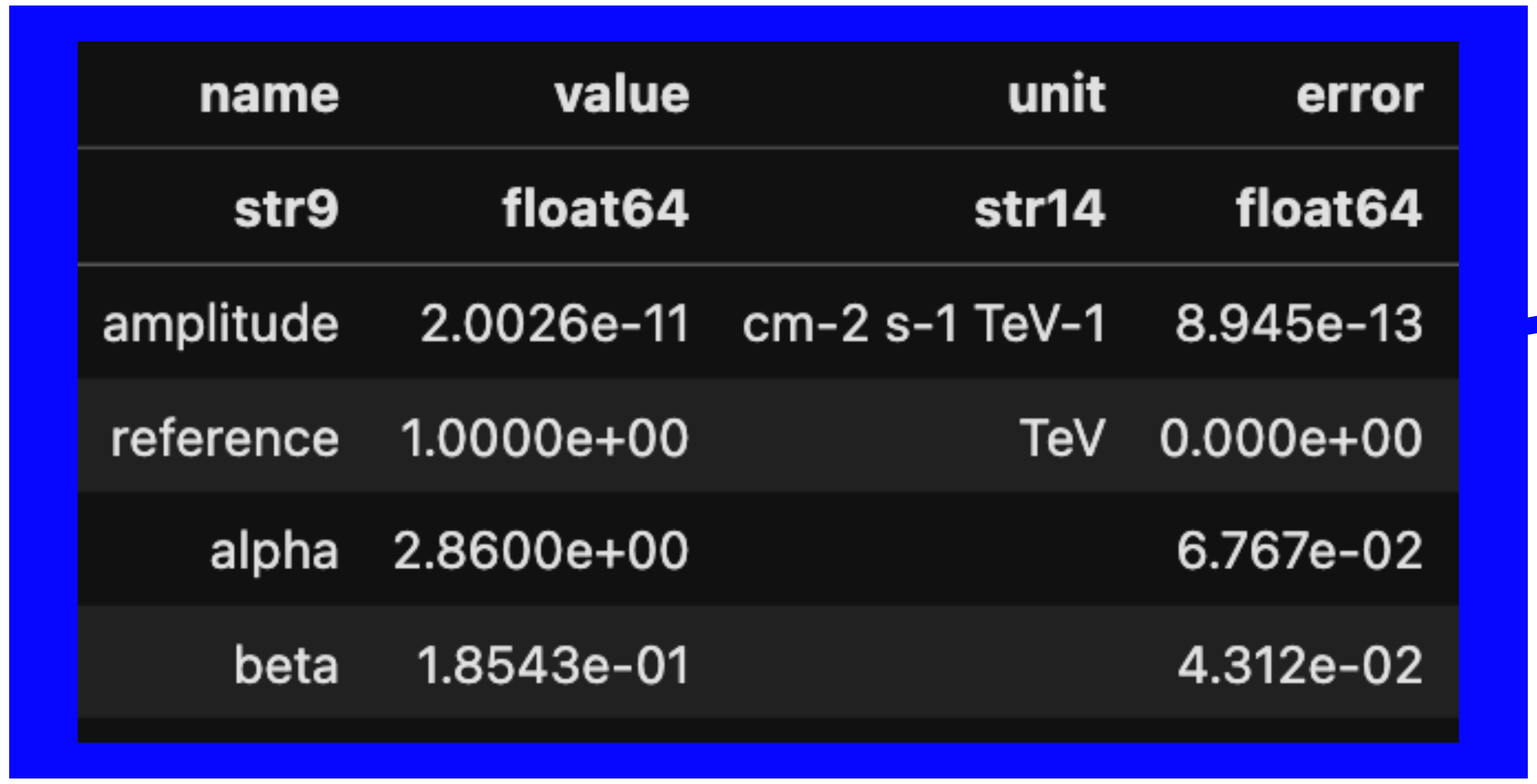
Plot obtained with gammapy

Forward folding

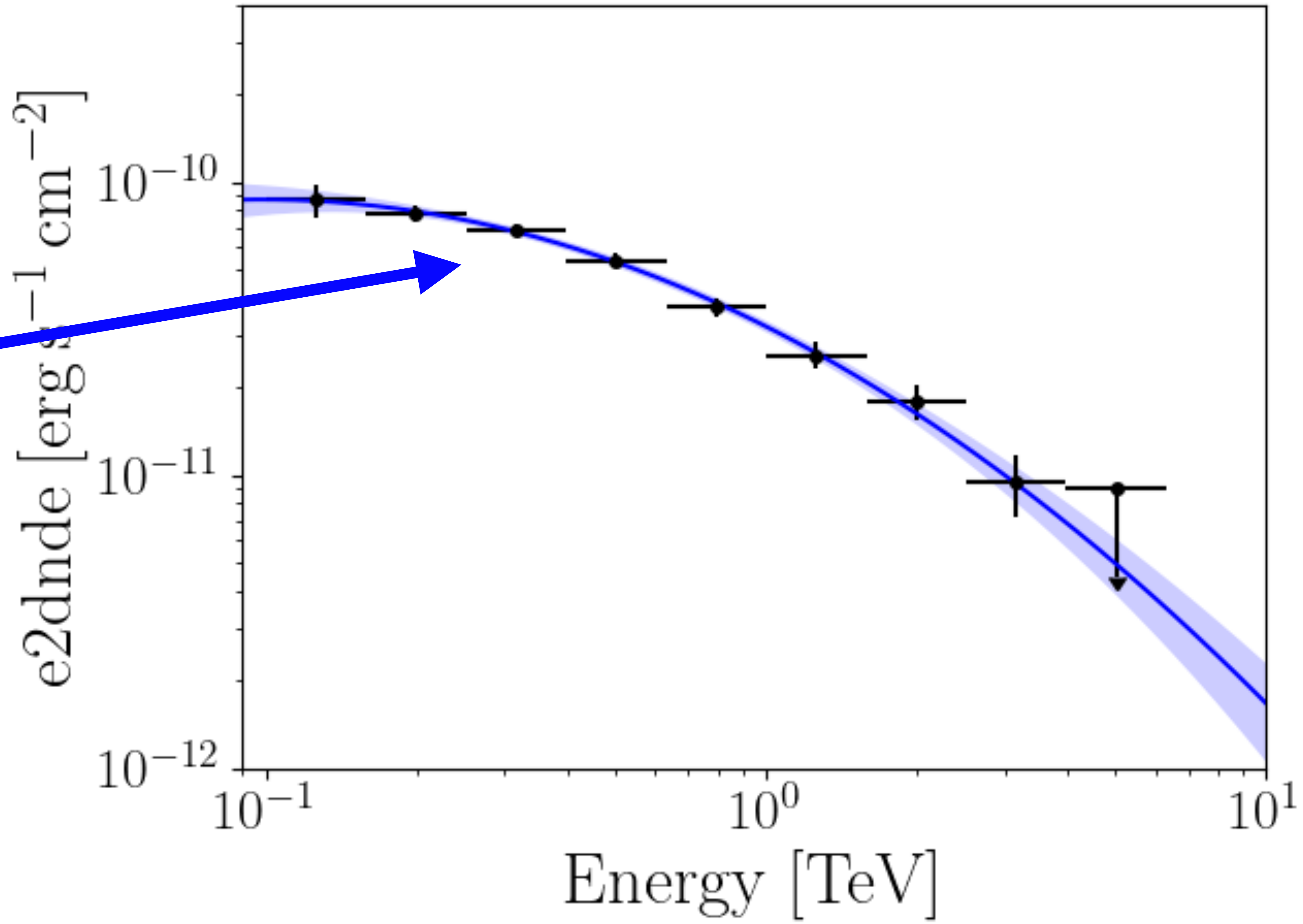
Example with Log Parabola model

$$\frac{d\phi(E)}{dE} = \phi_0 \left(\frac{E}{E_0}\right)^{-\alpha-\beta\log\left(\frac{E}{E_0}\right)}$$

