Introduction to Statistical Inference

Giacomo D'Amico

arQus
European University Alliance
TWINNING ACTIVITIES
2021/2022

ArQus School 2022, Bergen, Norway

5-9 Sep. 2022

What do we mean by inferring?

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

What do we mean by inferring?

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

An opinion that has to be *quantified* through the instrument of **probability** and **statistics**

A given theoretical model

The data we have collected

The Model → *All sheep are white*

The data →

The opinion → *The model is rejected*

The Model → *1% of the sheep are black*

The data →

The opinion → ?

We will come back later on this!

Two approaches are used to **quantify** an ***opinion*** about a **model** given an **observation**

- 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?*

- The **Frequentist approach** tries to answer the 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?*

The Bayesian approach

The Bayes theorem

- Marginalised probability

$$f(x|I) = \int f(x, y|I)dy$$

- Conditional probability

$$f(x, y|I) = f(x|y, I) \cdot f(y|I)$$

- I represents our prior knowledge
- $f()$ is for a generic probability distribution (or mass) function

The Bayes theorem

- Marginalised probability

$$f(x|I) = \int f(x, y|I)dy$$

- Conditional probability

$$f(x, y|I) = f(x|y, I) \cdot f(y|I)$$

$$\left.\begin{array}{l} \\ \\ \end{array}\right\} \quad f(x|y, I) = \frac{f(y|x, I)f(x|I)}{\int f(y|x, I)f(x|I)dx}$$

- I represents our prior knowledge
- $f()$ is for a generic probability distribution (or mass) function

The Bayes theorem

- Marginalised probability

$$f(x|I) = \int f(x,y|I)dy$$

- Conditional probability

$$f(x,y|I) = f(x|y,I) \cdot f(y|I)$$

Likelihood Prior

$$f(x|y,I) = \frac{f(y|x,I)f(x|I)}{\int f(y|x,I)f(x|I)dx}$$

Posterior Normalisation

- I represents our prior knowledge
- $f()$ is for a generic probability distribution (or mass) function

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

1

2

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

1

2

- 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

- 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

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

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

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

The Monty Hall problem

1

2

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

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

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

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

The Monty Hall problem

1

2

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

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

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

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

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

The Monty Hall problem

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

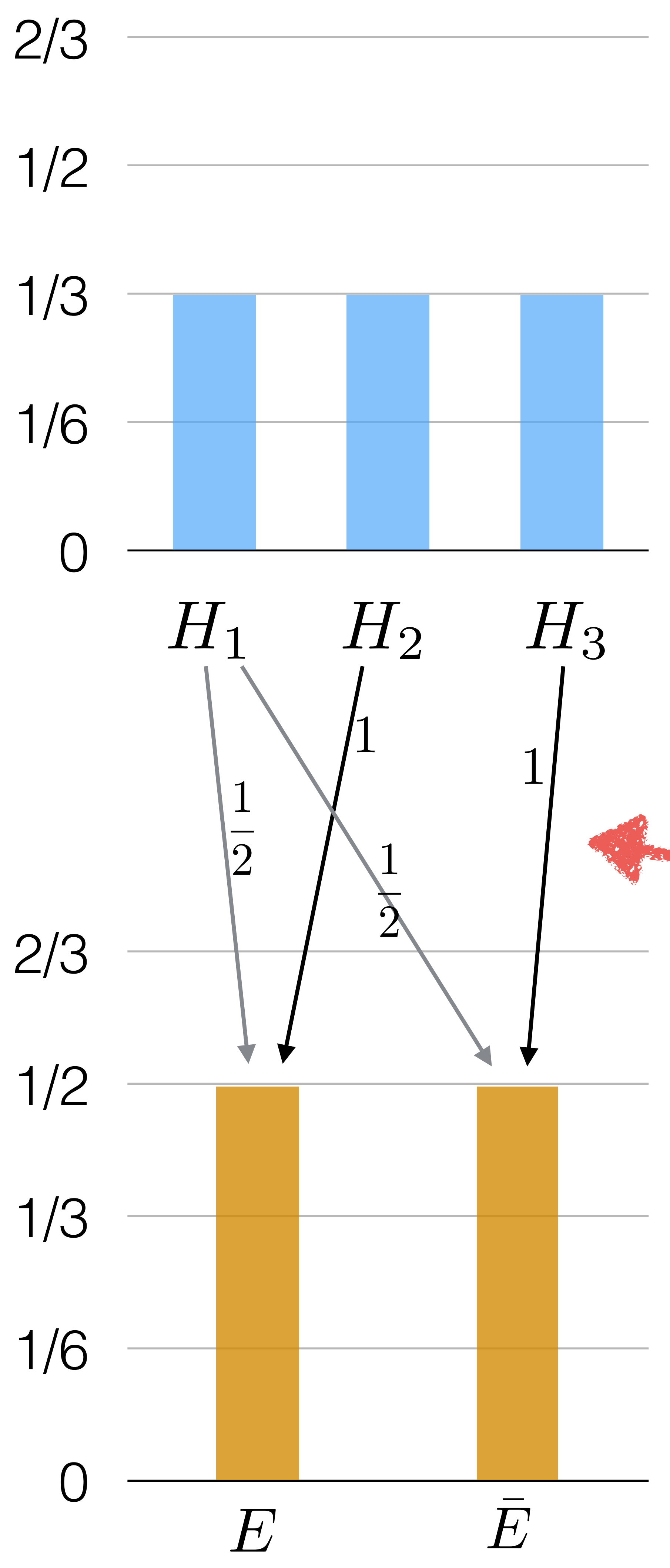
$$f(H_2|E,I) = \frac{f(E|H_2,I)f(H_2|I)}{f(E|I)} = \frac{1 \cdot 1/3}{1/2} = \frac{2}{3}$$

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

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

The Monty Hall problem

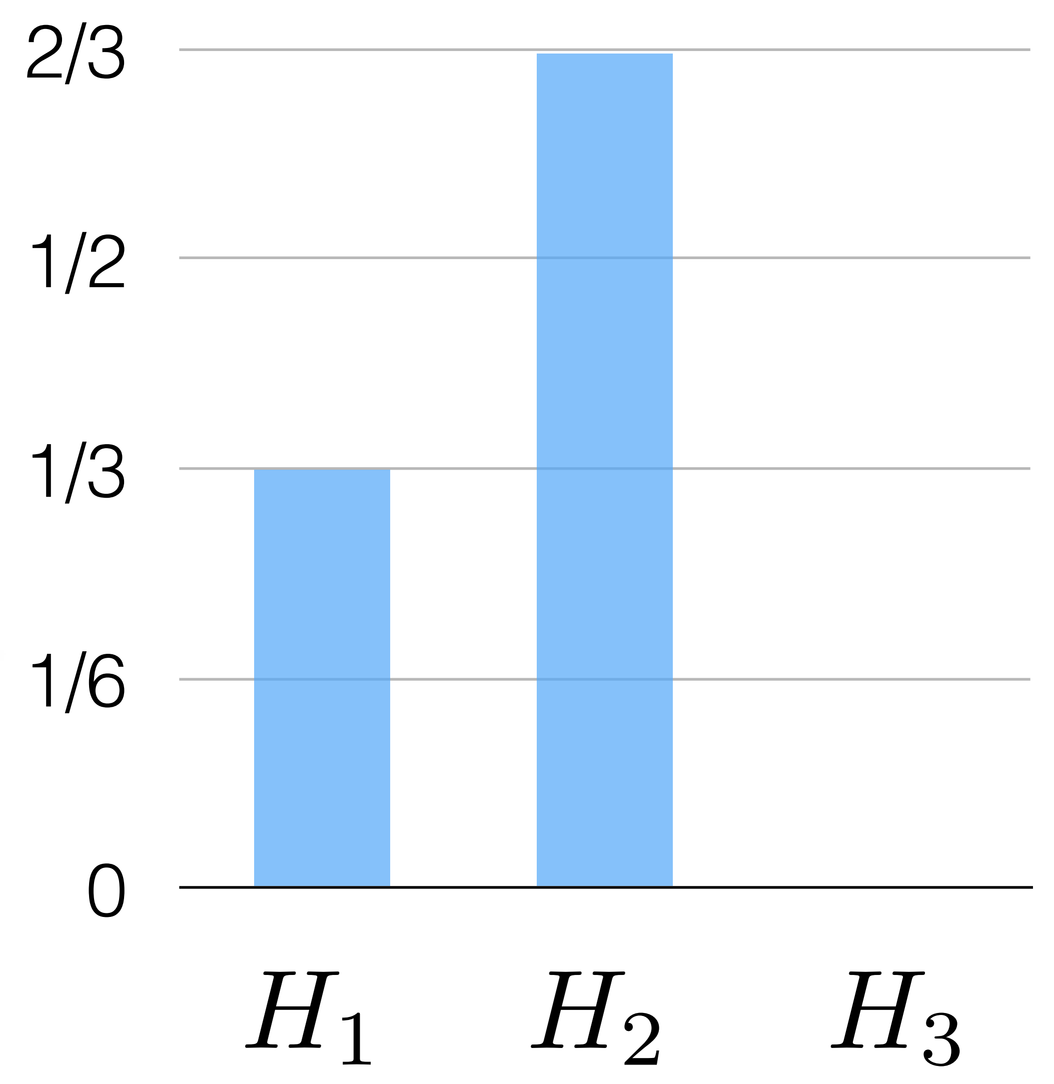
What if the TV-Show hoster did not know where the car is?



