Introduction to Statistics for Astronomy

Università degli Studi di Padova

Giacomo D'Amico

Astrophysics Data Camp at the University of Padova

Shaping a World-class University

Padova, 25-29 September 2023

What are we going to talk about?

1. The Bayesian approach: probability theory and the Bayes Theorem
2. The frequentist approach: p-values and sigmas
3. The likelihood
4. Statistical inference in On/Off measurement

Why is Statistic so important?

Why is Statistic so important?

Short answer: the experimental data on their own are useless!

The final goal is to **infer** from the observed data a given hypothesis

Definition of **infer verb** from the Oxford Learner's Dictionary of Academic English

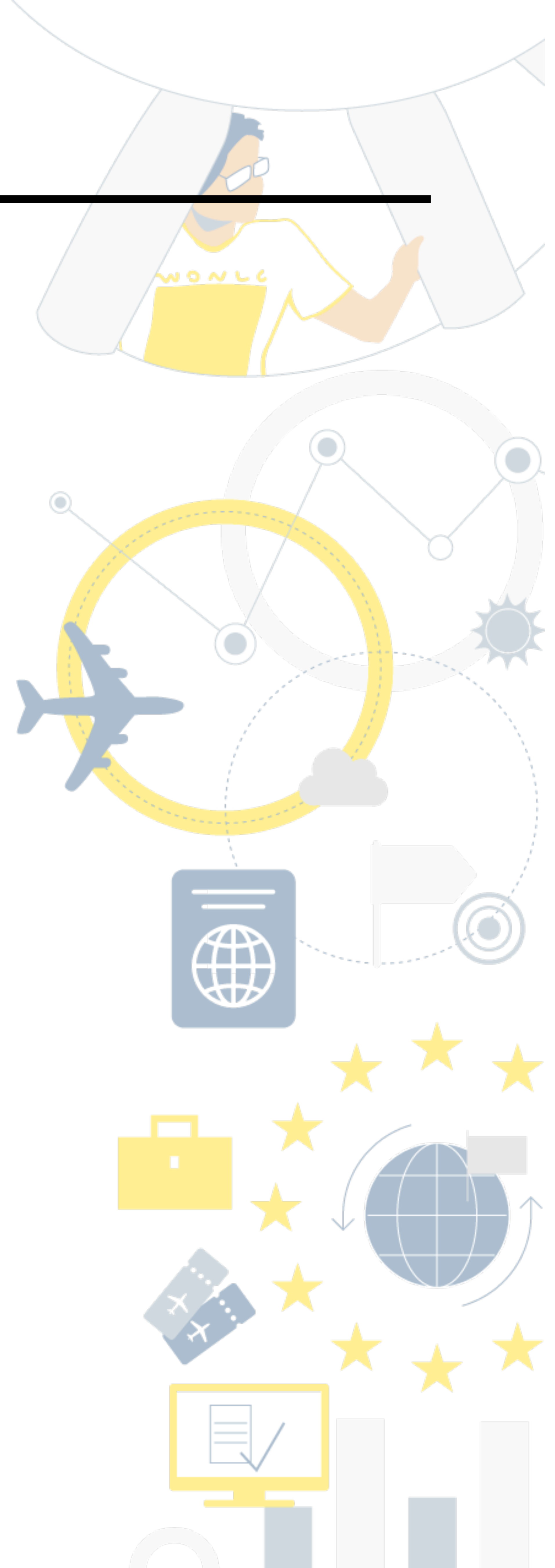
infer *verb*

OPAL written

BrE /ɪn'fɜː(r)/; NAmE /ɪn'fɜːr/

+ Verb Forms

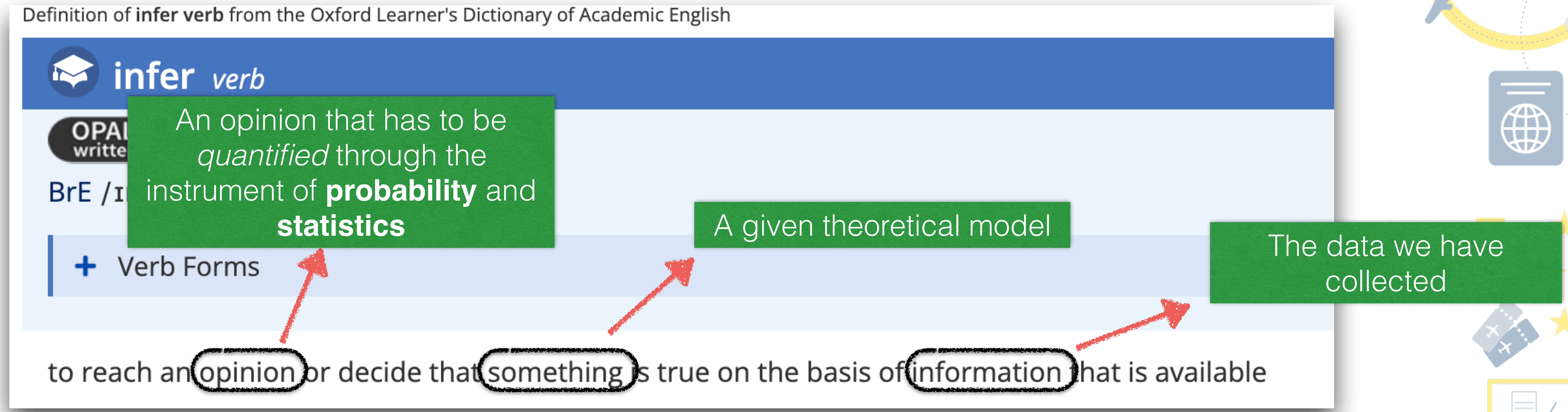
to reach an opinion or decide that something is true on the basis of information that is available



Why is Statistic so important?

Short answer: the experimental data on their own are useless!

The final goal is to **infer** from the observed data a given hypothesis



Why is Statistic so important?

The Model → *All Italians are good drivers*

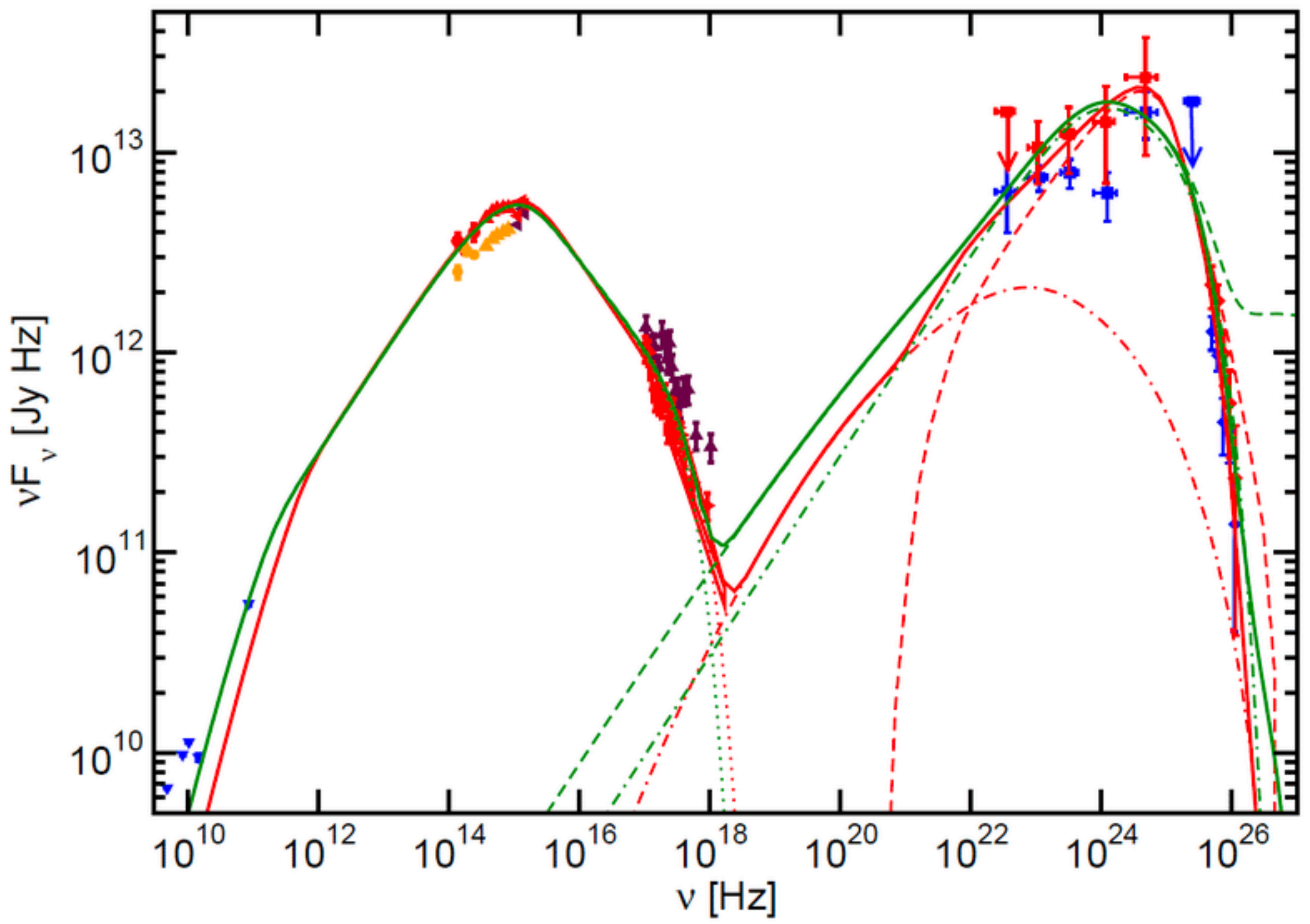
The data →

The opinion → *The model is rejected*

Why is Statistic so important?

The Model → *hadronic radiative processes are negligible in the emission of the source*

The data →



The opinion → *??*

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

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

- Conditional probability

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

Probability theory

- Marginalised probability

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

- Conditional probability

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

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

Probability theory

- Marginalised probability

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

- Conditional probability

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

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

Likelihood

Prior

Posterior

Normalisation

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

The Monty Hall problem

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

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

The Monty Hall problem

1

2

S/He then give us the opportunity to change our box (n.1) with the other (n. 2)

What would you do? Would you accept the opportunity?

The Monty Hall problem

- H_i The hypothesis “the car is in the i-th box”

The Monty Hall problem

1

2

- 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”

The Monty Hall 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”

The Monty Hall problem

1

2

- 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"

Posterior $f(H_i \,|\, E, I)$

The Monty Hall problem

1

2

$$f(H_1|E,I) = \frac{f(E|H_1,I)f(H_1|I)}{f(E|I)} = \dots$$

$$f(H_2|E,I) = \frac{f(E|H_2,I)f(H_2|I)}{f(E|I)} = \dots$$

$$f(H_3|E,I) = \frac{f(E|H_3,I)f(H_3|I)}{f(E|I)} = \dots$$

The Monty Hall problem

$$f(H_1|E,I) = \frac{f(E|H_1,I)f(H_1|I)}{f(E|I)} = \frac{\quad\cdot 1/3}{\quad}$$

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

$$f(H_3|E,I) = \frac{f(E|H_3,I)f(H_3|I)}{f(E|I)} = \frac{\quad\cdot 1/3}{\quad}$$

Priors → $f(H_1|I) = f(H_2|I) = f(H_3|I) = \frac{1}{3}$

The Monty Hall problem

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

$$f(H_2|E,I) = \frac{f(E|H_2,I)f(H_2|I)}{f(E|I)} = \frac{\cdot 1/3}{1/2}$$

$$f(H_3|E,I) = \frac{f(E|H_3,I)f(H_3|I)}{f(E|I)} = \frac{\cdot 1/3}{1/2}$$

Normalisation → $\sum_i f(E|H_i,I)f(H_i|I) = f(E|I) = \frac{1}{2}$

The Monty Hall problem

1

2

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

$$f(H_2|E,I) = \frac{f(E|H_2,I)f(H_2|I)}{f(E|I)} = \frac{1 \cdot 1/3}{1/2} =$$

$$f(H_3|E,I) = \frac{f(E|H_3,I)f(H_3|I)}{f(E|I)} = \frac{0 \cdot 1/3}{1/2} =$$

Likelihoods → $f(E|H_1,I) = \frac{1}{2}$ $f(E|H_2,I) = 1$ $f(E|H_3,I) = 0$

The Monty Hall problem

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

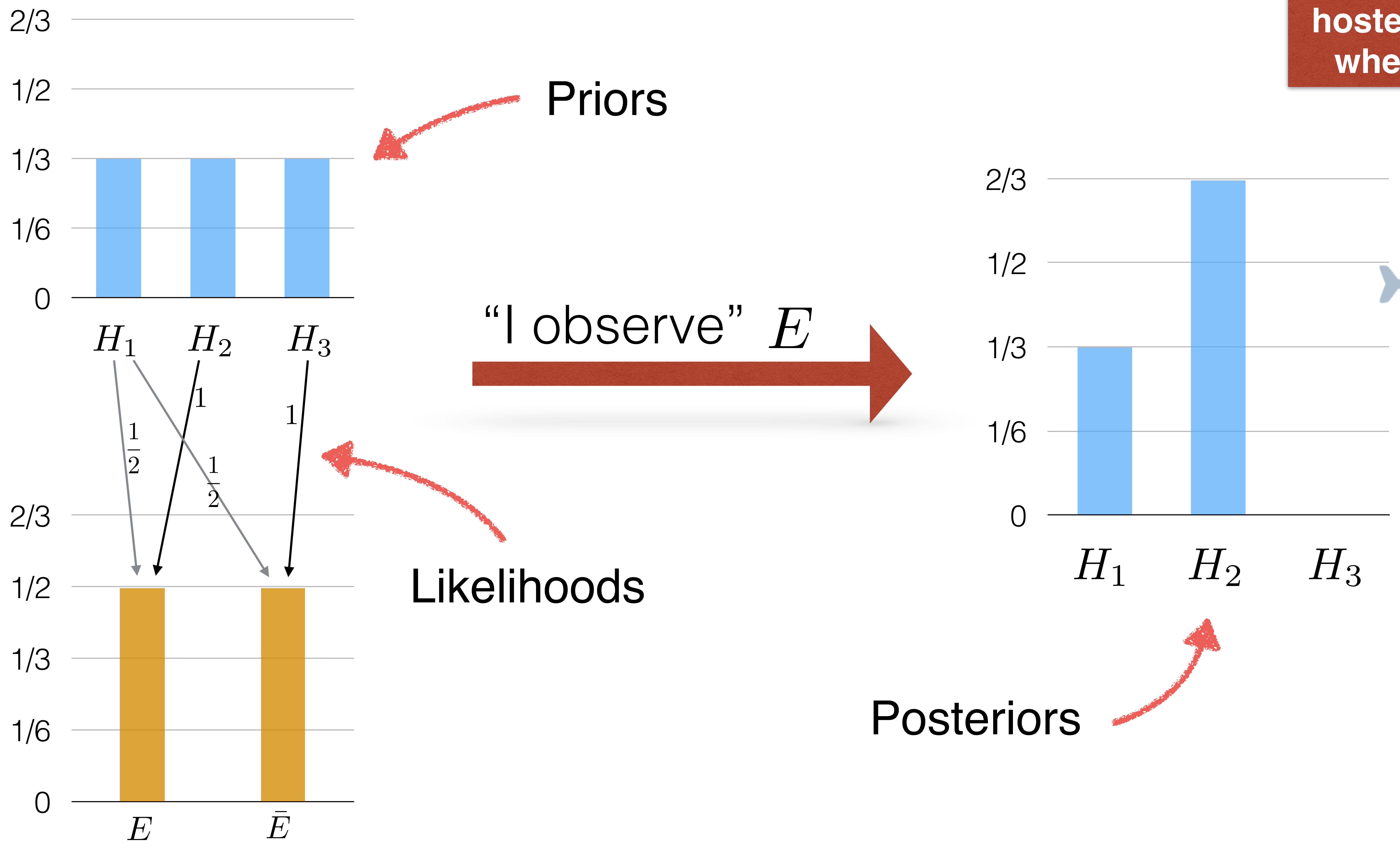
$$f(H_2|E,I) = \frac{f(E|H_2,I)f(H_2|I)}{f(E|I)} = \frac{1 \cdot 1/3}{1/2} = \frac{2}{3}$$

$$f(H_3|E,I) = \frac{f(E|H_3,I)f(H_3|I)}{f(E|I)} = \frac{0 \cdot 1/3}{1/2} = 0$$

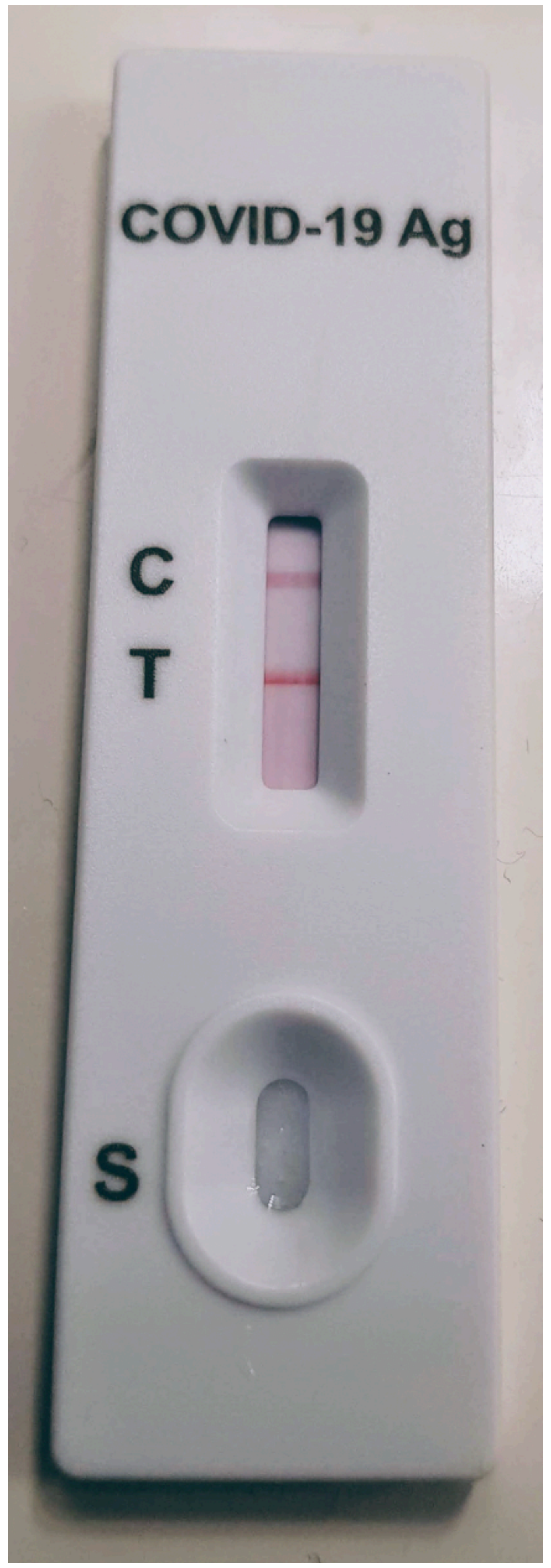
If we want to win the car,
we should change the box!

The Monty Hall problem

What if the TV-Show hoster did not know where the car is?



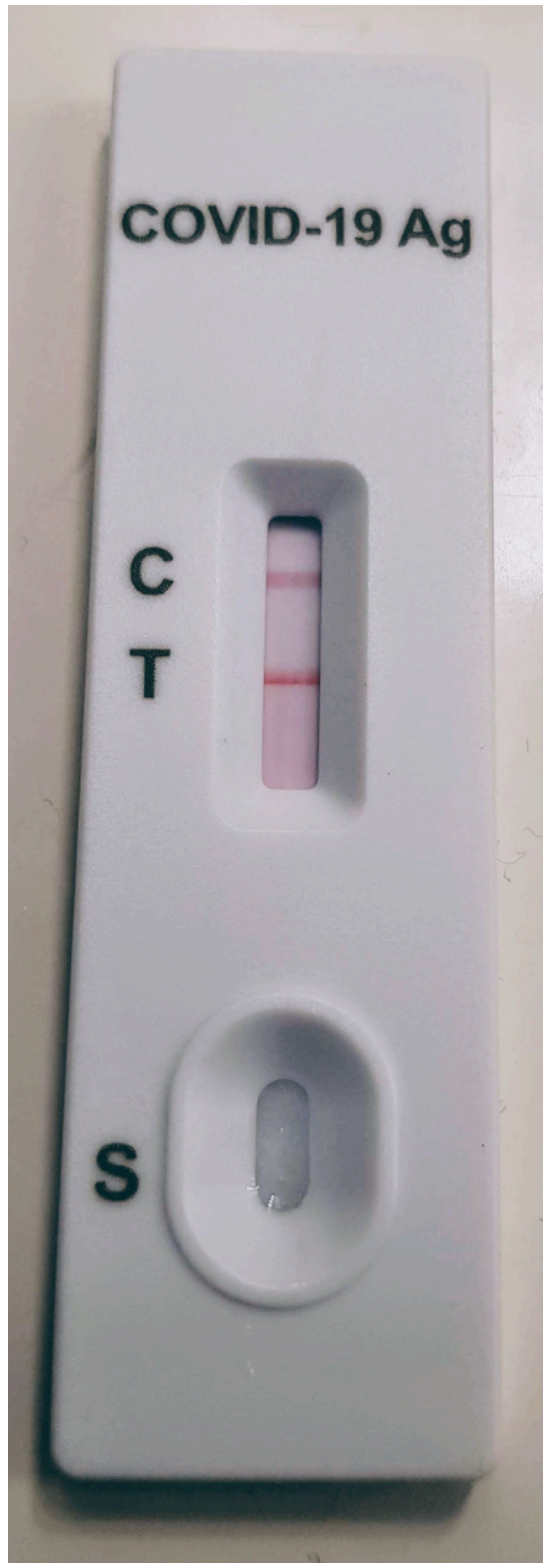
Covid-19 test



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

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

Covid-19 test



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

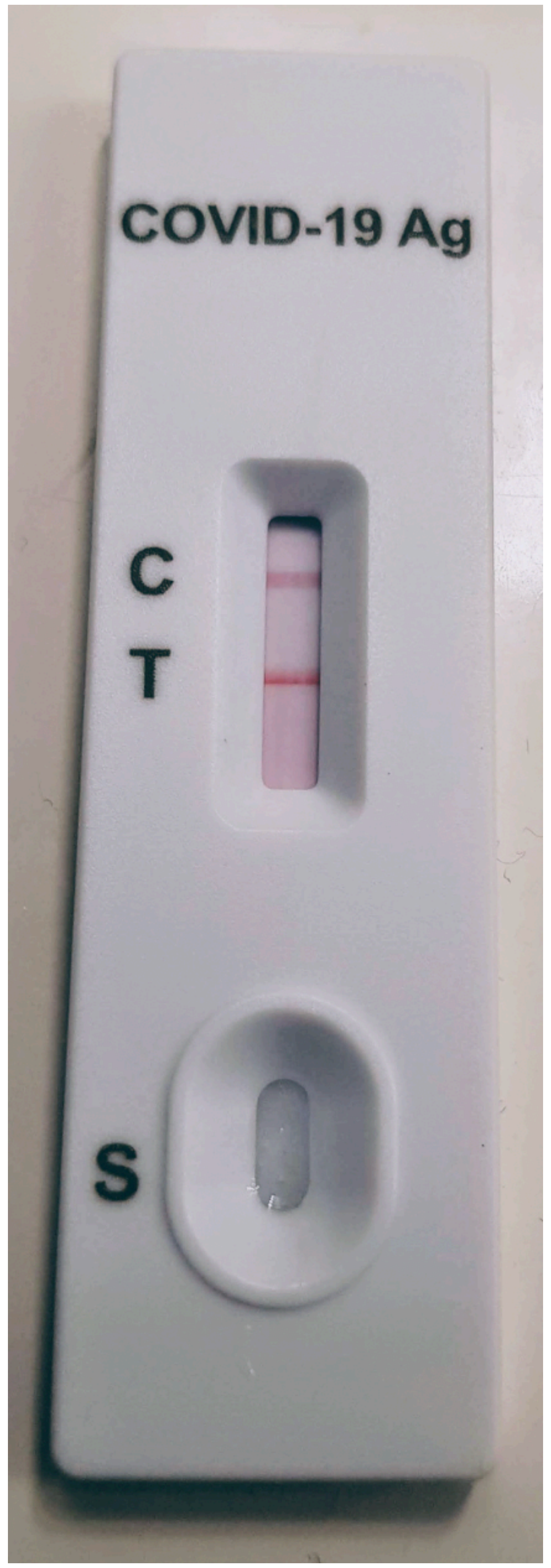
$$p(S\,|\,+) = \frac{P(+\,|\,S)P(S)}{P(+\,|\,S)P(S) + P(+\,|\,\bar{S})P(\bar{S})}$$

Priors

Probability of True positive

Probability of False positive

Covid-19 test



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

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

Priors

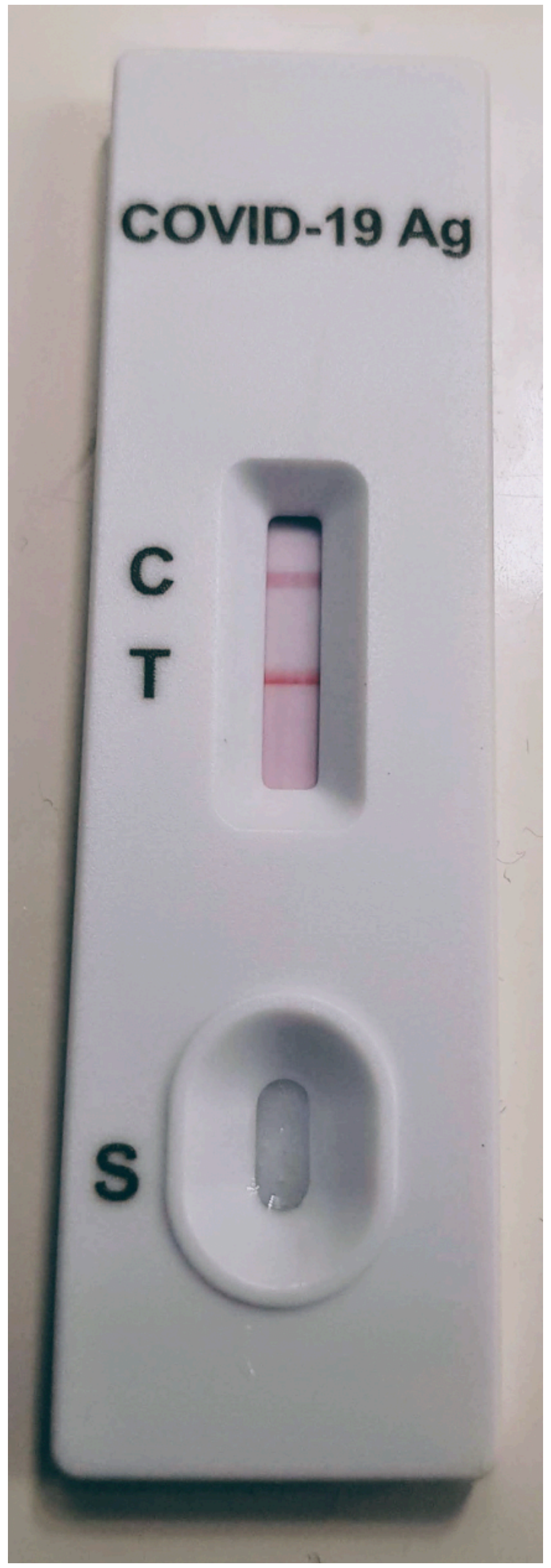
Probability of True positive

Probability of False positive

$$\text{Sensitivity} \equiv P(+\,|\,S)$$

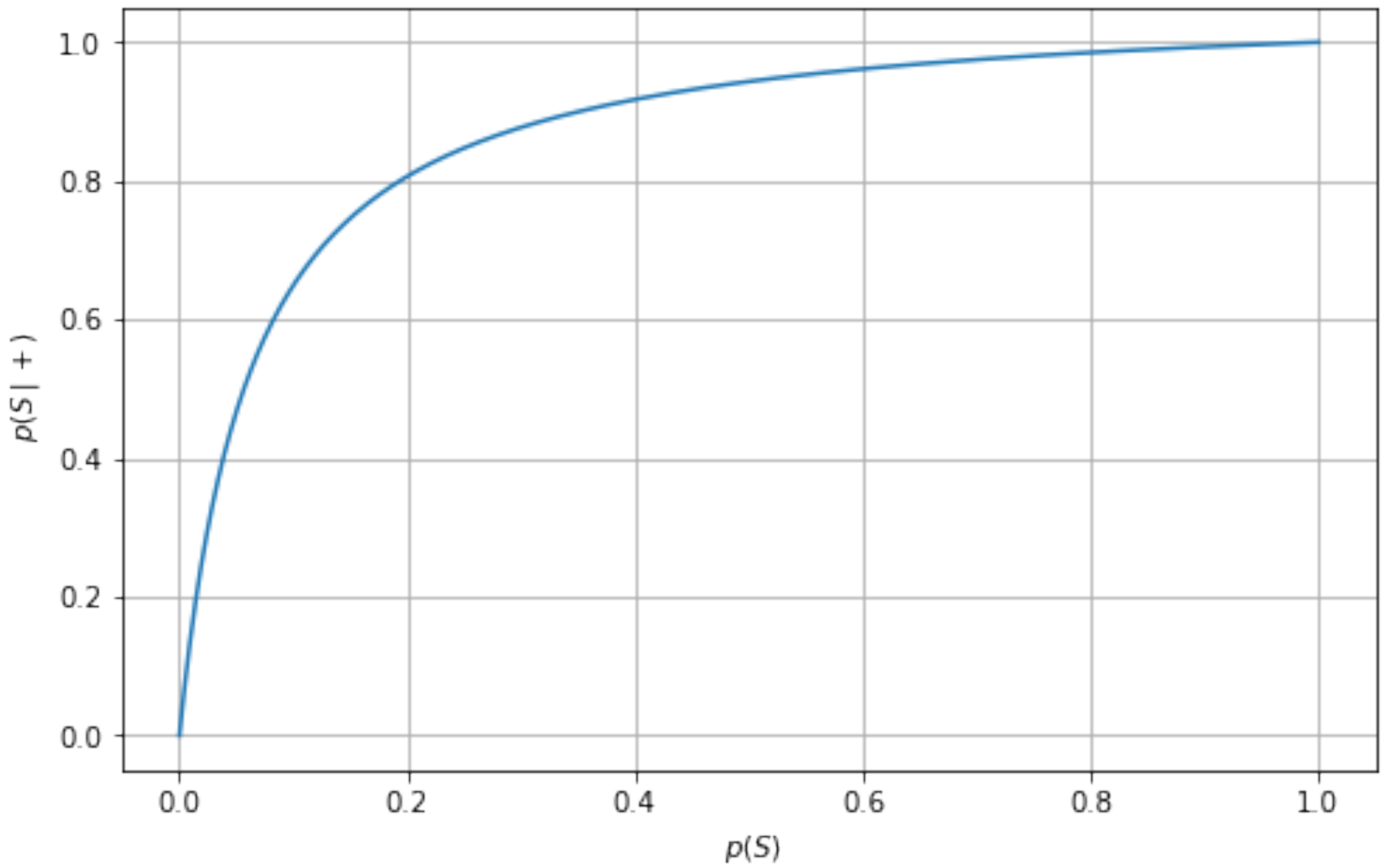
$$\text{Specificity} \equiv P(-\,|\,\bar{S})$$

Covid-19 test



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

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



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

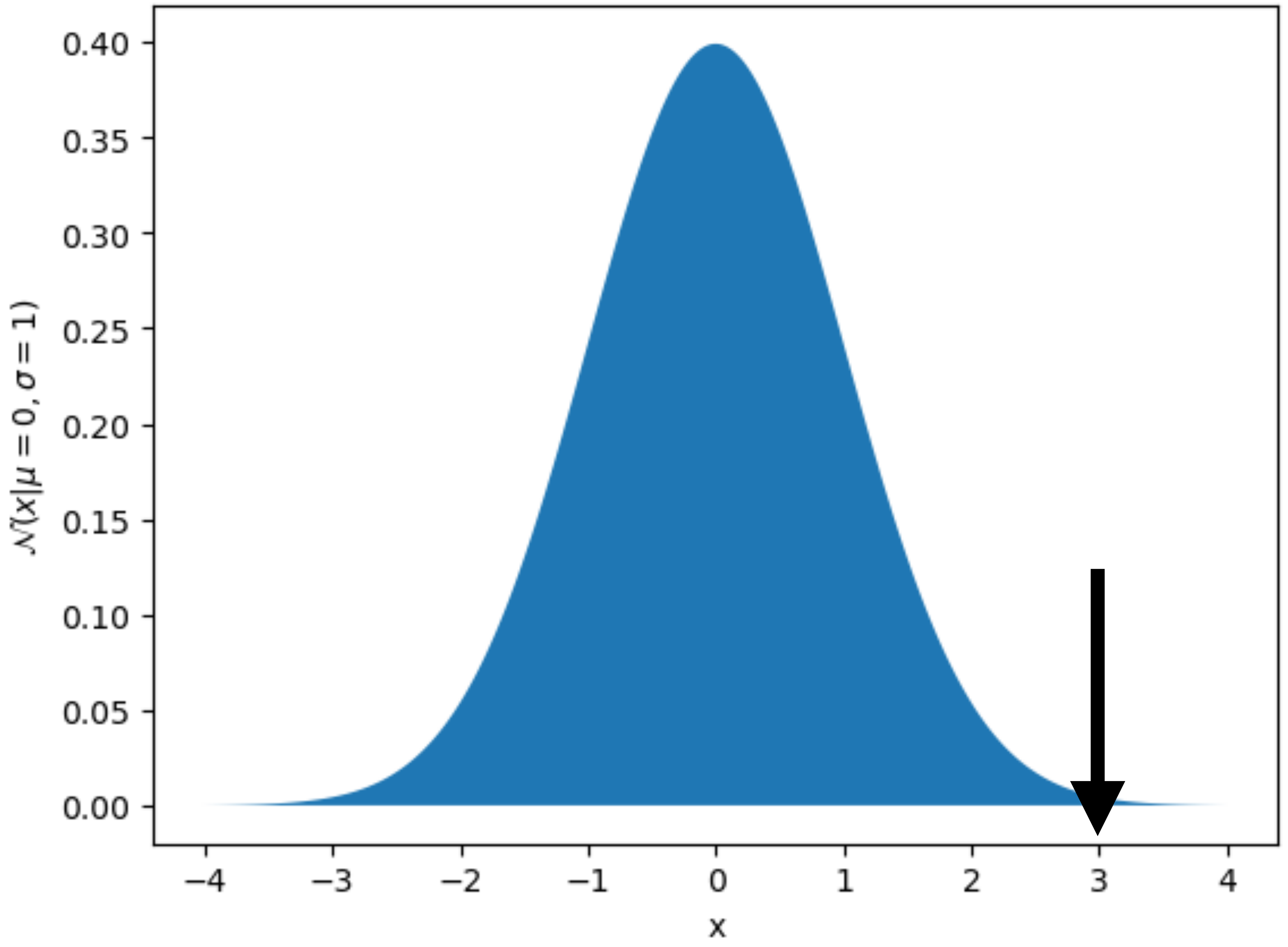
$$\mathrm{Se.} = 50\%$$

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

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

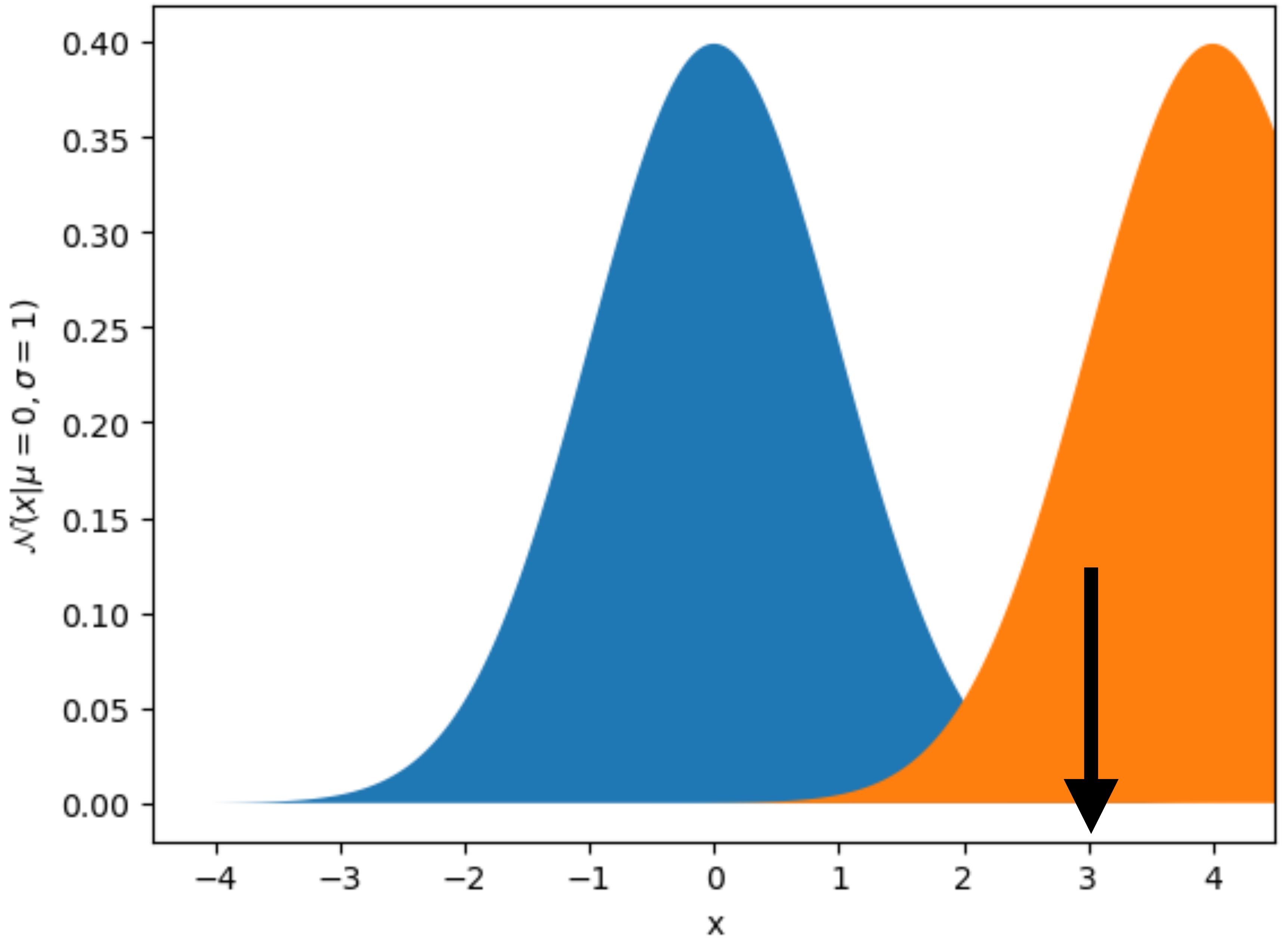


A trivial example

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

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



According to the alternative hypotheses $\bar{H}$, *D* is a random variable that follows a normal distribution centered in 4 and variance = 1

$$p(D|H) = \mathcal{N}(x = 3|\mu = 4, \sigma = 1)$$

A trivial example

Let's now assume that the observation D can only be explained by either H *or* $\bar{H}$ *and that we have no bias towards any of the two hypotheses*

$$p(H) = 1 - p(\bar{H}) = 0.5$$

By applying the Bayes theorem

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{p(D|H) \cdot p(H) + p(D|\bar{H}) \cdot p(\bar{H})} =$$

$$= \frac{p(D|H))}{p(D|H) + p(D|\bar{H})} = \frac{\mathcal{N}(x=3|\mu=0,\sigma=1))}{\mathcal{N}(x=3|\mu=0,\sigma=1) + \mathcal{N}(x=3|\mu=4,\sigma=1)}$$

A trivial example

Let's now assume that the observation D can only be explained by either H *or* $\bar{H}$ *and that we have no bias towards any of the two hypotheses*

$$p(H) = 1 - p(\bar{H}) = 0.5$$

By applying the Bayes theorem

$$p(H\,|\,D) = \frac{p(D\,|\,H)\cdot p(H)}{p(D\,|\,H)\cdot p(H) + p(D\,|\,\bar{H})\cdot p(\bar{H})} =$$

$$= \frac{e^{-(3-0)^2/2}}{e^{-(3-0)^2/2} + e^{-(3-4)^2/2}} = \frac{1}{1 + e^{9/2-1/2}} = \frac{1}{1+e^4} \sim 1.8\,\%$$

A trivial example

Conclusion of the inference analysis performed with the Bayesian approach:

Having observed D=3 and assuming **uniform** priors, the probability of the hypothesis H being true is **1.8%**

The Frequentist approach

The Frequentist approach

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

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

value ?

frequency

true

extreme

The Frequentist approach

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

*If I repeat the experiment an infinite time, assuming the model is **true**, with which **frequency** I would observe a **<u>value</u>** more **extreme** than the one actually observed?*

value ?

The **data** “D” itself or a function of them known as the **statistic**

$$\mathcal{S} = \mathcal{S}(D)$$

frequency

true

extreme

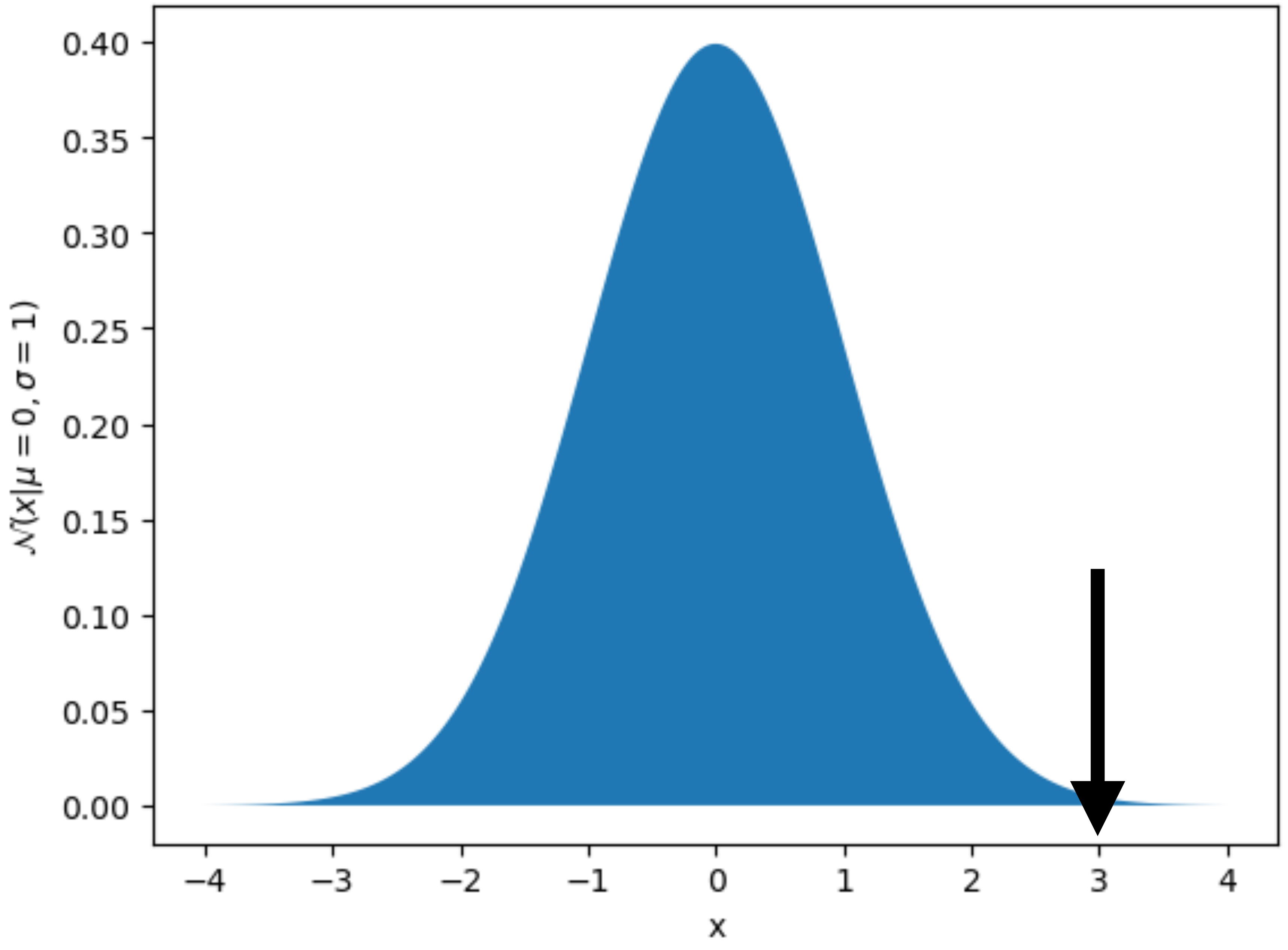
The Frequentist approach

Again the 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

Again the 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** time, assuming *H* to be **true**, I would have observed ***D > 3*** only **0.27%** of the times

The Frequentist approach

The **P-VALUE** is the frequency in which we would have observed "something" more extreme assuming the null hypothesis to be true

$$\text{p-value} = p(x \text{ more extreme than } x_{obs} | H_0)$$

The Frequentist approach

The **P-VALUE** is the frequency in which we would have observed "something" more extreme assuming the null hypothesis to be true

$$\text{p-value} = p(x \text{ more extreme than } x_{obs} | H_0)$$

… but then, what are all these "sigmas"?