Output of the optimization algorithm in gammapy



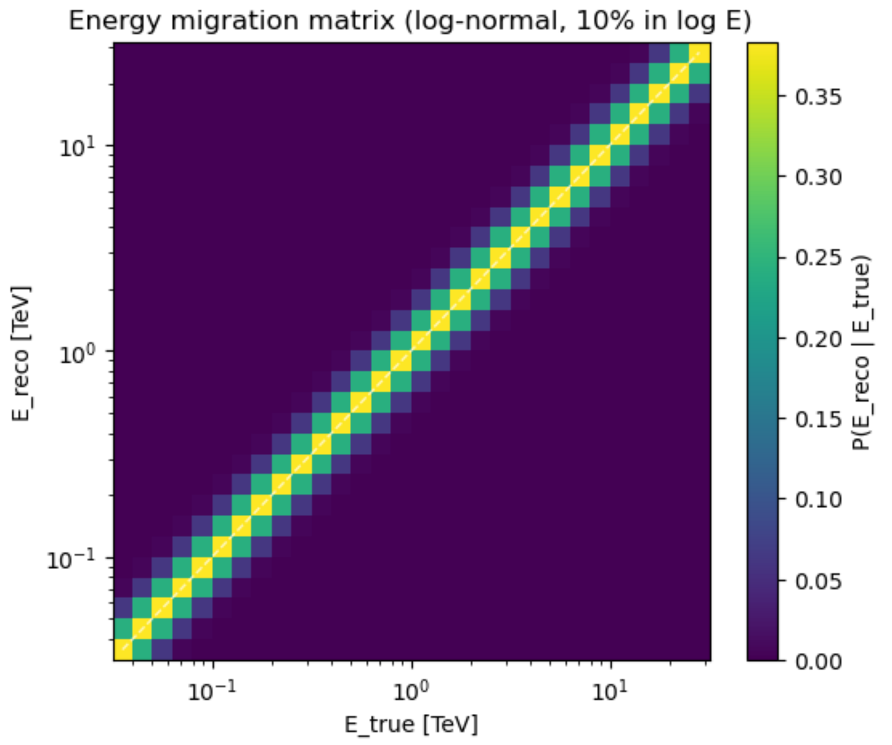
name	value	unit	error
str9	float64	str14	float64
amplitude	2.0026e-11	cm-2 s-1 TeV-1	8.945e-13
reference	1.0000e+00	TeV	0.000e+00
alpha	2.8600e+00		6.767e-02
beta	1.8543e-01		4.312e-02