Priors

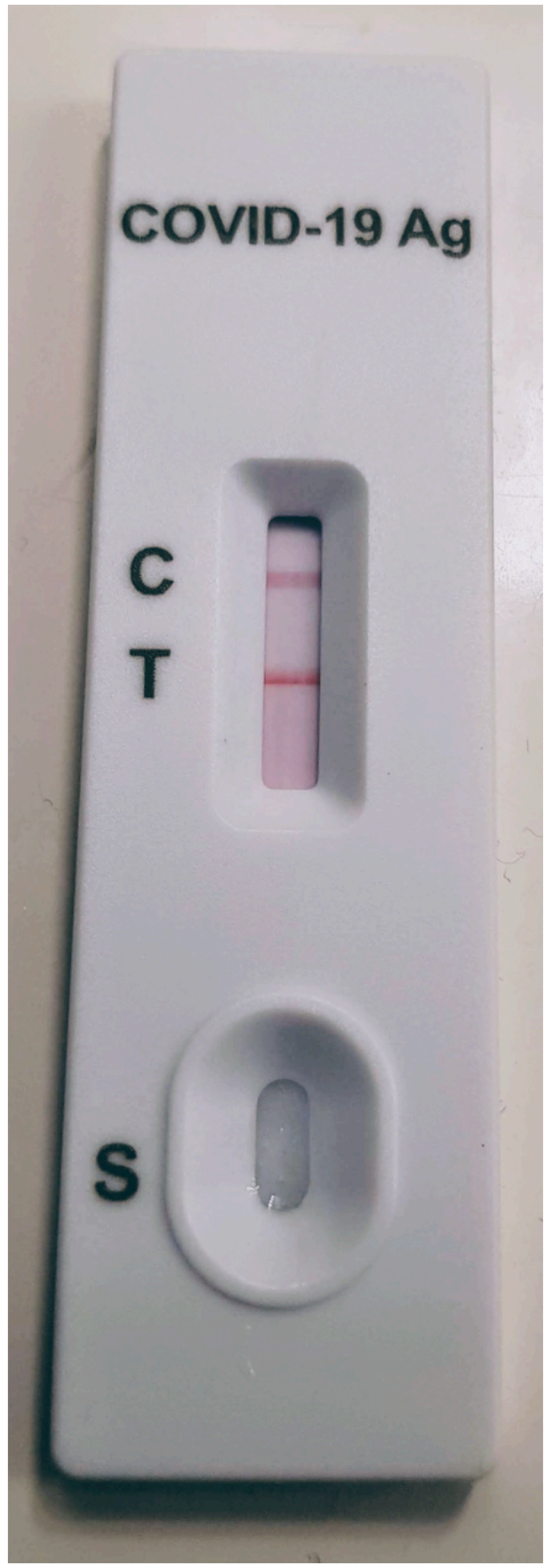
Likelihoods

"I observe" E



Posteriors

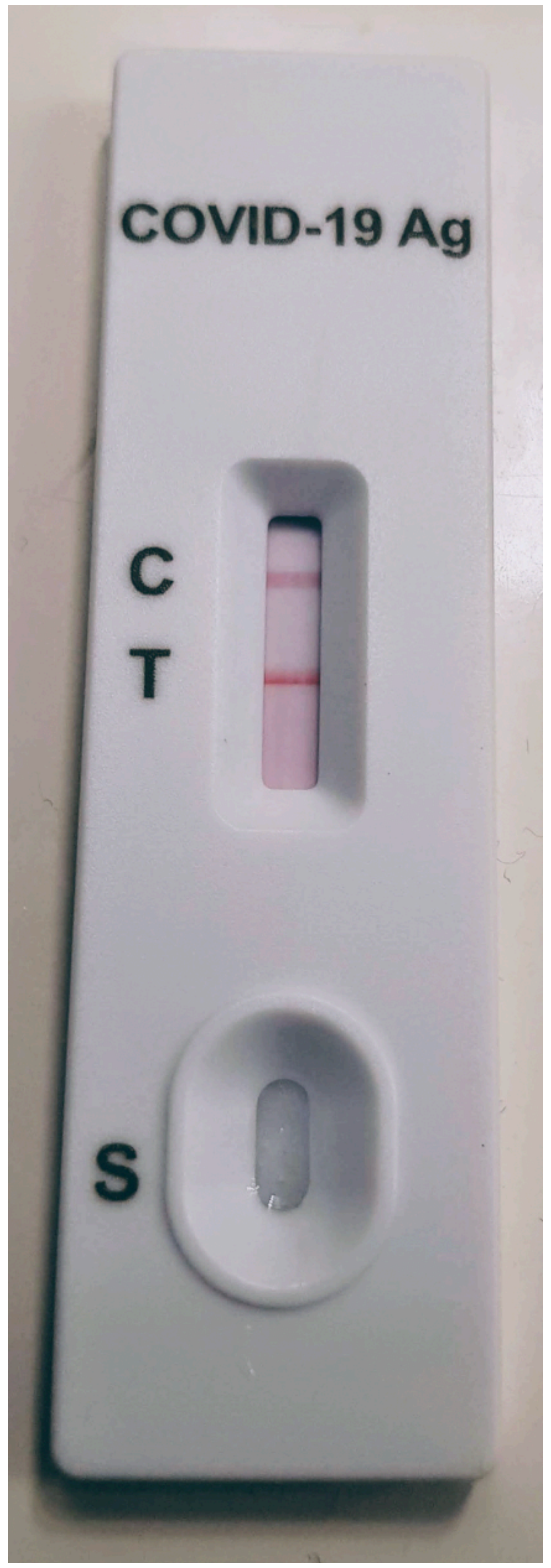
Another example: Covid-19 test



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

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

Another example: Covid-19 test



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

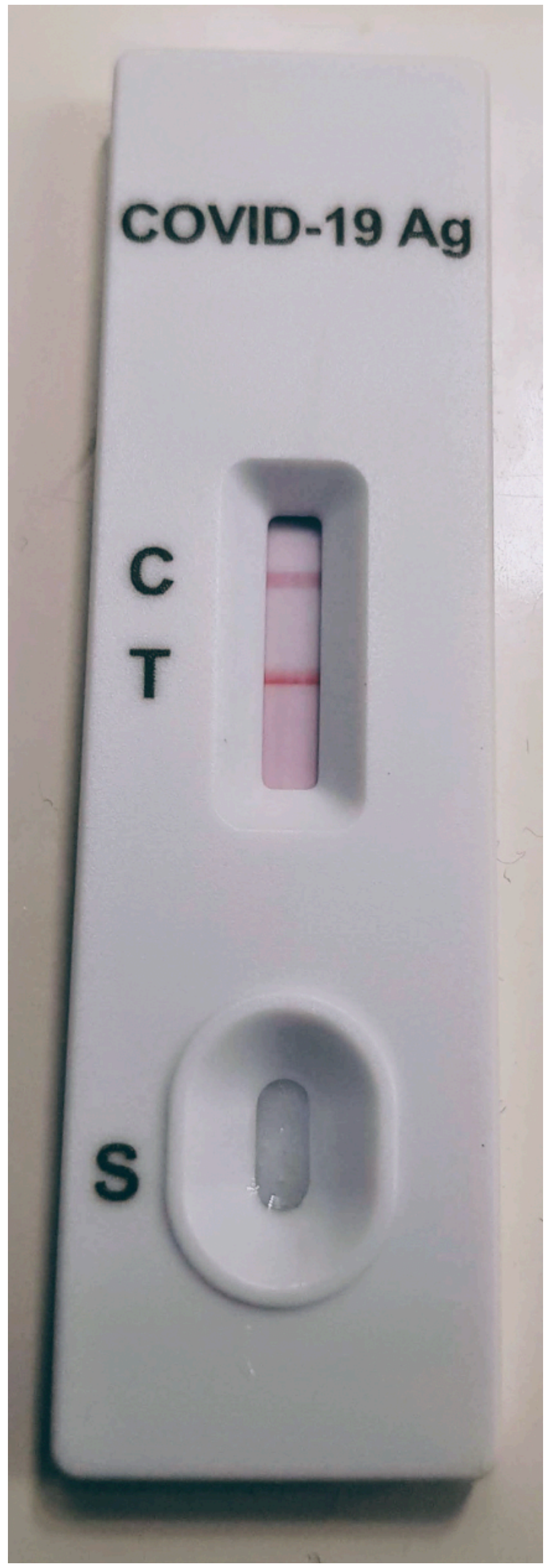
$$p(S \mid +) = \frac{P(+ \mid S)P(S)}{P(+ \mid S)P(S) + P(+ \mid \bar{S})P(\bar{S})}$$

Priors

Probability of True positive

Probability of False positive

Another example: 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

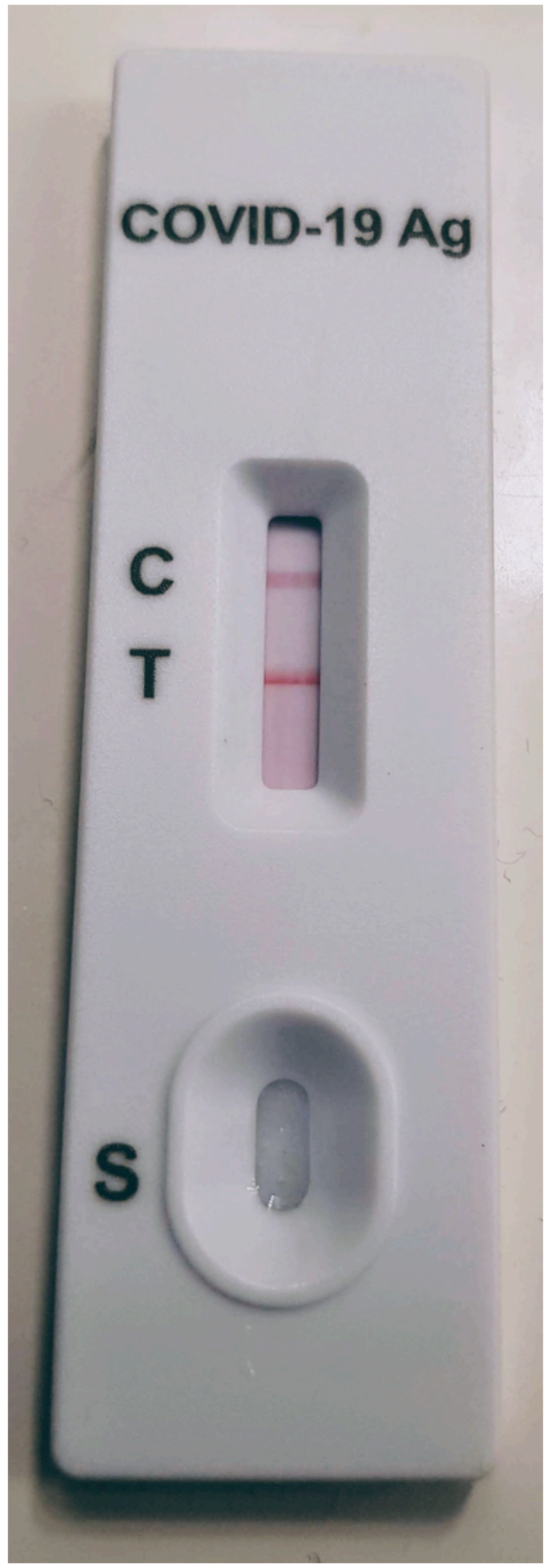
Probability of True positive

Probability of False positive