We therefore consider emission from the Sgr dSph as an alternative origin for the cocoon. In order to test this possibility, we fit the γ-ray emission observed by *Fermi*-LAT over a region of interest (ROI) containing the cocoon via template analysis. In our *baseline* model these templates include only known point sources and sources of Galactic diffuse γ-ray emission. We contrast the baseline with a *baseline + Sgr dSph* model that invokes these same templates plus an additional template constructed to be spatially coincident with the bright stars of the Sgr dSph (Extended Data (E.D.) Figure 1 and S.I. Figure 1); full details of the fitting procedure are provided in Methods and S.I. sec. 3. Using the best motivated choice of templates, we find that the baseline + Sgr dSph model is preferred at 8.1σ significance over the baseline model. We also repeat the analysis for a wide range of alternative templates for both Galactic diffuse emission and for the Sgr dSph (Table 1) and obtain $> 5\sigma$ detections for all combinations but one. Moreover, even this is an extremely conservative estimate, because our baseline model uses a structured template for the FBs that absorbs some of the signal that is spatially coincident with the Sgr dSph into a structure of unknown origin. If we follow the method recommended by the Fermi collaboration [2] and use a flat FB template in our analysis, the significance of our detection of the Sgr dSph is always $> 14\sigma$. Despite this, for the remainder of our analysis we follow the most conservative choice by using the structured template in our baseline model. In Methods, we also show that our analysis passes a series of validation tests: the residuals between our best-fitting model and the data are consistent with photon counting statistics (E.D. Figure 2 and Figure 3), our pipeline reliably recovers synthetic signals superimposed on a realistic background (E.D. Figure 4), fits using a template tracing the stars of the Sgr dSph yield significantly better results than fits using purely geometric templates (S.I. Table 1), and if we artificially rotate the Sgr dSph template on the sky, the best-fitting position angle is very close to the actual one (E.D. Figure 5). By contrast, if we displace the Sgr dSph template, we find moderate (4.5σ significance) evidence that the best-fitting position is $\sim 4°$ from the true position, in a direction very closely aligned with the dwarf galaxy's direction of travel (E.D. Figure 5); this plausibly represents a small, but real and expected (as explained below) physical offset between the stars and the γ-ray emission.

ro-ph.HE] 19 Jun 2022

PKS 1413+135: Bright GeV γ-ray Flares with Hard-spectrum and Hints for First Detection of TeV γ-rays from a Compact Symmetric Object

Ying-Ying Gan,[1] Jin Zhang†,[1] Su Yao,[2] Hai-Ming Zhang,[3] Yun-Feng Liang,[4] and En-Wei Liang[4]

[1] *School of Physics, Beijing Institute of Technology, Beijing 100081, People's Republic of China; j.zhang@bit.edu.cn*
[2] *Max-Planck-Institute für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany*
[3] *School of Astronomy and Space Science, Nanjing University, Nanjing 210023, People's Republic of China*
[4] *Guangxi Key Laboratory for Relativistic Astrophysics, School of Physical Science and Technology, Guangxi University, Nanning 530004, People's Republic of China*

ABSTRACT

PKS 1413+135, a typical compact symmetric object (CSO) with a two-side pc-scale structure in its miniature radio morphology, is spatially associated with the *Fermi*-LAT source 4FGL J1416.1+1320 and recently announced to be detected in the TeV γ-ray band with the MAGIC telescopes. We present the analysis of its X-ray and GeV γ-ray observations obtained with *Swift*-XRT, *XMM-Newton*, *Chandra*, and *Fermi*-LAT for revealing its high energy radiation physics. No significant variation trend is observed in the X-ray band. Its GeV γ-ray light curve derived from the *Fermi*-LAT 13.5-year observations shows that it is in a low γ-ray flux stage before MJD 58500 and experiences violent outbursts after MJD 58500. The confidence level of the flux variability is much higher than 5σ, and the flux at 10 GeV varies $\sim$ 3 orders of magnitude. The flux variation is accompanied by the clearly

The Frequentist approach

The **P-VALUE** is the frequency in which we would have observed "something" more extreme assuming the null hypothesis to be true

$$\text{p-value} = p(x \text{ more extreme than } x_{obs} | H_0)$$

… but then, what are all these "sigmas"?

We therefore consider emission from the Sgr dSph as an alternative origin for the cocoon. In order to test this possibility, we fit the γ-ray emission observed by *Fermi*-LAT over a region of interest (ROI) containing the cocoon via template analysis. In our *baseline* model these templates include only known point sources and sources of Galactic diffuse γ-ray emission. We contrast the baseline with a *baseline + Sgr dSph* model that invokes these same templates plus an additional template constructed to be spatially coincident with the bright stars of the Sgr dSph (Extended Data (E.D.) Figure 1 and S.I. Figure 1); full details of the fitting procedure are provided in Methods and S.I. sec. 3. Using the best motivated choice of templates, we find that the baseline + Sgr dSph model is preferred at 8.1σ significance over the baseline model. We also repeat the analysis for a wide range of alternative templates for both Galactic diffuse emission and for the Sgr dSph (Table 1) and obtain $> 5\sigma$ detections for all combinations but one. Moreover, even this is an extremely conservative estimate, because our baseline model uses a structured template for the FBs that absorbs some of the signal that is spatially coincident with the Sgr dSph into a structure of unknown origin. If we follow the method recommended by the Fermi collaboration [2] and use a flat FB template in our analysis, the significance of our detection of the Sgr dSph is always $> 14\sigma$. Despite this, for the remainder of our analysis we follow the most conservative choice by using the structured template in our baseline model. In Methods, we also show that our analysis passes a series of validation tests: the residuals between our best-fitting model and the data are consistent with photon counting statistics (E.D. Figure 2 and Figure 3), our pipeline reliably recovers synthetic signals superimposed on a realistic background (E.D. Figure 4), fits using a template tracing the stars of the Sgr dSph yield significantly better results than fits using purely geometric templates (S.I. Table 1), and if we artificially rotate the Sgr dSph template on the sky, the best-fitting position angle is very close to the actual one (E.D. Figure 5). By contrast, if we displace the Sgr dSph template, we find moderate (4.5σ significance) evidence that the best-fitting position is $\sim 4°$ from the true position, in a direction very closely aligned with the dwarf galaxy's direction of travel (E.D. Figure 5); this plausibly represents a small, but real and expected (as explained below) physical offset between the stars and the γ-ray emission.

PKS 1413+135: Bright

Ying-Ying Gan,[1] Jin

[1]School of Physics, Beij
[2]Max-Plan
[3]School of Astronom
[4]Guangxi Key Laboratory for Relat

ro-ph.HE] 19 Jun 2022

PKS 1413+135, a t
miniature radio morp
and recently announ
present the analysis o
Chandra, and *Fermi*-
trend is observed in
year observations sho
outbursts after MJD
the flux at 10 GeV v

ABSTRACT

It is usually thought that long-duration gamma-ray bursts (GRBs) are associated with massive star core collapse whereas short-duration GRBs are associated with mergers of compact stellar binaries. The discovery of a kilonova associated with a nearby (350 Mpc) long-duration GRB- GRB 211211A, however, indicates that the progenitor of this long-duration GRB is a compact object merger. Here we report the *Fermi*-LAT detection of gamma-ray (> 100 MeV) afterglow emission from GRB 211211A, which lasts $\sim$ 20000 s after the burst, the longest event for conventional short-duration GRBs ever detected. We suggest that this gamma-ray emission results mainly from afterglow synchrotron emission. The soft spectrum of GeV emission may arise from a limited maximum synchrotron energy of only a few hundreds of MeV at $\sim$ 20000 s. The usually long duration of the GeV emission could be due to the proximity of this GRB and the long deceleration time of the GRB jet that is expanding in a low density circumburst medium, consistent with the compact stellar merger scenario.

Keywords: Gamma-ray bursts (629) — High energy astrophysics (739)

1. INTRODUCTION

Gamma-ray bursts (GRBs) are usually divided into two populations (Kouveliotou et al. 1993; Norris et al. 1984): long GRBs that originate from the core-collapse of massive stars (Galama et al. 1998) and short GRBs formed in the merger of two compact objects (Abbott et al. 2017). While it is common to divide the two populations at a duration of 2 s for the prompt keV/MeV emission, classification based on duration only does not always correctly point to the progenitor. Growing observations (Ahumada et al. 2021; Gal-Yam et al. 2006; Gehrels et al. 2006; Zhang et al. 2021) have shown that multiple criteria (such as supernova/kilonova associations and host galaxy properties) rather than burst duration only are needed to classify GRBs physically.

GRB 211211A triggered the Burst Alert Telescope (Barthelmy et al. 2005) onboard The Neil Gehrels Swift Observatory at 13:09:59 UT (D'Ai et al. 2021), the Gamma-ray Burst Monitor (Meegan et al. 2009) onboard The Fermi Gamma-Ray Space Telescope at 13:09:59.651 UT (Mangan et al. 2021) and High energy X-ray Telescope onboard Insight-HXMT (Xiao et al. 2022) at 13:09:59 UT on 11 December 2021. The burst is characterized by a spiky main emission phase lasting $\sim$13 seconds, and a longer, weaker extended emission phase lasting $\sim$55 seconds (Yang et al. 2022). The prompt emission is suggested to be produced by the fast-cooling synchrotron emission (Gompertz et al. 2022). The discovery of a kilonova associated with this GRB indicates clearly that the progenitor is a compact object merger (Rastinejad et al. 2022). The event fluence (10-1000 keV) of the prompt emission is $(5.4 \pm 0.01) \times 10^{-4}$ erg cm^{-2}, making this GRB an exceptionally bright event. The host galaxy redshift of GRB 211211A is $z = 0.0763 \pm 0.0002$ (corresponding to a distance of $\approx$350 Mpc (Rastinejad et al. 2022)). At 350 Mpc, GRB 211211A is one of the closet GRBs, only a bit further than GRB 170817A, which is associated with the gravitational wave (GW)-detected binary neutron star (BNS) merger GW170817. For GRB 170817A, no GeV afterglow was detected by the LAT on timescales of minutes, hours, or days after the LIGO/Virgo detection (Ajello et al. 2018).

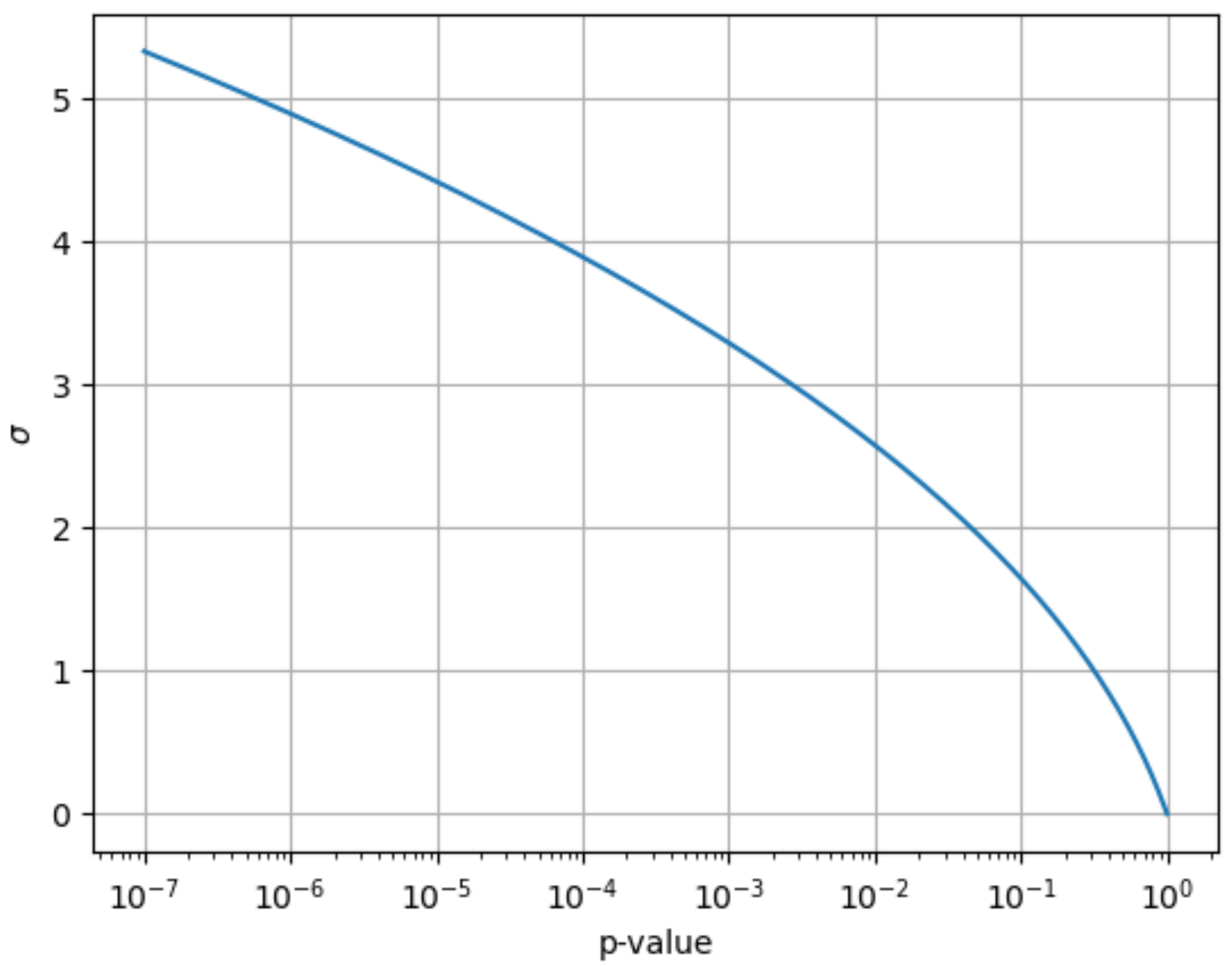
As the angle from the *Fermi*-LAT boresight at the GBM trigger time of GRB 211211A is 106.5 degrees (Mangan et al. 2021), LAT cannot place constraints on the existence of high-energy (E > 100 MeV) emission associated with the prompt GRB emission. We focus instead on constraining high-energy emission on the longer timescale. We analyze the late-time *Fermi*-LAT data when the GRB enters the field-of-view (FOV) of *Fermi*-LAT. We detect a transient source with a significance of $TS_{max} \simeq 51$, corresponding to a detection significance over 6σ. The result of the data analysis is shown in §2

The Frequentist approach

It is common to express such probability in multiples S of the standard deviations of a normal distribution via the inverse error function

$$S = \sqrt{2}\,\mathrm{erf}^{-1}\,(1 - \text{p-value})$$

Here the (in-)famous number of “sigma”




```
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()
```

The Frequentist approach

A bit of terminology…

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

p-value

The above frequentist conclusion can be rephrased as follow

The hypothesis H is rejected with a 99.73% C.L.

Confidence level = 1 - p-value

The hypothesis H is rejected with a significance of 3 "sigma"

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

The Frequentist approach

… when you read something like

" … we detected the source at 6 sigma … "

what they actually mean is:

If we repeat the experiment an **infinitely** time, assuming the "no-source" hypothesis to be **true**, we would have observed the statistic $\mathcal{S} > \mathcal{S}_{\rm obs.}$ only $1.97 \cdot 10^{-7}\,\%$ of the times

Recap:

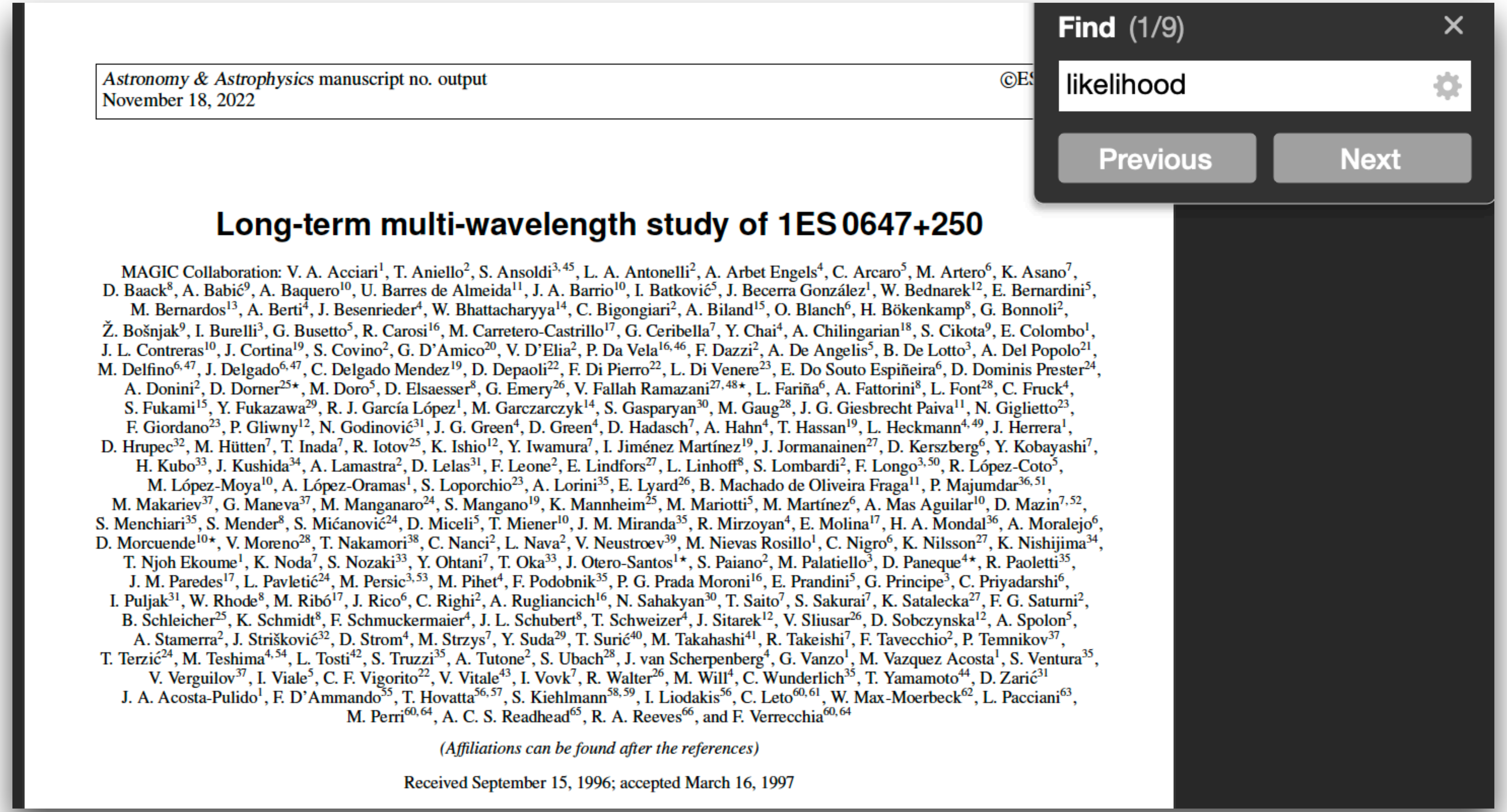
1. The **Bayesian** approach allows us to quantify our “opinion” on a given model from the observed data using the rules of **probability theory**
    - **Pros:** Alternative hypotheses are taken into account. No need to define a statistic and to know its distribution.
    - **Cons:** One needs a prior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

The Likelihood

The Likelihood

Let's take a look at some recent papers



Astronomy & Astrophysics manuscript no. output
November 18, 2022 ©ES

Long-term multi-wavelength study of 1ES 0647+250

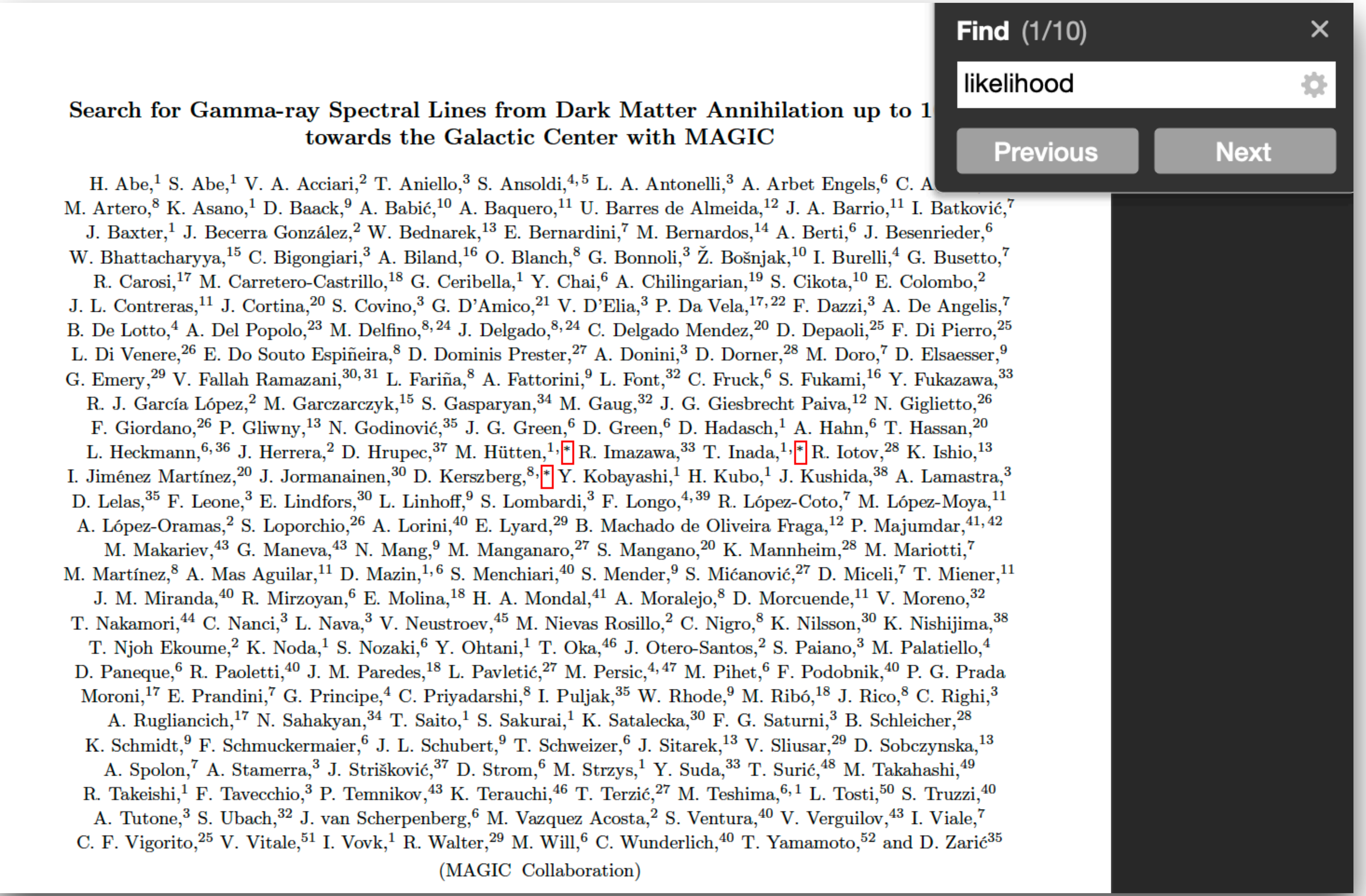
MAGIC Collaboration: V. A. Acciari[1], T. Aniello[2], S. Ansoldi[3,45], L. A. Antonelli[2], A. Arbet Engels[4], C. Arcaro[5], M. Artero[6], K. Asano[7], D. Baack[8], A. Babić[9], A. Baquero[10], U. Barres de Almeida[11], J. A. Barrio[10], I. Batković[5], J. Becerra González[1], W. Bednarek[12], E. Bernardini[5], M. Bernardos[13], A. Berti[4], J. Besenrieder[4], W. Bhattacharyya[14], C. Bigongiari[2], A. Biland[15], O. Blanch[6], H. Bökenkamp[8], G. Bonnoli[2], Ž. Bošnjak[9], I. Burelli[3], G. Busetto[5], R. Carosi[16], M. Carretero-Castrillo[17], G. Ceribella[7], Y. Chai[4], A. Chilingarian[18], S. Cikota[9], E. Colombo[1], J. L. Contreras[10], J. Cortina[19], S. Covino[2], G. D'Amico[20], V. D'Elia[2], P. Da Vela[16,46], F. Dazzi[2], A. De Angelis[5], B. De Lotto[3], A. Del Popolo[21], M. Delfino[6,47], J. Delgado[6,47], C. Delgado Mendez[19], D. Depaoli[22], F. Di Pierro[22], L. Di Venere[23], E. Do Souto Espiñeira[6], D. Dominis Prester[24], A. Donini[2], D. Dorner[25★], M. Doro[5], D. Elsaesser[8], G. Emery[26], V. Fallah Ramazani[27,48★], L. Fariña[6], A. Fattorini[8], L. Font[28], C. Fruck[4], S. Fukami[15], Y. Fukazawa[29], R. J. García López[1], M. Garczarczyk[14], S. Gasparyan[30], M. Gaug[28], J. G. Giesbrecht Paiva[11], N. Giglietto[23], F. Giordano[23], P. Gliwny[12], N. Godinović[31], J. G. Green[4], D. Green[4], D. Hadasch[7], A. Hahn[4], T. Hassan[19], L. Heckmann[4,49], J. Herrera[1], D. Hrupec[32], M. Hütten[7], T. Inada[7], R. Iotov[25], K. Ishio[12], Y. Iwamura[7], I. Jiménez Martínez[19], J. Jormanainen[27], D. Kerszberg[6], Y. Kobayashi[7], H. Kubo[33], J. Kushida[34], A. Lamastra[2], D. Lelas[31], F. Leone[2], E. Lindfors[27], L. Linhoff[8], S. Lombardi[2], F. Longo[3,50], R. López-Coto[5], M. López-Moya[10], A. López-Oramas[1], S. Loporchio[23], A. Lorini[35], E. Lyard[26], B. Machado de Oliveira Fraga[11], P. Majumdar[36,51], M. Makariev[37], G. Maneva[37], M. Manganaro[24], S. Mangano[19], K. Mannheim[25], M. Mariotti[5], M. Martínez[6], A. Mas Aguilar[10], D. Mazin[7,52], S. Menchiari[35], S. Mender[8], S. Mićanović[24], D. Miceli[5], T. Miener[10], J. M. Miranda[35], R. Mirzoyan[4], E. Molina[17], H. A. Mondal[36], A. Moralejo[6], D. Morcuende[10★], V. Moreno[28], T. Nakamori[38], C. Nanci[2], L. Nava[2], V. Neustroev[39], M. Nievas Rosillo[1], C. Nigro[6], K. Nilsson[27], K. Nishijima[34], T. Njoh Ekoume[1], K. Noda[7], S. Nozaki[33], Y. Ohtani[7], T. Oka[33], J. Otero-Santos[1★], S. Paiano[2], M. Palatiello[3], D. Paneque[4★], R. Paoletti[35], J. M. Paredes[17], L. Pavletić[24], M. Persic[3,53], M. Pihet[4], F. Podobnik[35], P. G. Prada Moroni[16], E. Prandini[5], G. Principe[3], C. Priyadarshi[6], I. Puljak[31], W. Rhode[8], M. Ribó[17], J. Rico[6], C. Righi[2], A. Rugliancich[16], N. Sahakyan[30], T. Saito[7], S. Sakurai[7], K. Satalecka[27], F. G. Saturni[2], B. Schleicher[25], K. Schmidt[8], F. Schmuckermaier[4], J. L. Schubert[8], T. Schweizer[4], J. Sitarek[12], V. Sliusar[26], D. Sobczynska[12], A. Spolon[5], A. Stamerra[2], J. Strišković[32], D. Strom[4], M. Strzys[7], Y. Suda[29], T. Surić[40], M. Takahashi[41], R. Takeishi[7], F. Tavecchio[2], P. Temnikov[37], T. Terzić[24], M. Teshima[4,54], L. Tosti[42], S. Truzzi[35], A. Tutone[2], S. Ubach[28], J. van Scherpenberg[4], G. Vanzo[1], M. Vazquez Acosta[1], S. Ventura[35], V. Verguilov[37], I. Viale[5], C. F. Vigorito[22], V. Vitale[43], I. Vovk[7], R. Walter[26], M. Will[4], C. Wunderlich[35], T. Yamamoto[44], D. Zarić[31] J. A. Acosta-Pulido[1], F. D'Ammando[55], T. Hovatta[56,57], S. Kiehlmann[58,59], I. Liodakis[56], C. Leto[60,61], W. Max-Moerbeck[62], L. Pacciani[63], M. Perri[60,64], A. C. S. Readhead[65], R. A. Reeves[66], and F. Verrecchia[60,64]

(Affiliations can be found after the references)

Received September 15, 1996; accepted March 16, 1997

The Likelihood

Let’s take a look at some recent papers

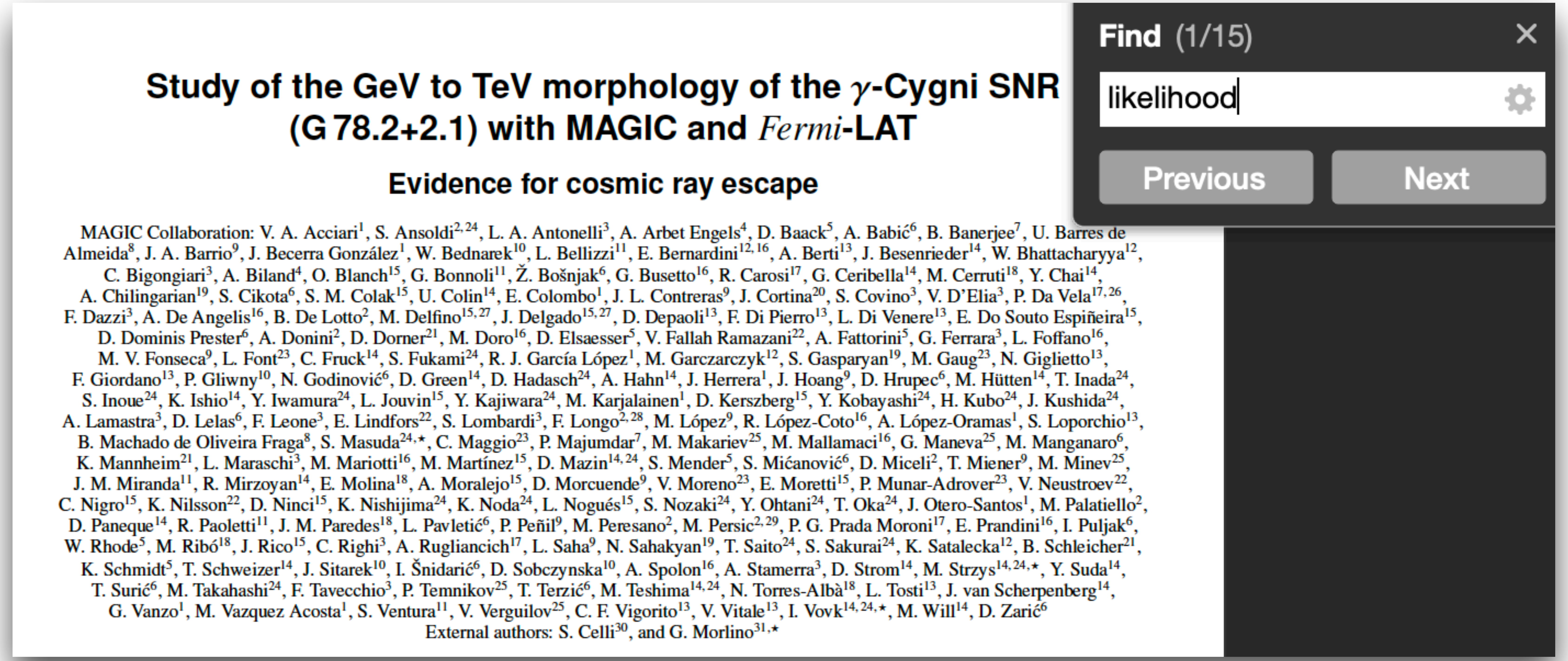


Search for Gamma-ray Spectral Lines from Dark Matter Annihilation up to 1
towards the Galactic Center with MAGIC

H. Abe,[1] S. Abe,[1] V. A. Acciari,[2] T. Aniello,[3] S. Ansoldi,[4,5] L. A. Antonelli,[3] A. Arbet Engels,[6] C. A
M. Artero,[8] K. Asano,[1] D. Baack,[9] A. Babić,[10] A. Baquero,[11] U. Barres de Almeida,[12] J. A. Barrio,[11] I. Batković,[7]
J. Baxter,[1] J. Becerra González,[2] W. Bednarek,[13] E. Bernardini,[7] M. Bernardos,[14] A. Berti,[6] J. Besenrieder,[6]
W. Bhattacharyya,[15] C. Bigongiari,[3] A. Biland,[16] O. Blanch,[8] G. Bonnoli,[3] Ž. Bošnjak,[10] I. Burelli,[4] G. Busetto,[7]
R. Carosi,[17] M. Carretero-Castrillo,[18] G. Ceribella,[1] Y. Chai,[6] A. Chilingarian,[19] S. Cikota,[10] E. Colombo,[2]
J. L. Contreras,[11] J. Cortina,[20] S. Covino,[3] G. D’Amico,[21] V. D’Elia,[3] P. Da Vela,[17,22] F. Dazzi,[3] A. De Angelis,[7]
B. De Lotto,[4] A. Del Popolo,[23] M. Delfino,[8,24] J. Delgado,[8,24] C. Delgado Mendez,[20] D. Depaoli,[25] F. Di Pierro,[25]
L. Di Venere,[26] E. Do Souto Espiñeira,[8] D. Dominis Prester,[27] A. Donini,[3] D. Dorner,[28] M. Doro,[7] D. Elsaesser,[9]
G. Emery,[29] V. Fallah Ramazani,[30,31] L. Fariña,[8] A. Fattorini,[9] L. Font,[32] C. Fruck,[6] S. Fukami,[16] Y. Fukazawa,[33]
R. J. García López,[2] M. Garczarczyk,[15] S. Gasparyan,[34] M. Gaug,[32] J. G. Giesbrecht Paiva,[12] N. Giglietto,[26]
F. Giordano,[26] P. Gliwny,[13] N. Godinović,[35] J. G. Green,[6] D. Green,[6] D. Hadasch,[1] A. Hahn,[6] T. Hassan,[20]
L. Heckmann,[6,36] J. Herrera,[2] D. Hrupec,[37] M. Hütten,[1,*] R. Imazawa,[33] T. Inada,[1,*] R. Iotov,[28] K. Ishio,[13]
I. Jiménez Martínez,[20] J. Jormanainen,[30] D. Kerszberg,[8,*] Y. Kobayashi,[1] H. Kubo,[1] J. Kushida,[38] A. Lamastra,[3]
D. Lelas,[35] F. Leone,[3] E. Lindfors,[30] L. Linhoff,[9] S. Lombardi,[3] F. Longo,[4,39] R. López-Coto,[7] M. López-Moya,[11]
A. López-Oramas,[2] S. Loporchio,[26] A. Lorini,[40] E. Lyard,[29] B. Machado de Oliveira Fraga,[12] P. Majumdar,[41,42]
M. Makariev,[43] G. Maneva,[43] N. Mang,[9] M. Manganaro,[27] S. Mangano,[20] K. Mannheim,[28] M. Mariotti,[7]
M. Martínez,[8] A. Mas Aguilar,[11] D. Mazin,[1,6] S. Menchiari,[40] S. Mender,[9] S. Mićanović,[27] D. Miceli,[7] T. Miener,[11]
J. M. Miranda,[40] R. Mirzoyan,[6] E. Molina,[18] H. A. Mondal,[41] A. Moralejo,[8] D. Morcuende,[11] V. Moreno,[32]
T. Nakamori,[44] C. Nanci,[3] L. Nava,[3] V. Neustroev,[45] M. Nievas Rosillo,[2] C. Nigro,[8] K. Nilsson,[30] K. Nishijima,[38]
T. Njoh Ekoume,[2] K. Noda,[1] S. Nozaki,[6] Y. Ohtani,[1] T. Oka,[46] J. Otero-Santos,[2] S. Paiano,[3] M. Palatiello,[4]
D. Paneque,[6] R. Paoletti,[40] J. M. Paredes,[18] L. Pavletić,[27] M. Persic,[4,47] M. Pihet,[6] F. Podobnik,[40] P. G. Prada
Moroni,[17] E. Prandini,[7] G. Principe,[4] C. Priyadarshi,[8] I. Puljak,[35] W. Rhode,[9] M. Ribó,[18] J. Rico,[8] C. Righi,[3]
A. Rugliancich,[17] N. Sahakyan,[34] T. Saito,[1] S. Sakurai,[1] K. Satalecka,[30] F. G. Saturni,[3] B. Schleicher,[28]
K. Schmidt,[9] F. Schmuckermaier,[6] J. L. Schubert,[9] T. Schweizer,[6] J. Sitarek,[13] V. Sliusar,[29] D. Sobczynska,[13]
A. Spolon,[7] A. Stamerra,[3] J. Strišković,[37] D. Strom,[6] M. Strzys,[1] Y. Suda,[33] T. Surić,[48] M. Takahashi,[49]
R. Takeishi,[1] F. Tavecchio,[3] P. Temnikov,[43] K. Terauchi,[46] T. Terzić,[27] M. Teshima,[6,1] L. Tosti,[50] S. Truzzi,[40]
A. Tutone,[3] S. Ubach,[32] J. van Scherpenberg,[6] M. Vazquez Acosta,[2] S. Ventura,[40] V. Verguilov,[43] I. Viale,[7]
C. F. Vigorito,[25] V. Vitale,[51] I. Vovk,[1] R. Walter,[29] M. Will,[6] C. Wunderlich,[40] T. Yamamoto,[52] and D. Zarić[35]
(MAGIC Collaboration)

The Likelihood

Let’s take a look at some recent papers



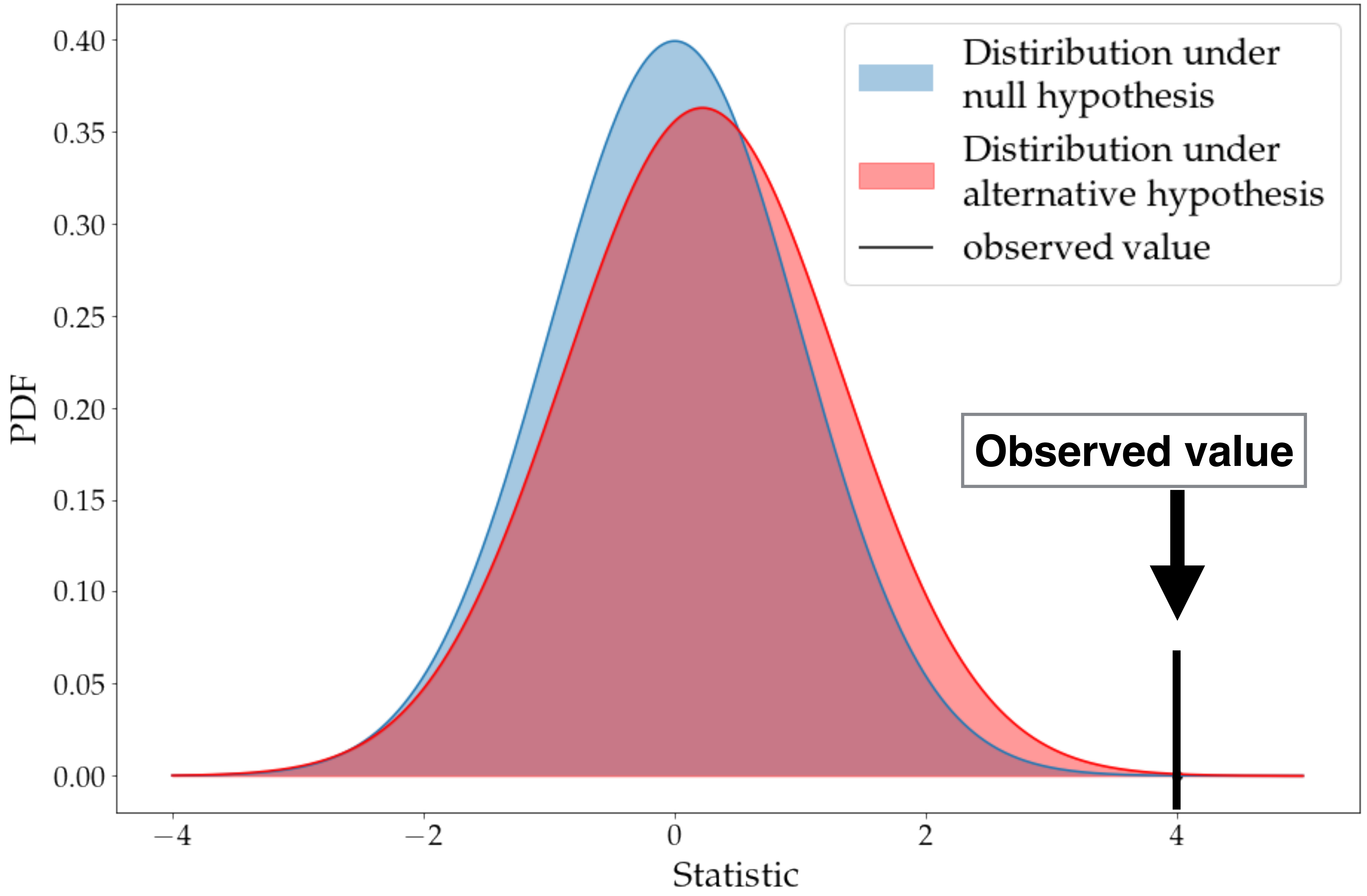
Study of the GeV to TeV morphology of the γ-Cygni SNR (G 78.2+2.1) with MAGIC and *Fermi*-LAT