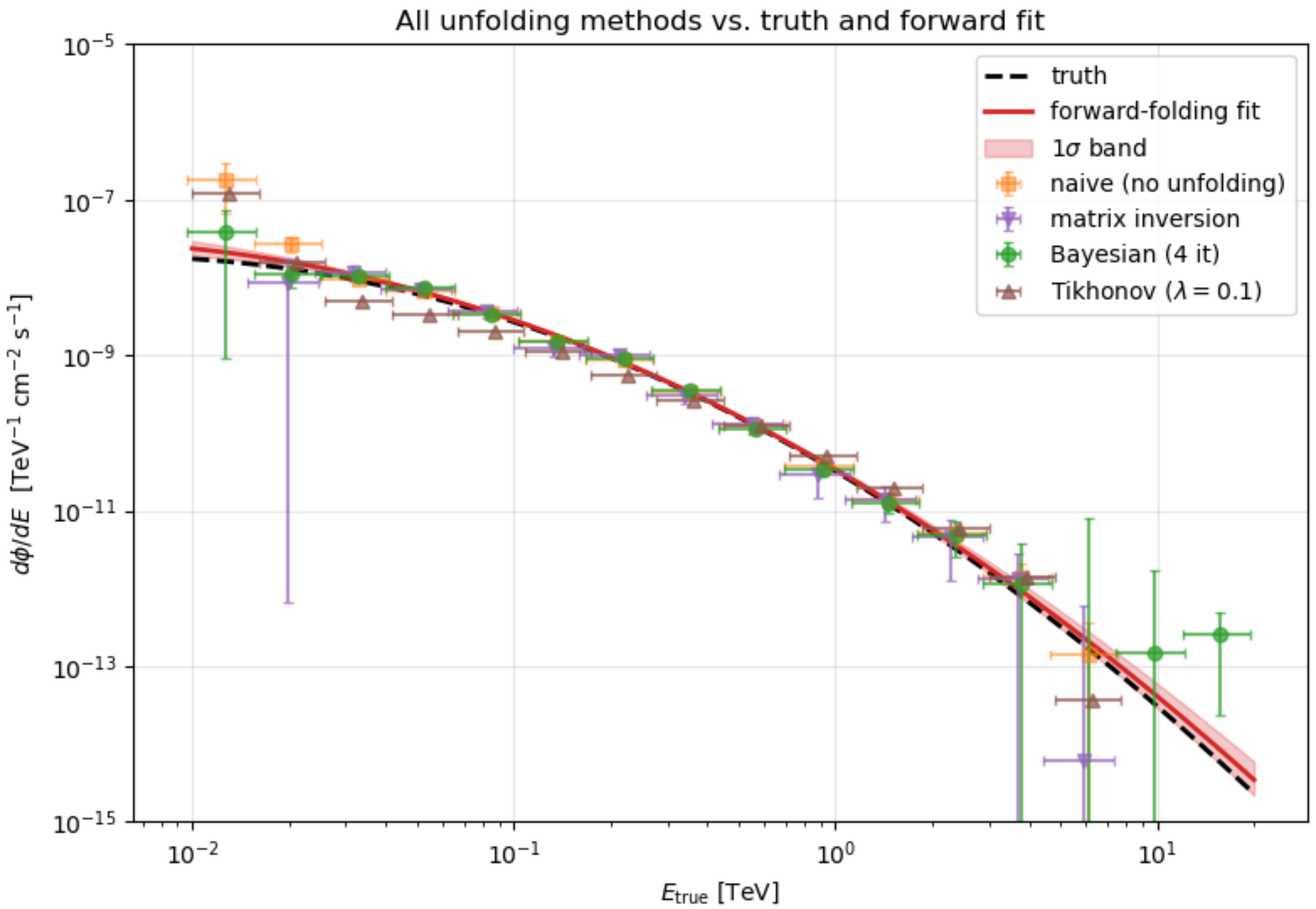


Plot obtained with gammapy

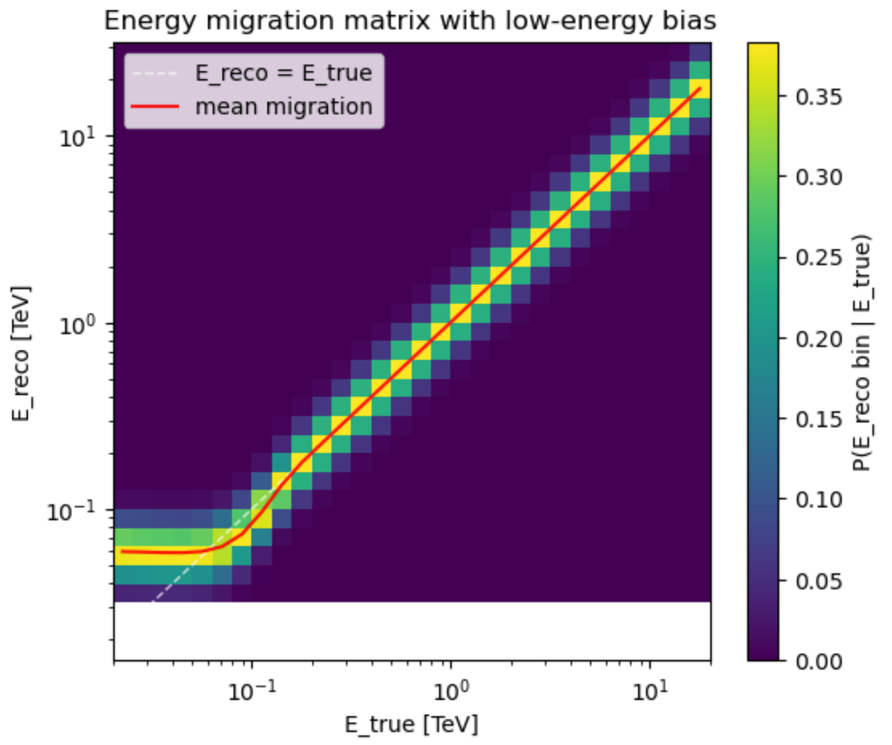
Summary

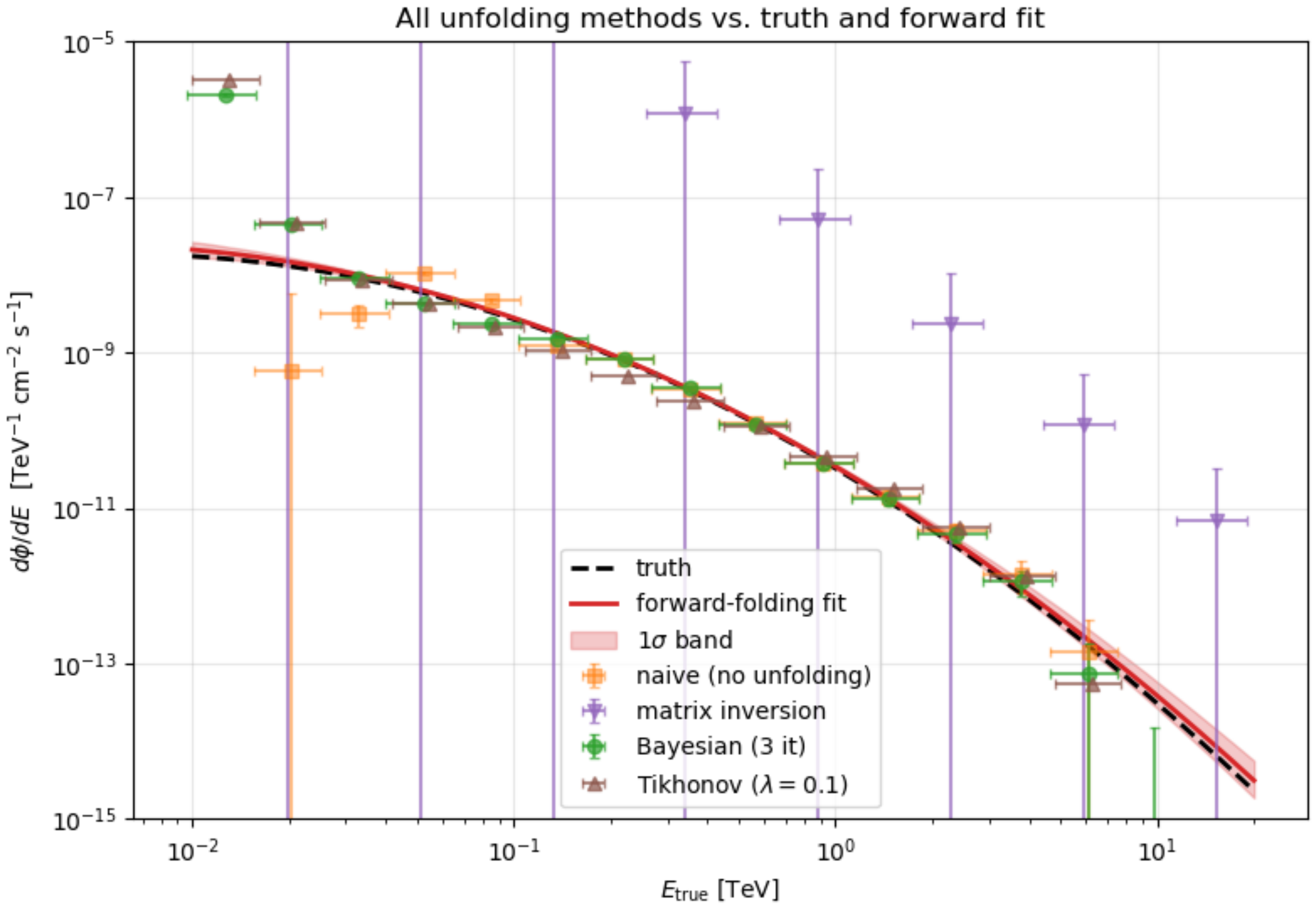
No energy bias





Extreme energy bias





Forward folding - 3D case

1D

$$s_i = \int_{\Delta E_i} dE_{est} \int dE_{true} \frac{d\phi(E_{true}, \vec{\theta})}{dE_{true}} \, A_{eff}(E_{true}) \, D(E_{est} \,|\, E_{true}) \, t_{eff}$$

3D