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

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

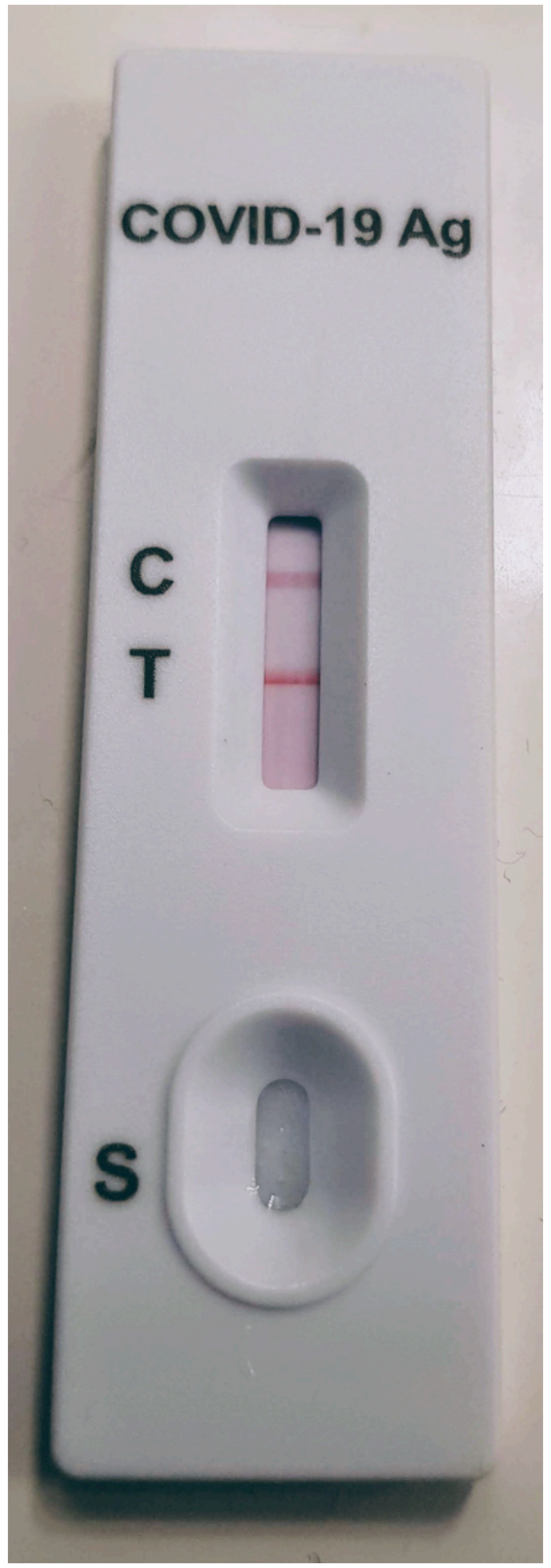
Another example: Covid-19 test



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

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

Another example: Covid-19 test



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

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

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

April 29, 2022

Comparison of Home Antigen Testing With RT-PCR and Viral Culture During the Course of SARS-CoV-2 Infection

Victoria T. Chu, MD, MPH[1,2]; Noah G. Schwartz, MD[1,2]; Marisa A. P. Donnelly, PhD[1,2]; et al