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

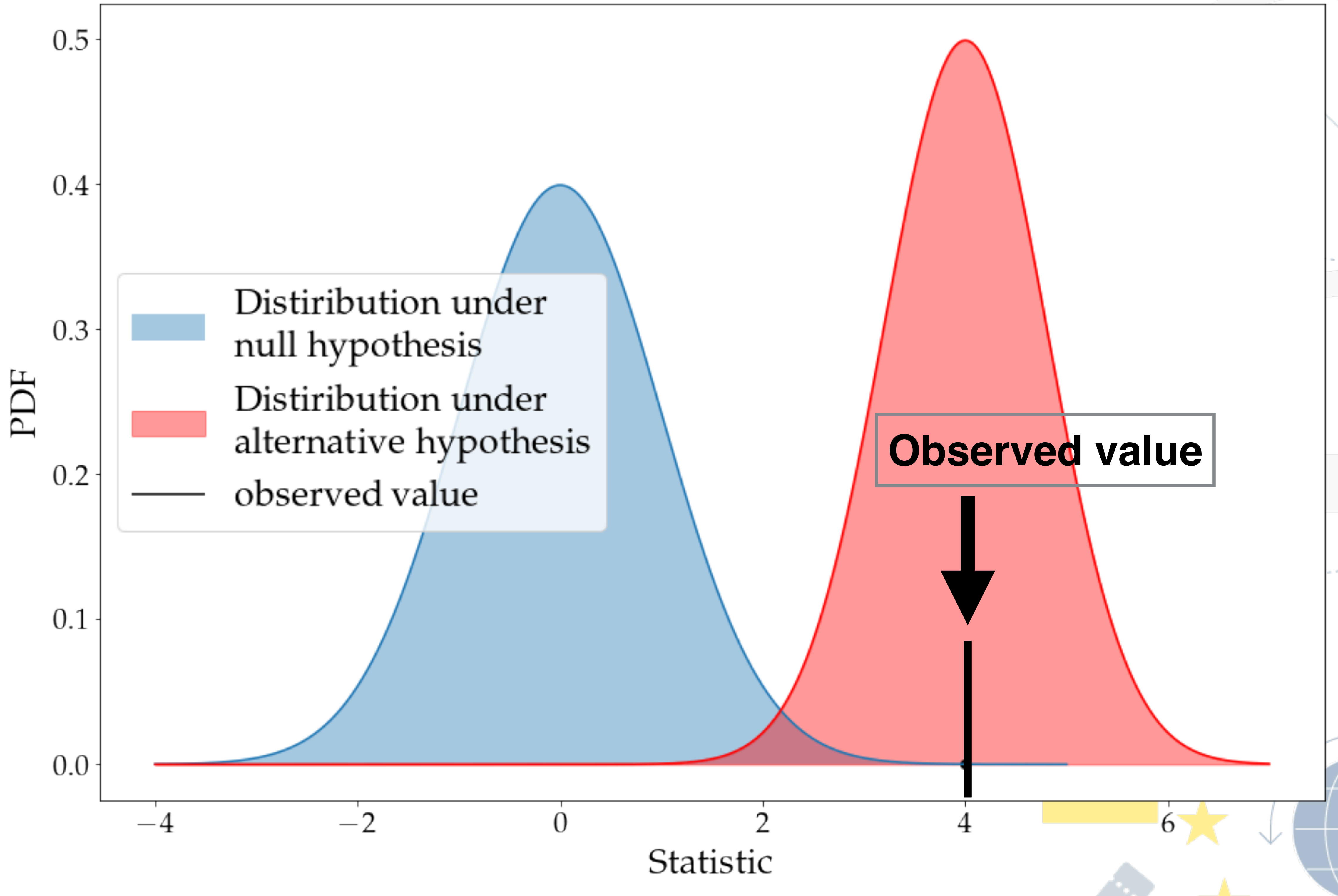
SCENARIO 1



Frequentist conclusion:

The null hypothesis is rejected at 4 sigma level

SCENARIO 2

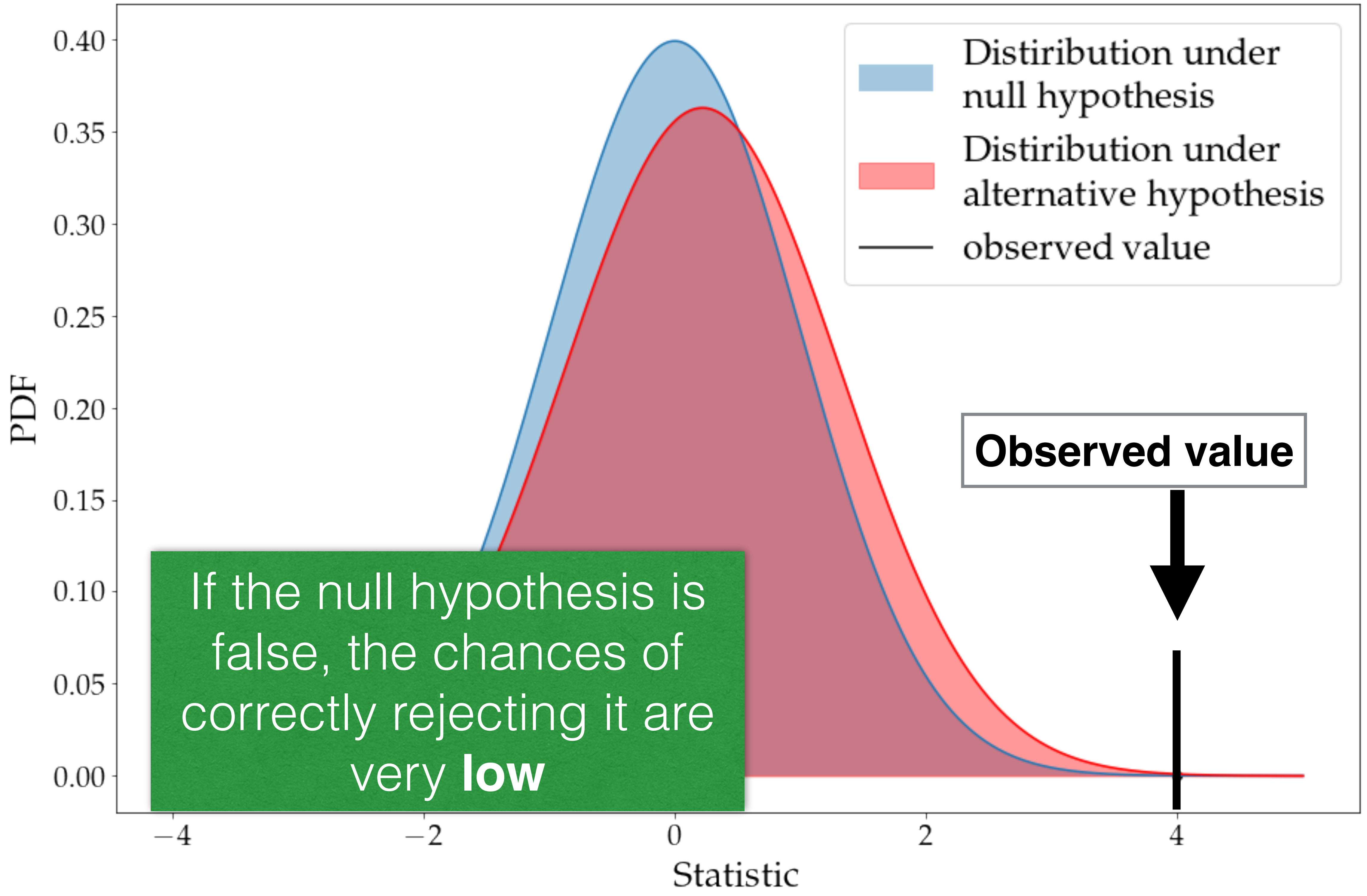


Frequentist conclusion:

The null hypothesis is rejected at 4 sigma level

The Likelihood

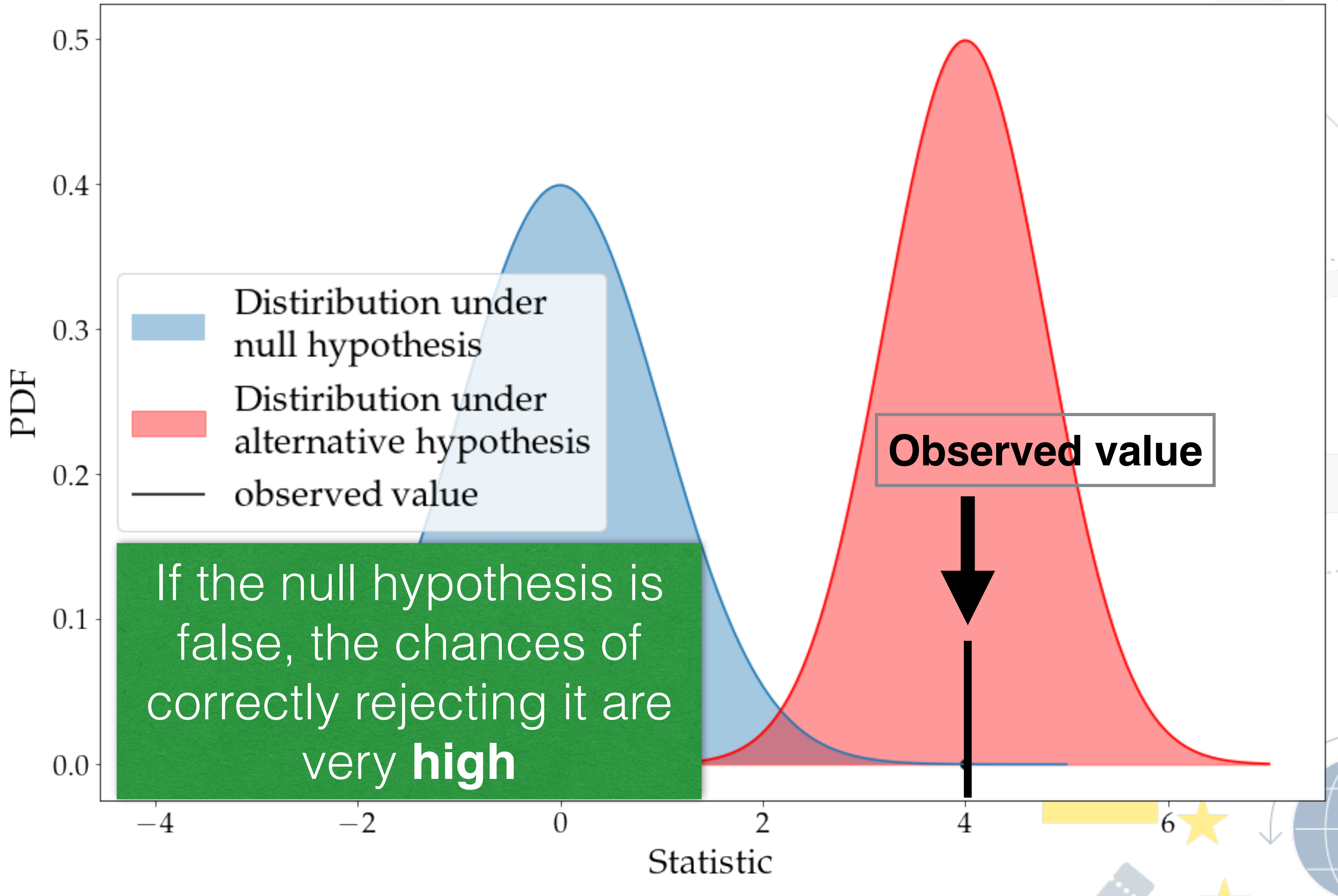
SCENARIO 1



Frequentist conclusion:

The null hypothesis is rejected at 4 sigma level

SCENARIO 2



Frequentist conclusion:

The null hypothesis is rejected at 4 sigma level

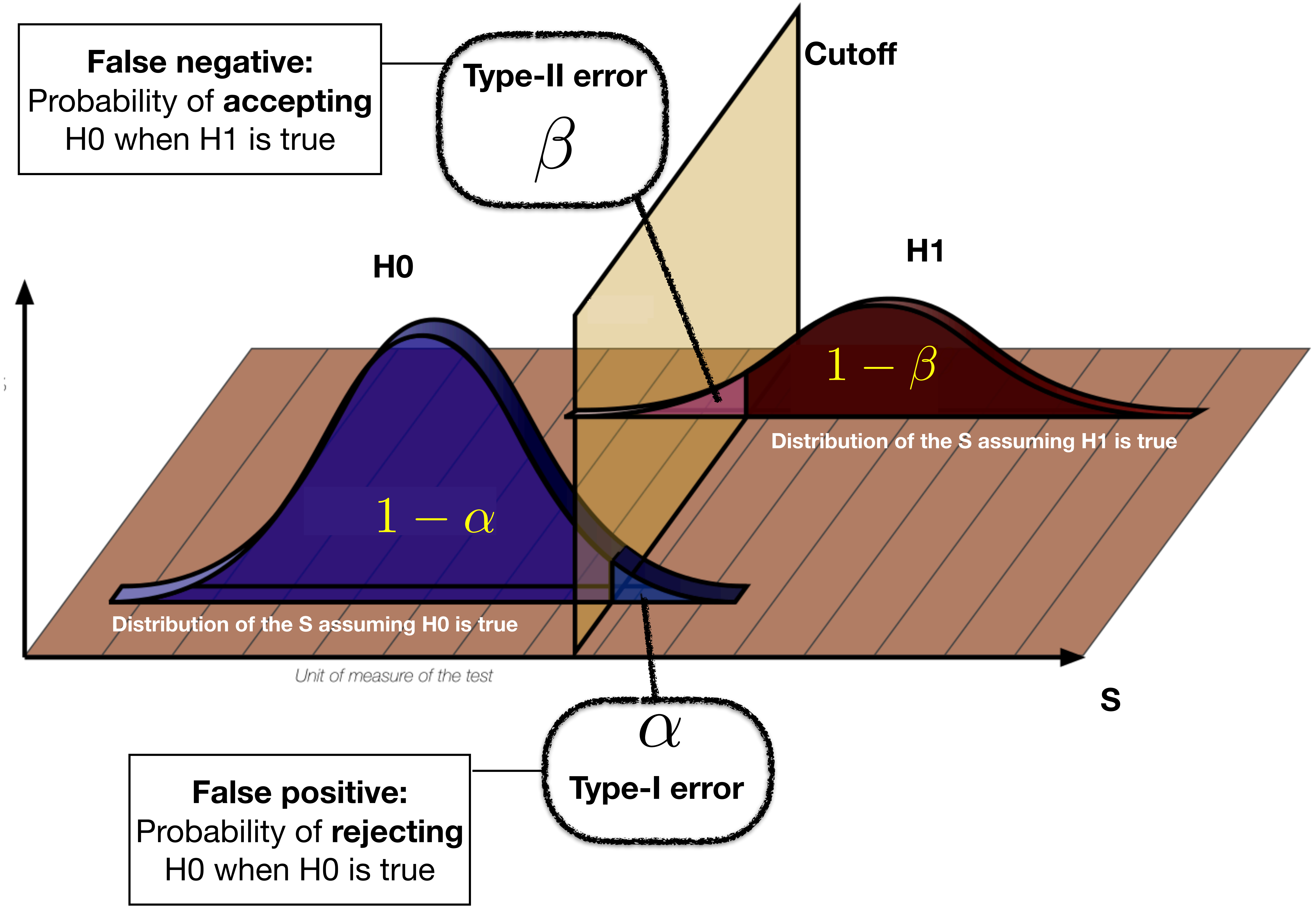
The Likelihood

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

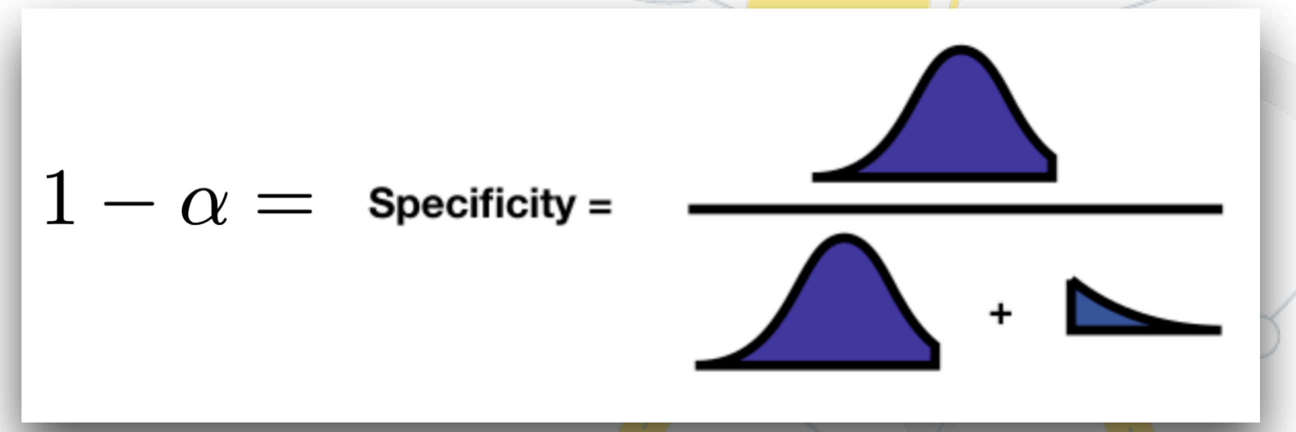
You want your statistic to be

POWERFUL!

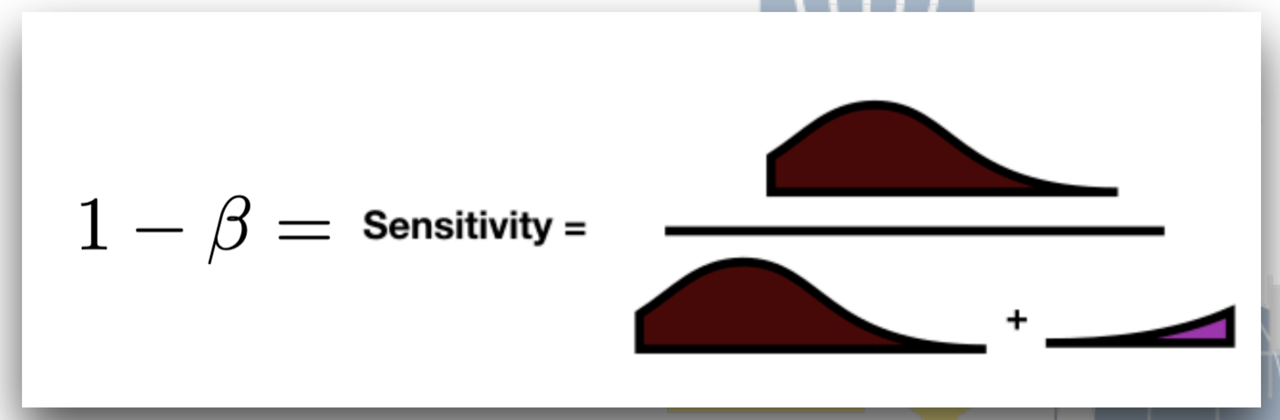
The Likelihood



Confidence level

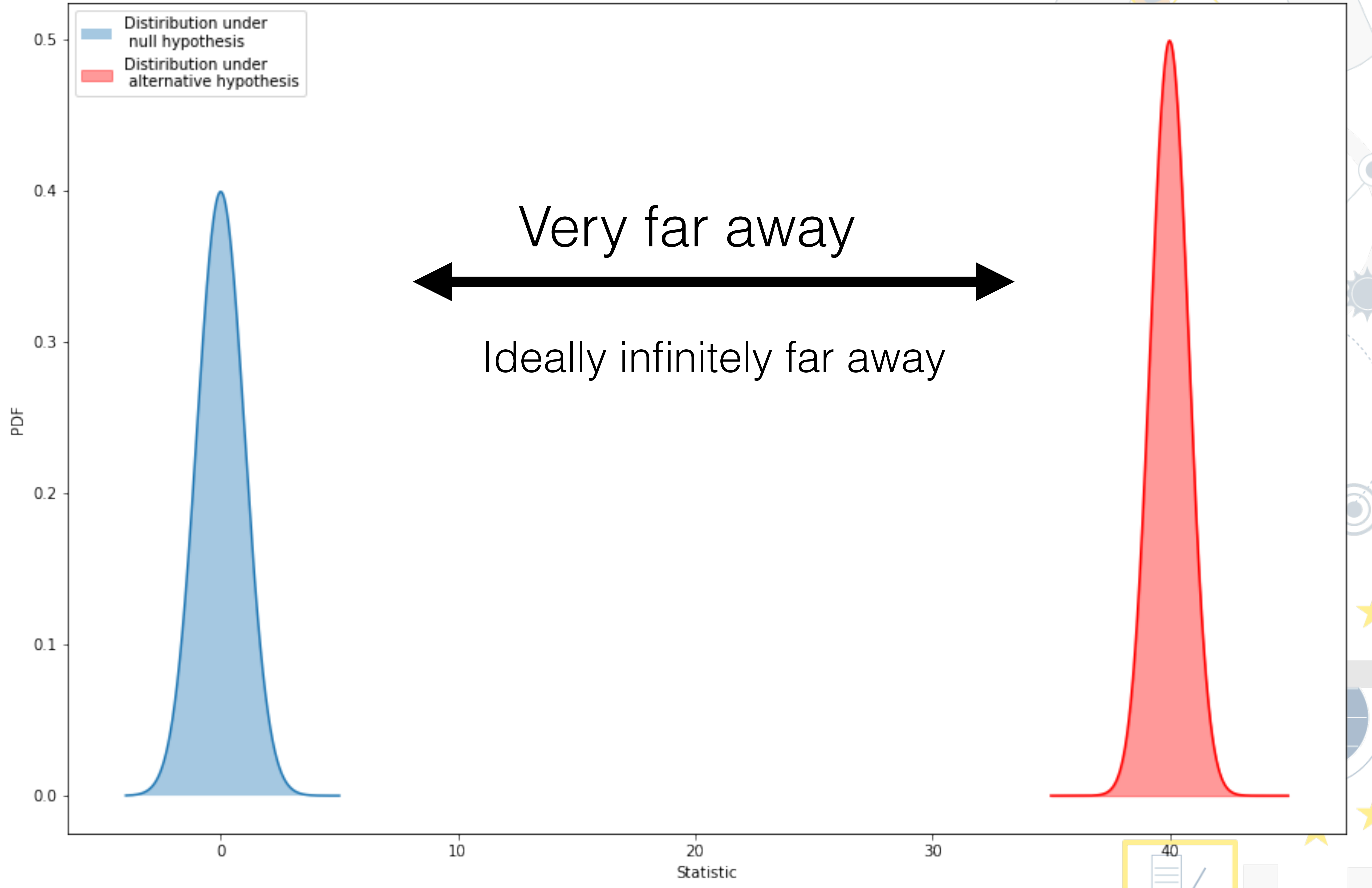


Power of the test



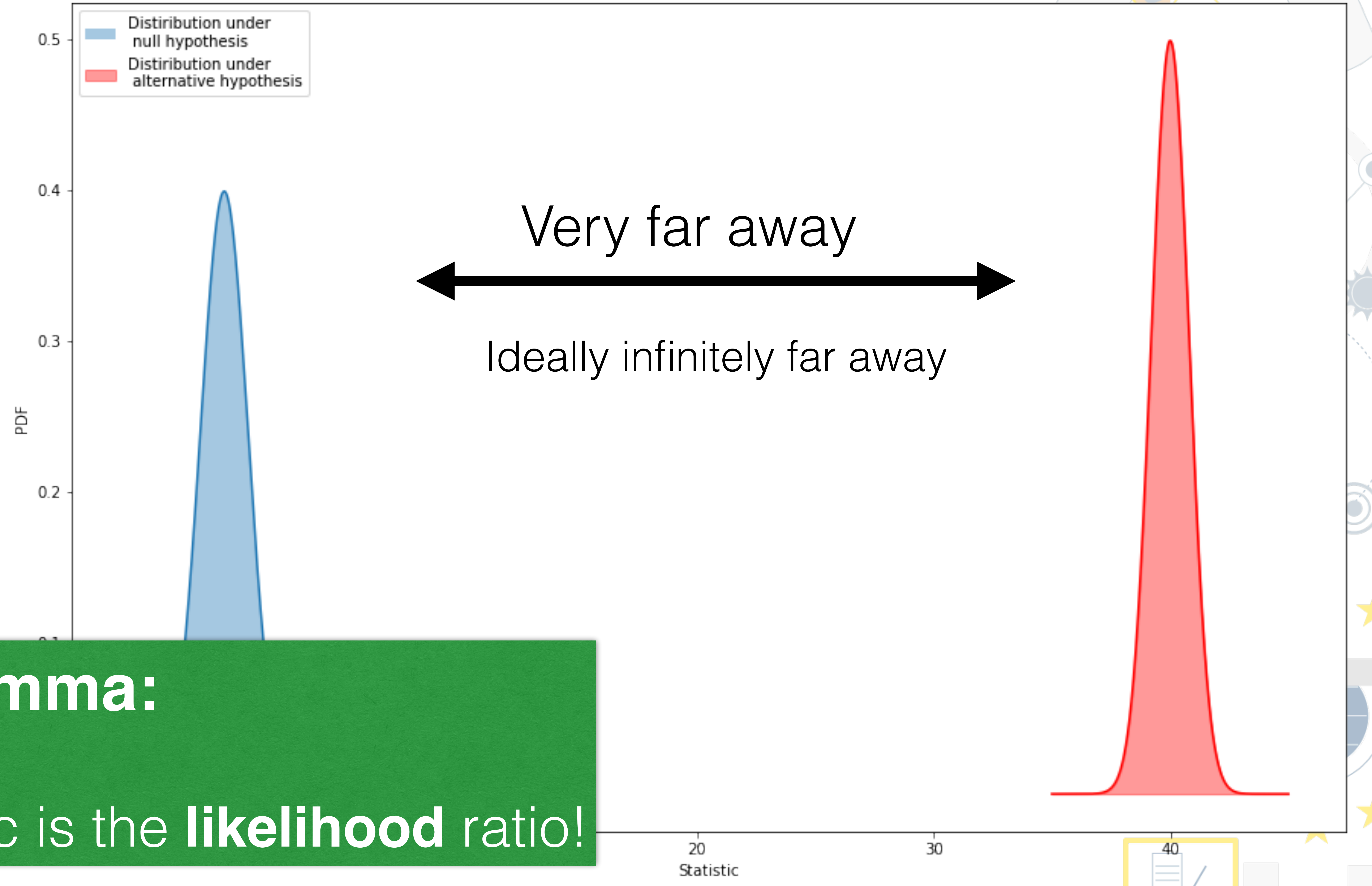
The Likelihood

Ideal case:



The Likelihood

Ideal case:



In reality:

Neyman–Pearson lemma:

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

The Likelihood

Definition of likelihood:

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

Definition of 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

Another very useful property of the likelihood ratio

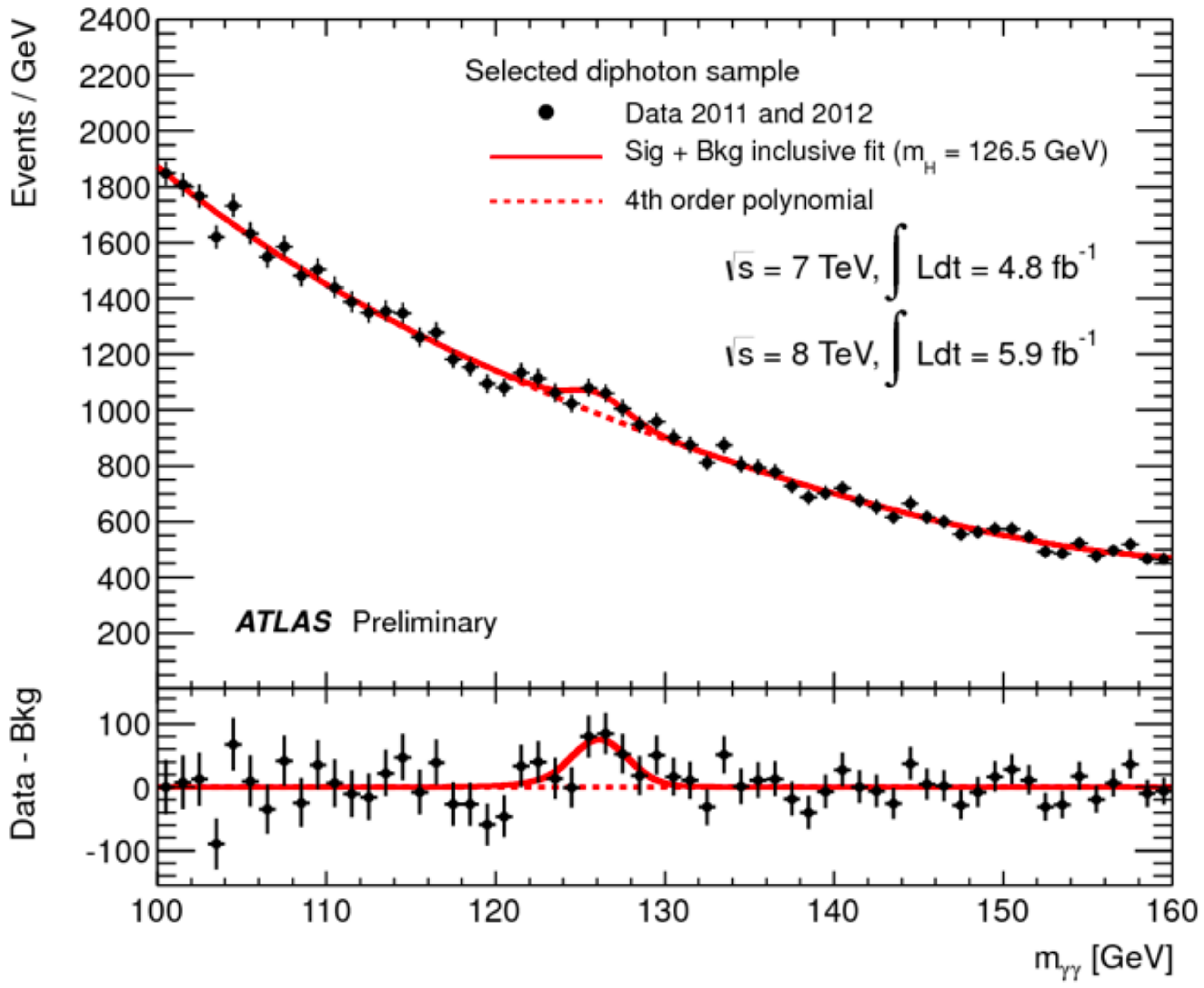
THE WILKS' THEOREM

$$-2 \cdot \log \Delta \mathscr{L} \sim \chi^2$$

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

The Likelihood

Example:



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

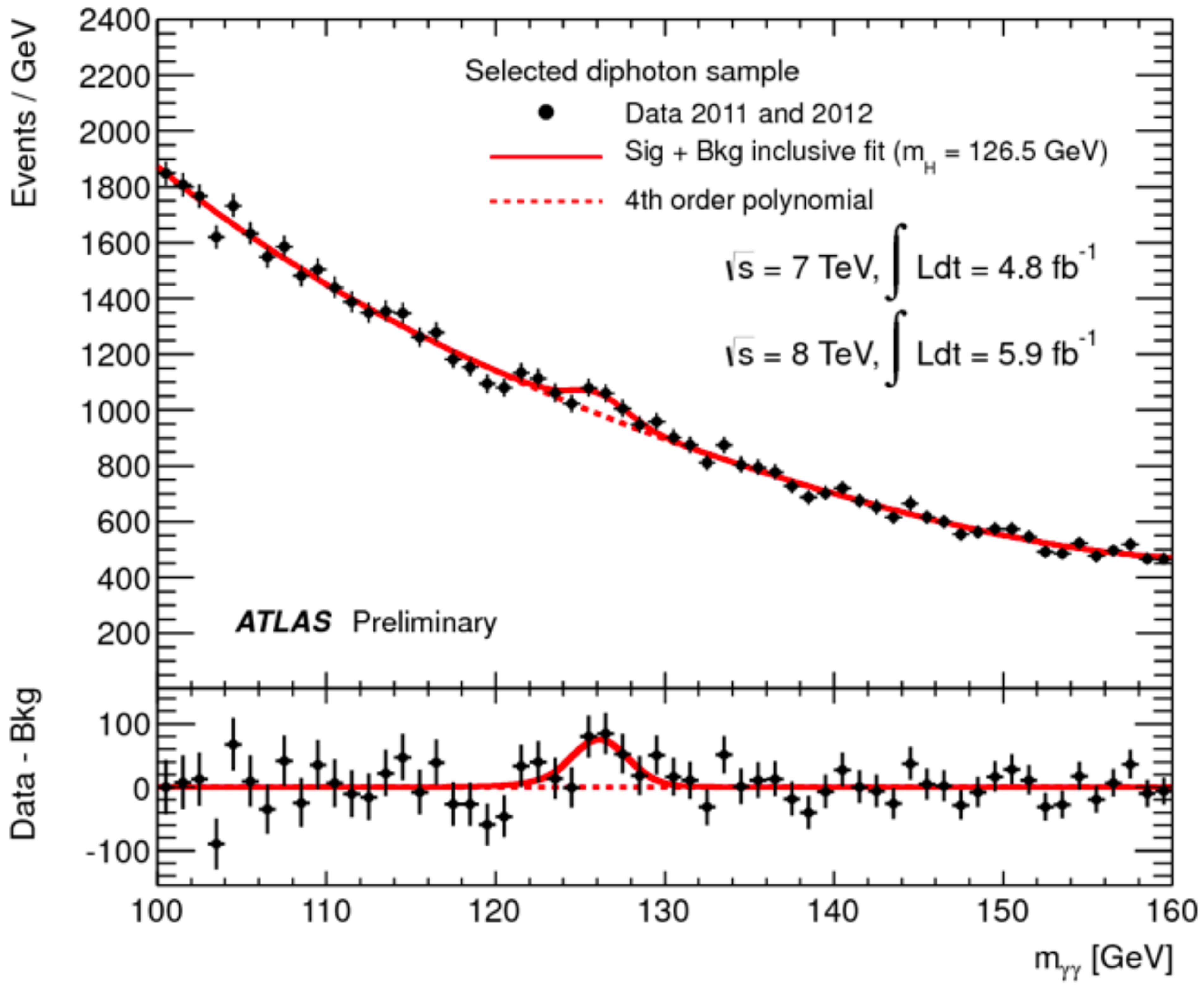
Let's figure out how they were able to make such a claim with a **Toy Model** and with the **theory** we have learned so far

The Likelihood

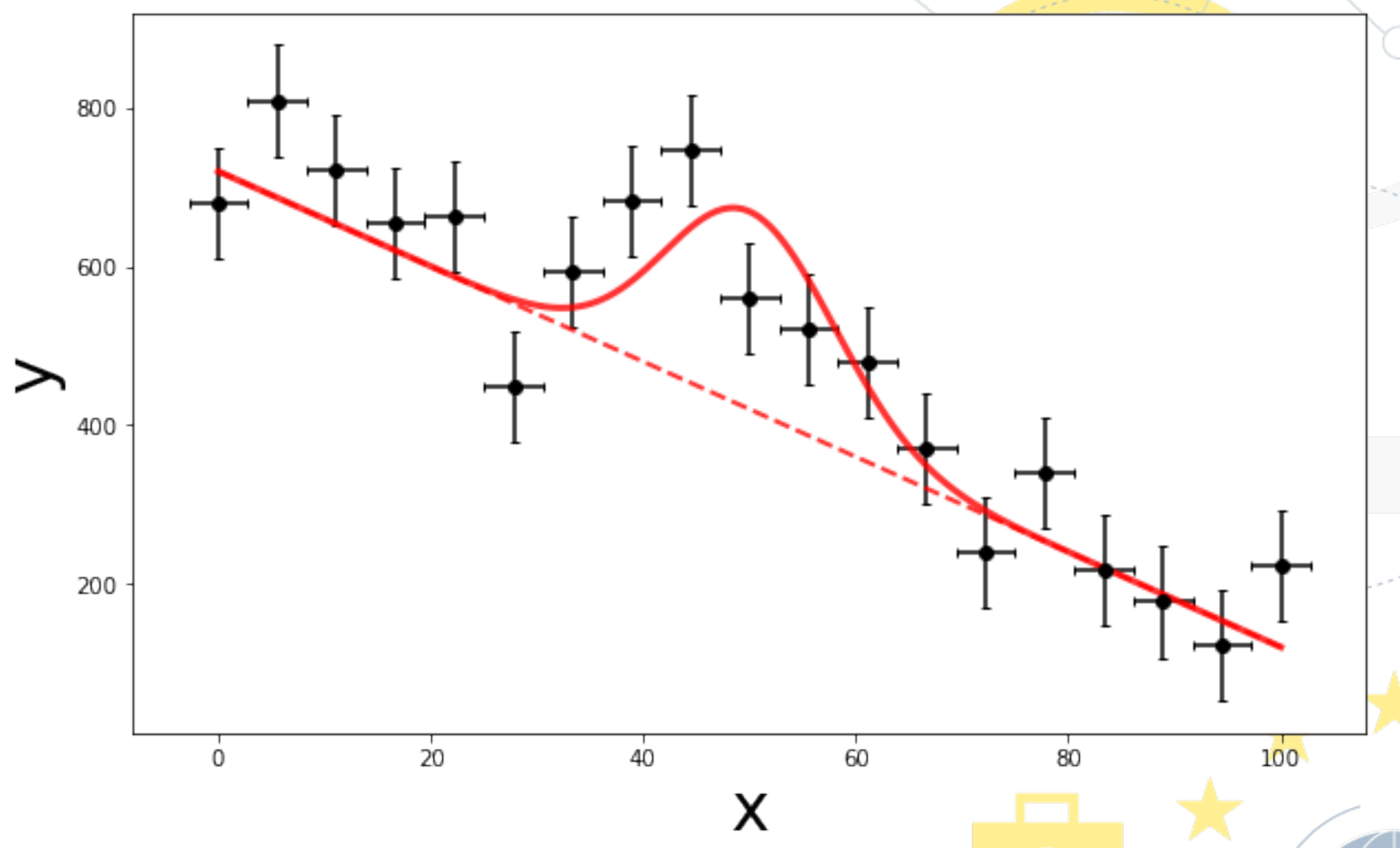
Example:

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

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



Toy Model



Null hypothesis H0
a = 0

Alternative hyp. H1
a = 5

The Likelihood

Example:

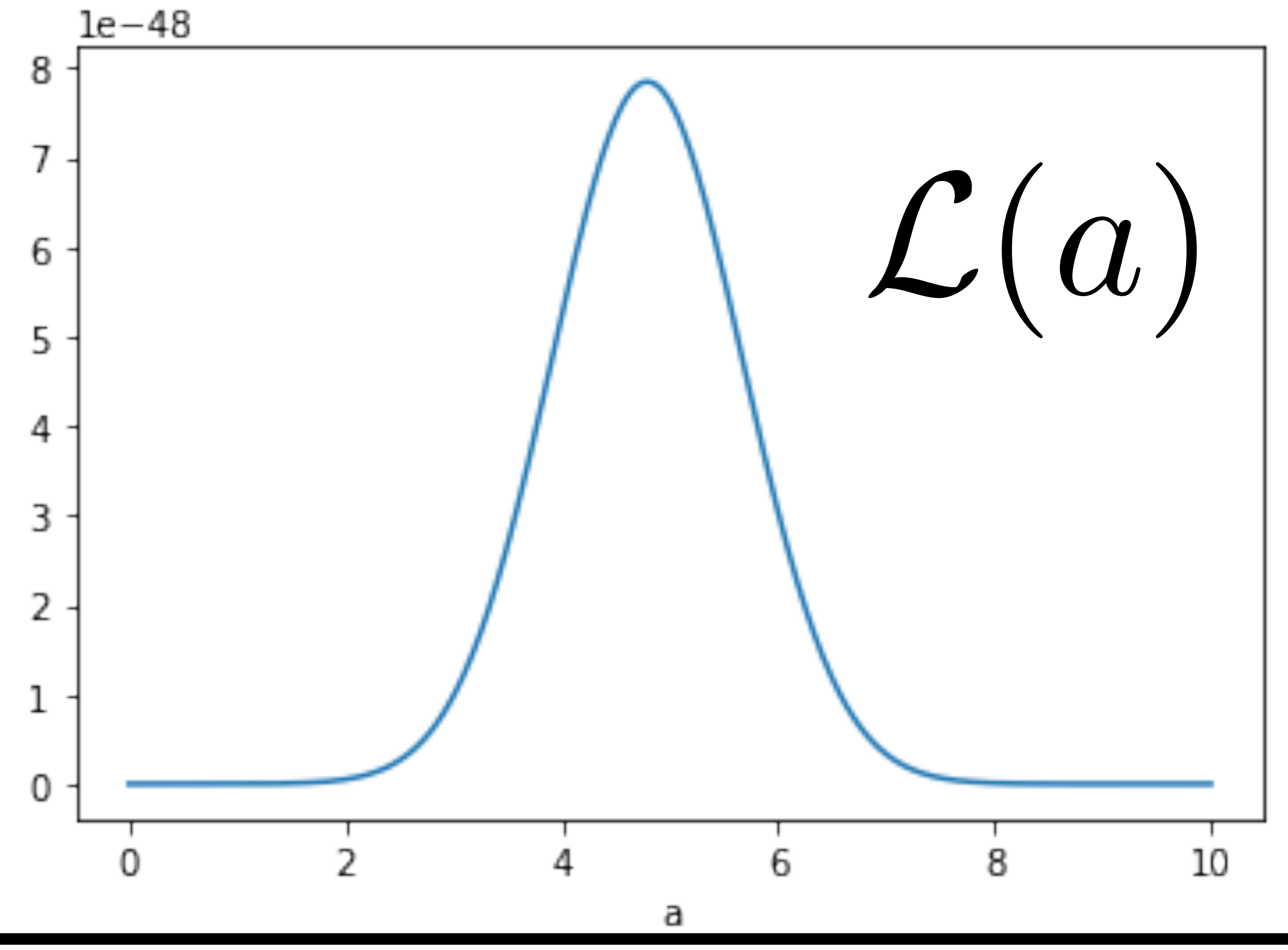
$$y' = mx + q + a \cdot \mathcal{G}(x; \mu = 50, \sigma = 8)$$

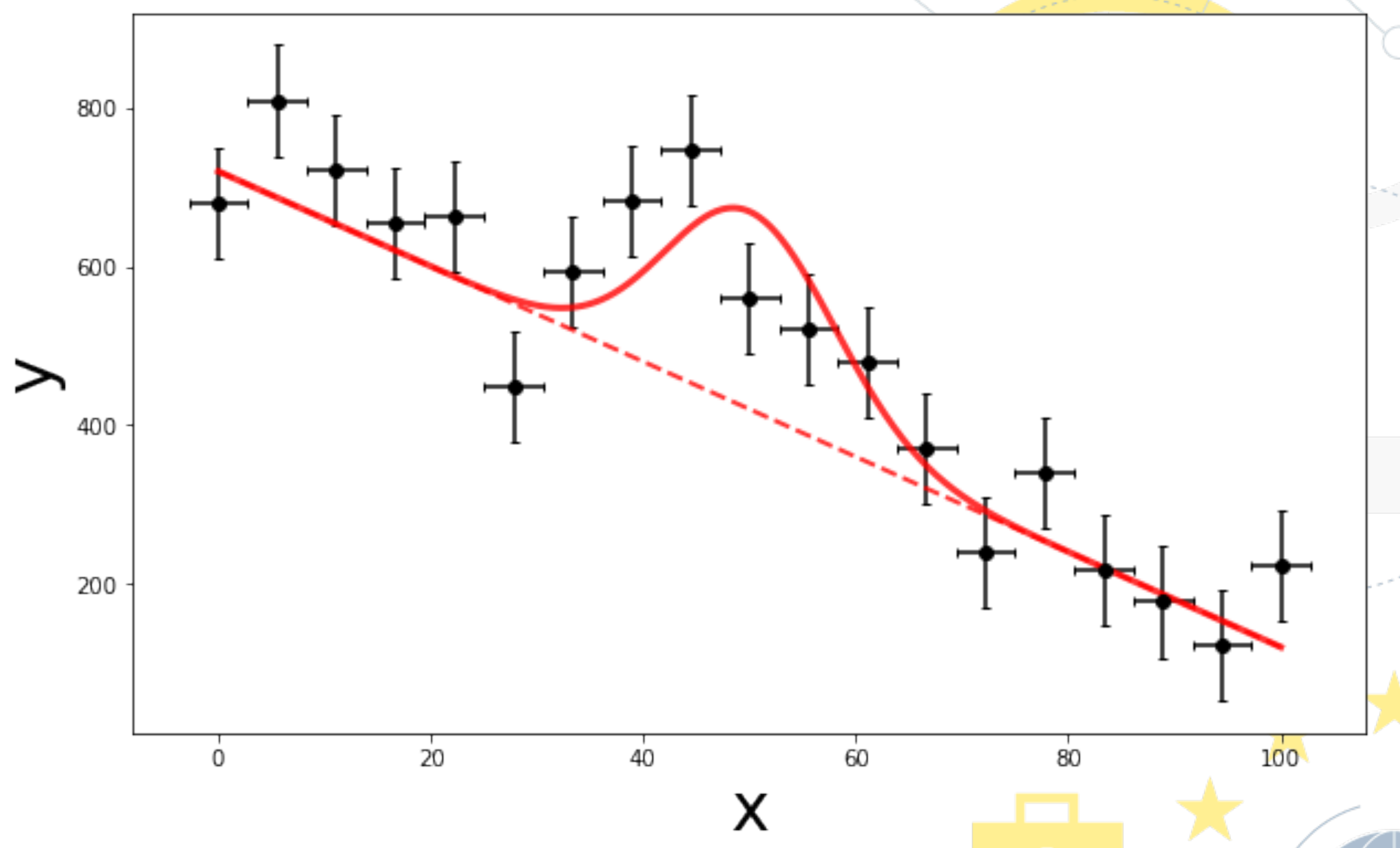
$$y \sim \mathcal{N}(\mu = y', \sigma = 70)$$

Likelihood

$$\mathcal{L}(a) \equiv p(\vec{x}, \vec{y}|a) = \prod_i p(x_i, y_i|a)$$

$$p(x_i, y_i|a) \propto e^{-\frac{1}{2}\left(\frac{y_i'(a) - y_i}{\sigma}\right)^2}$$





Null hypothesis H0
a = 0

Alternative hyp. H1
a = 5

The Likelihood

Example:

Likelihood

$$\mathcal{L}(a) \equiv p(\vec{x}, \vec{y}|a) = \prod_i p(x_i, y_i|a)$$

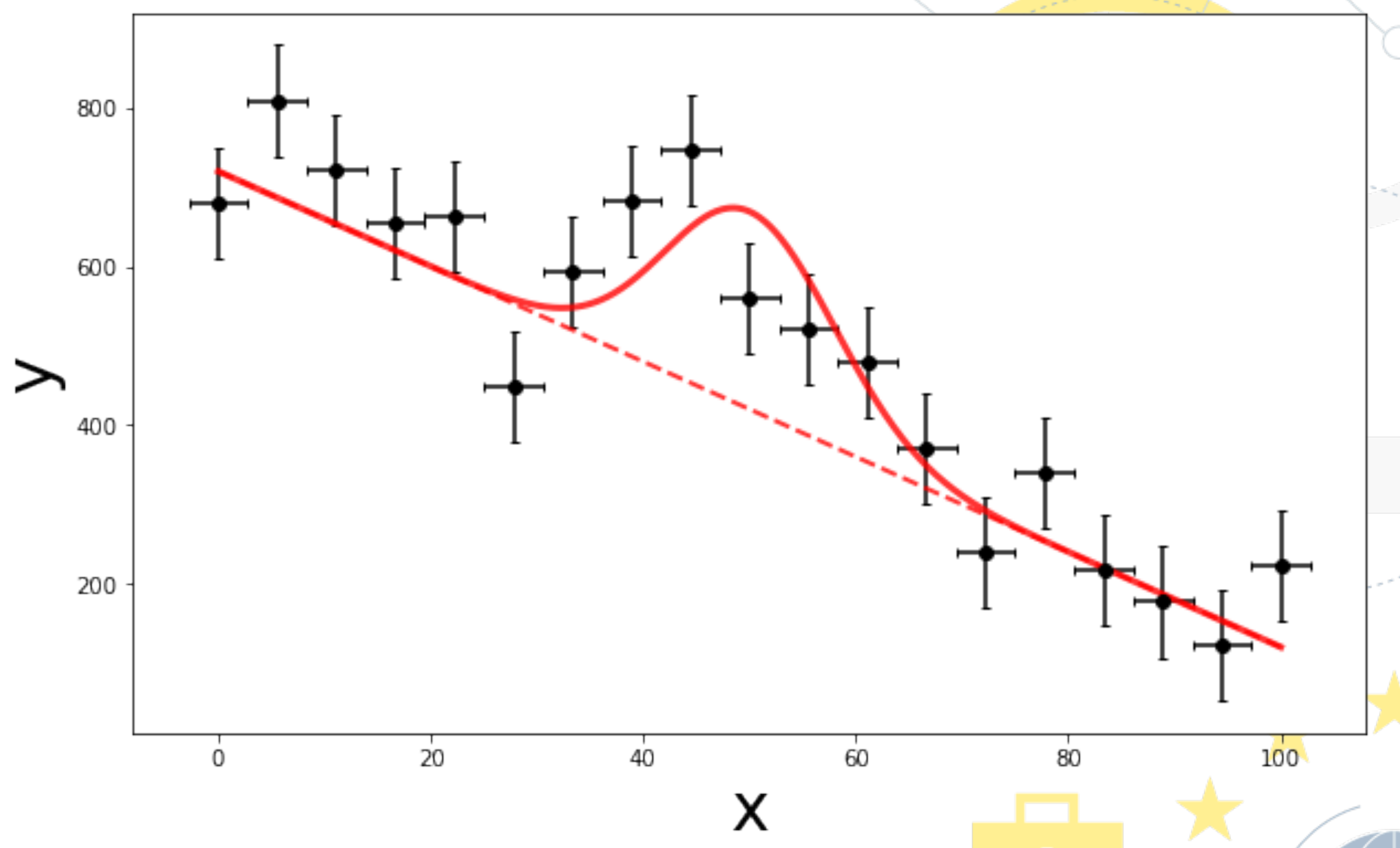
$$p(x_i, y_i|a) \propto e^{-\frac{1}{2}\left(\frac{y_i'(a)-y_i}{\sigma}\right)^2}$$

$$\mathcal{S} = \frac{\mathcal{L}(a=0)}{\mathcal{L}(a=\hat{a})} = 3.52 \cdot 10^{-7}$$

How do we interpret this value of the **statistic**?