$$s_{ij} = \int_{\Delta \vec{x}_j} d\vec{x}_{est} \int_{\Delta E_i} dE_{est} \int dE_{true} \int d\vec{x}_{true} \frac{d\phi(E_{true}, \vec{x}_{true}, \vec{\theta})}{dE_{true}} \, A_{eff}(E_{true}, \vec{x}_{true}) \, D(E_{est} \,|\, E_{true}, \vec{x}_{true}) P(\vec{x}_{est} \,|\, \vec{x}_{true}, E_{true},) \, t_{eff}$$

Estimated direction

ϕ, A_{eff} , and D can now depend also on $\vec{x}_{true}$

Point Spread Function of the instrument

True direction

Expected γ counts in the i-th bin in **estimated** energy and j-th pixel

Forward folding - 3D case

$$s_{ij} = \int_{\Delta\vec{x}_j} d\vec{x}_{est} \int_{\Delta E_i} dE_{est} \int dE_{true} \int d\vec{x}_{true} \frac{d\phi(E_{true}, \vec{x}_{true}, \vec{\theta})}{dE_{true}} \; A_{eff}(E_{true}, \vec{x}_{true}) \; D(E_{est} \,|\, E_{true}, \vec{x}_{true}) P(\vec{x}_{est} \,|\, \vec{x}_{true}, E_{true},) \; t_{eff}$$

$$\mathscr{L}(\vec{\theta}) = \sum_{ij} \left(\frac{\left(s_{ij} + b_{ij}\right)^{n_{ij}}}{n_{ij}!} e^{-(s_{ij}+b_{ij})} \right)$$