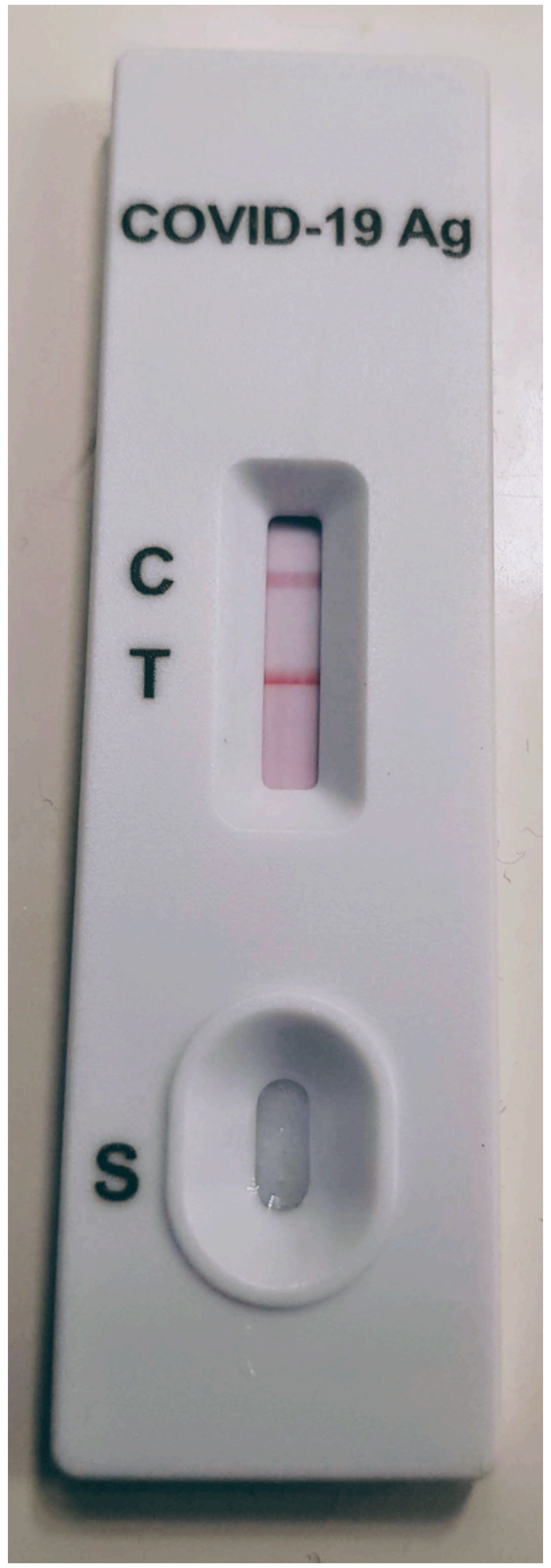
Author Affiliations | Article Information

JAMA Intern Med. 2022;182(7):701-709. doi:10.1001/jamainternmed.2022.1827

COVID-19 Resource Center

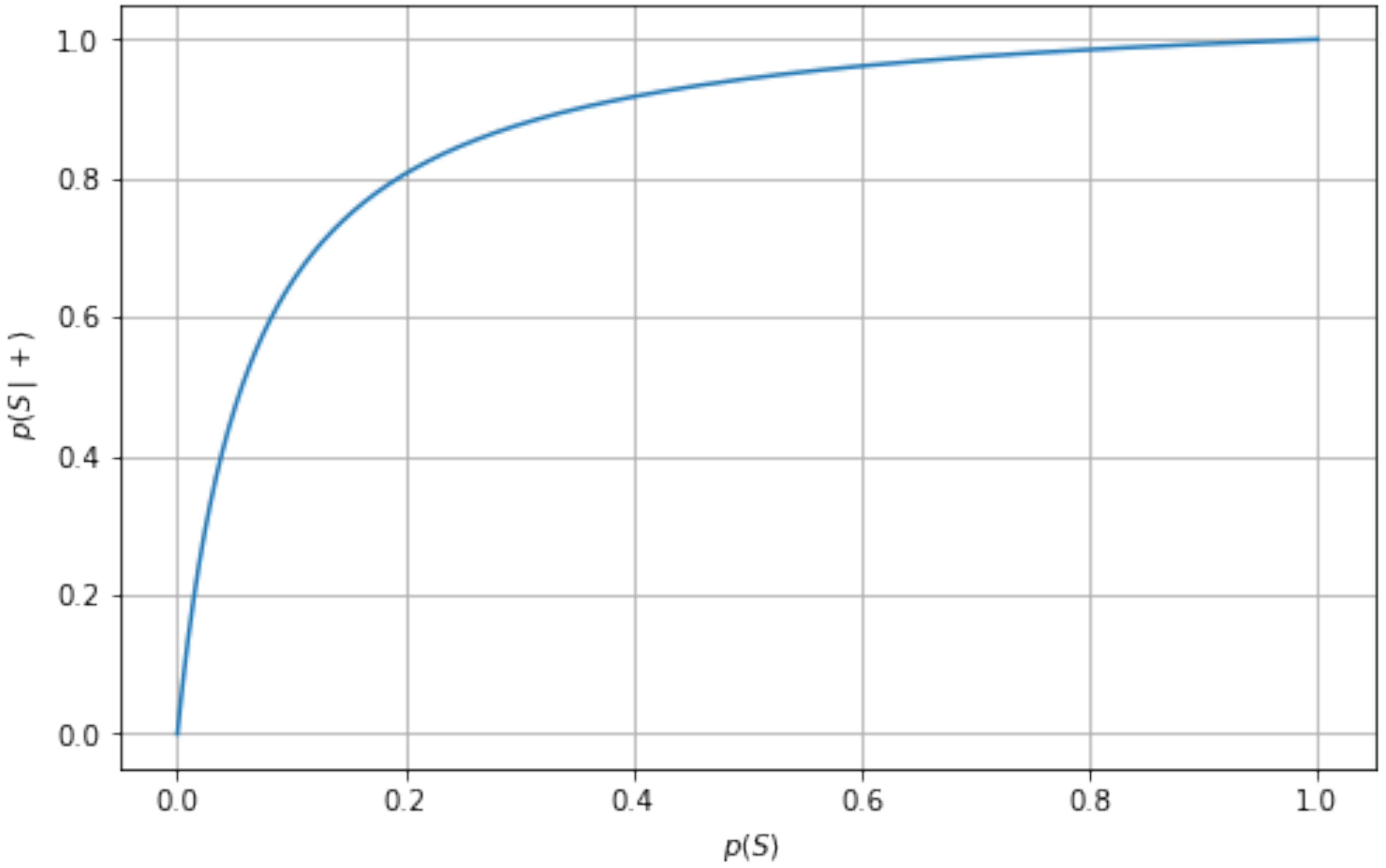
Overall sensitivity of home antigen tests for detecting cases was 50% (95% CI, 45%-55%) (Figure 3), whereas specificity was 97% (95% CI, 95%-98%). Sensitivity was higher for symptomatic cases (53%; 95% CI, 48%-57%) compared with asymptomatic cases (20%; 95% CI,

Another example: Covid-19 test



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

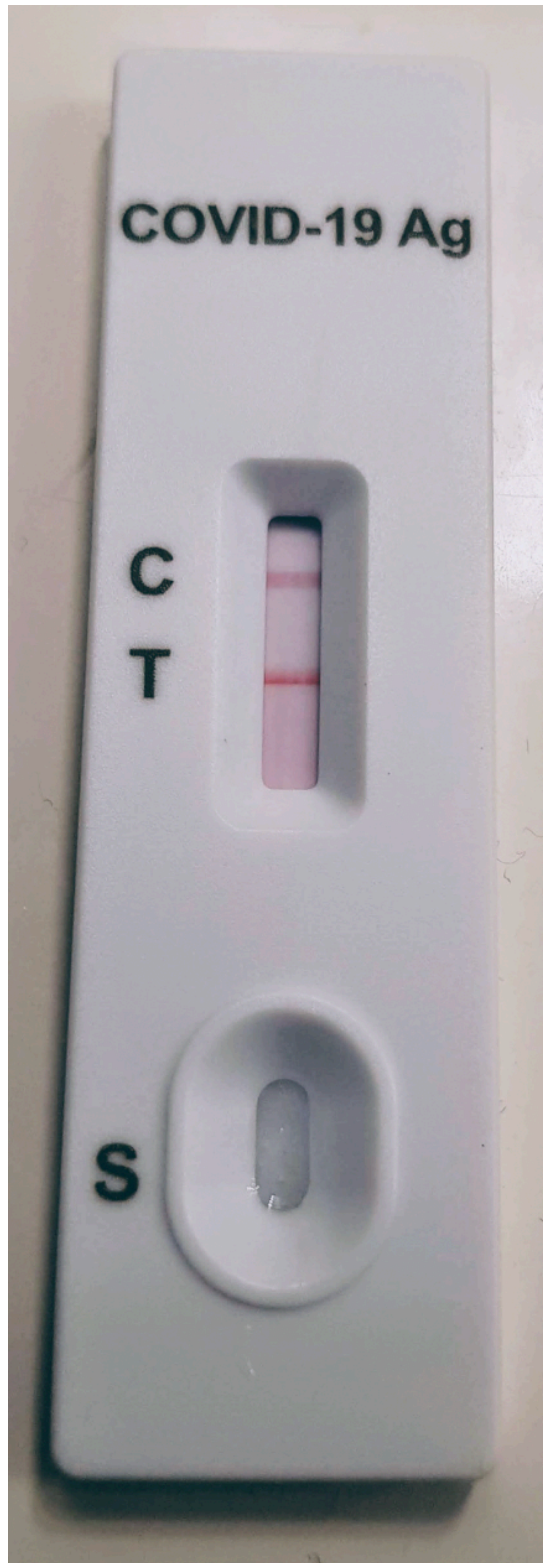
$$p(S\,|\,+) = \left(1 + \frac{1 - \text{Sp.}}{\text{Se.}} \cdot \frac{p(\bar{S})}{p(S)}\right)^{-1}$$