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

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



Null hypothesis H0
a = 0

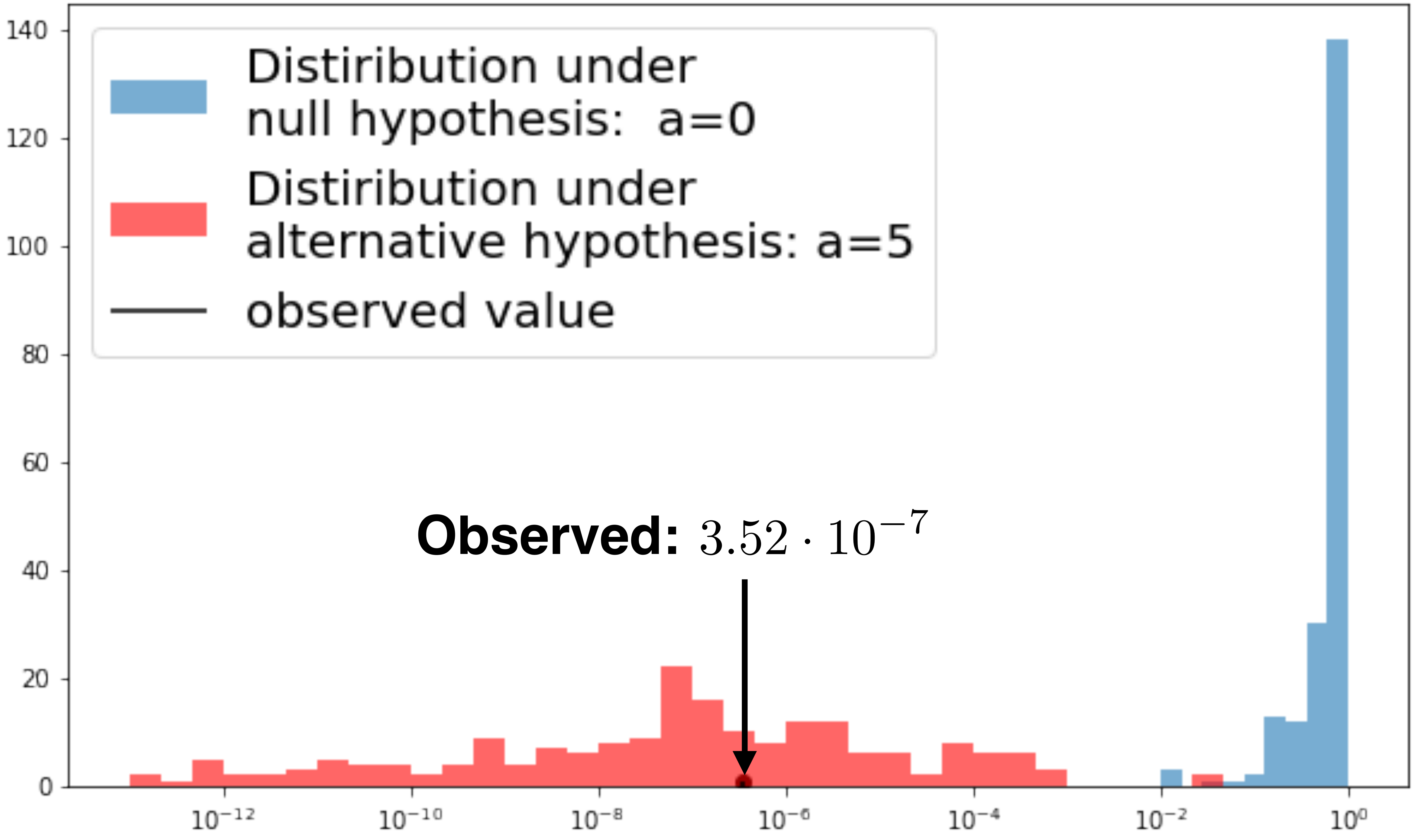
Alternative hyp. H1
a = 5

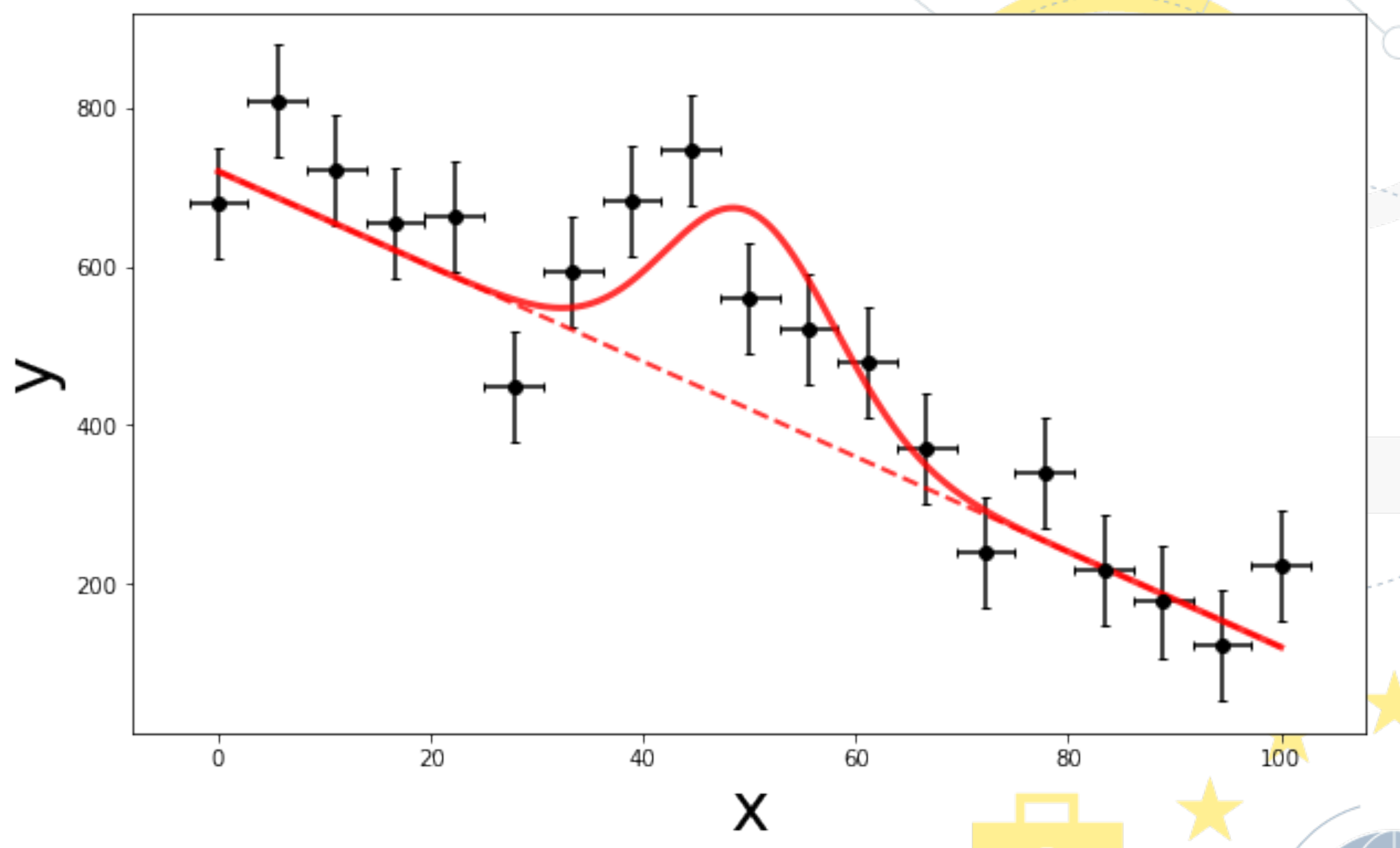
The Likelihood

Example:

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

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





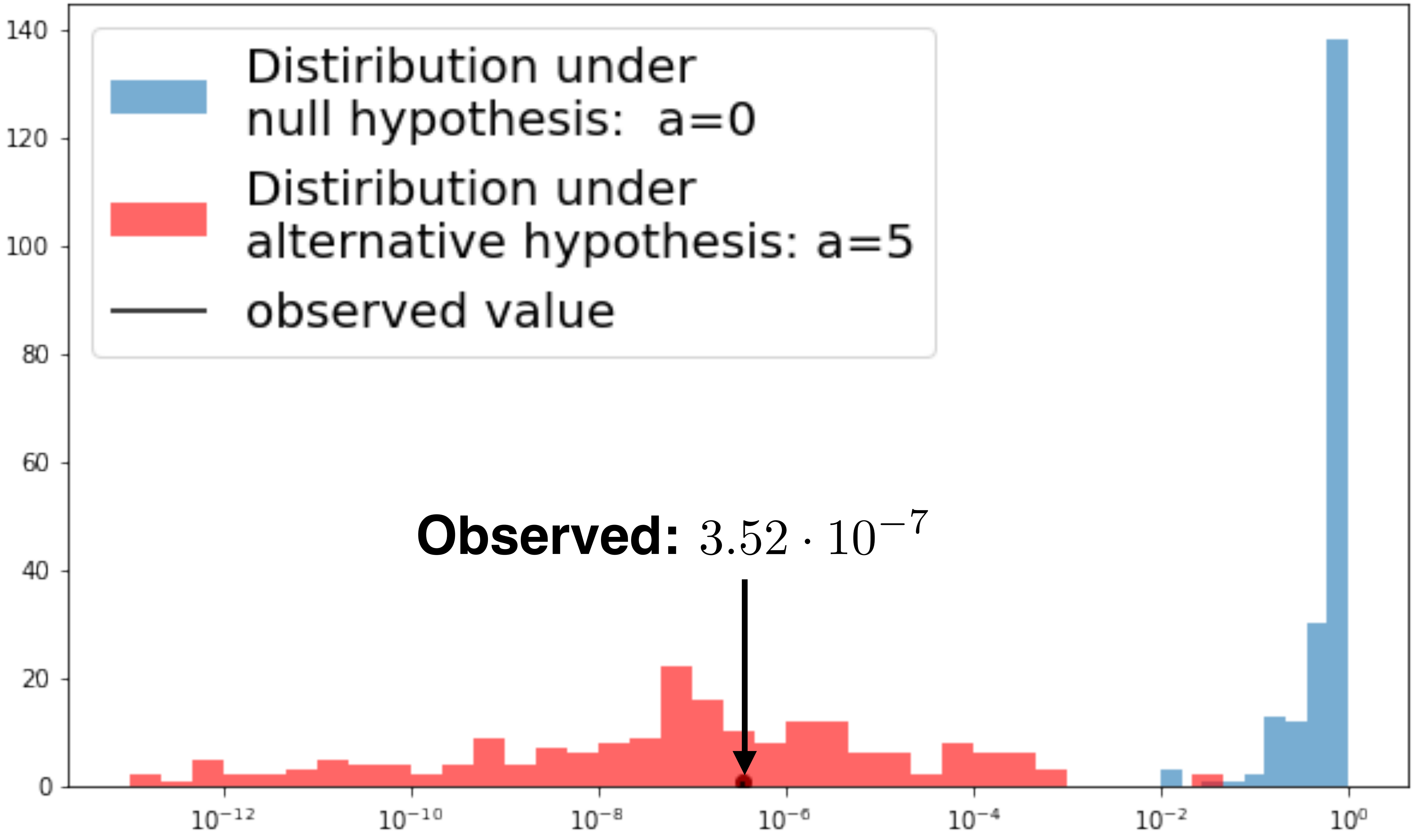
$$\mathcal{S} = \frac{\mathcal{L}(a=0)}{\mathcal{L}(a=\hat{a})}$$

Null hypothesis H0
a = 0

Alternative hyp. H1
a = 5

The Likelihood

Example:



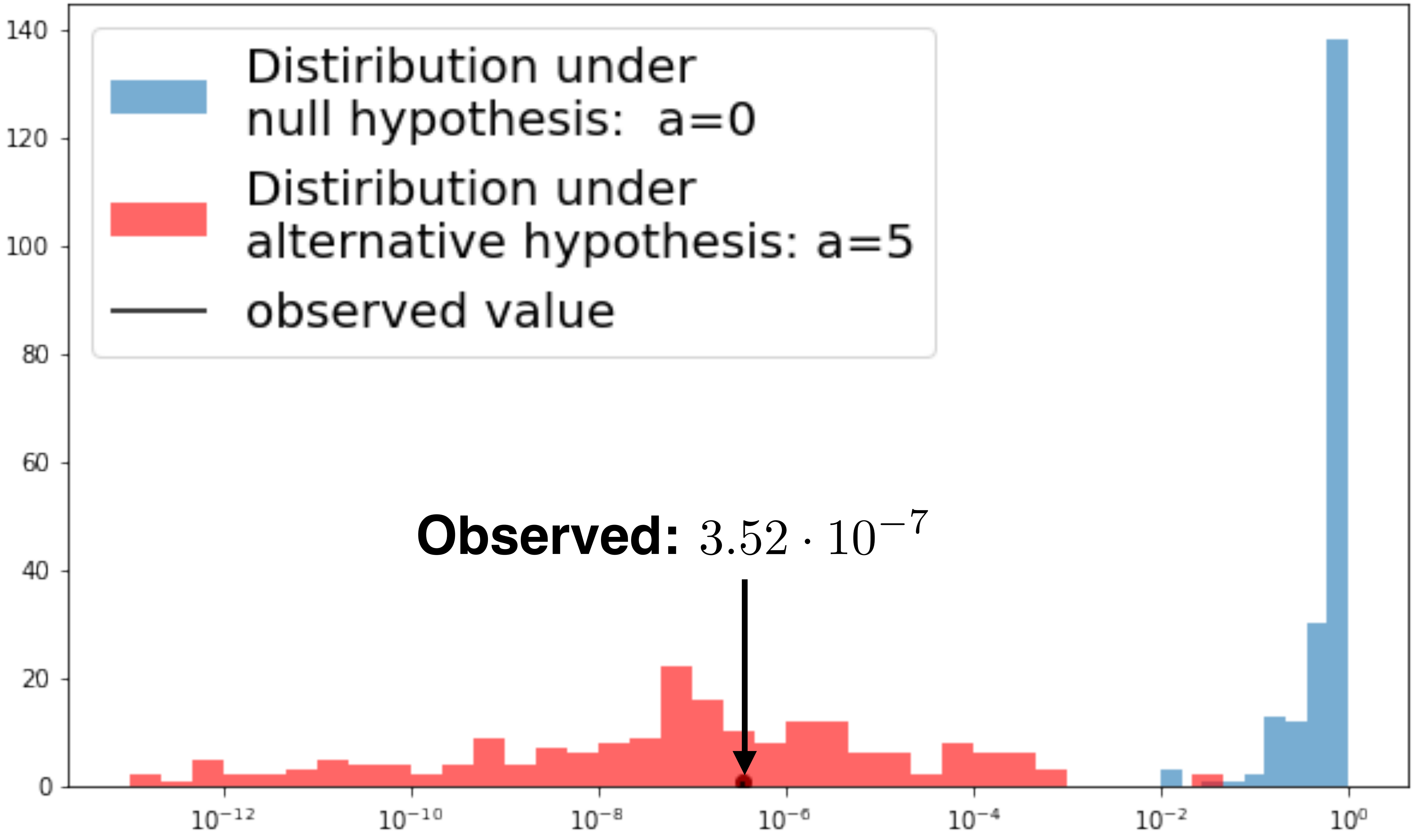
$$\mathcal{S} = \frac{\mathcal{L}(a=0)}{\mathcal{L}(a=\hat{a})}$$

Such a value of the **statistic** is more luckily to have been produced by the **alternative** hypothesis rather than by the **null hypothesis**!

Therefore, we can exclude the null hypothesis and be quite sure of avoiding a type I error.

But with what confidence?

The Likelihood

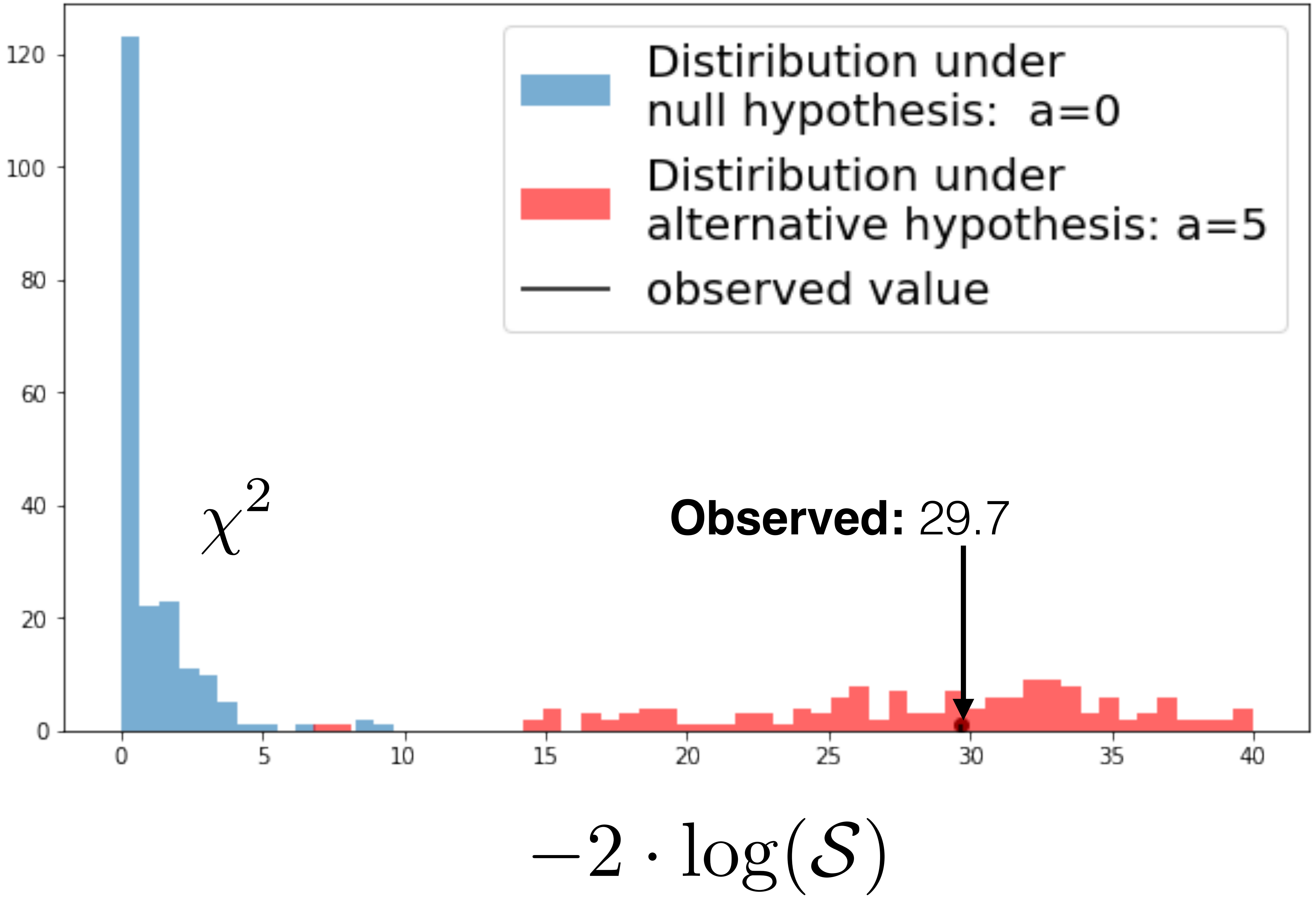


$$\mathcal{S} = \frac{\mathcal{L}(a=0)}{\mathcal{L}(a=\hat{a})}$$

Taking the $-2 \cdot \log(\mathcal{S})$ the **blu** distribution becomes a χ^2 distribution

This is known as the **Wilks' theorem**

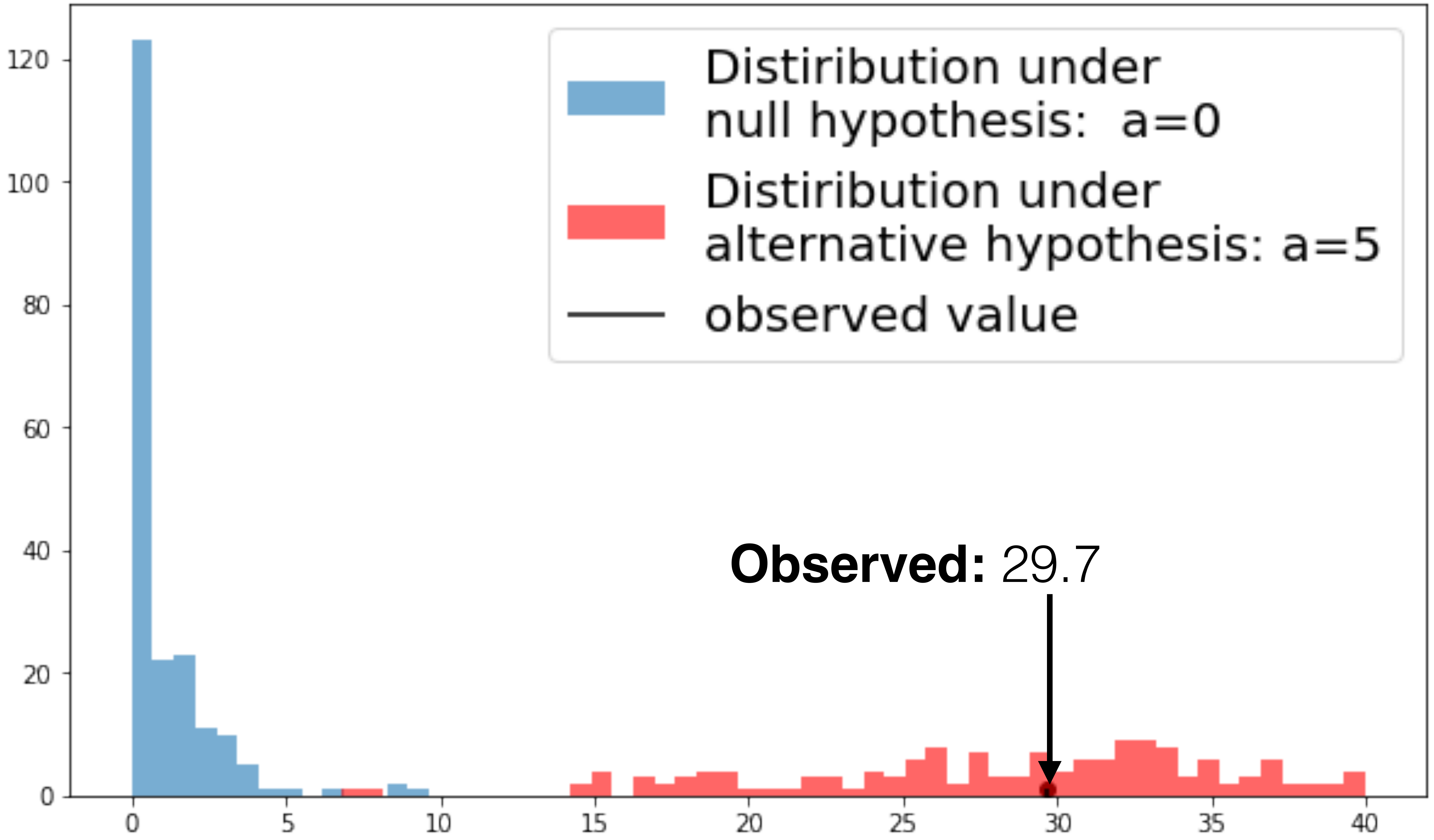
The Likelihood



Taking the $-2 \cdot \log(\mathcal{S})$ the **blu** distribution becomes a χ^2 distribution

This is known as the **Wilks' theorem**

The Likelihood



$$-2 \cdot \log(\mathcal{S})$$

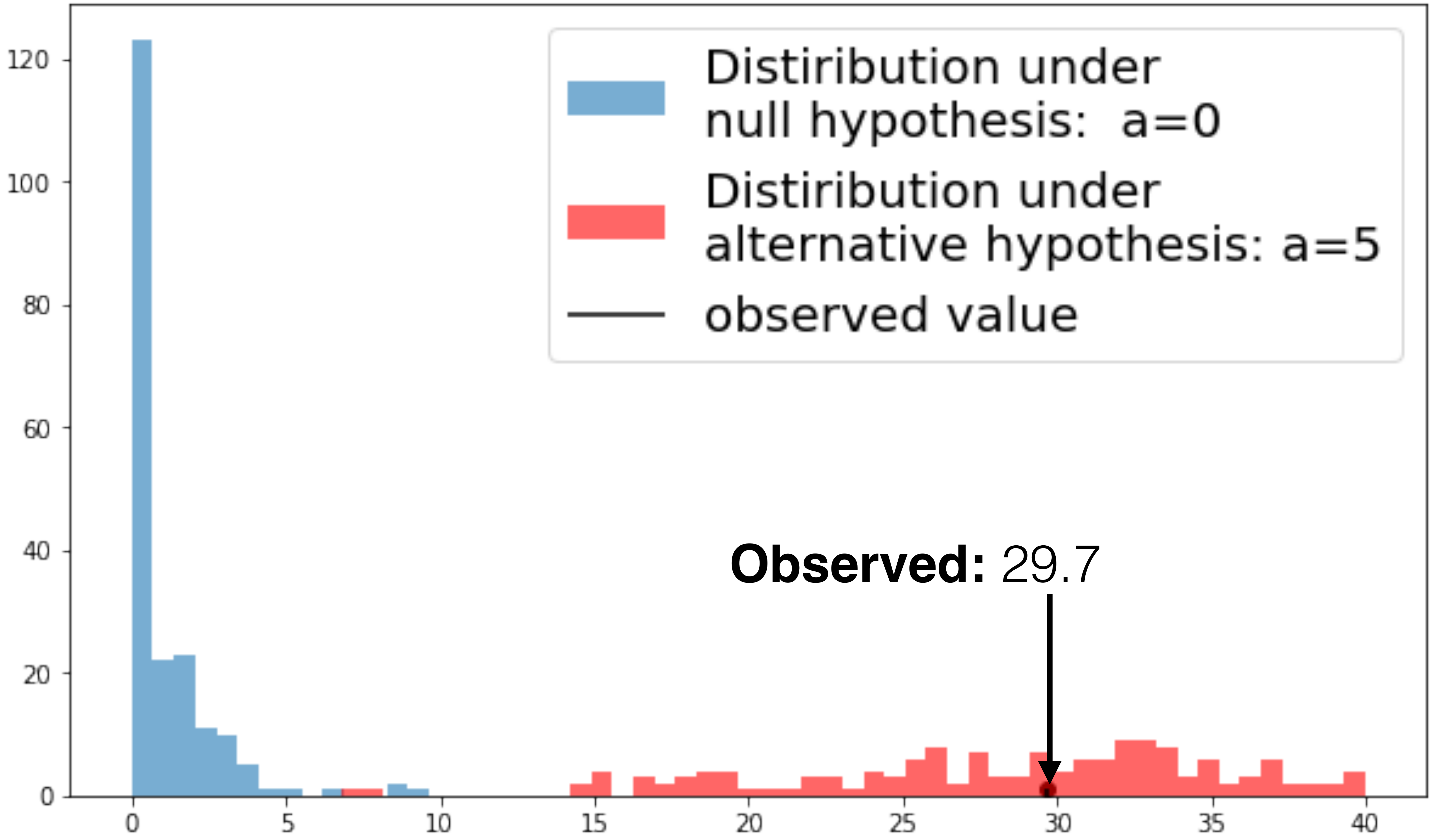
$$\text{p-value} = \int_{29.7}^{\infty} dx\ \chi^2(x) \simeq 5 \cdot 10^{-8}$$

Converting the p-value to a "sigma"

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

We are above the 5 sigmas, we can therefore claim a **discovery**!

The Likelihood



$$-2 \cdot \log(\mathcal{S})$$

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

Converting the p-value to a “sigma”

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

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

Why?

Statistical inference in On/Off measurement

Statistical inference in On/Off measurement

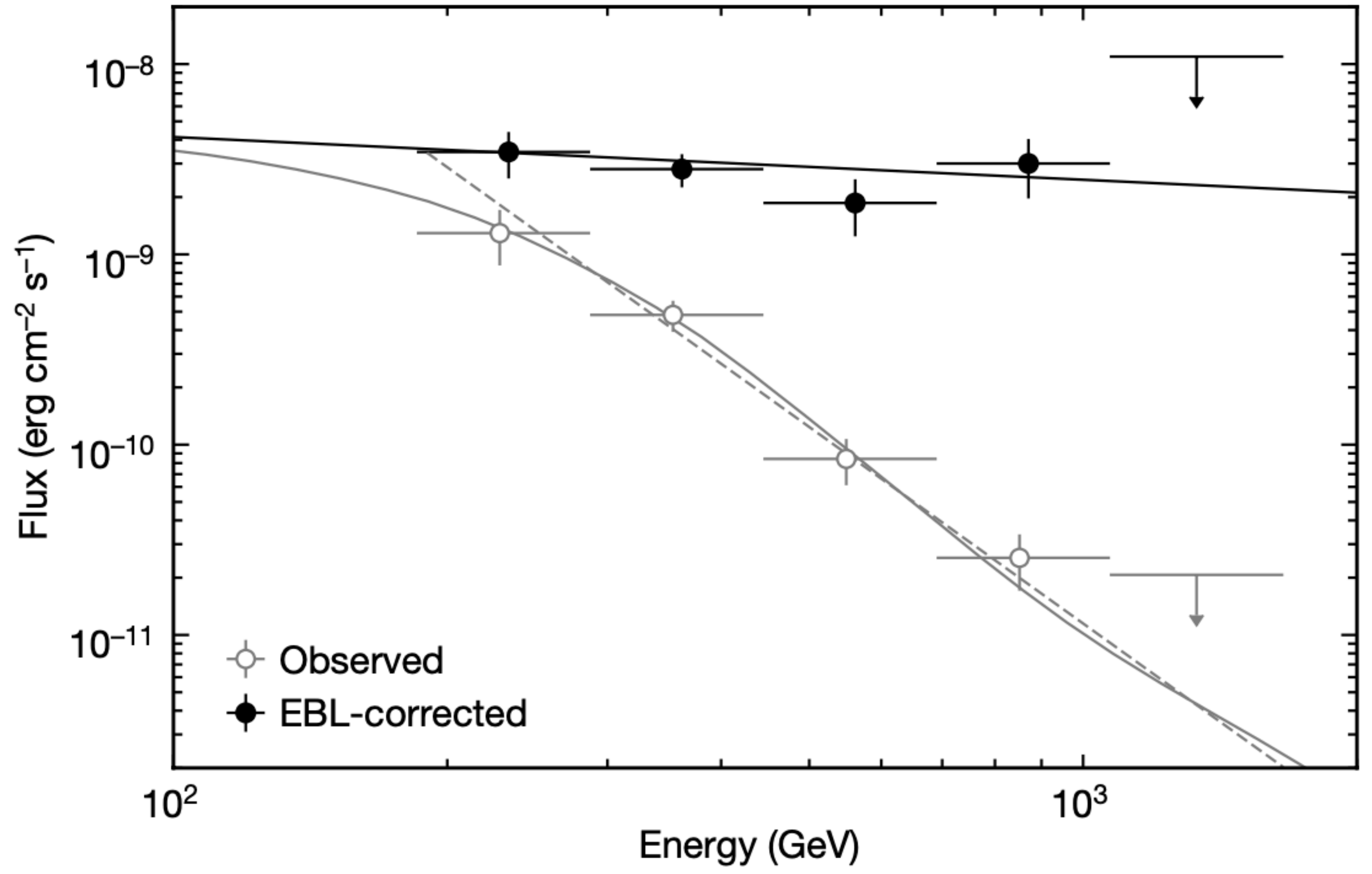
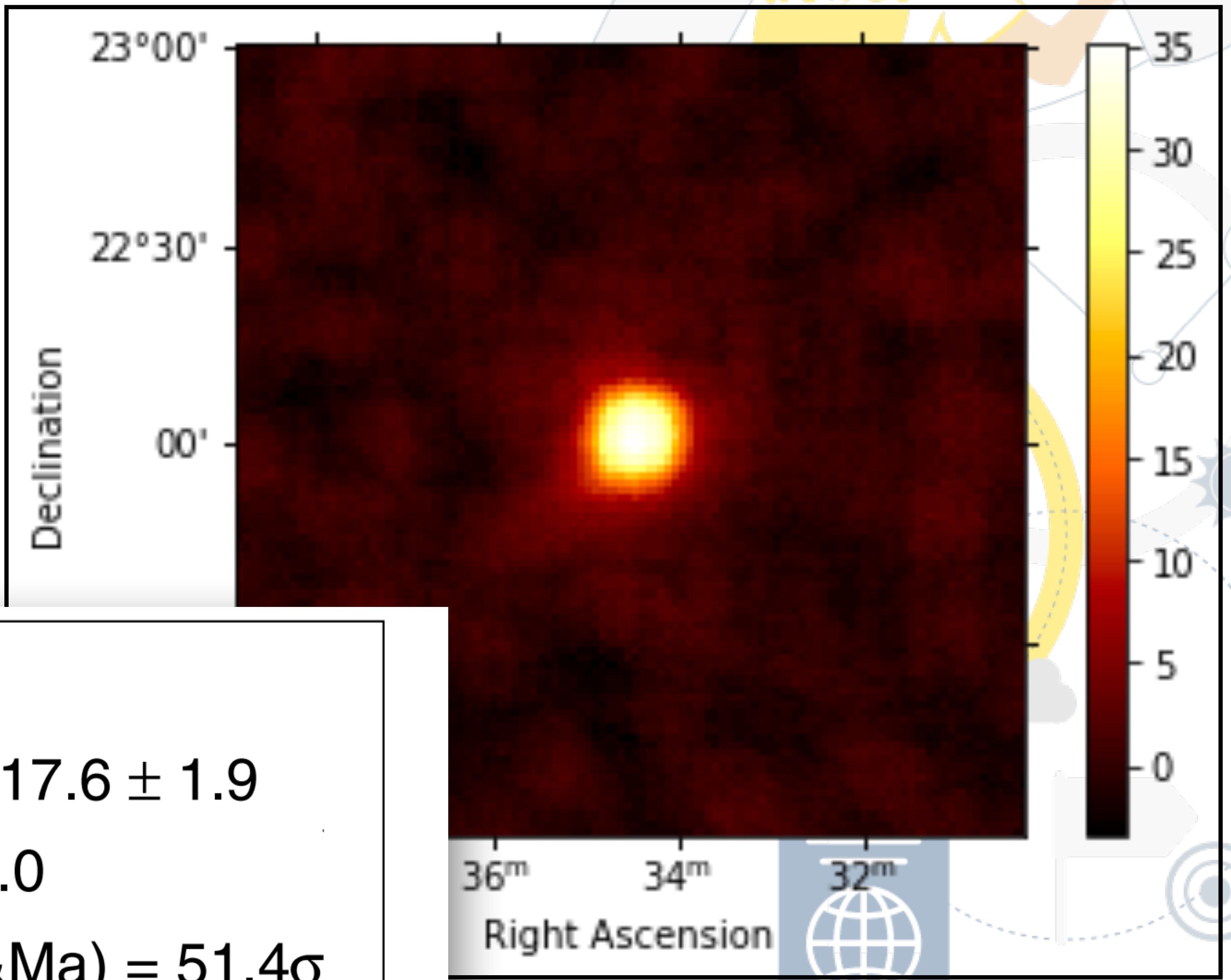
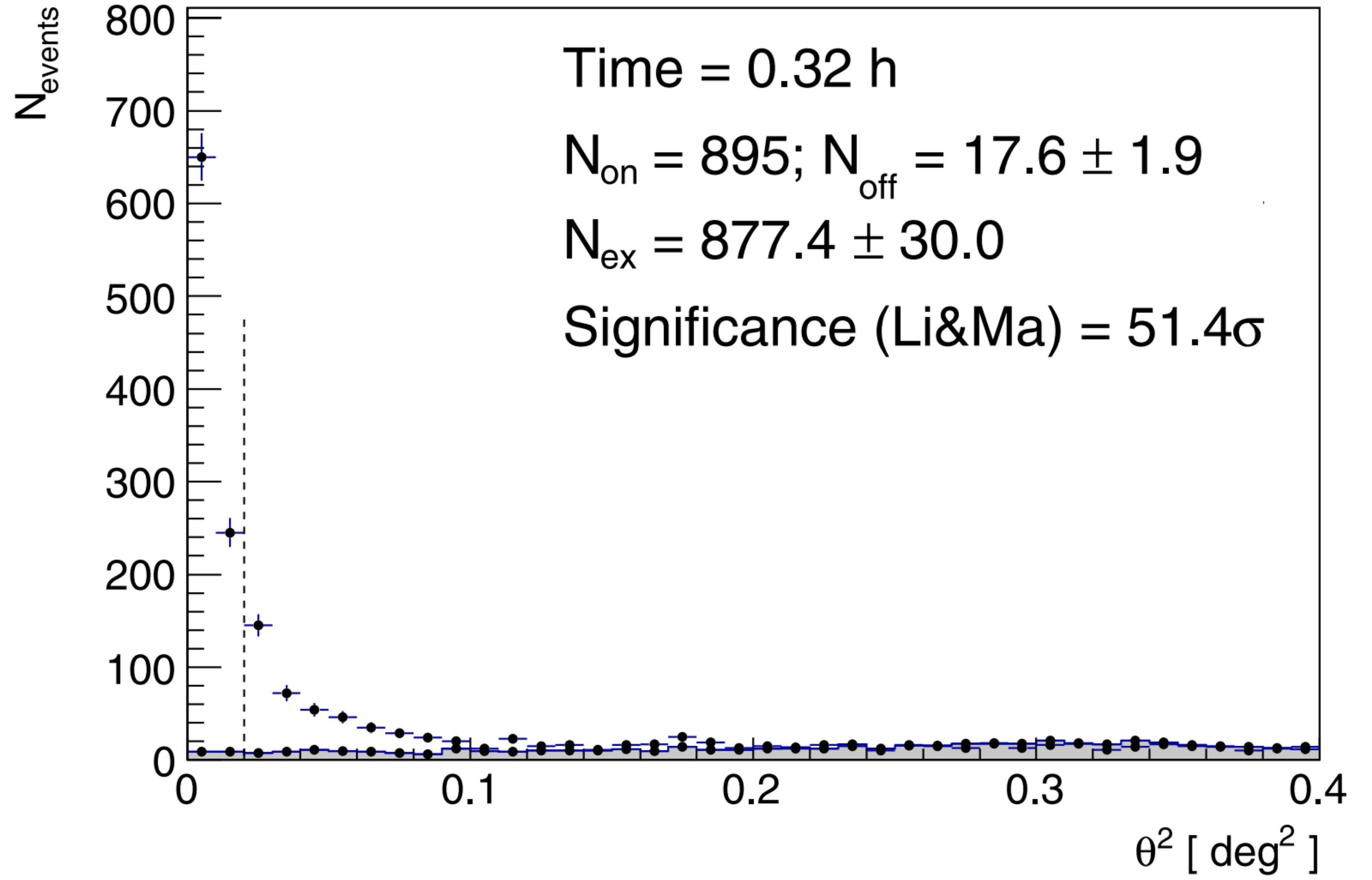


Fig. 2 | Spectrum above 0.2 TeV averaged over the period between $T_0 + 62$ s and $T_0 + 2{,}454$ s for GRB 190114C. Spectral-energy distributions for the spectrum observed by MAGIC (grey open circles) and the intrinsic spectrum corrected for EBL attenuation[25] (blue filled circles). The errors on the flux correspond to one standard deviation. The upper limits at 95% confidence level are shown for the first non-significant bin at high energies. Also shown is the best-fit model for the intrinsic spectrum (black curve) when assuming a power-law function. The grey solid curve for the observed spectrum is obtained by convolving this curve with the effect of EBL attenuation. The grey dashed curve is the forward-folding fit to the observed spectrum with a power-law function (Methods).