Counts observed in each i-th energy bin and j-th pixel

Your background can be, for instance, the Galactic diffuse background

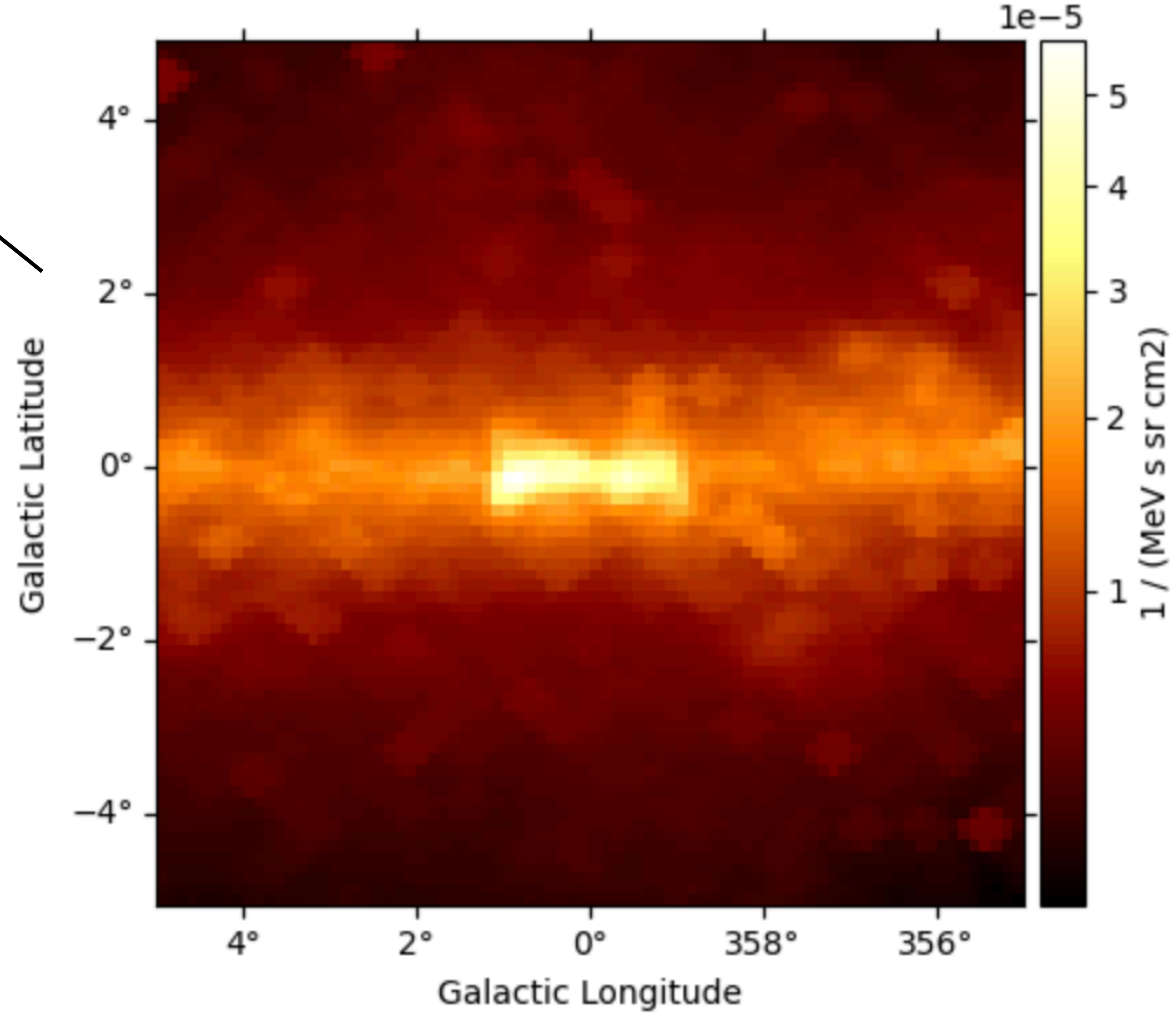


image from https://docs.gammapy.org/2.0/tutorials/data/fermi_lat.html

Conclusions Part II

- In this part we have **introduced** the **basic concepts** of a γ-ray analysis
    - Inference in an On/Off measurement
    - Flux estimation (effective area, energy dispersion, etc…)
    - Unfolding and forward folding

- for more details:
    - MAGIC performance paper: arXiv:1409.5594
    - LST-1 performace paper: arXiv:2306.12960
    - …

- Things we did not cover:
    - Night Sky Background, Proton, Muons, Image Cleaning and Parametrization, Random Forest, Particle Classification, Gamma-hadron separation, and more …