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

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

Another example: Covid-19 test

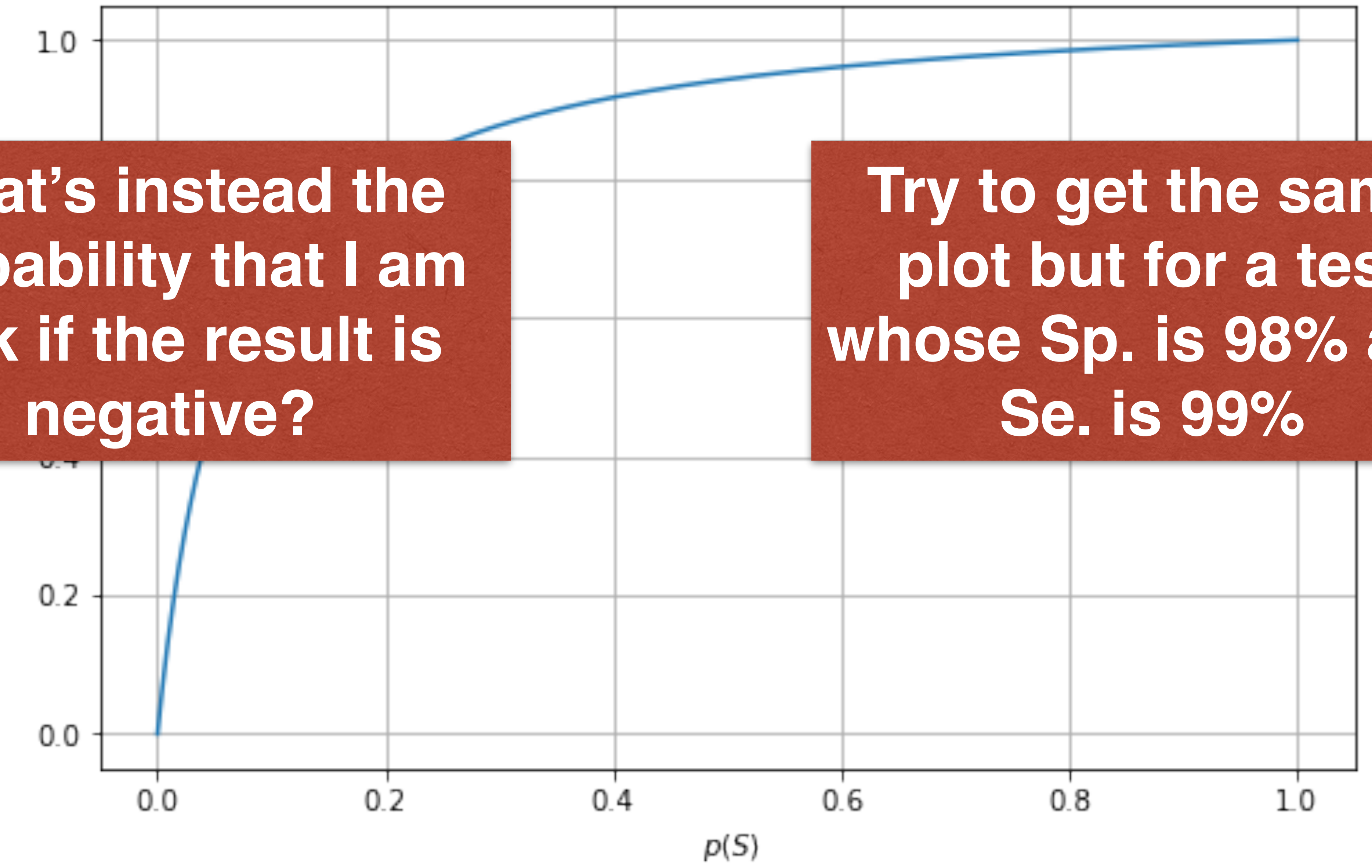


What's the probability that I am sick (S) ?

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

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

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



What's instead the probability that I am sick if the result is negative?

Try to get the same plot but for a test whose Sp. is 98% and Se. is 99%

Let’s go back to the “sheep” example

The Model → *1% of the sheep are black* **= M**

Out of 1 thousand sheep 20 are black

The data →  **= D**

The opinion → $p(M|D)$

Let's go back to the "sheep" example

The Model → *1% of the sheep are black* **= M**

The data → *Out of 1 thousand sheep 20 are black*

= D

The opinion →

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

Let's go back to the "sheep" example

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

Our prior knowledge:

- How much do you believe in your model before the observation?
- Are there other models/hypotheses that might explain the observation? How likely are they?

Let's go back to the "sheep" example

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

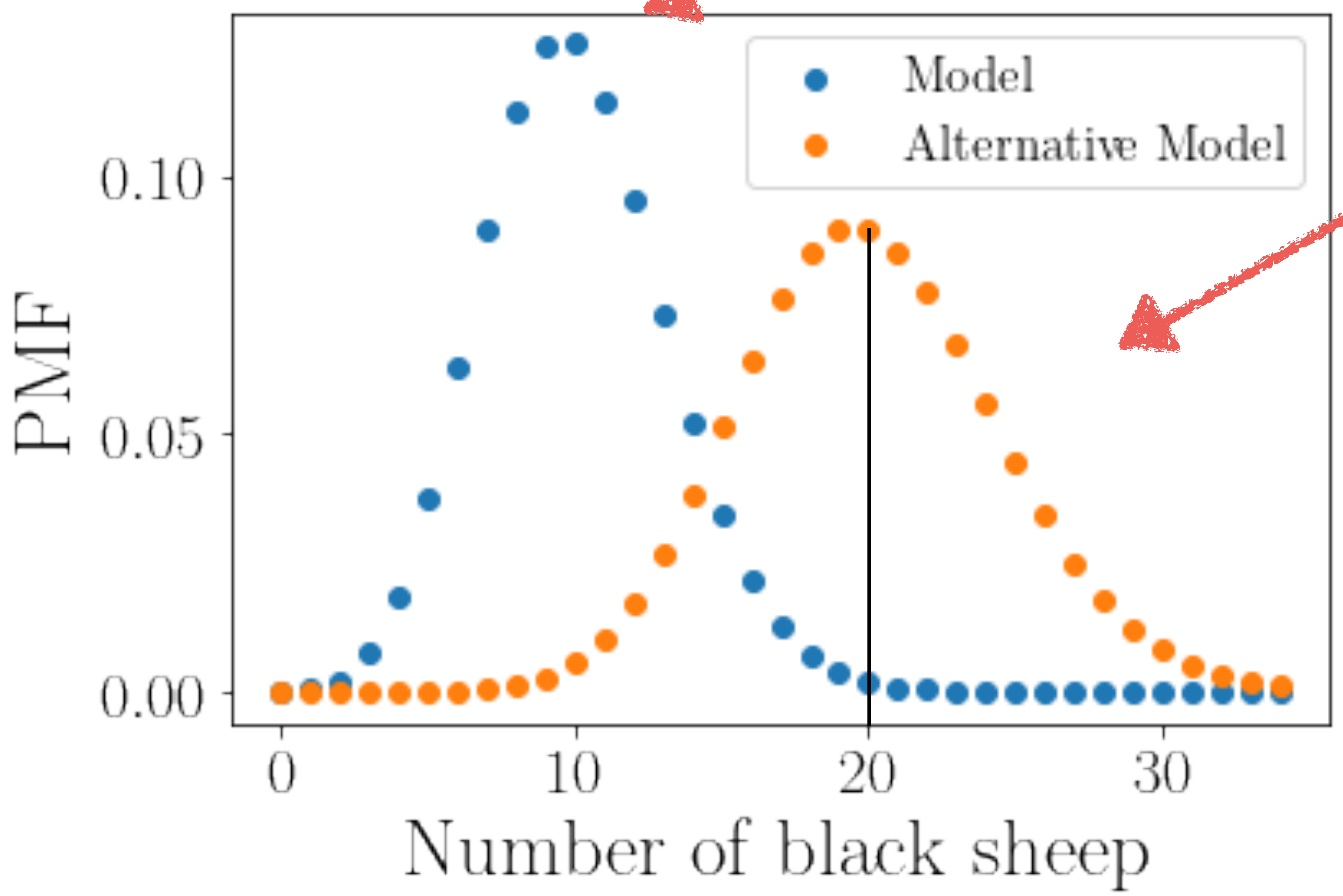
Our prior knowledge:

- We will assume for simplicity that there is only one alternative model "2% of the sheep are black"
- Both models are equally probable

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

Let's go back to the "sheep" example

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



Try to reproduce this plot and change the values used in the the model

Let's go back to the "sheep" example

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

1% of the sheep are black

$$p(D|M) = \mathcal{B}(20 \mid p = 0.01,\ N = 10^3)$$

$$\simeq 0.0018$$

2% of the sheep are black

$$p(D|\bar{M}) = \mathcal{B}(20 \mid p = 0.02,\ N = 10^3)$$

$$\simeq 0.090$$

Let’s go back to the “sheep” example

$$p(M|D) = \frac{p(D|M)p(M)}{p(D|M)p(M) + p(D|\bar{M})p(\bar{M})} \simeq 2\%$$

$$p(\bar{M}|D) = 1 - p(M|D) \simeq 98\%$$

The alternative model is much more likely of being true and the Bayesian approach let us quantify this “likeliness”

Let's go back to the "sheep" example

The Model → *1% of the sheep are black* **= M**

The data → *Out of 1 thousand sheep 20 are black*

= D

The opinion → $p(M|D) \simeq 2\%$

… but what if we do not know the priors of the models?

$$\frac{p(M|D)}{p(\bar{M}|D)} = \frac{p(D|M)}{p(D|\bar{M})} \times \frac{p(M)}{p(\bar{M})}$$

… but what if we do not know the priors of the models?

$$\frac{p(M|D)}{p(\bar{M}|D)} = \frac{p(D|M)}{p(D|\bar{M})} \times \frac{p(M)}{p(\bar{M})}$$

Bayes Factor

Bayes factor BF_{12}			Interpretation
	>	100	Extreme evidence for M_1
30	-	100	Very Strong evidence for M_1
10	-	30	Strong evidence for M_1
3	-	10	Moderate evidence for M_1
1	-	3	Anecdotal evidence for M_1
	1		No evidence
1/3	-	1	Anecdotal evidence for M_2
1/10	-	1/3	Moderate evidence for M_2
1/30	-	1/10	Strong evidence for M_2
1/100	-	1/30	Very Strong evidence for M_2
	<	1/100	Extreme evidence for M_2

What is the BF in our example?

The Frequentist approach

- The **Frequentist approach** tries to answer the 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?*

- The **Frequentist approach** tries to answer the 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?*

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

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

Let's go back to the "sheep" example

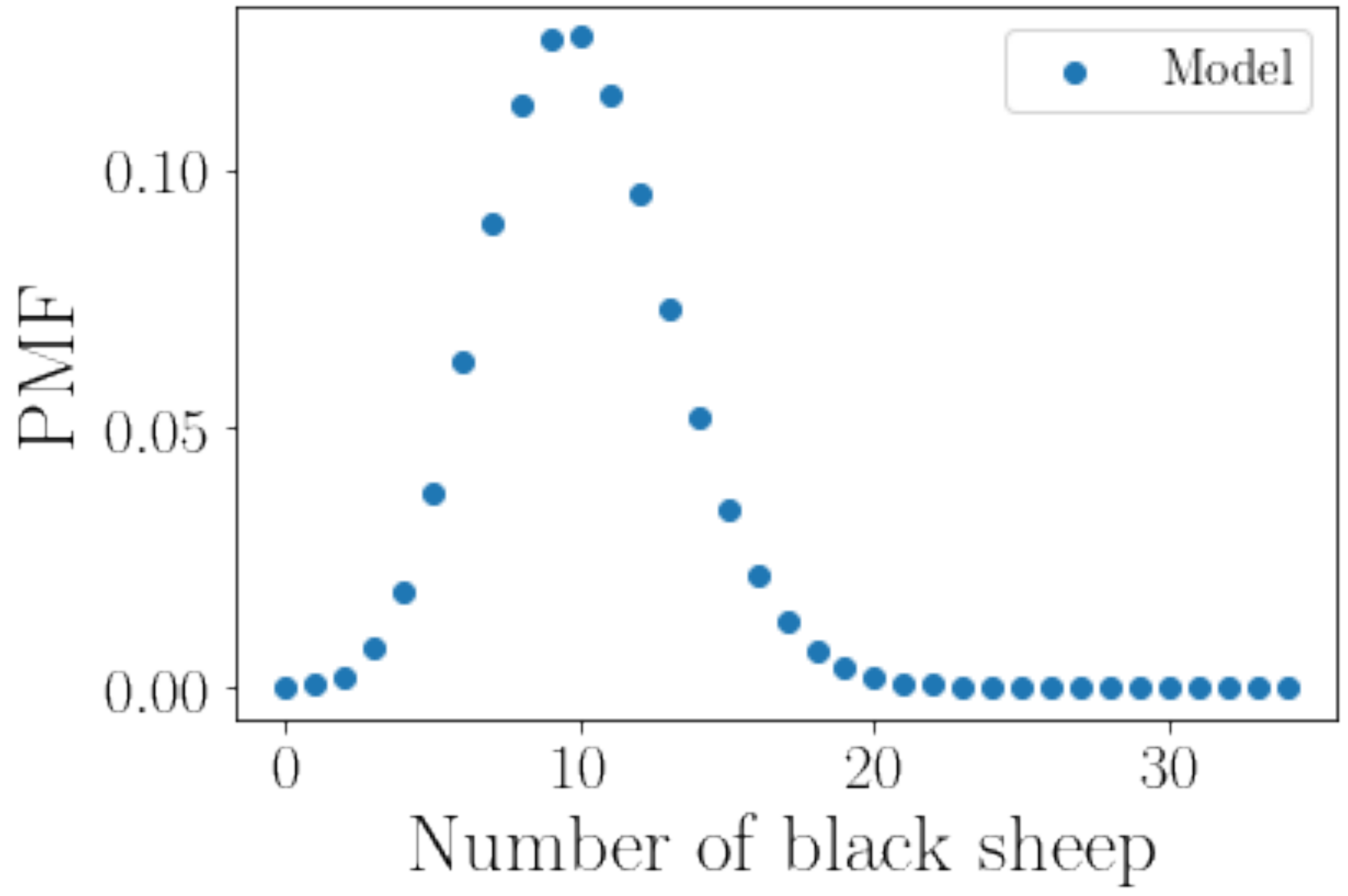
We can use the number of sheep observed as statistics and ask ourselves:

If I repeat the observation an infinity of time, how frequently would I have observed 20 or more sheep?

Let's go back to the "sheep" example

We can use the number of sheep observed as statistics and ask ourselves:

If I repeat the observation an infinity of time, how frequently would I have observed 20 or more sheep?



The answer is only 0.1% of the time!

Therefore the frequentist conclusion is that our model is excluded with a 99.9% confidence level.