Extended Data Fig. 2 | Significance of the γ-ray signal between $T_0 + 62$ s and $T_0 + 1{,}227$ s for GRB 190114C. Distribution of the squared angular distance, θ^2, for the MAGIC data (points) and background events (grey shaded area). θ^2 is defined as the squared angular distance between the nominal position of the source and the reconstructed arrival direction of the events. The dashed vertical line represents the value of the cut on θ^2. This defines the signal region, where the number of events coming from the source (N_{on}) and from the background (N_{off}) are computed. The errors for 'on' events are derived from Poissonian statistics. From N_{on} and N_{off}, the number of excess events (N_{ex}) is computed. The significance is calculated using the Li & Ma method[42].

Statistical inference in On/Off measurement

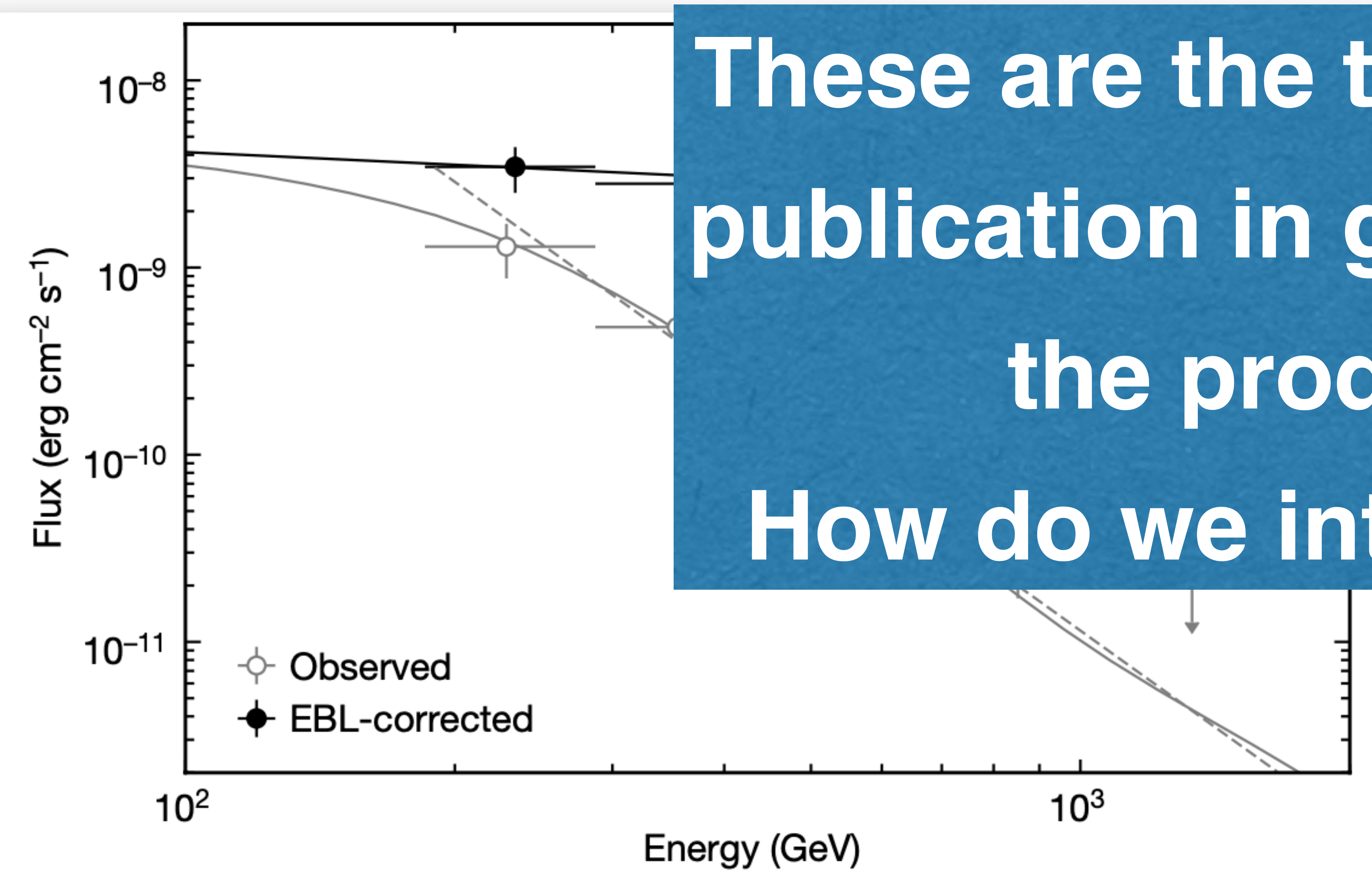
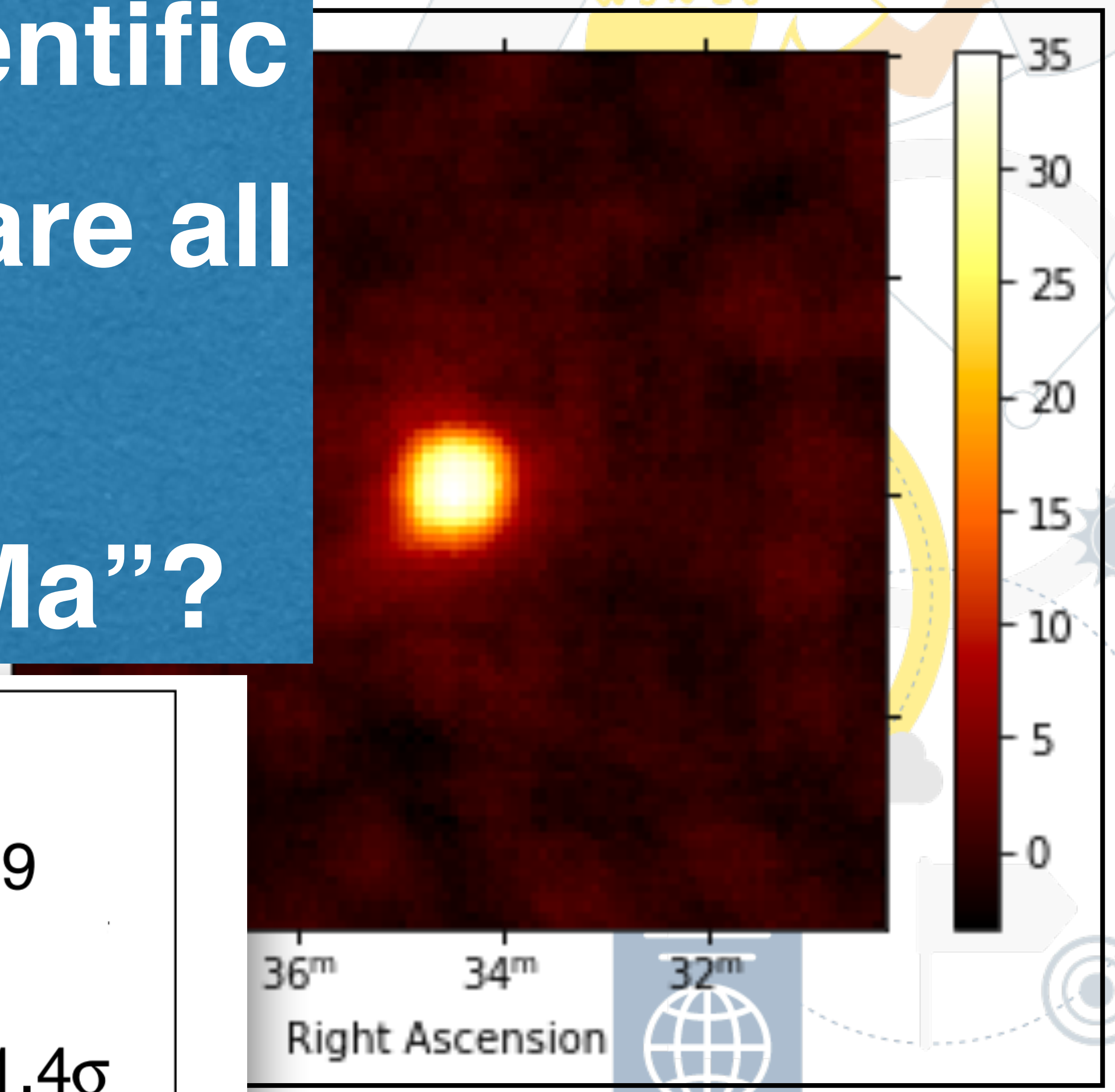
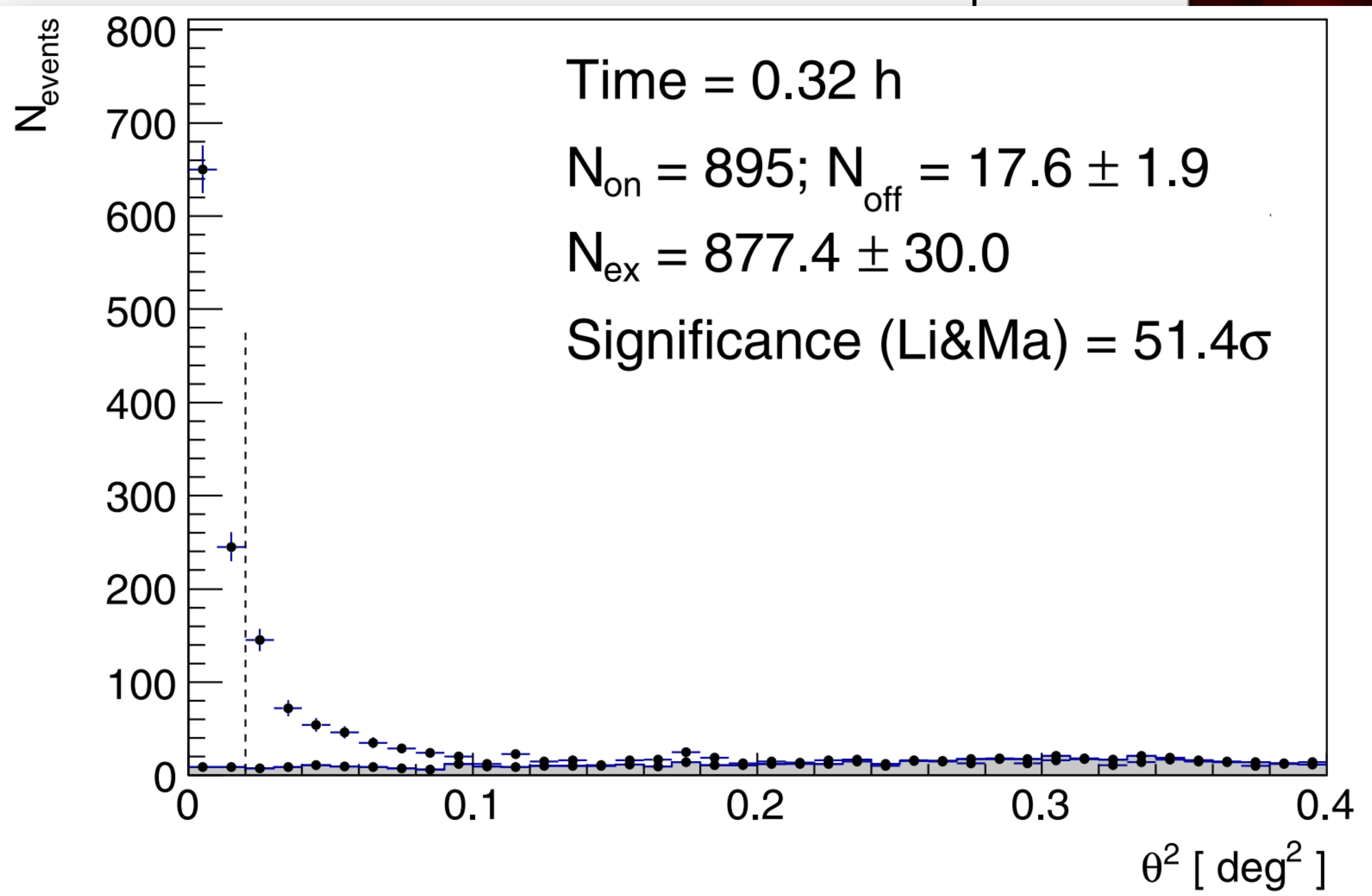


Fig. 2 | Spectrum above 0.2 TeV averaged over the period between $T_0 + 62$ s and $T_0 + 2{,}454$ s for GRB 190114C. Spectral-energy distributions for the spectrum observed by MAGIC (grey open circles) and the intrinsic spectrum corrected for EBL attenuation[25] (blue filled circles). The errors on the flux correspond to one standard deviation. The upper limits at 95% confidence level are shown for the first non-significant bin at high energies. Also shown is the best-fit model for the intrinsic spectrum (black curve) when assuming a power-law function. The grey solid curve for the observed spectrum is obtained by convolving this curve with the effect of EBL attenuation. The grey dashed curve is the forward-folding fit to the observed spectrum with a power-law function (Methods).

These are the typical plots shown in a scientific publication in gamma-ray astronomy and are all the product of statistical analysis.

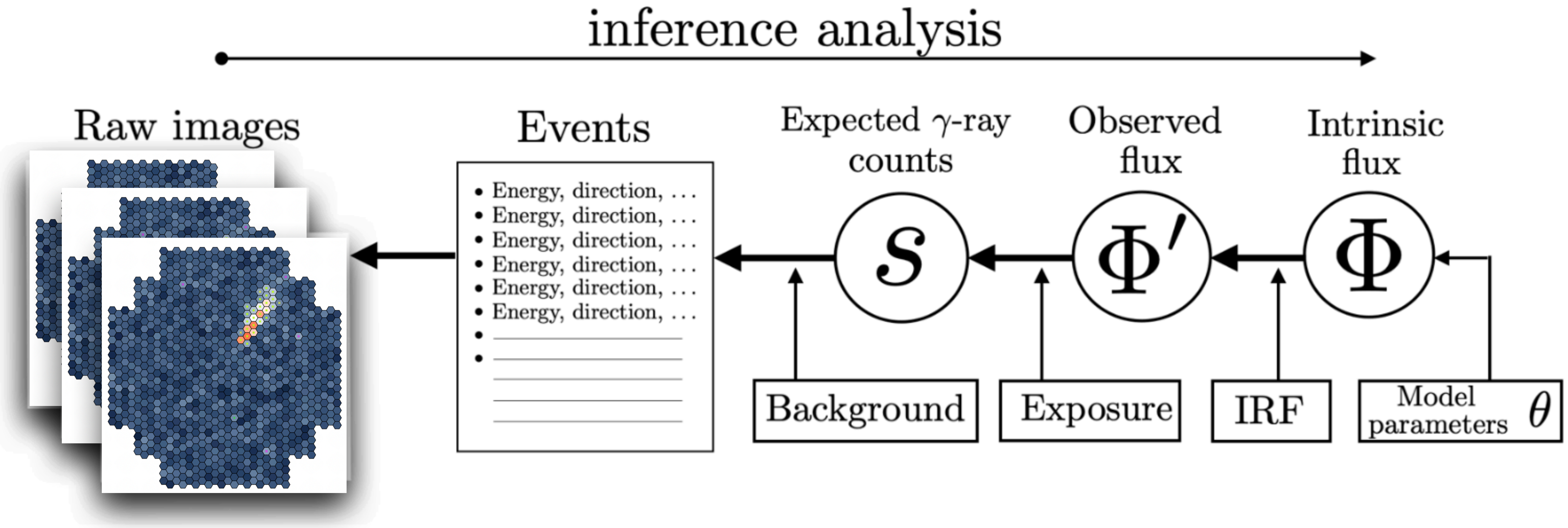
How do we interpret them? What is "Li&Ma"?



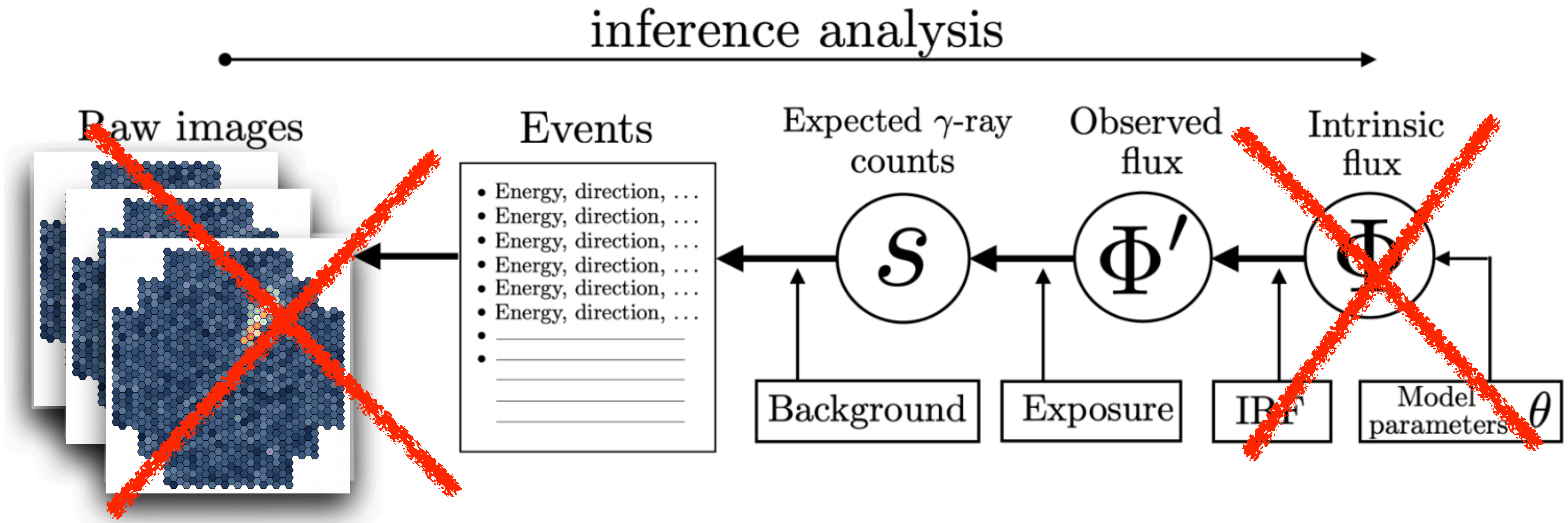


Extended Data Fig. 2 | Significance of the γ-ray signal between $T_0 + 62$ s and $T_0 + 1{,}227$ s for GRB 190114C. Distribution of the squared angular distance, θ^2, for the MAGIC data (points) and background events (grey shaded area). θ^2 is defined as the squared angular distance between the nominal position of the source and the reconstructed arrival direction of the events. The dashed vertical line represents the value of the cut on θ^2. This defines the signal region, where the number of events coming from the source (N_{on}) and from the background (N_{off}) are computed. The errors for 'on' events are derived from Poissonian statistics. From N_{on} and N_{off}, the number of excess events (N_{ex}) is computed. The significance is calculated using the Li & Ma method[42].

Statistical inference in On/Off measurement



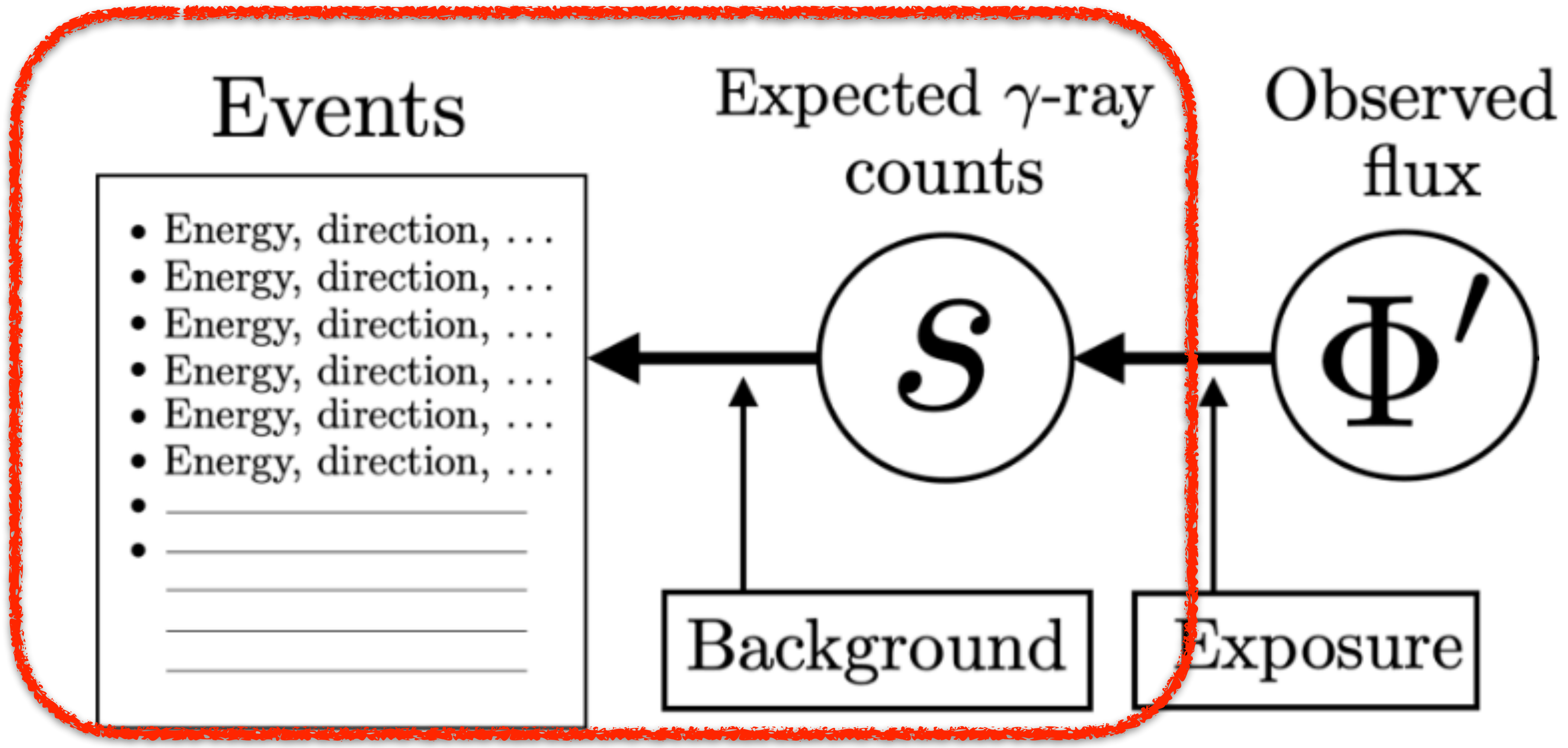
Statistical inference in On/Off measurement



We will skip the first and last part (being too technical and too instrument dependent) and focus on the remaining part:

given a list of events how do we **reconstruct the flux** and with which **confidence** can we claim that there is indeed a flux of gamma-ray?

Statistical inference in On/Off measurement

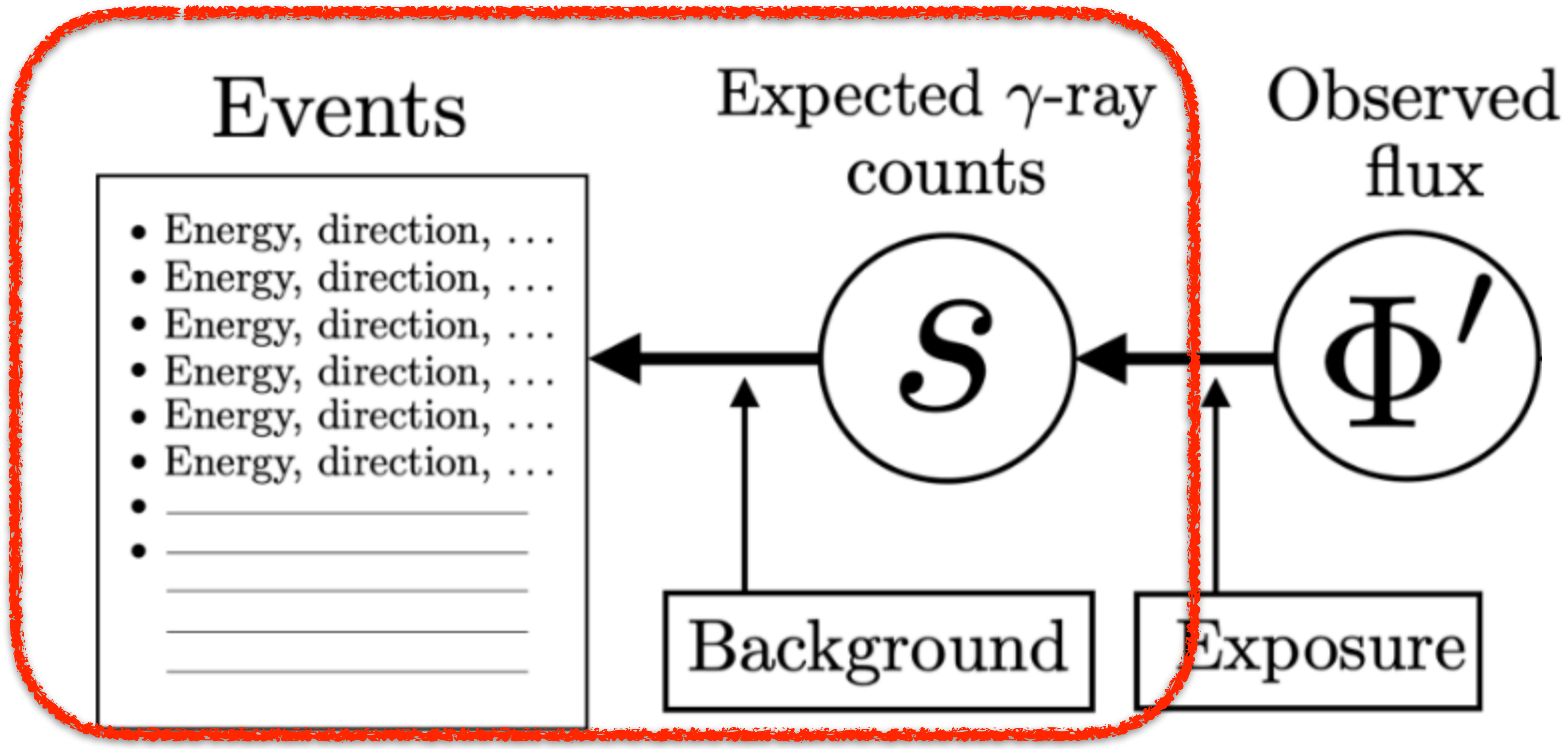


Given your event list what's the expected number of gamma-ray?

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

Statistical inference in On/Off measurement



Given your event list what's the expected number of gamma-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

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

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

Statistical inference in On/Off measurement

The "ingredients"

The flux

number N_γ of expected photons per unit energy (E), time (t), and area (A)

$$\Phi(E,t,\hat{\mathbf{n}}) = \frac{dN_\gamma(E,t,\hat{\mathbf{n}})}{dEdAdt}$$

Expected background events "b"

- can be assumed to be known
- can be estimated from an OFF measurement (see next slide)

Expected signal events "s"

Taking into account the exposure of the observation given by the energetic (E), temporal (t) and solid angle (Ω) range (hereafter denote by Δ) in which the events have been collected we have

$$s = \int_\Delta \Phi\,(E\,,\hat{\mathbf{n}}\,,t)dE\;d\hat{\mathbf{n}}\;dt$$

Total number of observed events "on source"

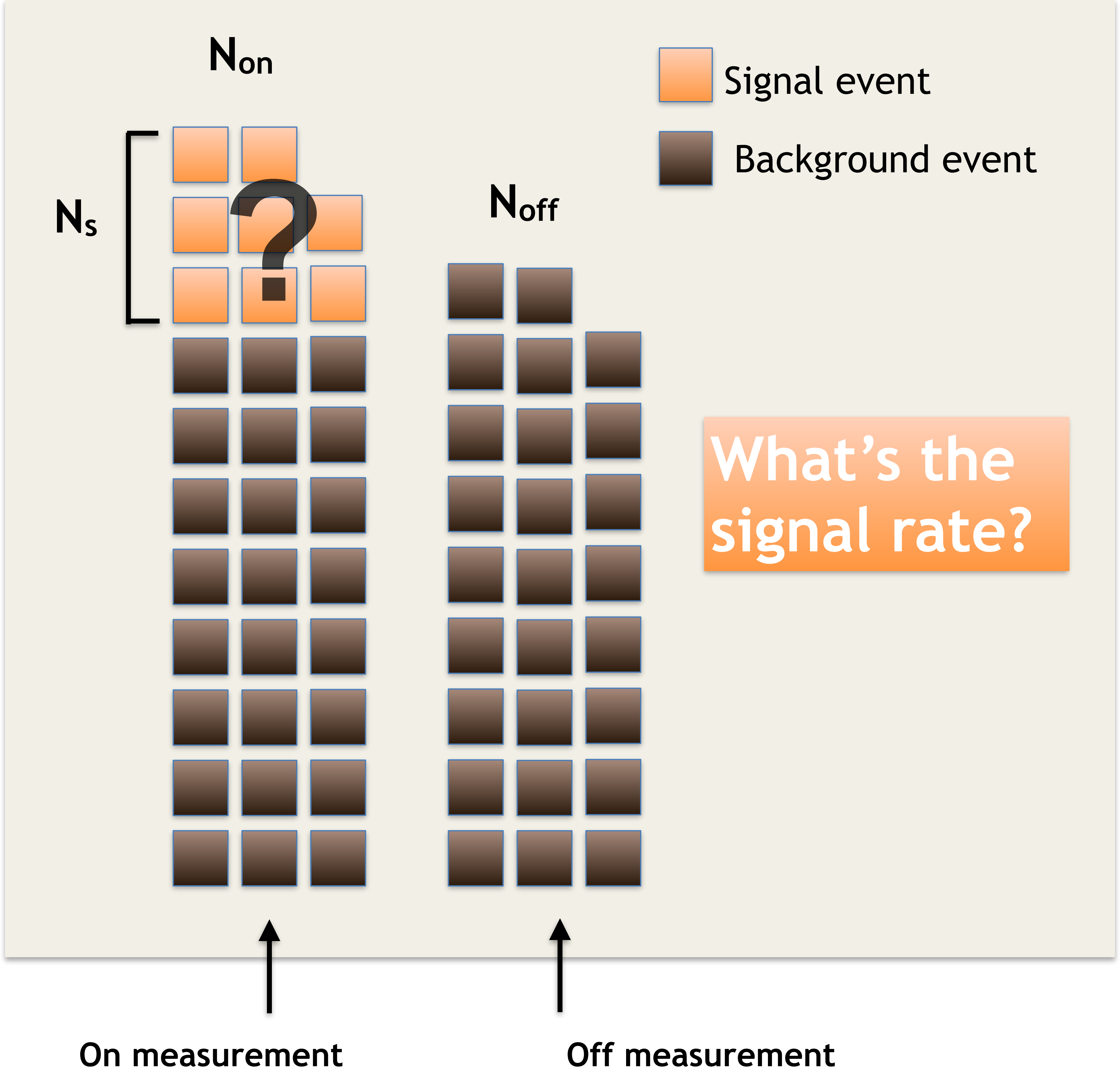
$$N_{ON} \;\sim\; \mathcal{P}(N_{ON}|s+b) \;=\; \frac{(s+b)^{N_{ON}}}{N_{ON}!}e^{-(s+b)}$$

Total number of observed events "off source"

$$N_{OFF} \;\sim\; \mathcal{P}(N_{OFF}|b) \;=\; \frac{b^{N_{OFF}}}{N_{OFF}!}\,e^{-b}$$

Statistical inference in On/Off measurement

On/Off measurement



$$\frac{(s+\alpha b)^{N_{on}}}{N_{on}!}e^{-(s+\alpha b)} \times \frac{b^{N_{off}}}{N_{off}!}e^{-b}$$

variable	description	property
N_{on}	number of events in the On region	measured
N_{off}	number of events in the Off region	measured
α	exposure in the On region over the one in the Off regions	measured
b	expected rate of occurrences of background events in the Off regions	unknown
s	expected rate of occurrences of signal events in the On region	unknown
N_s	number of signal events in the On region	unknown

Statistical inference in On/Off measurement

On/Off measurement

Signal estimation in the **frequentist approach:**

Likelihood function:

$$p(N_{on}, N_{off} \mid s, b; \alpha) = p(N_{on} \mid s, \alpha b) \cdot p(N_{off} \mid b) = \frac{(s+\alpha b)^{N_{on}}}{N_{on}!} e^{-(s+\alpha b)} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b}$$

Statistical inference in On/Off measurement

On/Off measurement

Signal estimation in the **frequentist approach:**

Likelihood function:

$$p(N_{on}, N_{off} \mid s, b; \alpha) = p(N_{on} \mid s, \alpha b) \cdot p(N_{off} \mid b) = \frac{(s+\alpha b)^{N_{on}}}{N_{on}!} e^{-(s+\alpha b)} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b}$$

↓

Likelihood ratio:

$$\lambda(s) \equiv \frac{p(N_{on}, N_{off} \mid s, b = \hat{b}\ ;\ \alpha)}{p(N_{on}, N_{off} \mid s = N_{on} - \alpha N_{off}\ ,\ b = N_{off}\ ;\ \alpha)}$$

value of b that **maximizes** the likelihood for a given s

$$\hat{b} = \frac{N^2 + \sqrt{N^2 + 4(1+1/\alpha) s N_{off}}}{2(1+\alpha)}$$

$$N = N_{on} + N_{off} - s(1+1/\alpha)$$

Statistical inference in On/Off measurement

On/Off measurement

Signal estimation in the **frequentist approach:**

Likelihood function:

$$p(N_{on}, N_{off} \mid s, b; \alpha) = p(N_{on} \mid s, \alpha b) \cdot p(N_{off} \mid b) = \frac{(s+\alpha b)^{N_{on}}}{N_{on}!} e^{-(s+\alpha b)} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b}$$

Likelihood ratio:

$$\lambda(s) \equiv \frac{p(N_{on}, N_{off} \mid s, b = \hat{b} \; ; \; \alpha)}{p(N_{on}, N_{off} \mid s = N_{on} - \alpha N_{off} \; , \; b = N_{off} \; ; \; \alpha)}$$

value of b that **maximizes** the likelihood for a given s

$$\hat{b} = \frac{N^2 + \sqrt{N^2 + 4(1+1/\alpha) s N_{off}}}{2(1+\alpha)}$$

$$N = N_{on} + N_{off} - s(1+1/\alpha)$$

$$-2 \log \lambda(s) = 2 \left[N_{on} \log \left(\frac{N_{on}}{s + \alpha \hat{b}} \right) + N_{off} \log \left(\frac{N_{off}}{\hat{b}} \right) + s + (1+\alpha)\hat{b} - N_{on} - N_{off} \right]$$

Statistical inference in On/Off measurement

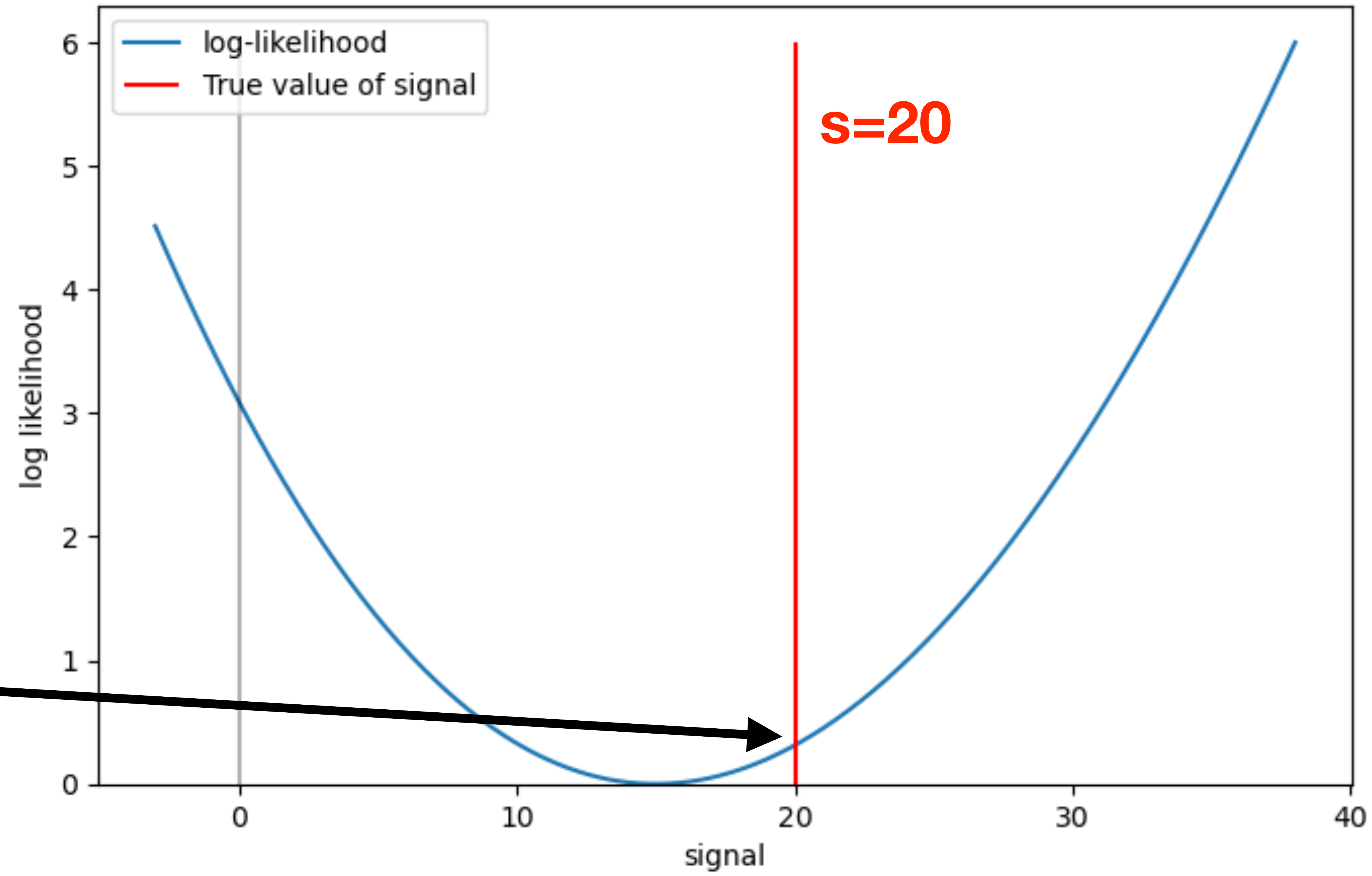
On/Off measurement

Signal estimation in the **frequentist approach:**

$$-2\log\lambda(s) = 2\left[N_{on}\log\left(\frac{N_{on}}{s+\alpha\hat{b}}\right) + N_{off}\log\left(\frac{N_{off}}{\hat{b}}\right) + s + (1+\alpha)\hat{b} - N_{on} - N_{off}\right]$$

Example with:

Non = 57
Noff = 85
α = 0.5



Let's assume we want to test the hypothesis s=20

Statistical inference in On/Off measurement

On/Off measurement

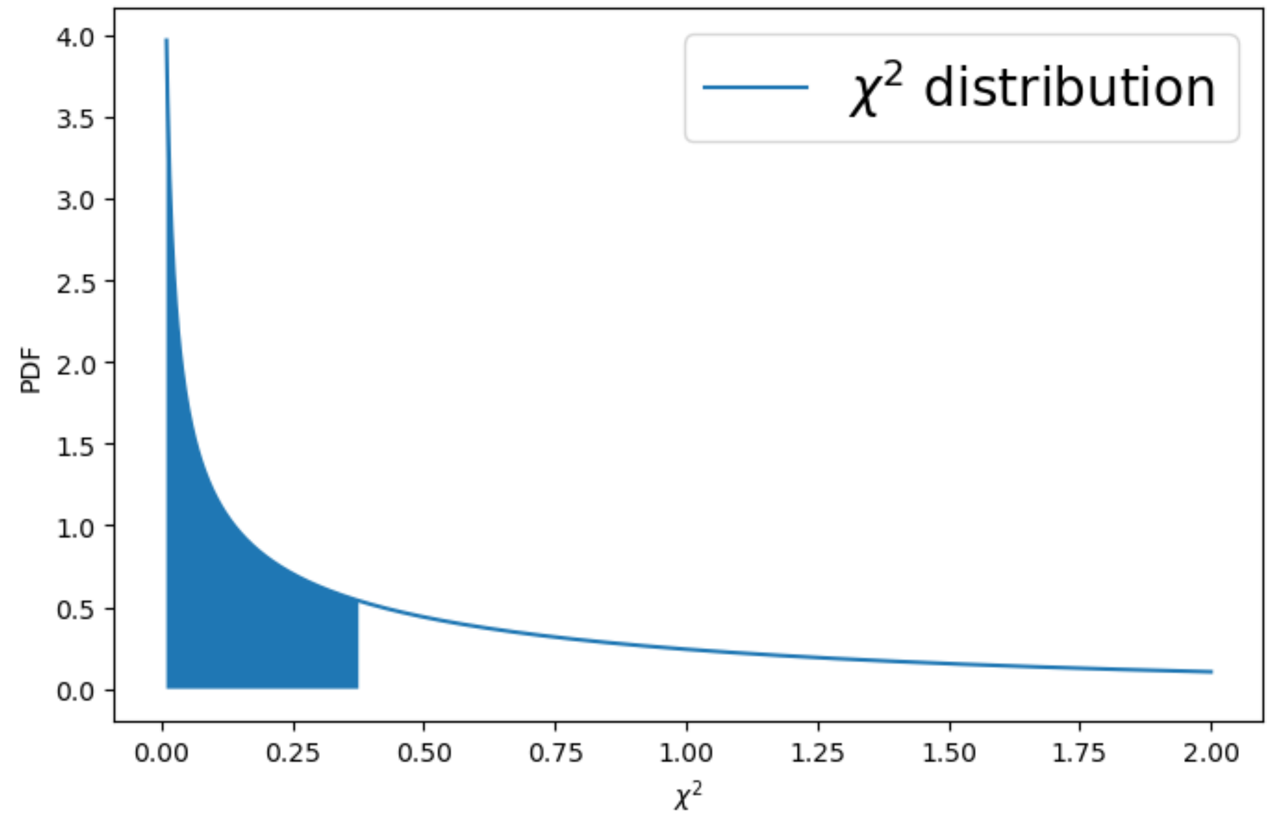
Signal estimation in the **frequentist approach:**

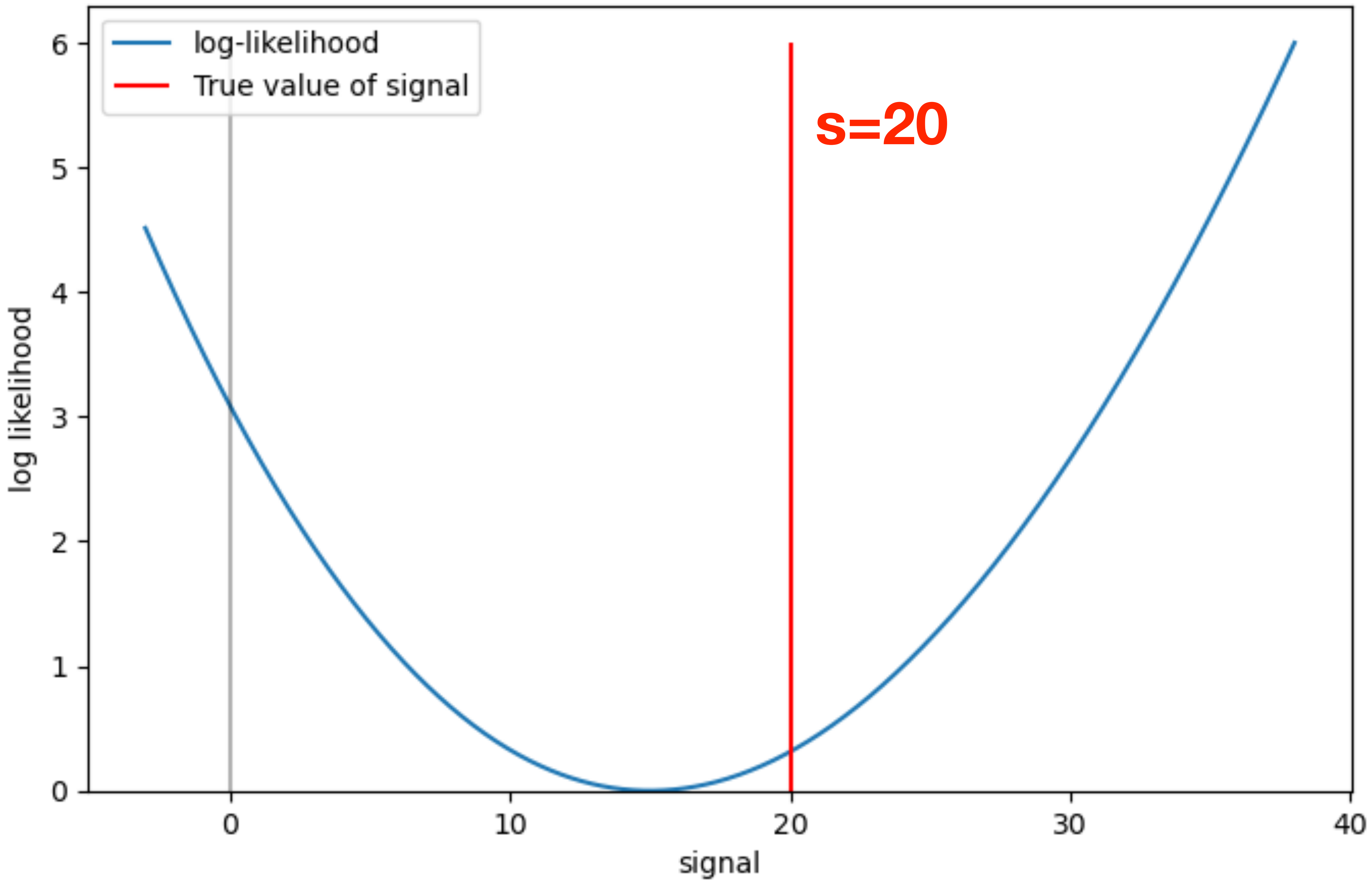
$$-2\log\lambda(s) = 2\left[N_{on}\log\left(\frac{N_{on}}{s+\alpha\hat{b}}\right) + N_{off}\log\left(\frac{N_{off}}{\hat{b}}\right) + s + (1+\alpha)\hat{b} - N_{on} - N_{off}\right]$$

Our statistic is

$$-2\log\lambda(s=20) \simeq 0.38$$

Values more extreme than 0.38 would have been observed ~ 54% of the times





Statistical inference in On/Off measurement

On/Off measurement

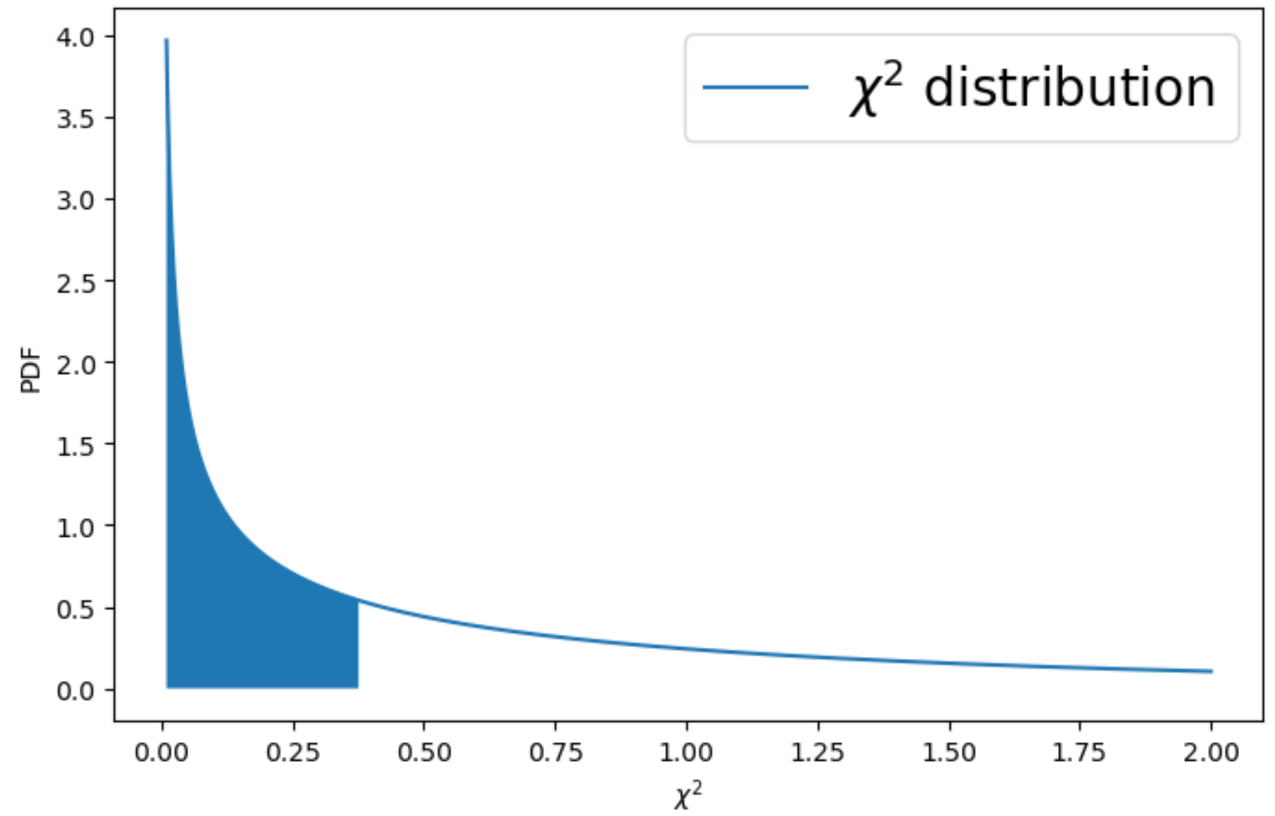
Signal estimation in the **frequentist approach:**

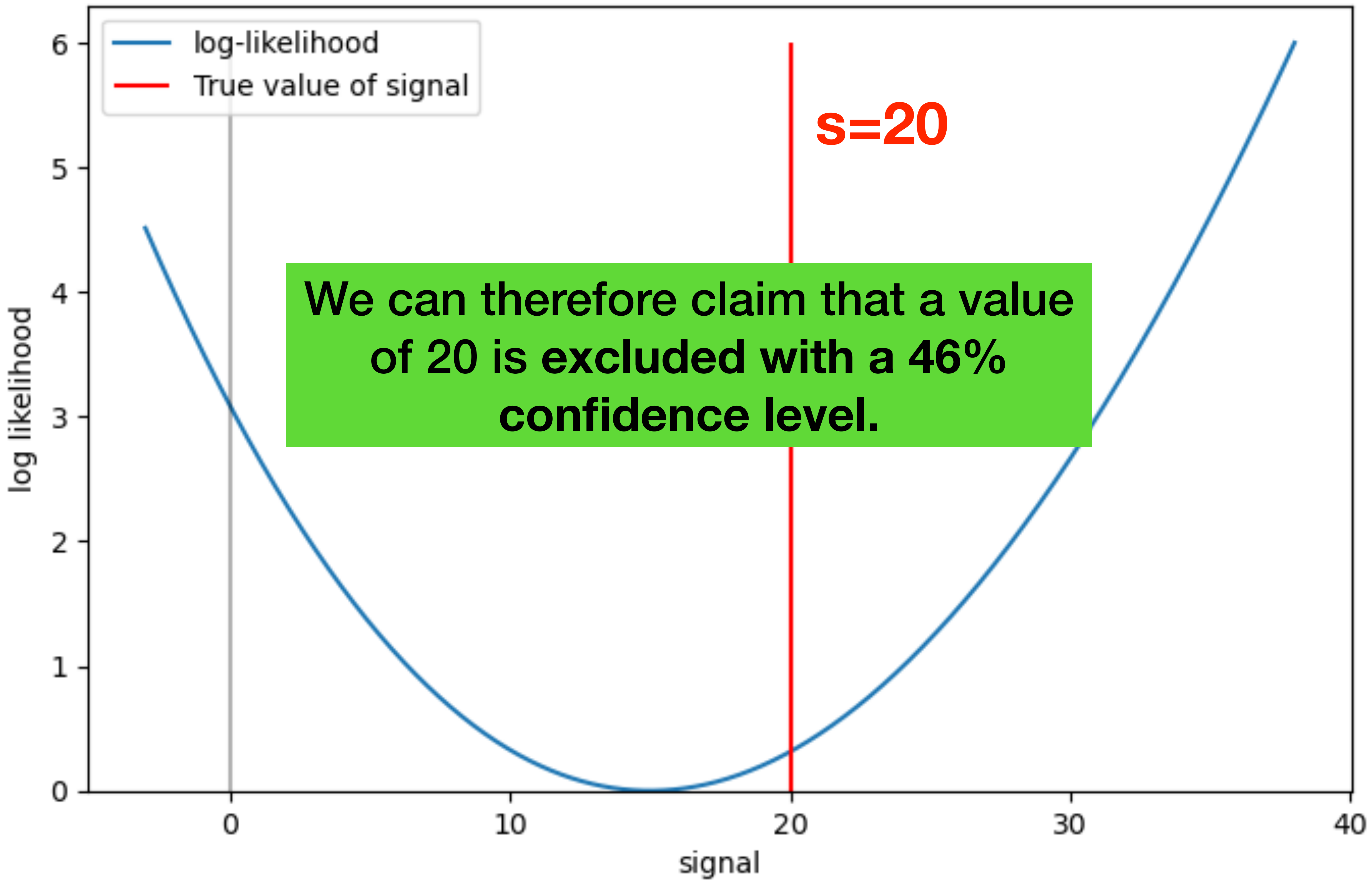
$$-2\log\lambda(s) = 2\left[N_{on}\log\left(\frac{N_{on}}{s+\alpha\hat{b}}\right) + N_{off}\log\left(\frac{N_{off}}{\hat{b}}\right) + s + (1+\alpha)\hat{b} - N_{on} - N_{off}\right]$$

Our statistic is

$$-2\log\lambda(s=20) \simeq 0.38$$

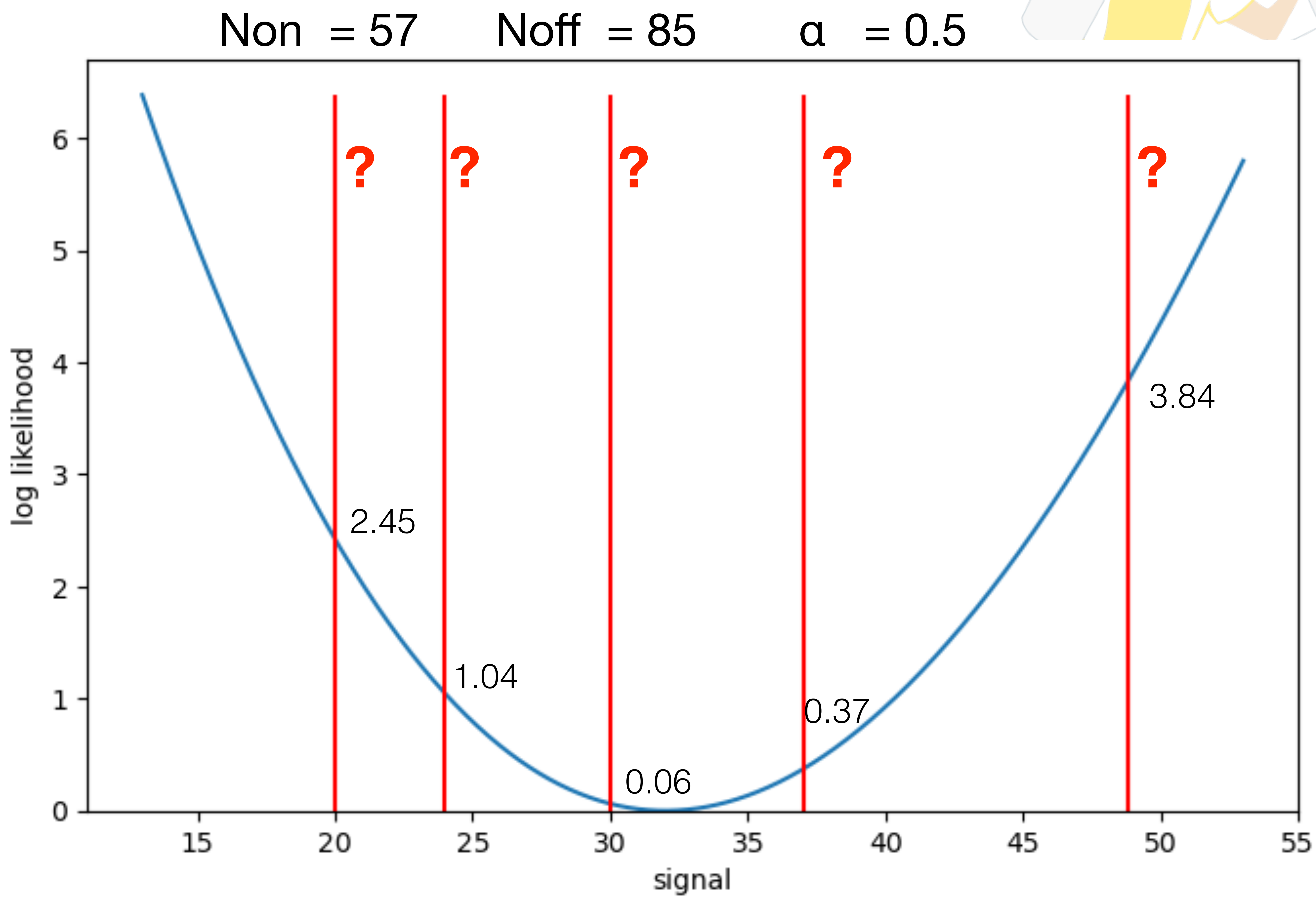
Values more extreme than 0.38 would have been observed ~ 54% of the times





Statistical inference in On/Off measurement

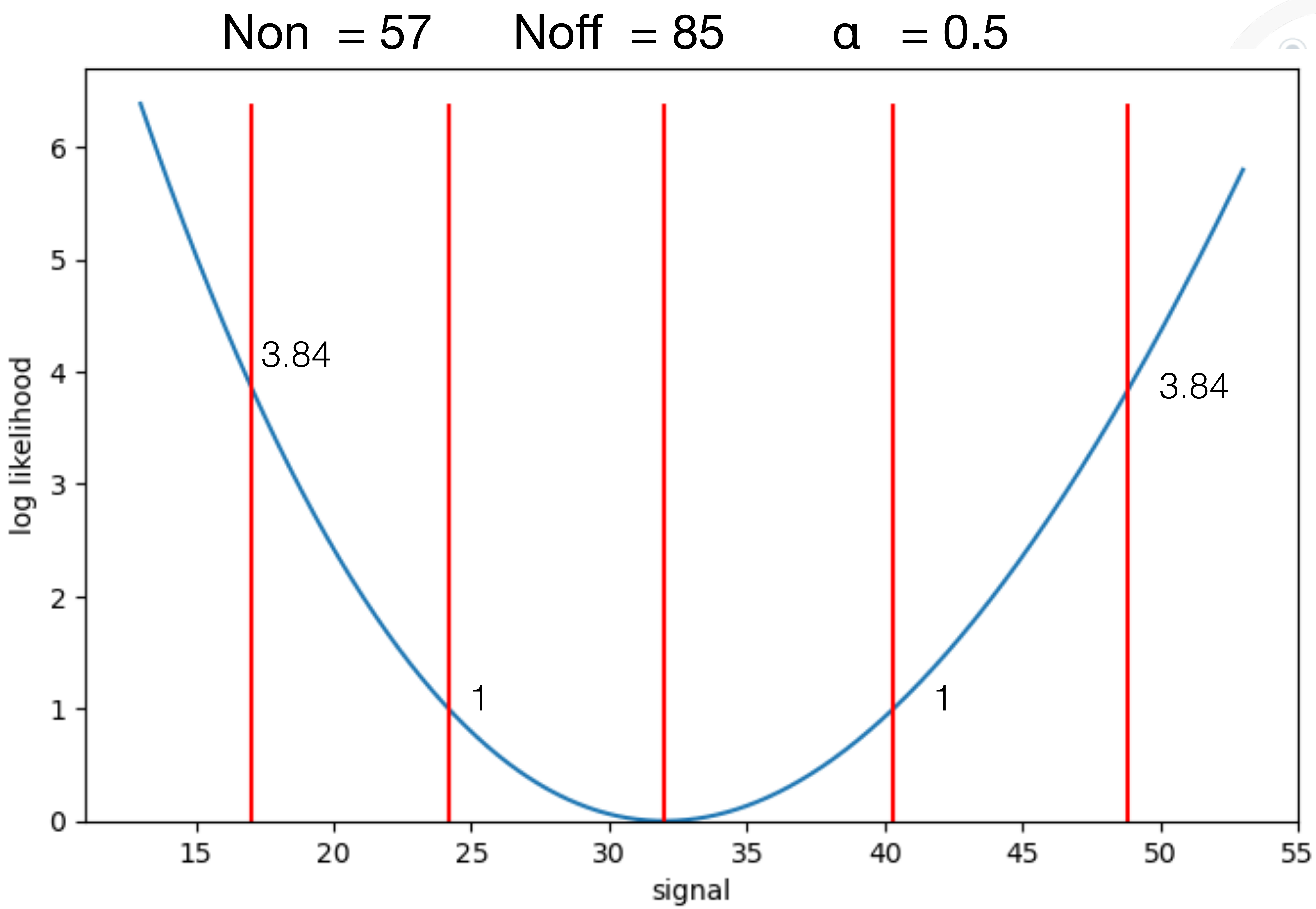
Does it mean that we have to repeat this for all possible values of the signal 's'?



Statistical inference in On/Off measurement

Conventionally 3 confidence levels are reported:

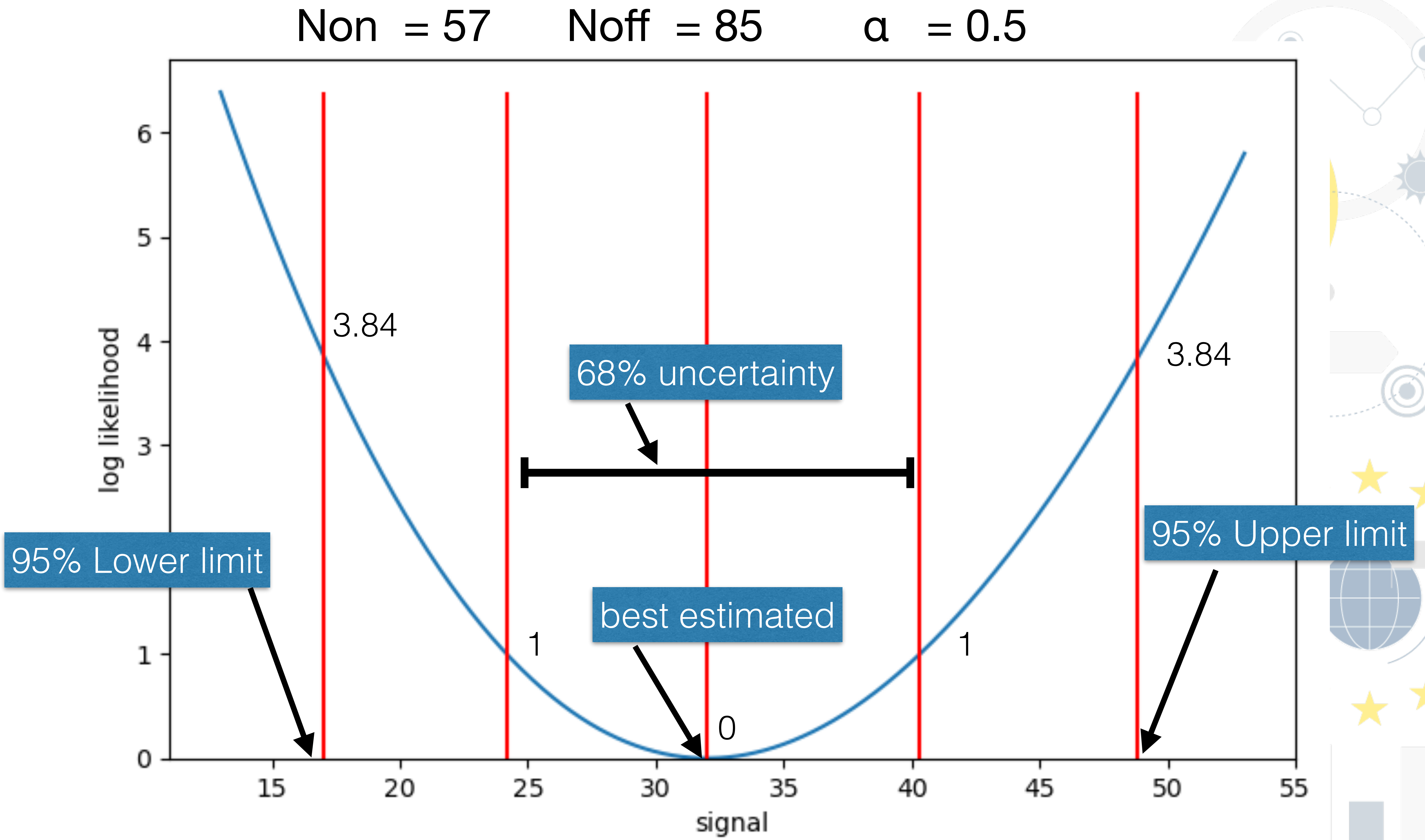
- **0% CL** : which is by definition when the chi-squared is **zero**
- **68% CL** : which is when the chi-squared is **1**
- **95% CL** : which is when the chi-squared is **3.84**



Statistical inference in On/Off measurement

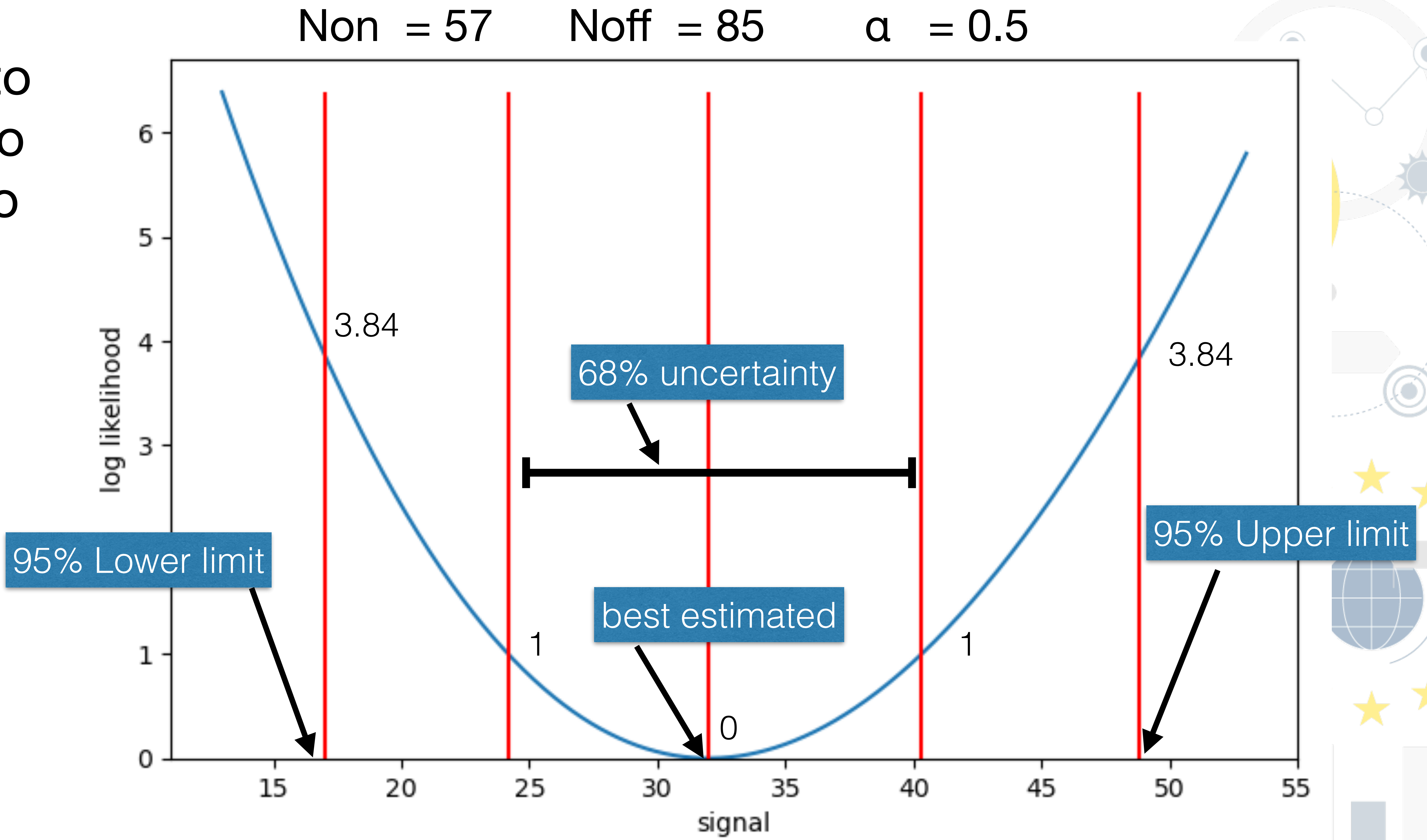
Conventionally 3 confidence levels are reported:

- **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**



Statistical inference in On/Off measurement

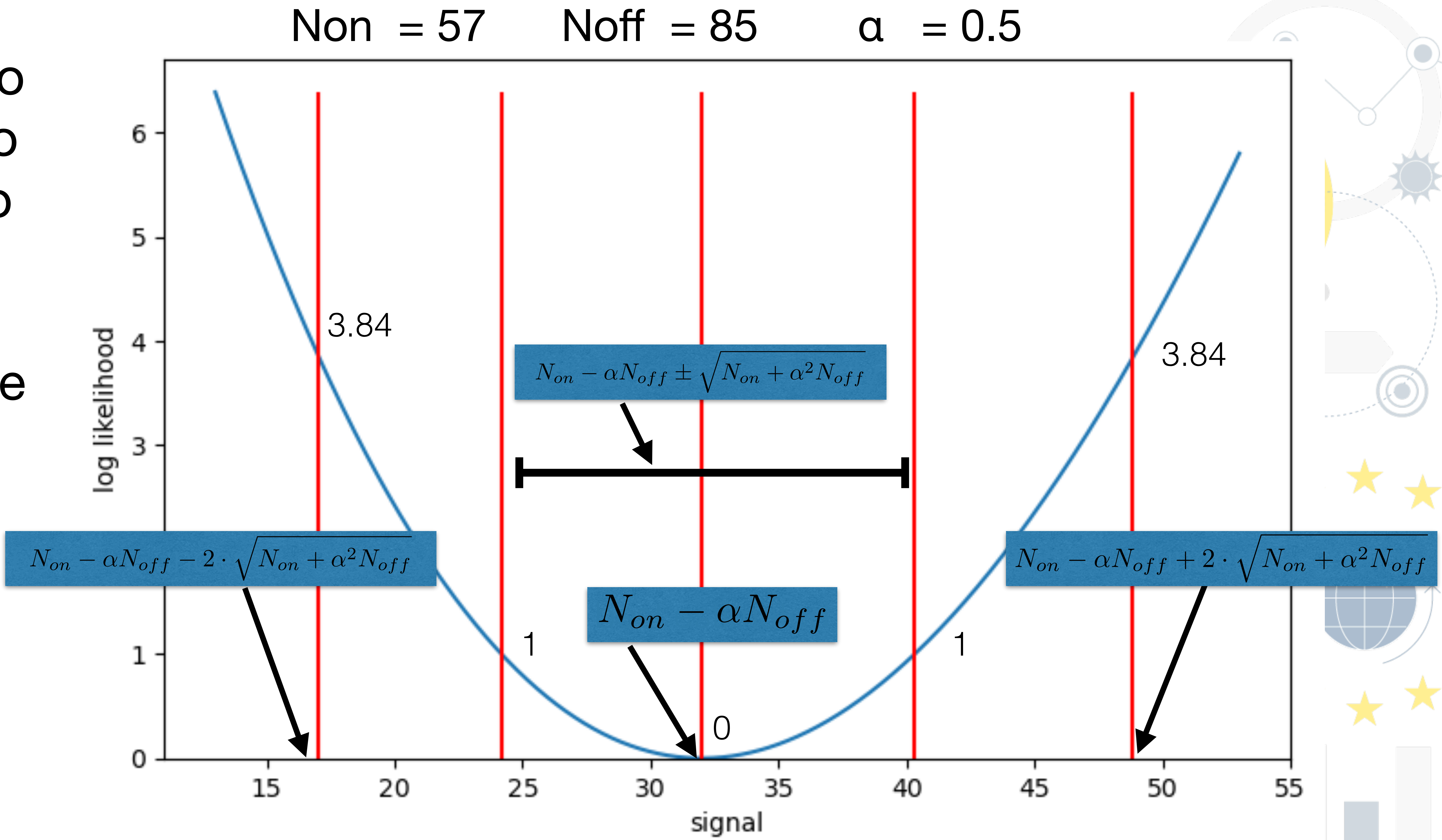
So do we need each time to compute the likelihood ratio and find where it is equal to zero, one, and 3.84?



Statistical inference in On/Off measurement

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



Statistical inference in On/Off measurement

$$N_{on} - \alpha N_{off} = 32$$

$$\sqrt{N_{on} + \alpha^2 N_{off}} = 8.06$$

↓

- The signal estimation is:

32 ± 8

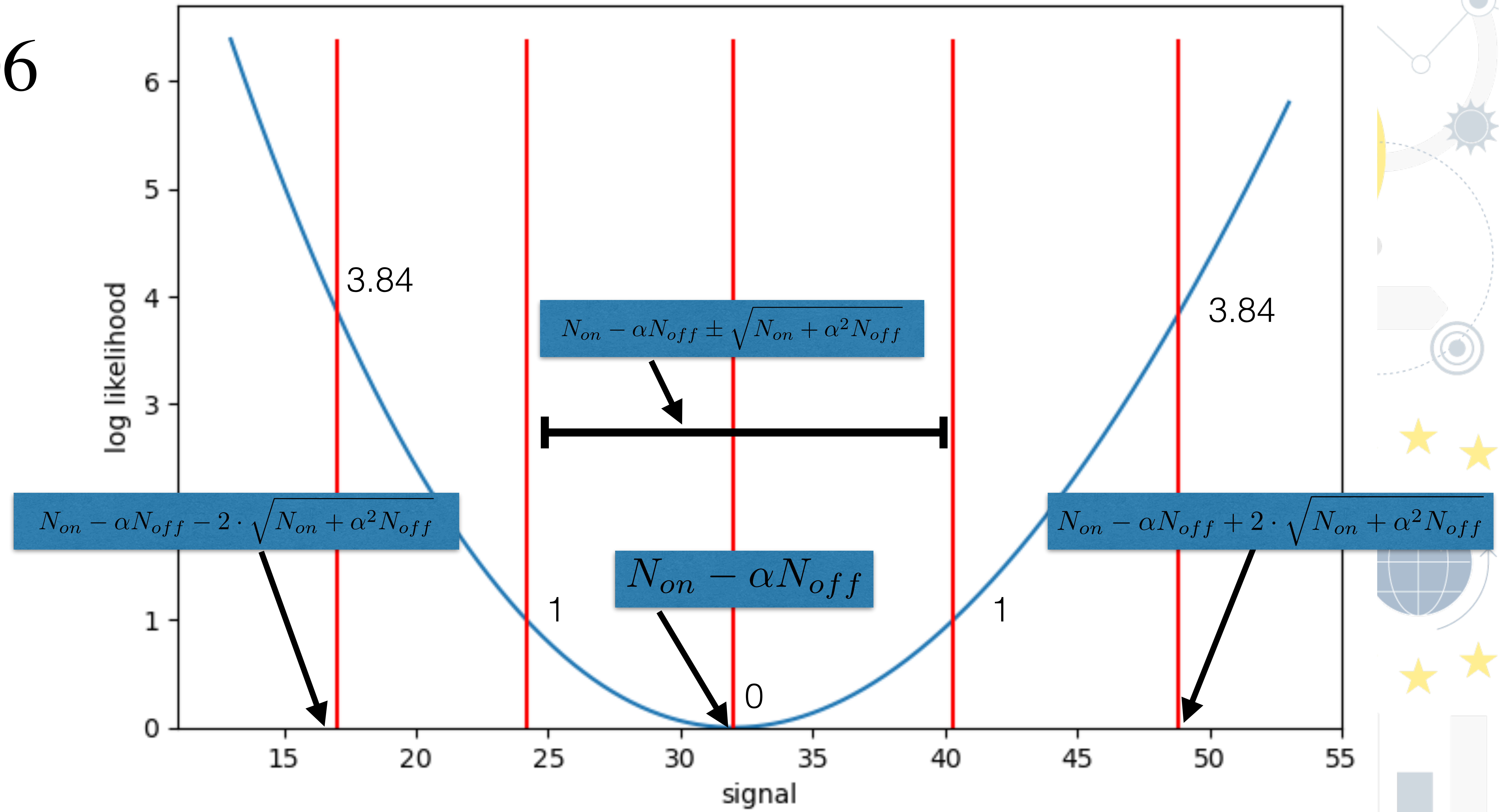
- with upper limit

48.1

- and lower limit

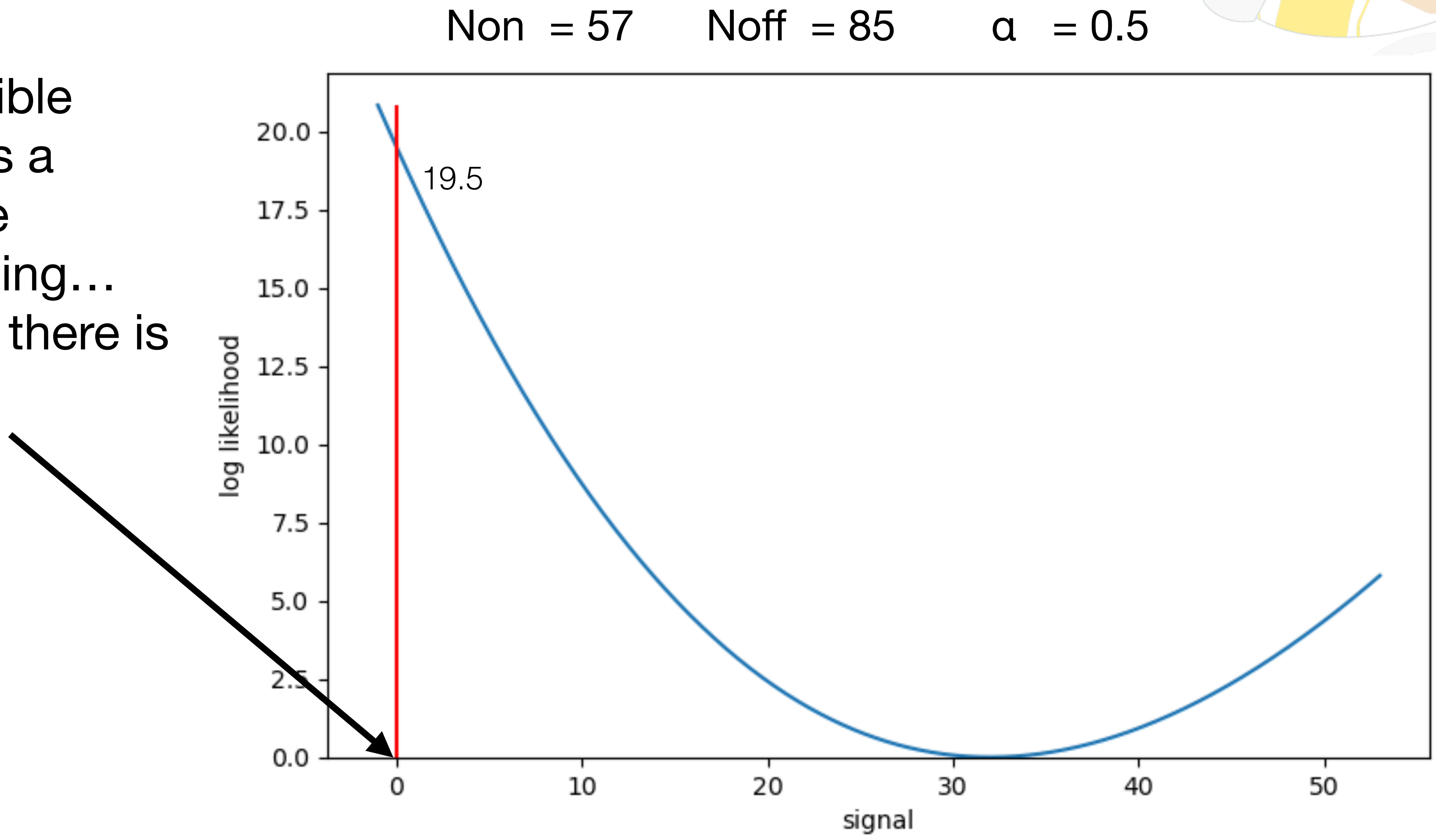
15.9

Non = 57 Noff = 85 α = 0.5



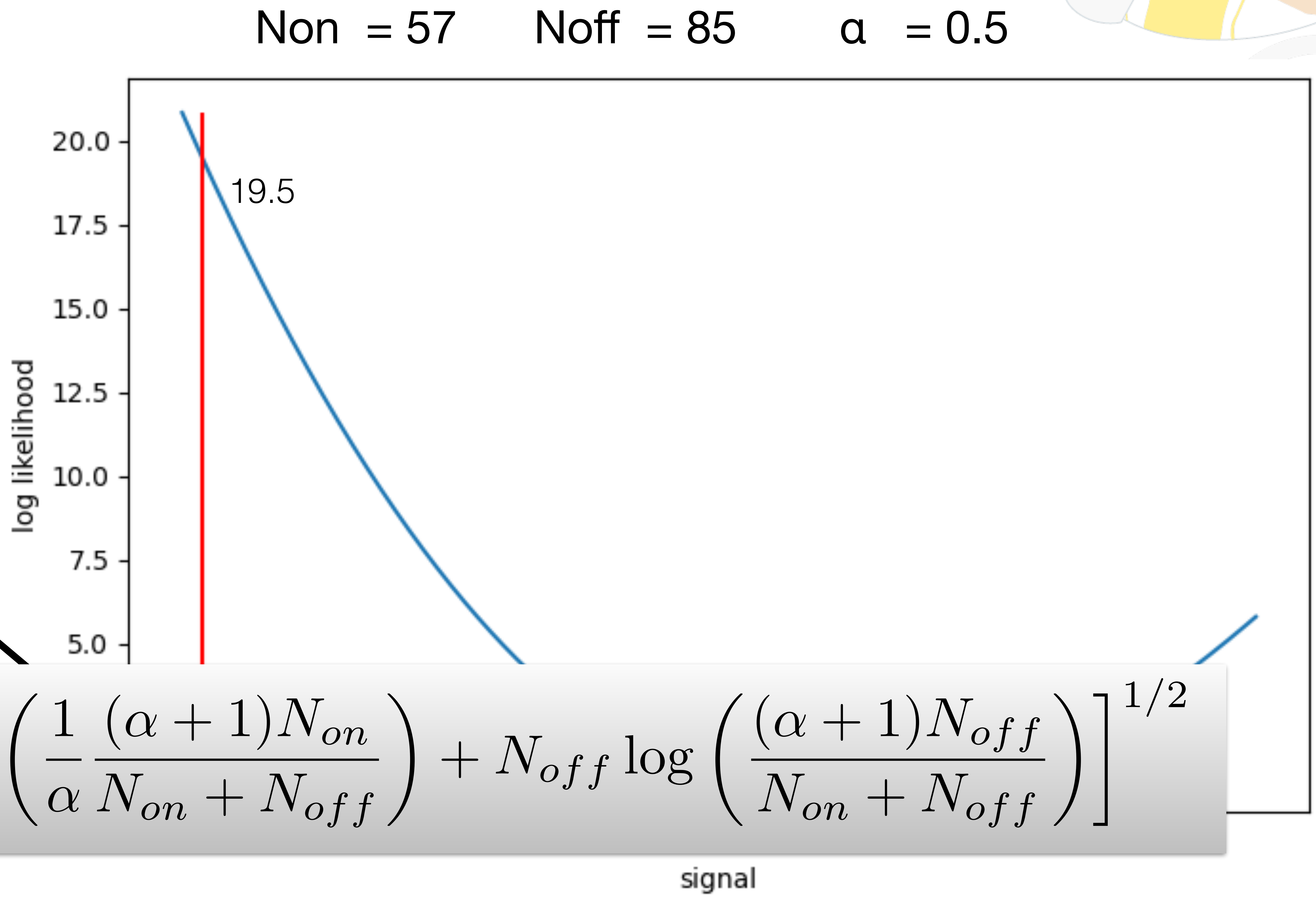
The Li&Ma formula

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**



The Li&Ma formula

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**



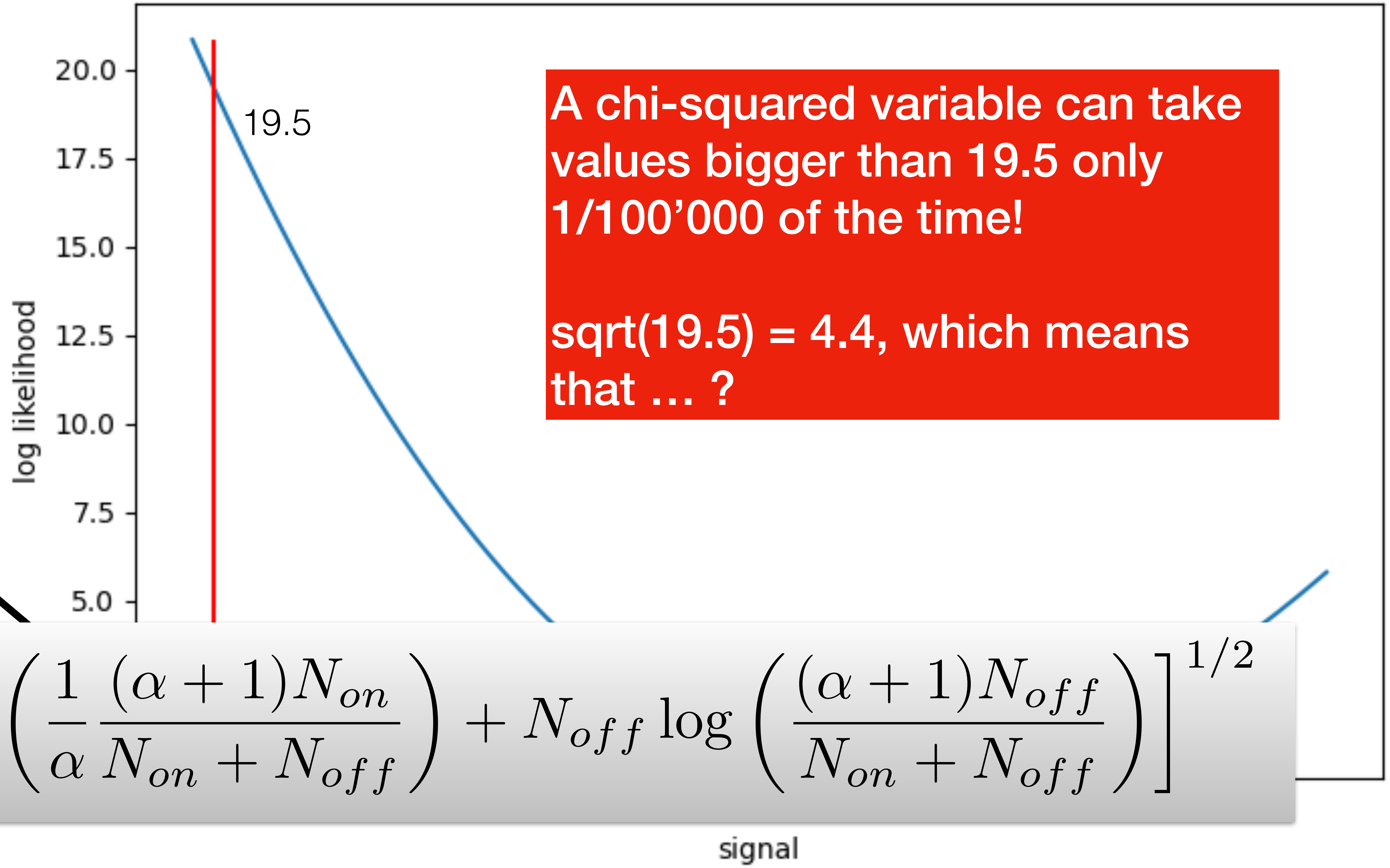
Li&Ma

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

The Li&Ma formula

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**

Non = 57 Noff = 85 α = 0.5



A chi-squared variable can take values bigger than 19.5 only 1/100'000 of the time!

sqrt(19.5) = 4.4, which means that … ?

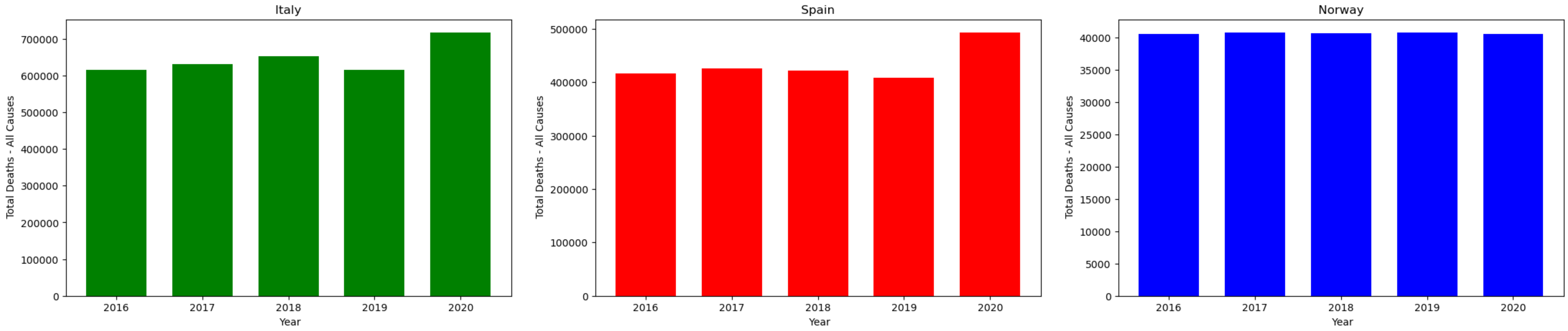
Li&Ma

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

The Li&Ma formula

Applying the Li&Ma significance in the '*real*' world

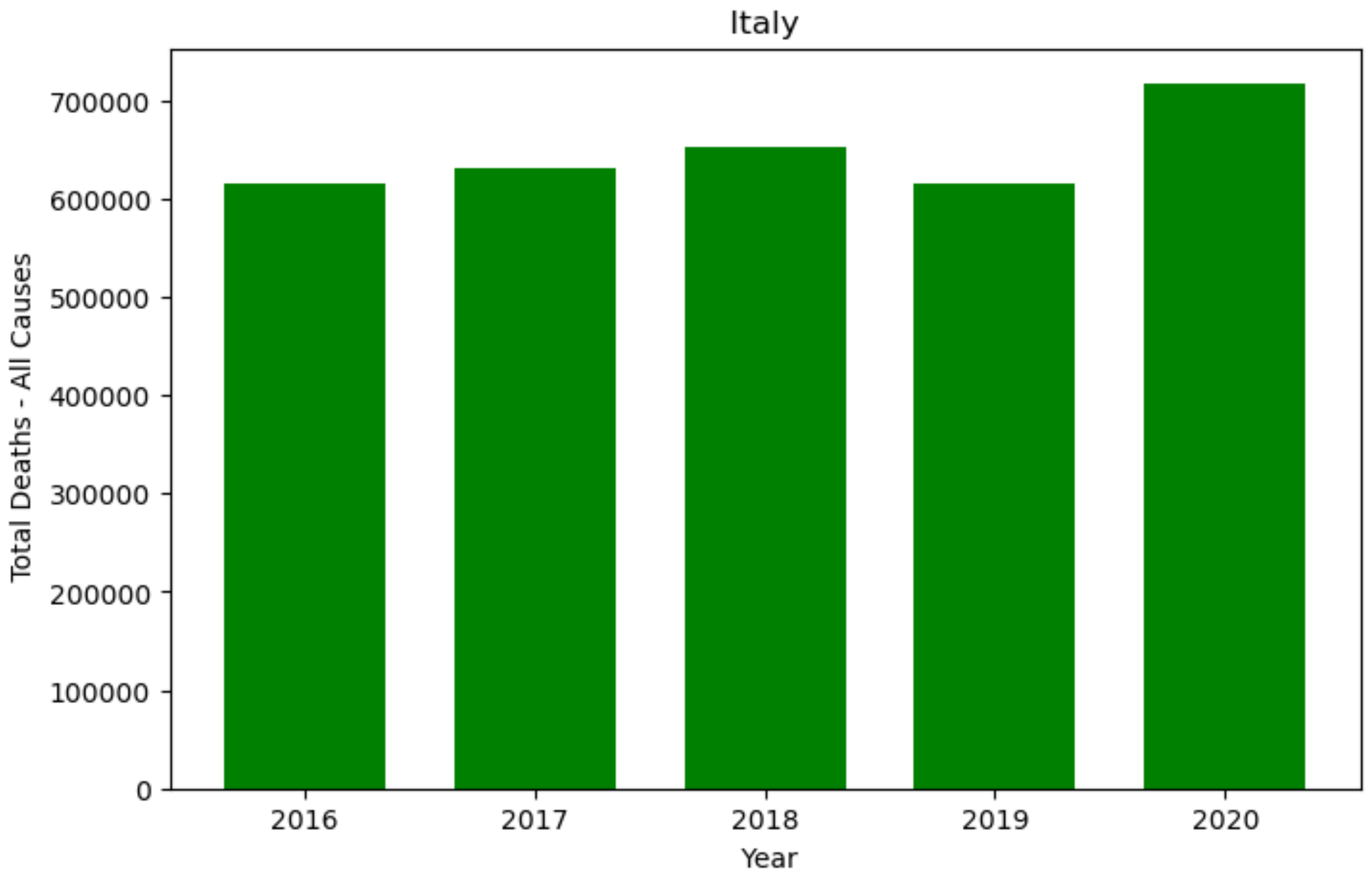
Source: UN World Population Prospects

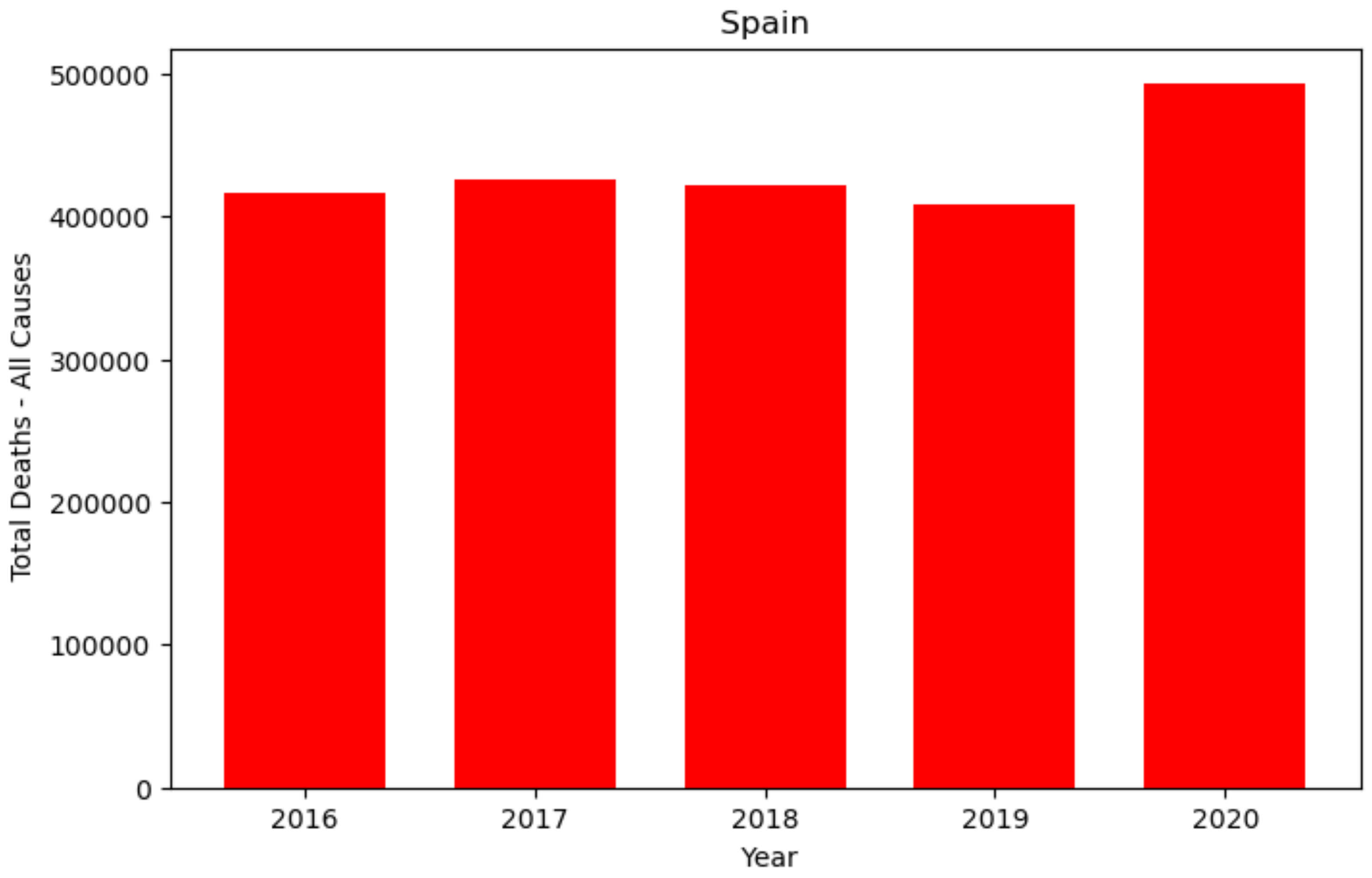


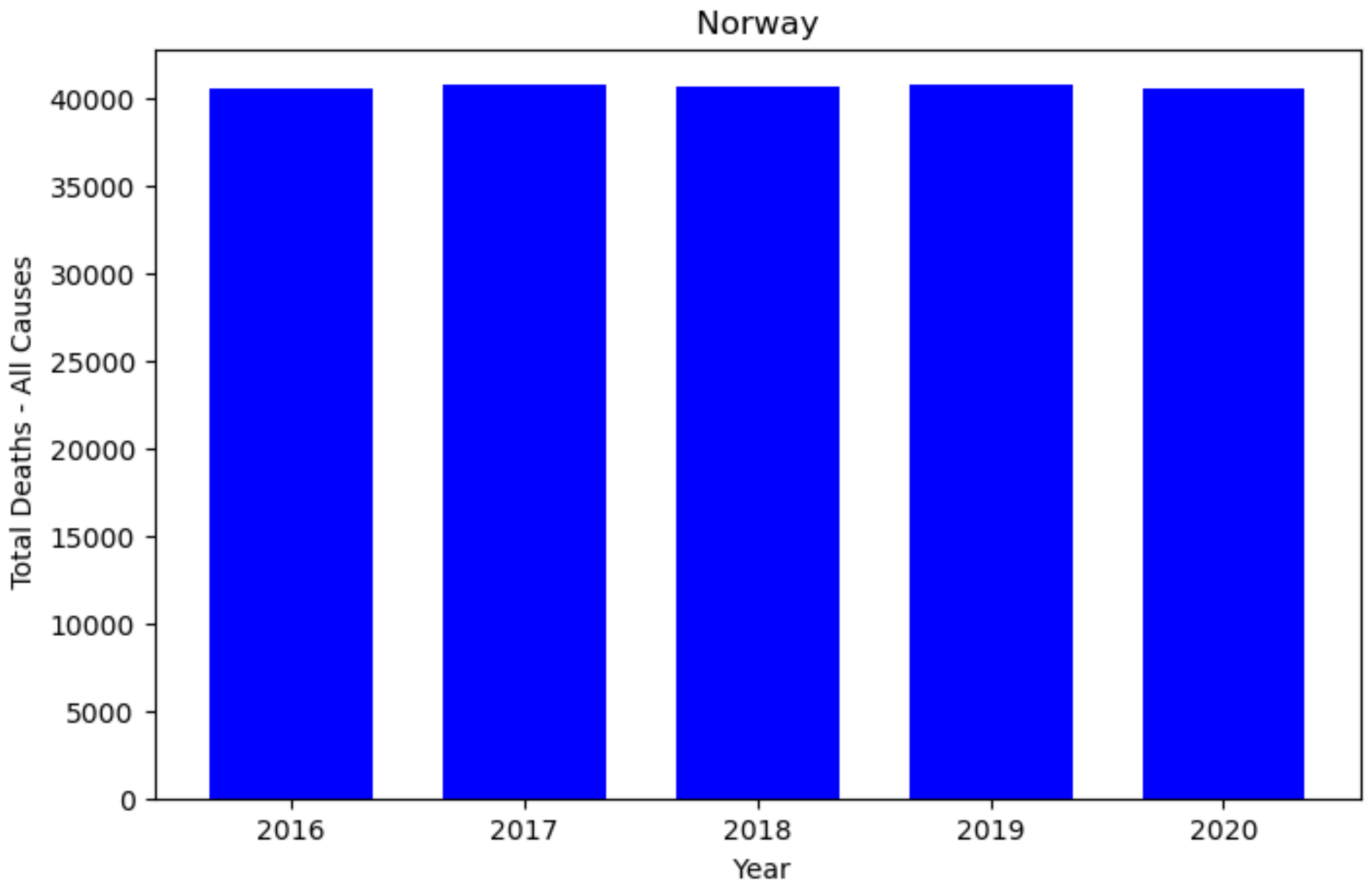
The Li&Ma formula

Applying the Li&Ma significance in the '*real*' world

Source: UN World Population Prospects







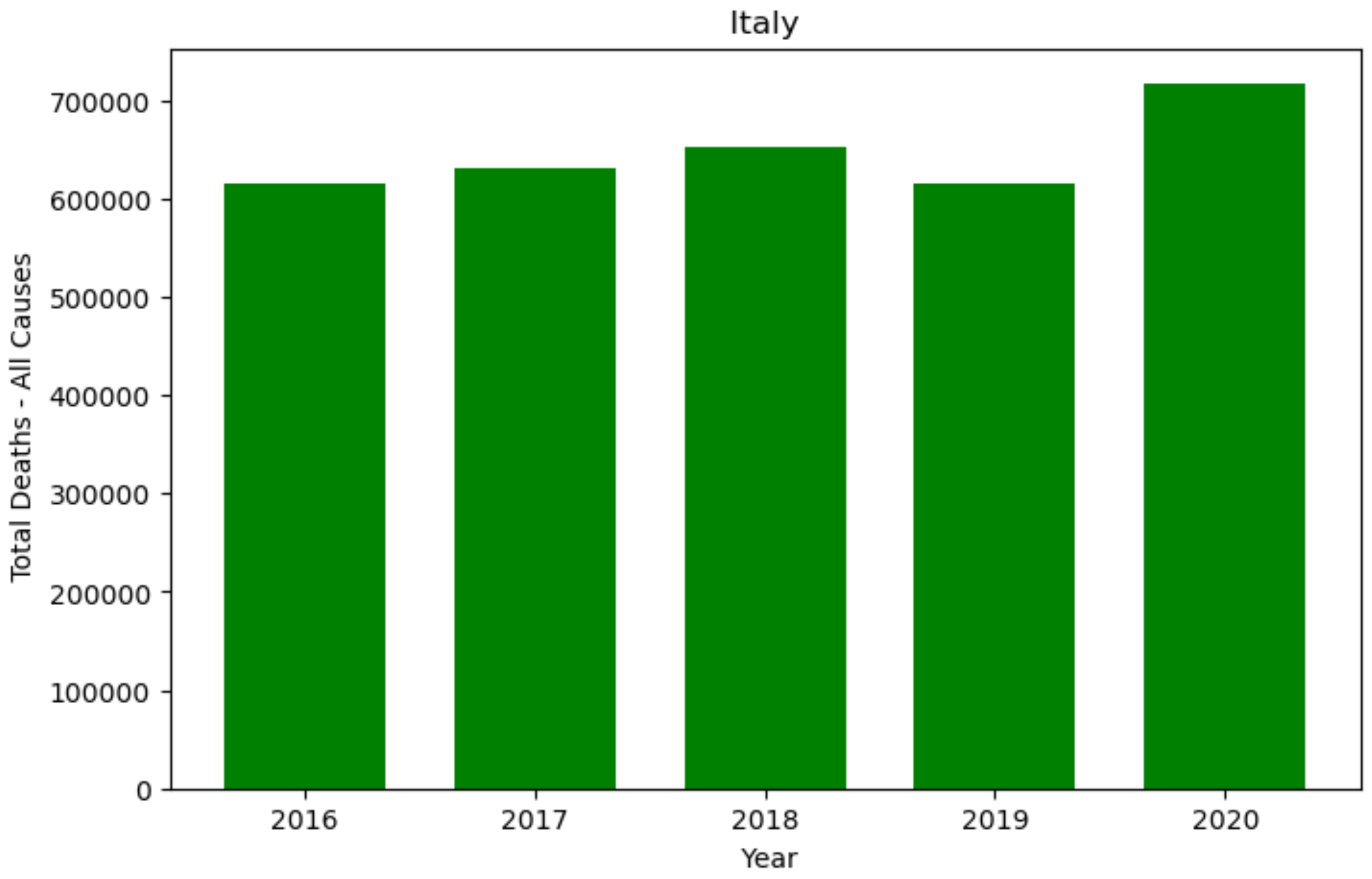
$$N_{off} = 2{,}512{,}853 \qquad N_{off} = 1{,}672{,}737 \qquad N_{off} = 162{,}809$$

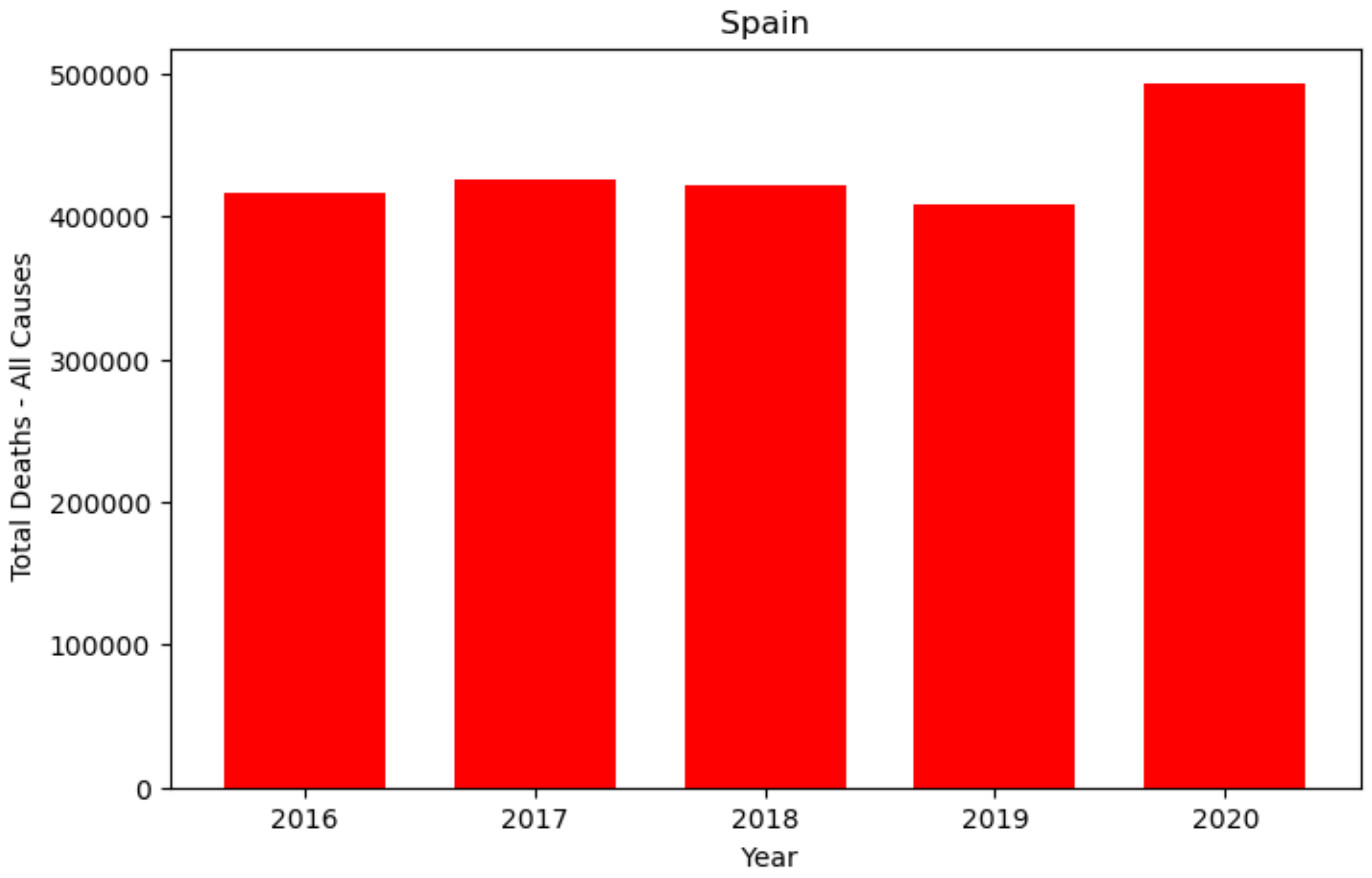
$$N_{on} = 716{,}753 \qquad N_{on} = 493{,}075 \qquad N_{on} = 40{,}578$$

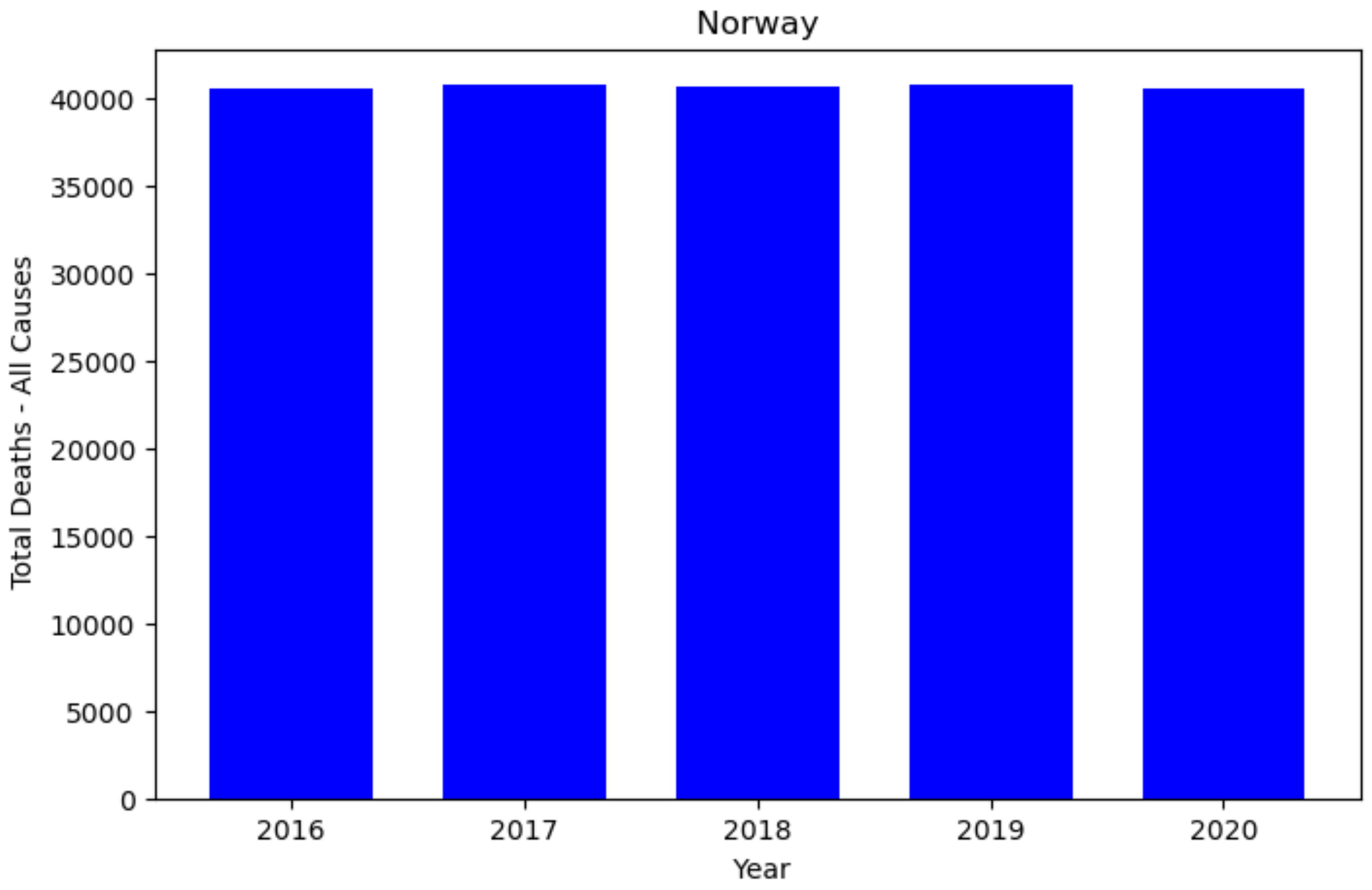
The Li&Ma formula

Applying the Li&Ma significance in the '*real*' world

Source: UN World Population Prospects







$$N_{off} = 2{,}512{,}853$$
$$N_{on} = 716{,}753$$
$$\alpha = 1/4$$

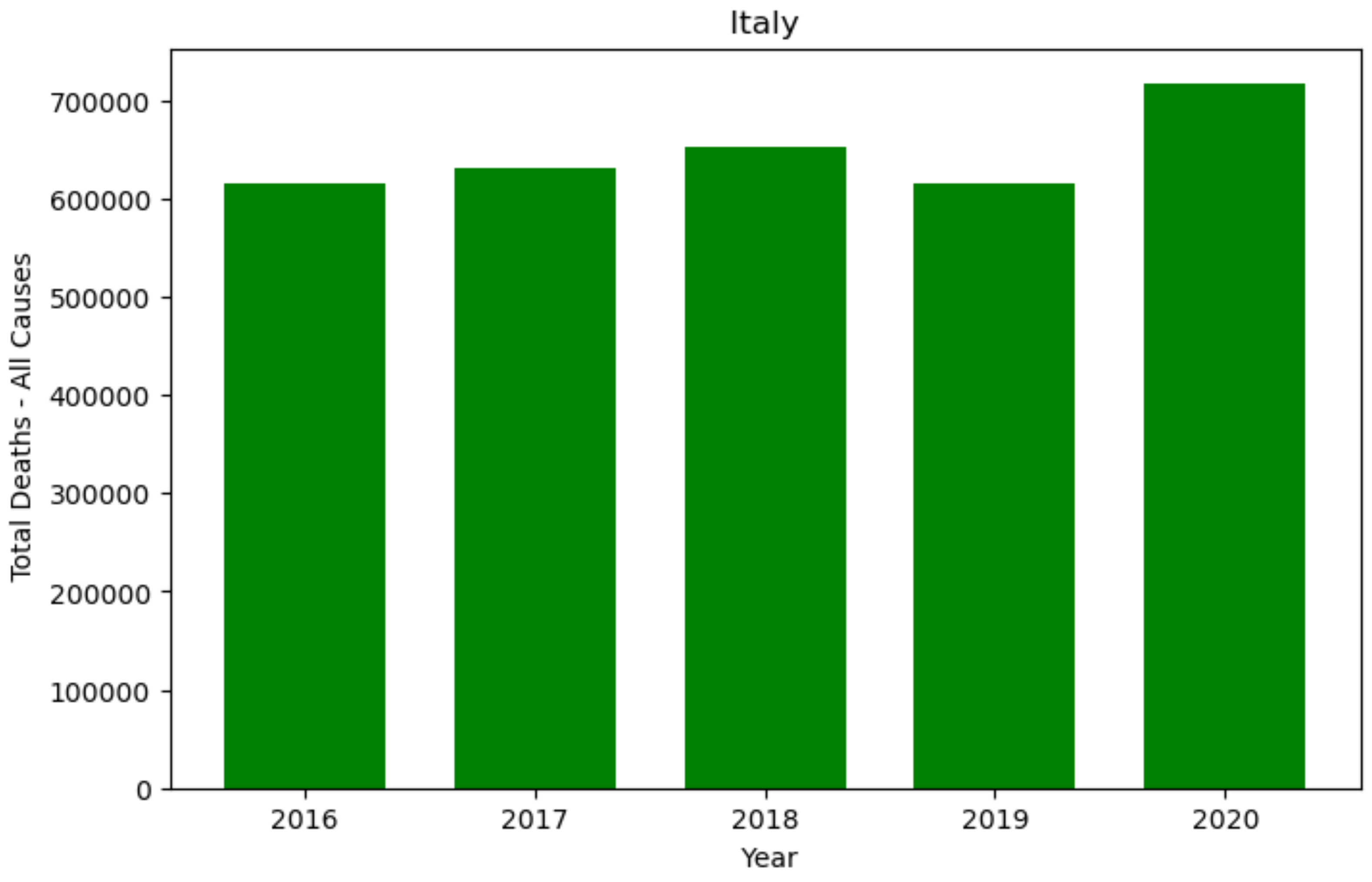
$$N_{off} = 1{,}672{,}737$$
$$N_{on} = 493{,}075$$
$$\alpha = 1/4$$

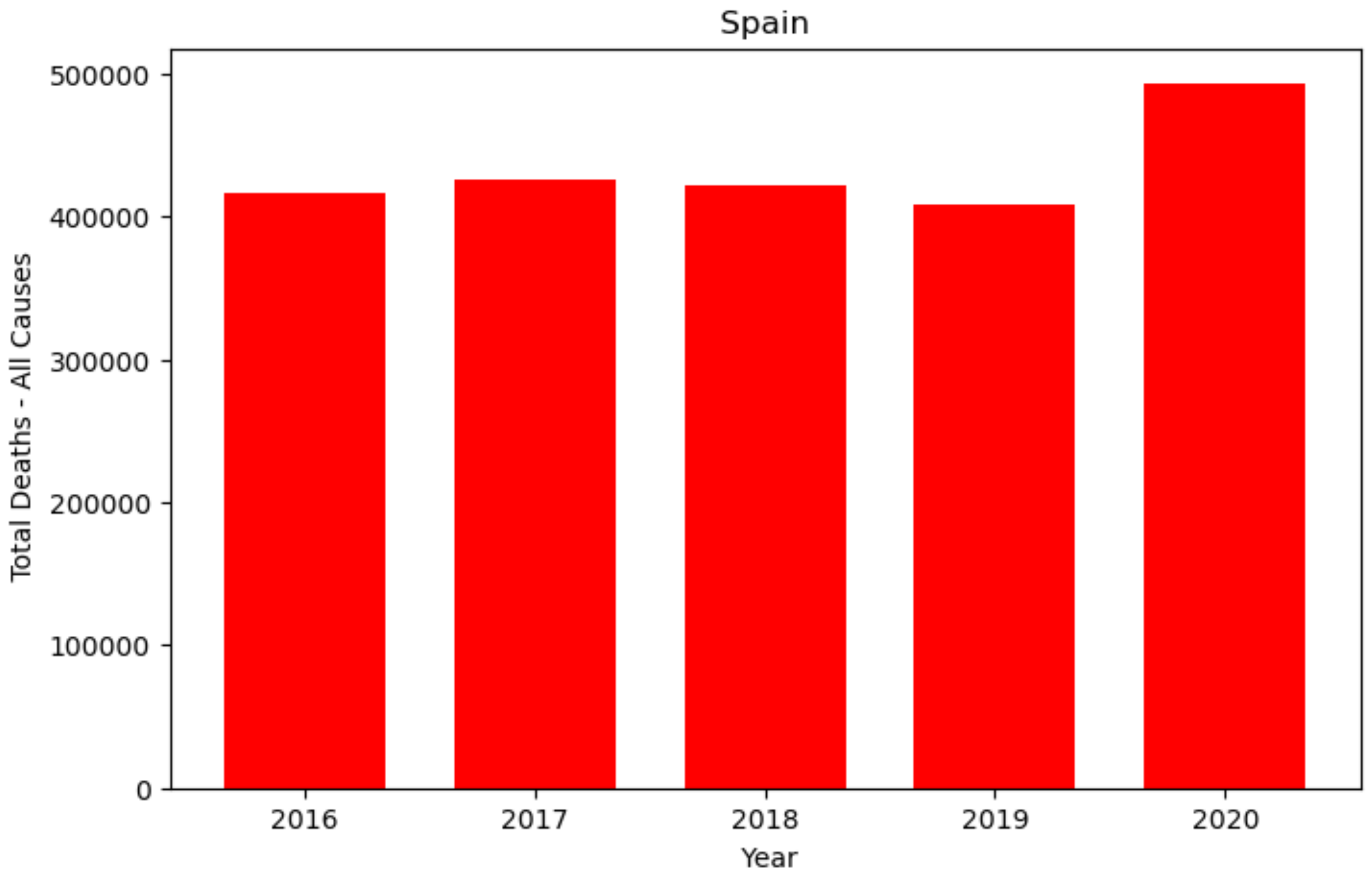
$$N_{off} = 162{,}809$$
$$N_{on} = 40{,}578$$
$$\alpha = 1/4$$

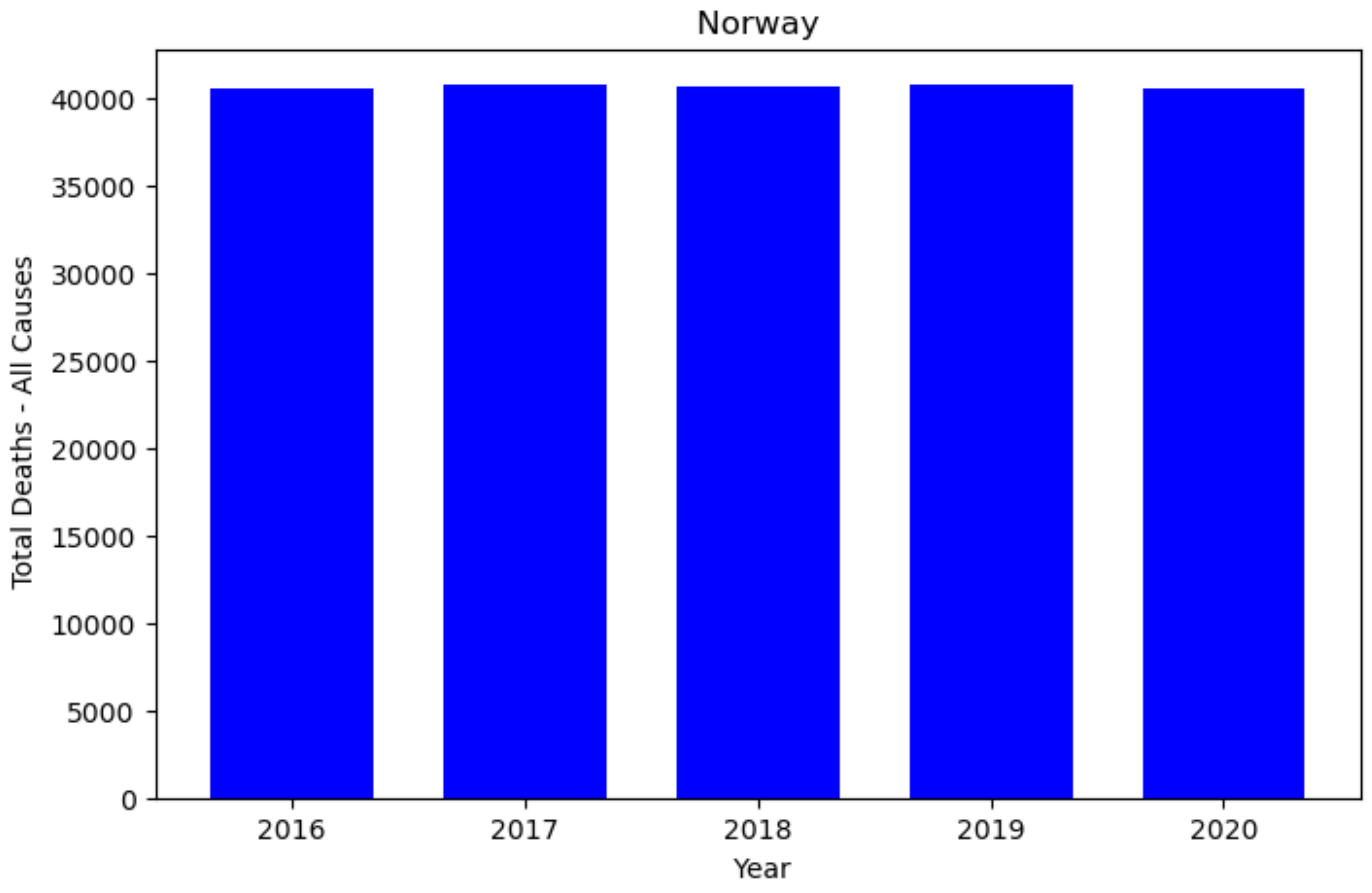
The Li&Ma formula

Applying the Li&Ma significance in the '*real*' world

Source: UN World Population Prospects







Italy:

$$N_{off} = 2{,}512{,}853$$
$$N_{on} = 716{,}753$$
$$\alpha = 1/4$$
$$\sigma = 97.3$$

Spain:

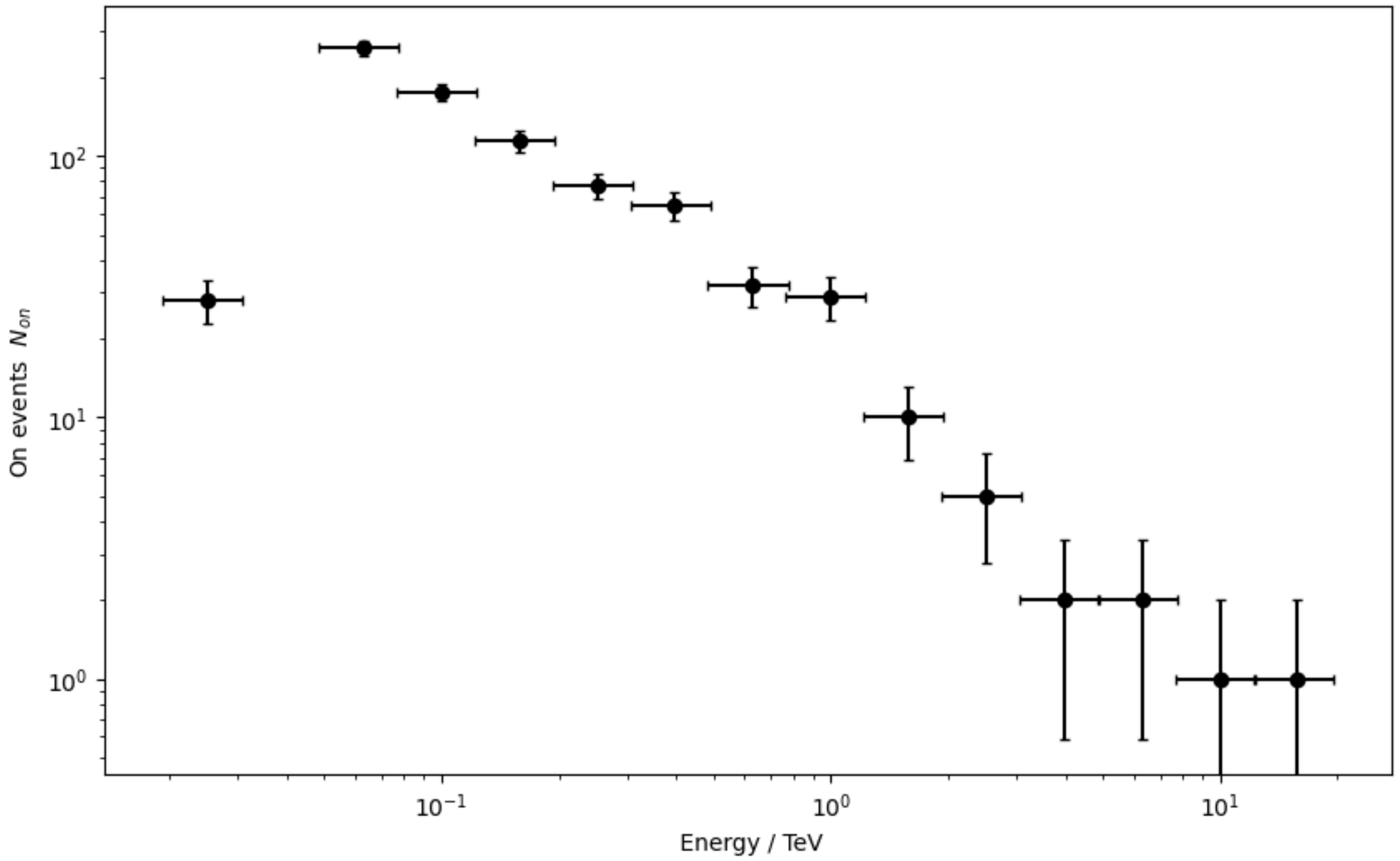
$$N_{off} = 1{,}672{,}737$$
$$N_{on} = 493{,}075$$
$$\alpha = 1/4$$
$$\sigma = 100$$

Norway:

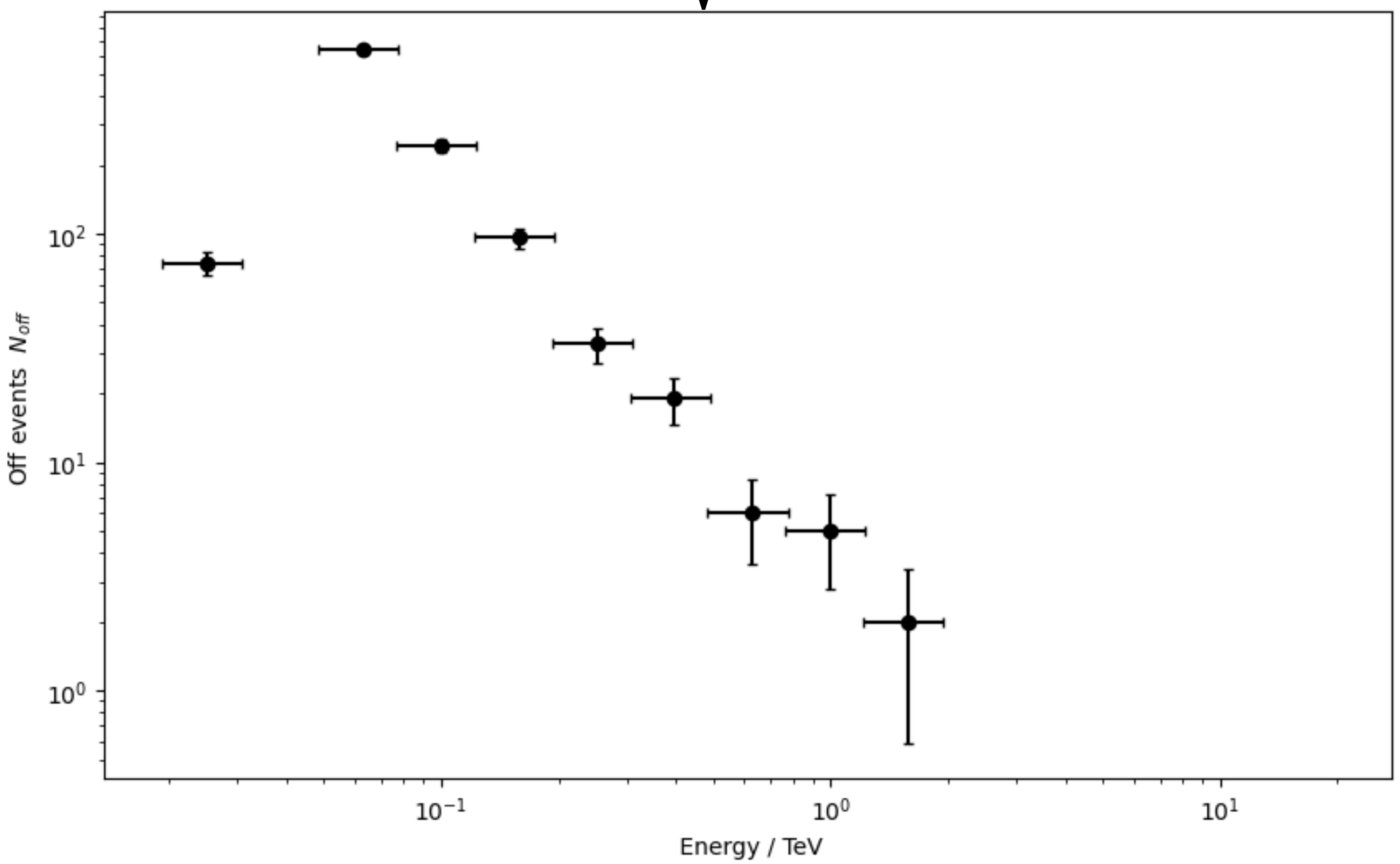
$$N_{off} = 162{,}809$$
$$N_{on} = 40{,}578$$
$$\alpha = 1/4$$
$$\sigma = -0.5$$

From excess to flux

$$N_{on} \pm \sqrt{N_{on}}$$

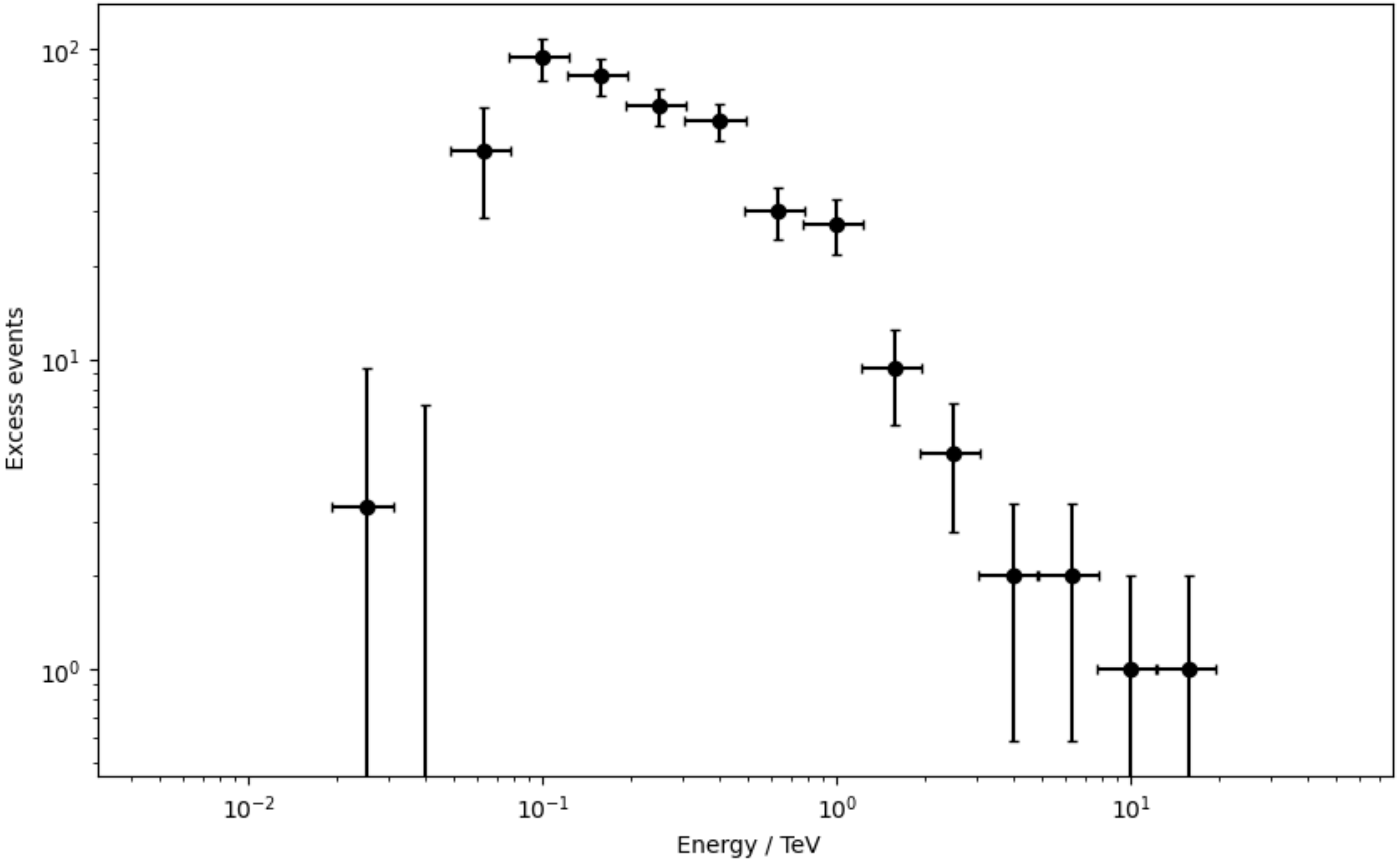


$$N_{off} \pm \sqrt{N_{off}}$$



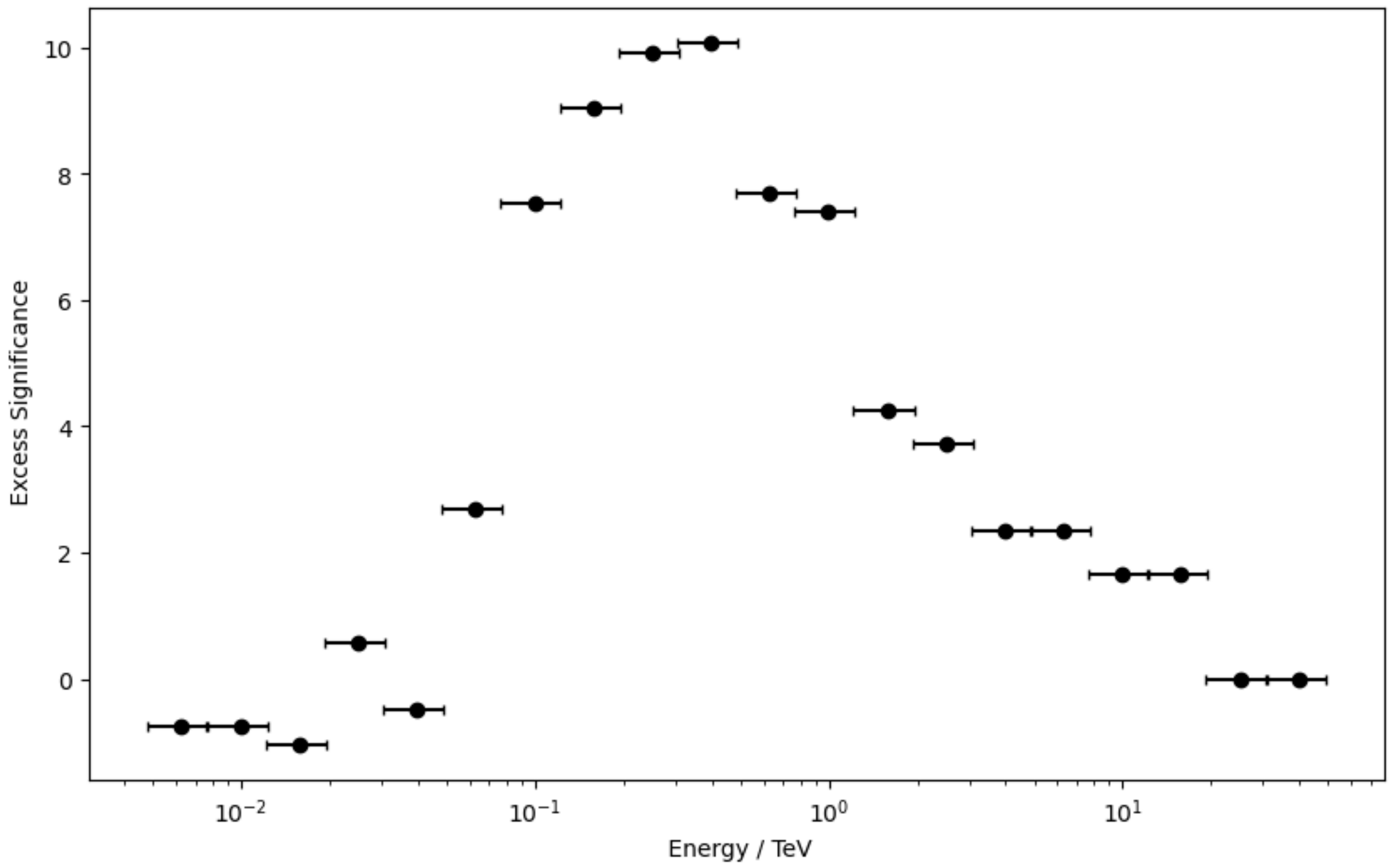
From excess to flux

$$N_{on} - \alpha N_{off} \pm \sqrt{N_{on} + \alpha^2 N_{off}}$$



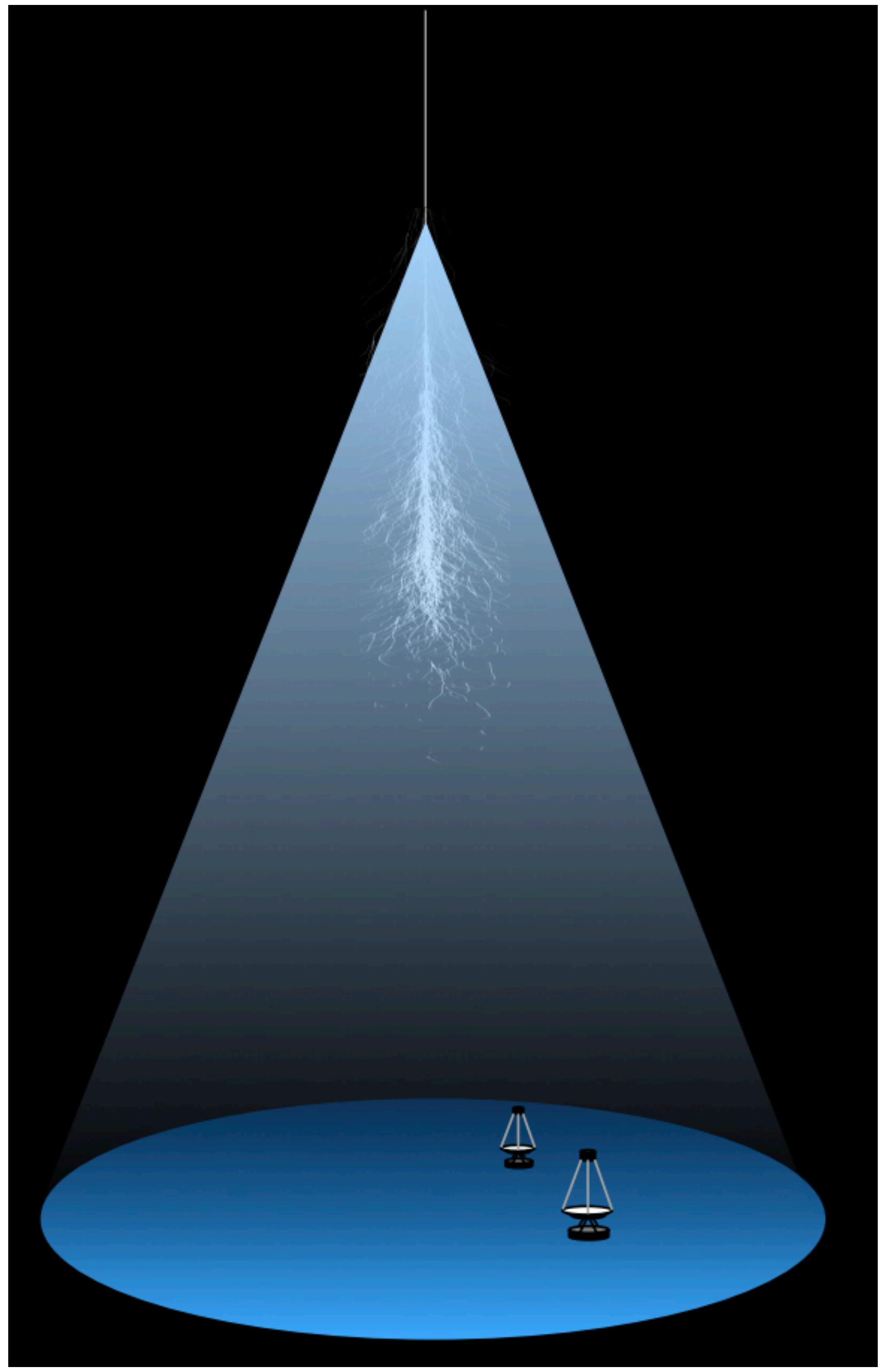
From excess to flux

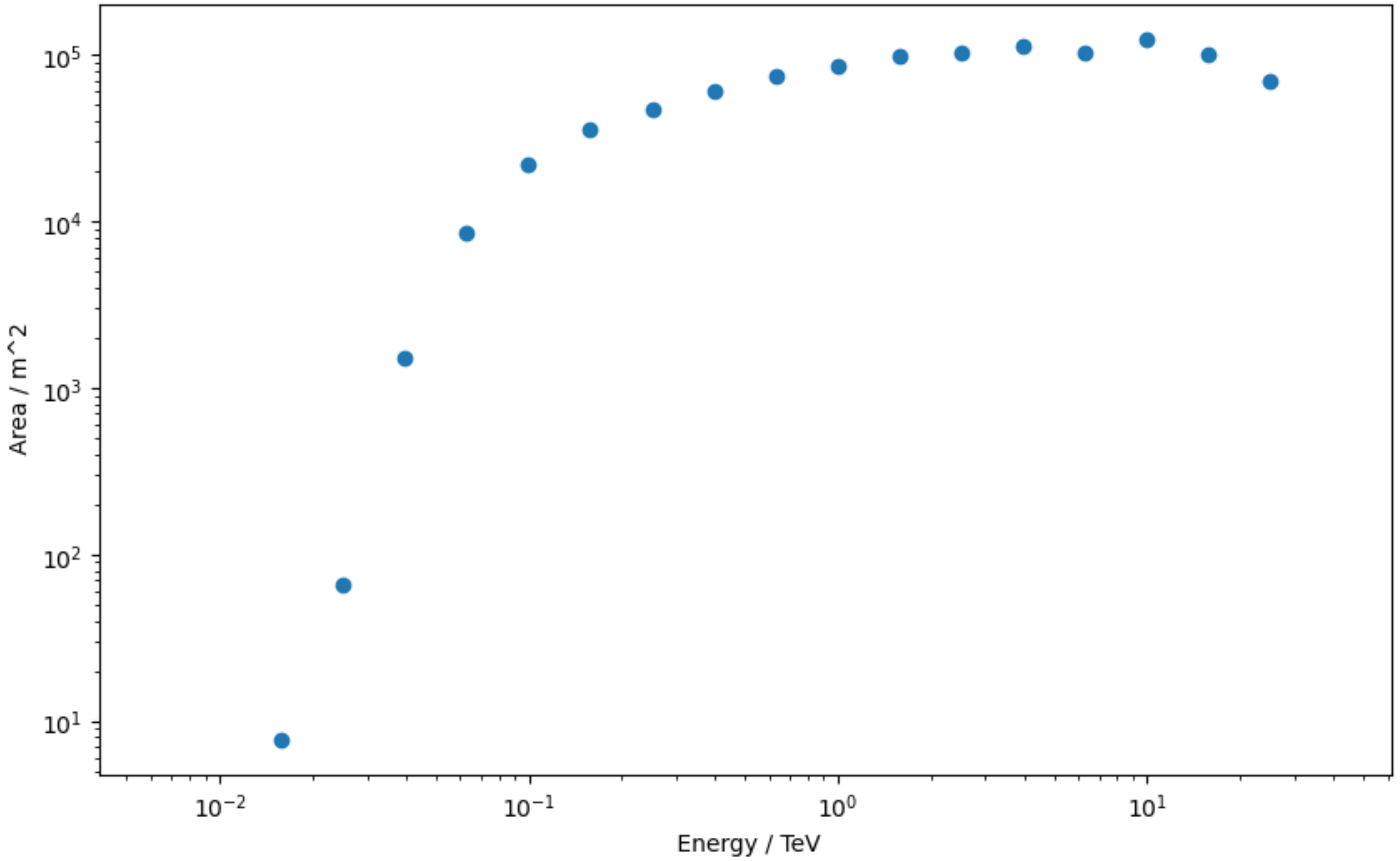
$$\pm\sqrt{2}\left[N_{on}\log\left(\frac{1}{\alpha}\frac{(\alpha+1)N_{on}}{N_{on}+N_{off}}\right)+N_{off}\log\left(\frac{(\alpha+1)N_{off}}{N_{on}+N_{off}}\right)\right]^{1/2}$$



From excess to flux

Collection Area of the telescope (per each energy bin)

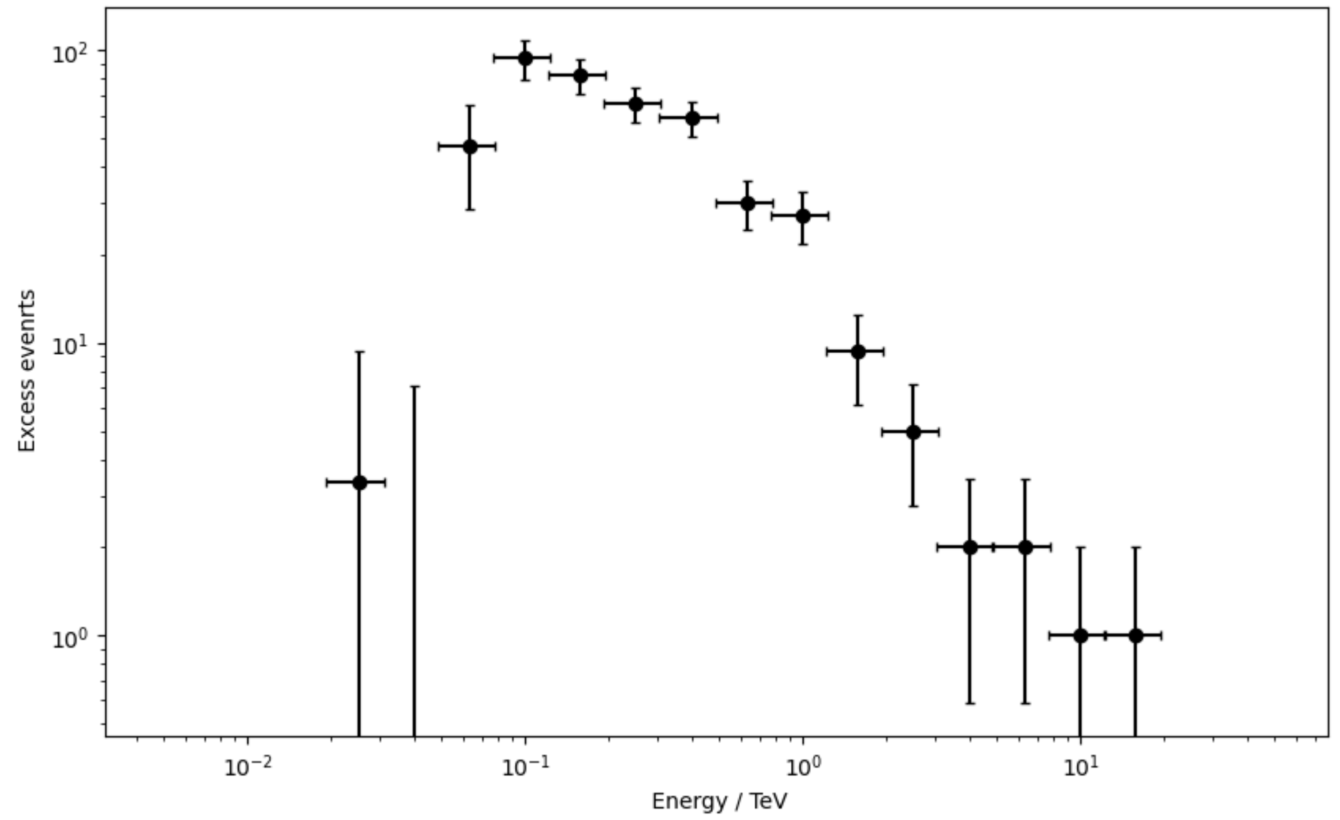


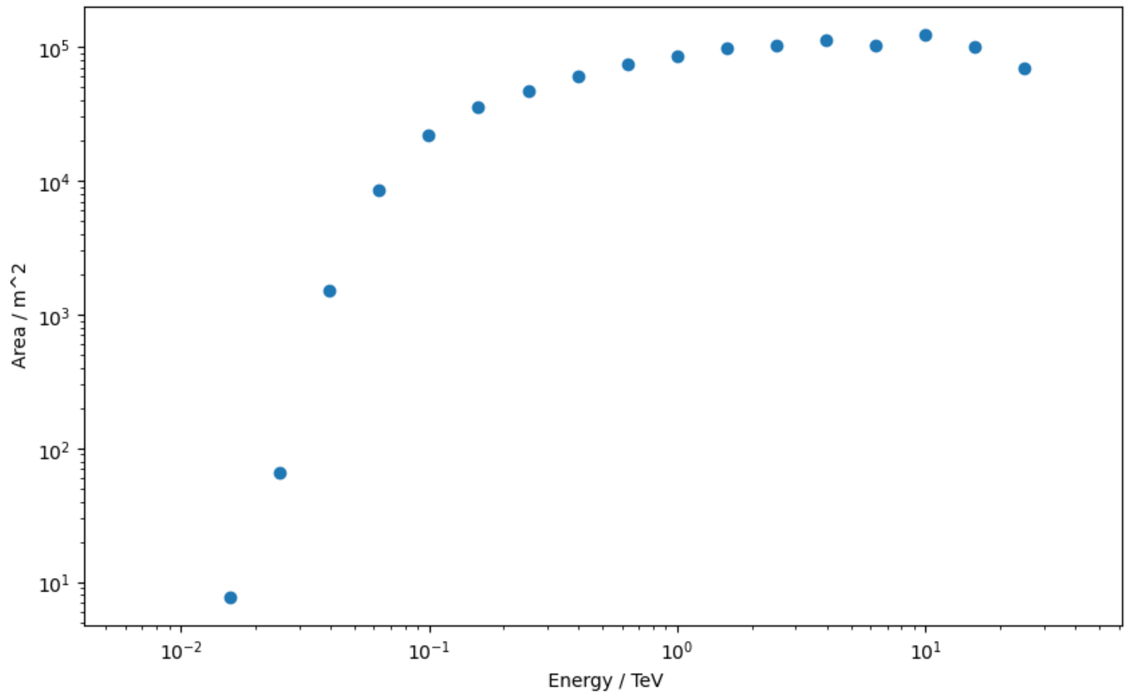


From excess to flux

At a first approximation (there are some caveats that won't be discussed here)

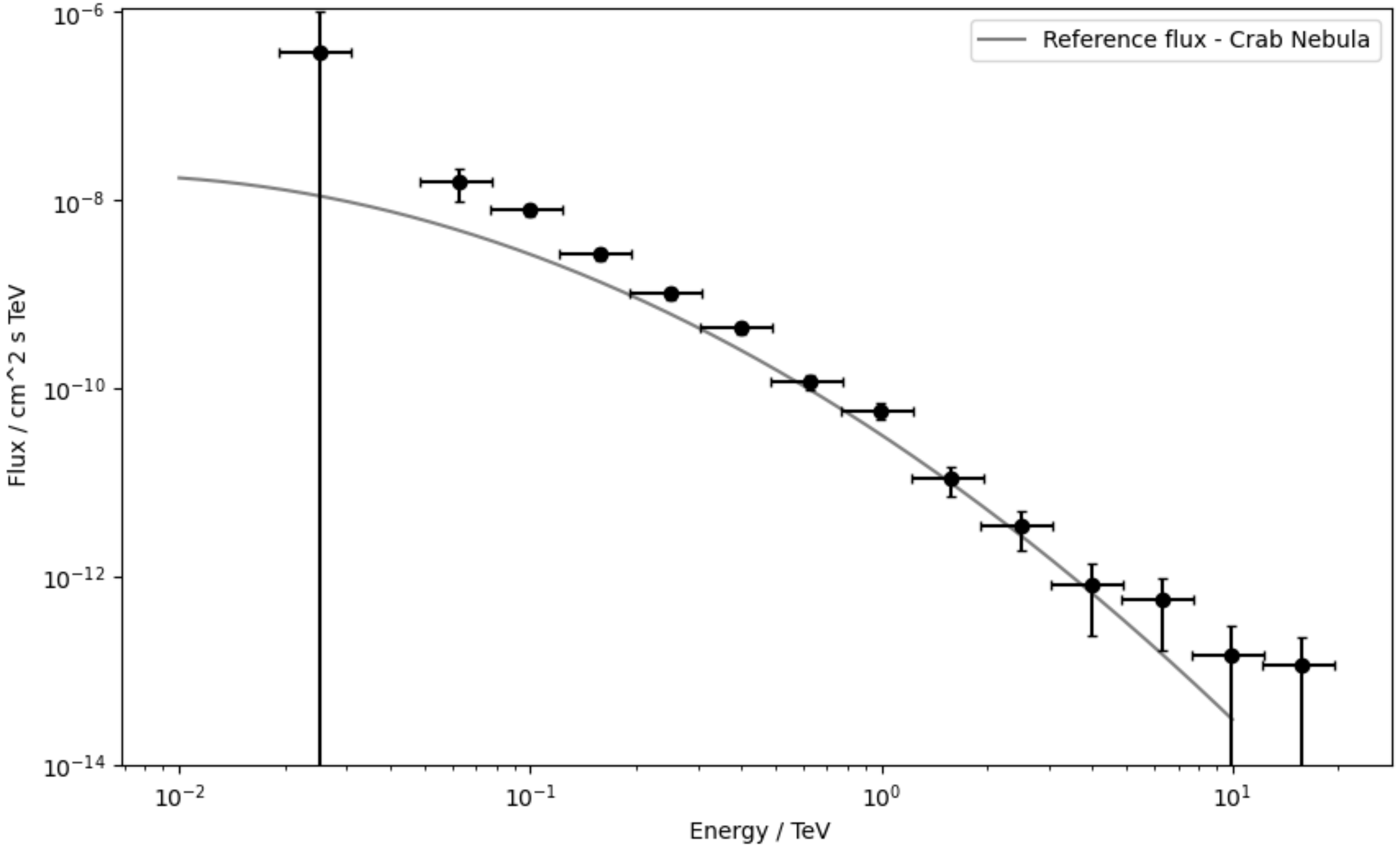
FLUX = EXCESS / (Collection Area X Observation Time)





$$\times \Delta T$$

From excess to flux



From excess to flux

With many more observation hours and after much more refined work:

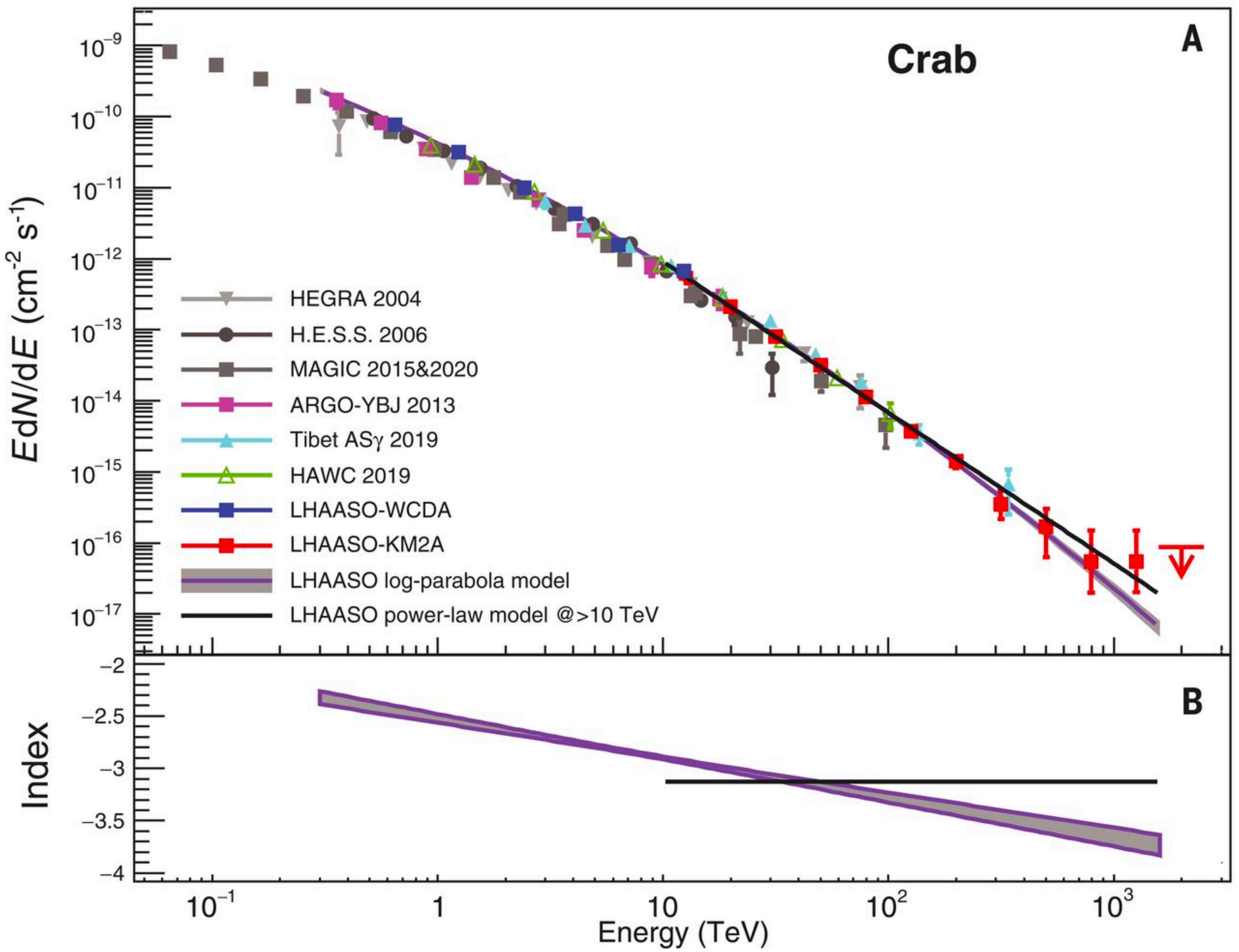


Image from: The LHAASO Collaboration et al. , Peta–electron volt gamma-ray emission from the Crab Nebula.Science373,425-430(2021).DOI:10.1126/science.abg5137

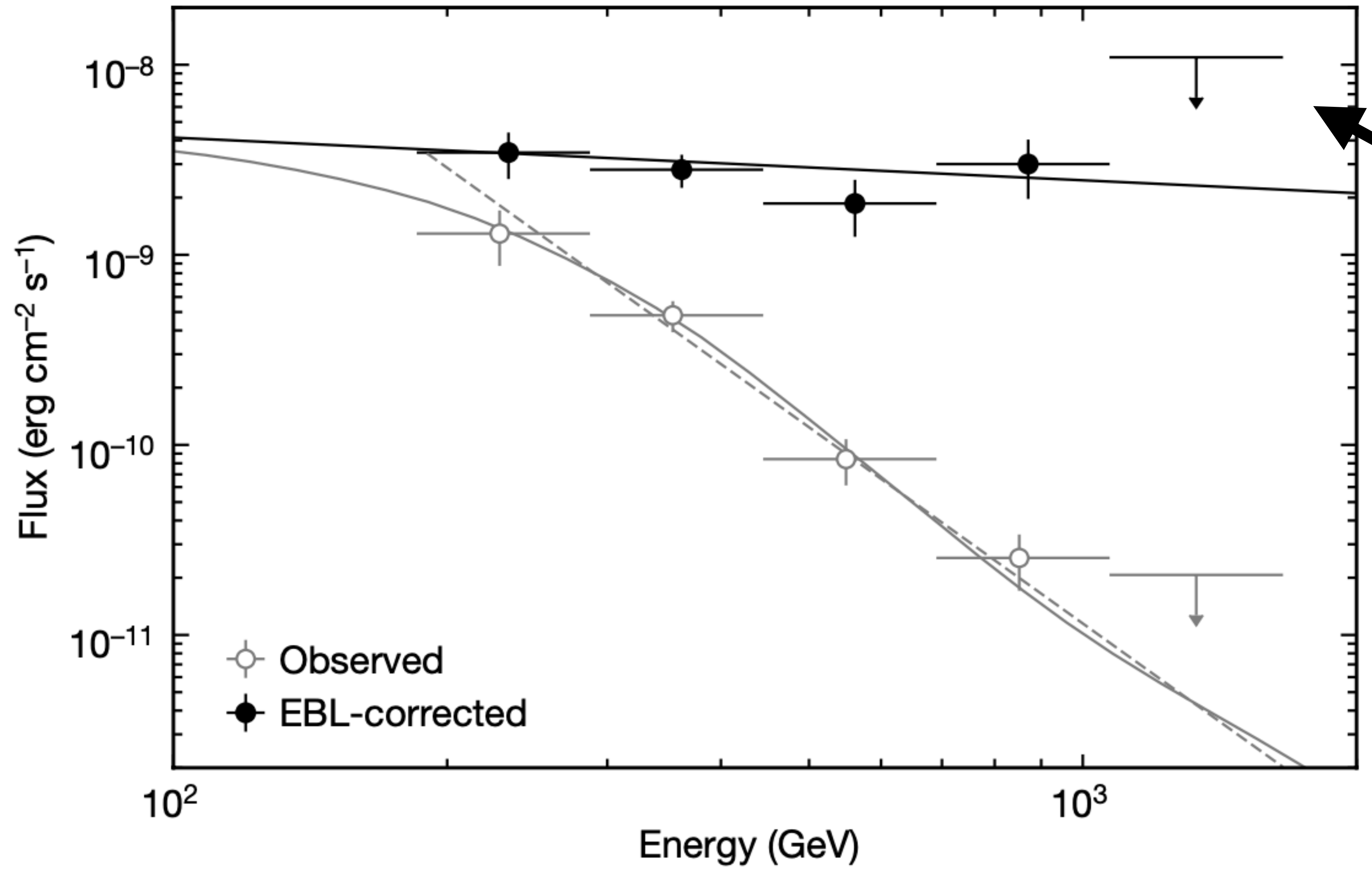
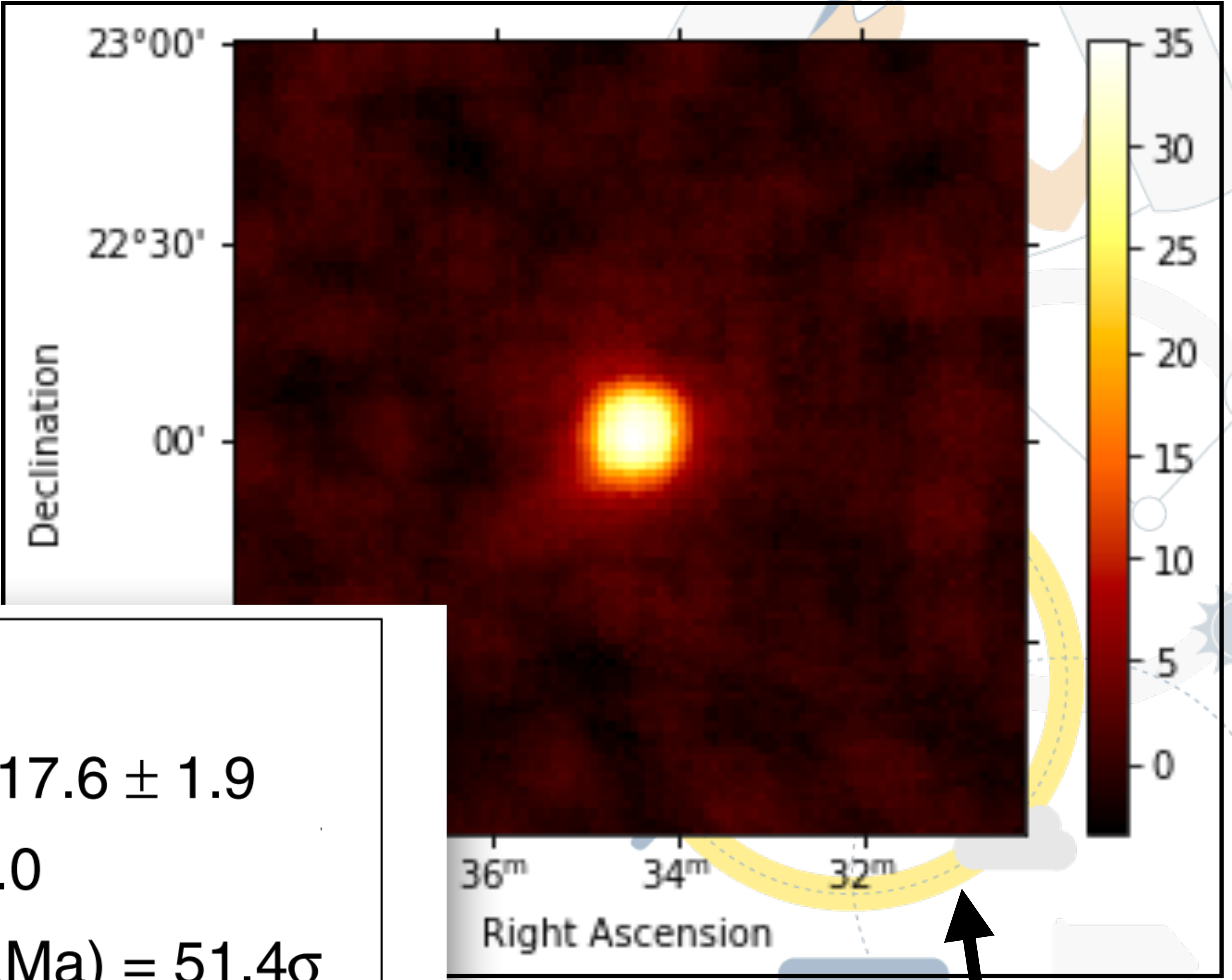
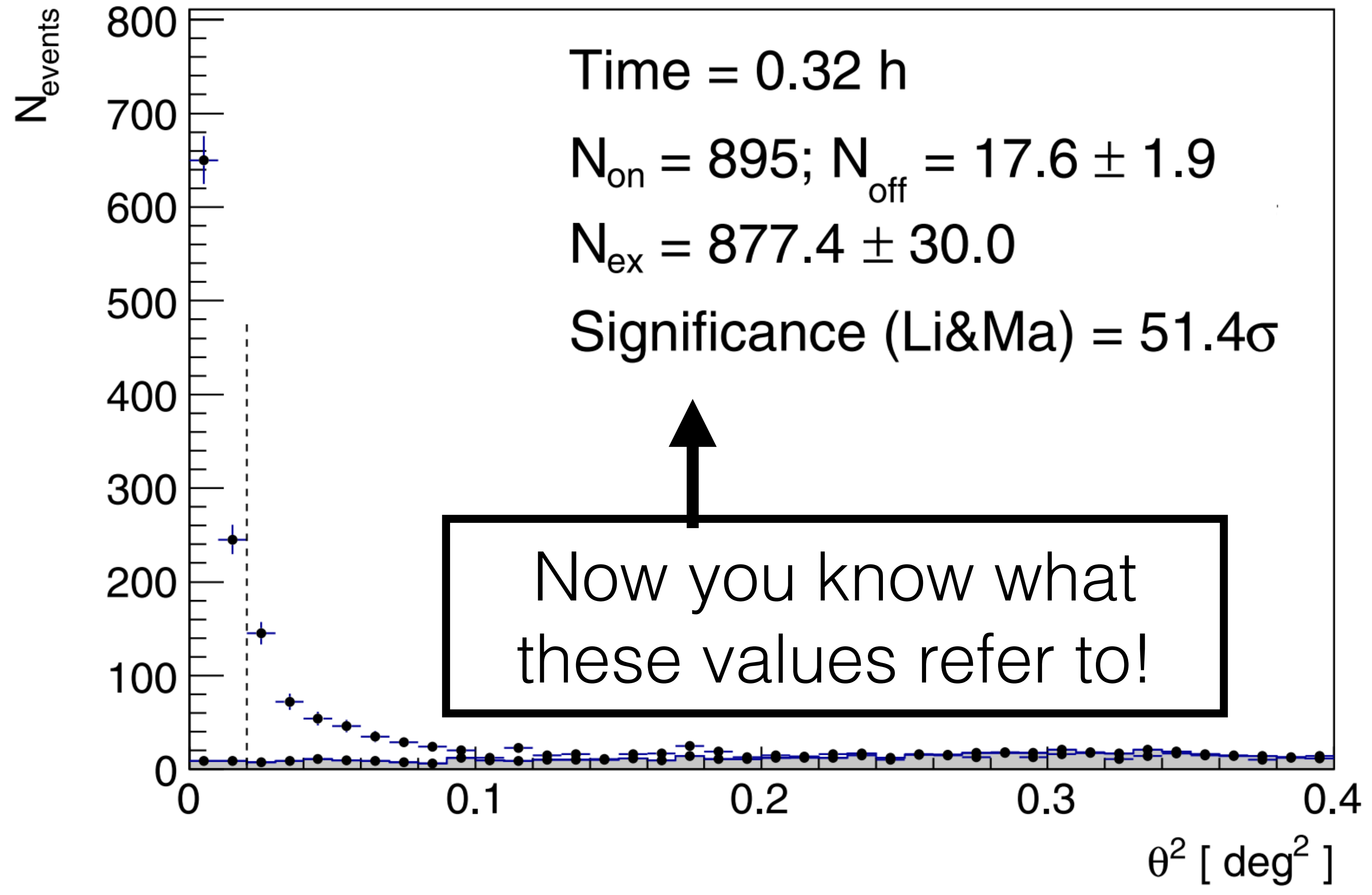


Fig. 2 | Spectrum above 0.2 TeV averaged over the period between T_0 + 62 s and T_0 + 2,454 s for GRB 190114C. Spectral-energy distributions for the spectrum observed by MAGIC (grey open circles) and the intrinsic spectrum corrected for EBL attenuation[25] (blue filled circles). The errors on the flux correspond to one standard deviation. The upper limits at 95% confidence level are shown for the first non-significant bin at high energies. Also shown is the best-fit model for the intrinsic spectrum (black curve) when assuming a power-law function. The grey solid curve for the observed spectrum is obtained by convolving this curve with the effect of EBL attenuation. The grey dashed curve is the forward-folding fit to the observed spectrum with a power-law function (Methods).

This arrow here indicates an upper limit, and now you know what it means!





Extended Data Fig. 2 | Significance of the γ-ray signal between T_0 + 62 s and T_0 + 1,227 s for GRB 190114C. Distribution of the squared angular distance, θ^2, for the MAGIC data (points) and background events (grey shaded area). θ^2 is defined as the squared angular distance between the nominal position of the source and the reconstructed arrival direction of the events. The dashed vertical line represents the value of the cut on θ^2. This defines the signal region, where the number of events coming from the source (N_{on}) and from the background (N_{off}) are computed. The errors for 'on' events are derived from Poissonian statistics. From N_{on} and N_{off}, the number of excess events (N_{ex}) is computed. The significance is calculated using the Li & Ma method[42].

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

Recap

1. We have defined an **On/Off measurement**, which is the most common type of measurement in gamma-ray astronomy when dealing with an unknown **background**

2. We have seen how to **estimate the excess** from an On and Off measurement in both the frequentist and bayesian approaches and how to put **confidence/credible intervals** on such estimates

3. The **frequentist** approach allows us to exclude the null hypothesis with given confidence via the usage of the **Li&Ma expression**

We will apply this knowledge in the hands-on sessions on the **spectra** and **light curve** analysis!