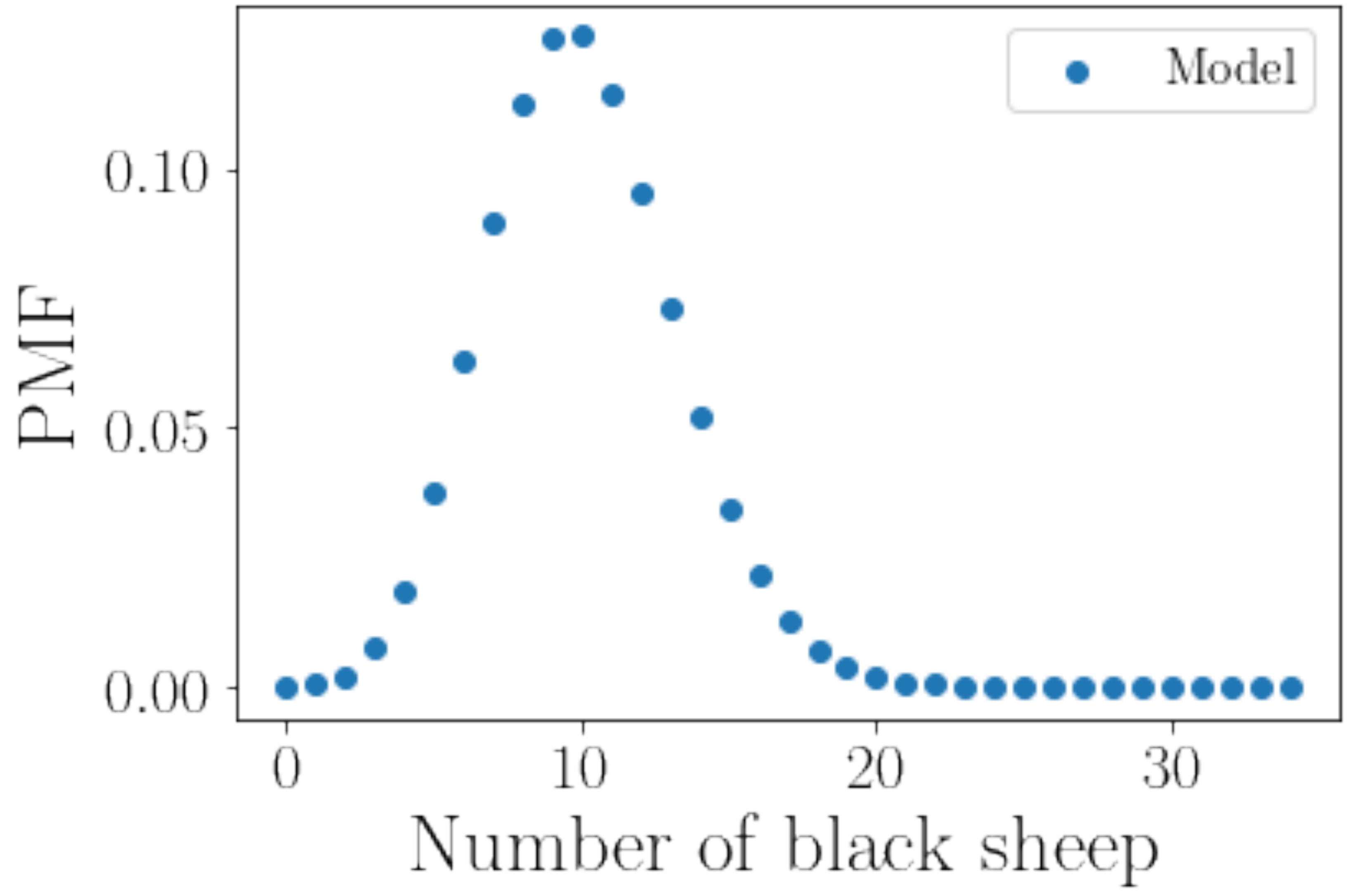
Let’s go back to the “sheep” example

We can use the number of sheep observed as statistics and ask ourselves:

If I repeat the observation an infinity of time, how frequently would I have observed 20 or more sheep?

Try to get this value

P-VALUE



The answer is only 0.1% of the time!

Therefore the frequentist conclusion is that our model is excluded with a 99.9% confidence level.

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 **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 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 ~ 3 orders of magnitude. The flux variation is accompanied by the clearly

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-Pla*

[3]*School of Astronom*

[4]*Guangxi Key Laboratory for Relat*

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

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 $\mathrm{TS_{max}} \simeq 51$, corresponding to a detection significance over 6σ. The result of the data analysis is shown in §2

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"?

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

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

Here the (in-)famous number of "sigma"

Let's go back to the "sheep" example

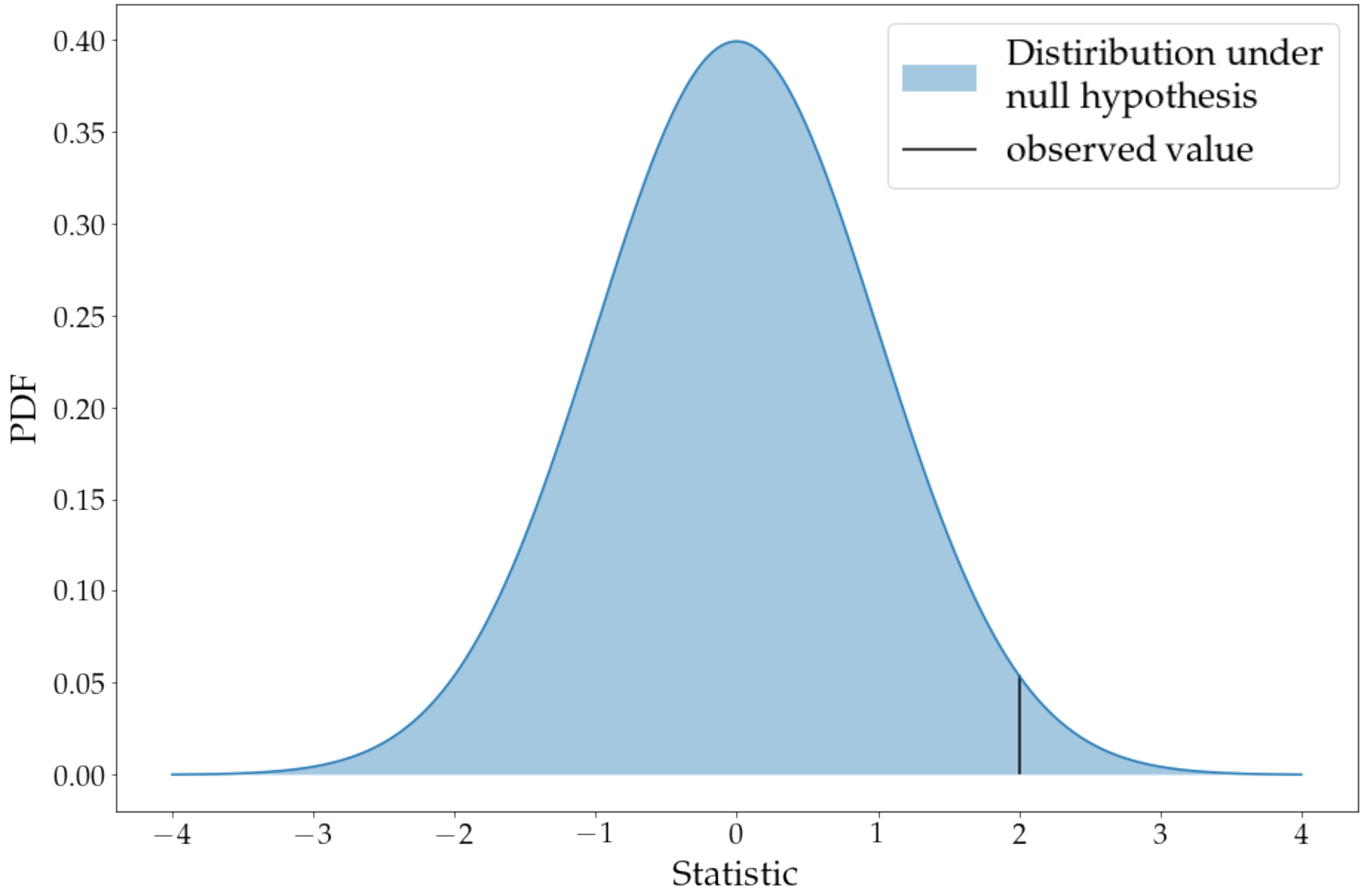
The Model → *1% of the sheep are black*

The data → *Out of 1 thousand sheep 20 are black*

The opinion → *The model is excluded at 3.2 sigma*

Issues of the frequentist approach:

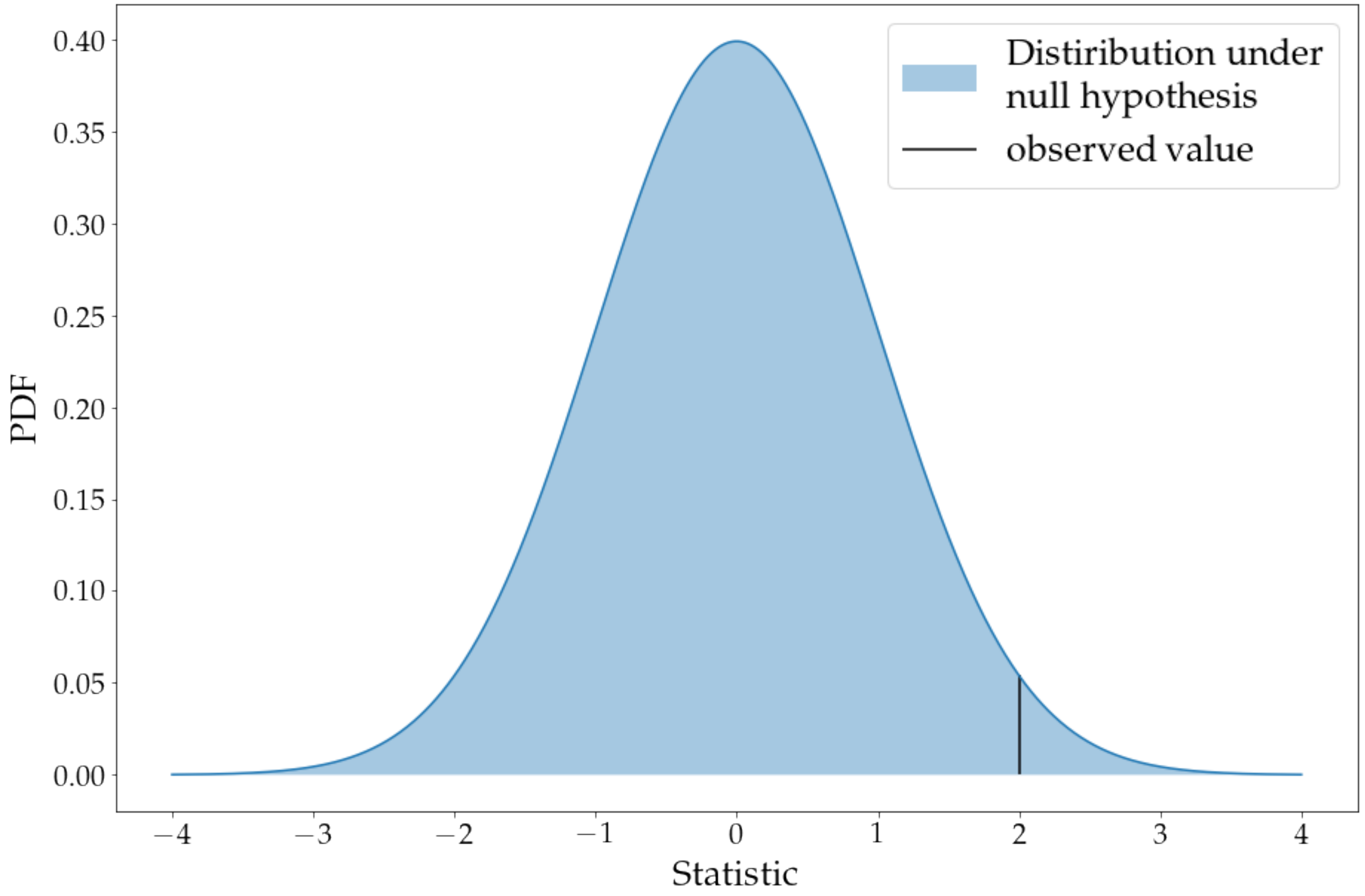
It does not take into account the **alternative hypothesis** that might explain the outcome of an event



Conclusion:
The null hypothesis is rejected with a **2 sigma** significance

Issues of the frequentist approach:

It does not take into account the **alternative hypothesis** that might explain the outcome of an event

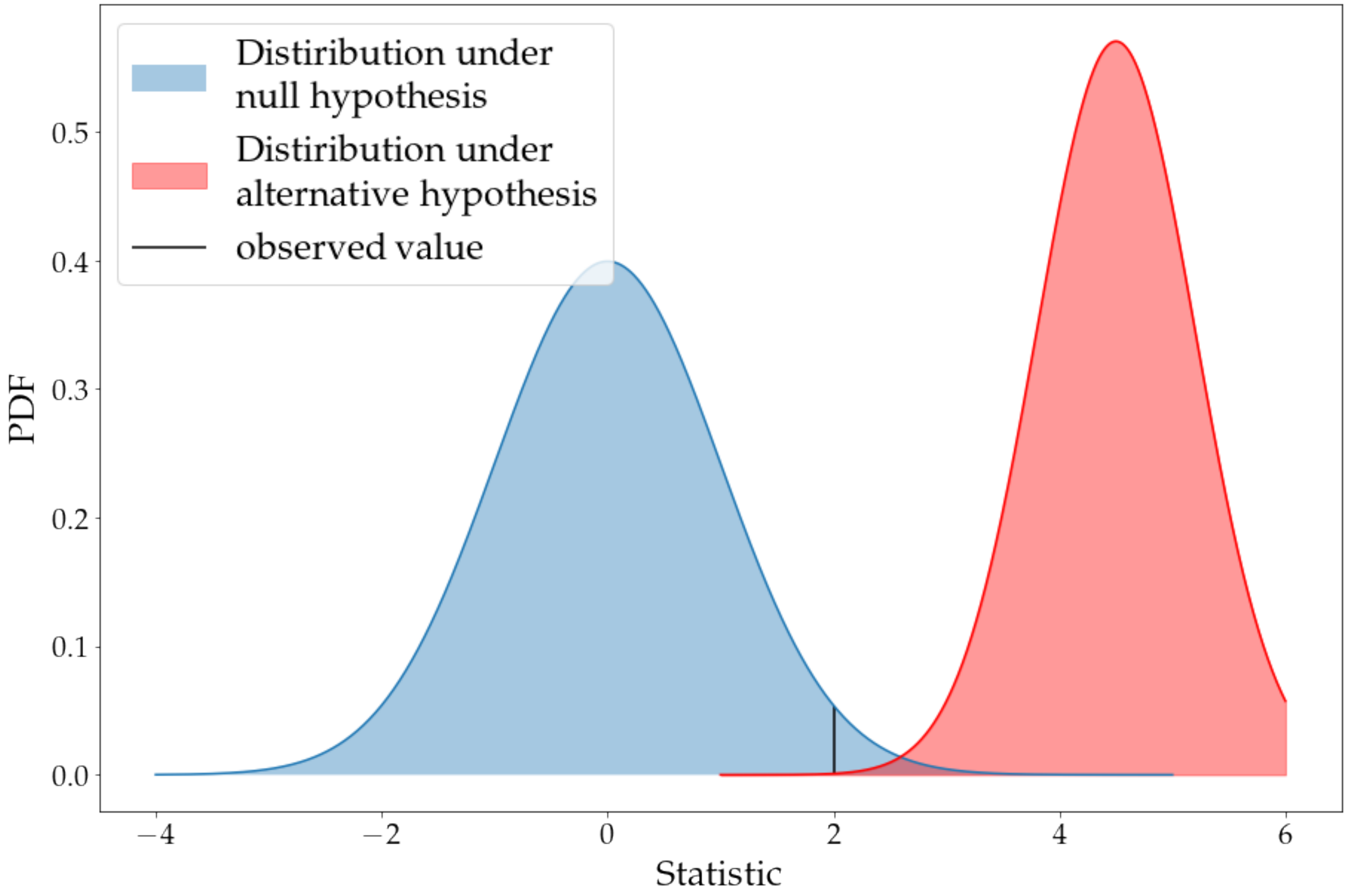


Conclusion:
The null hypothesis is rejected with a **2 sigma** significance

But what about the alternative hypothesis?

Issues of the frequentist approach:

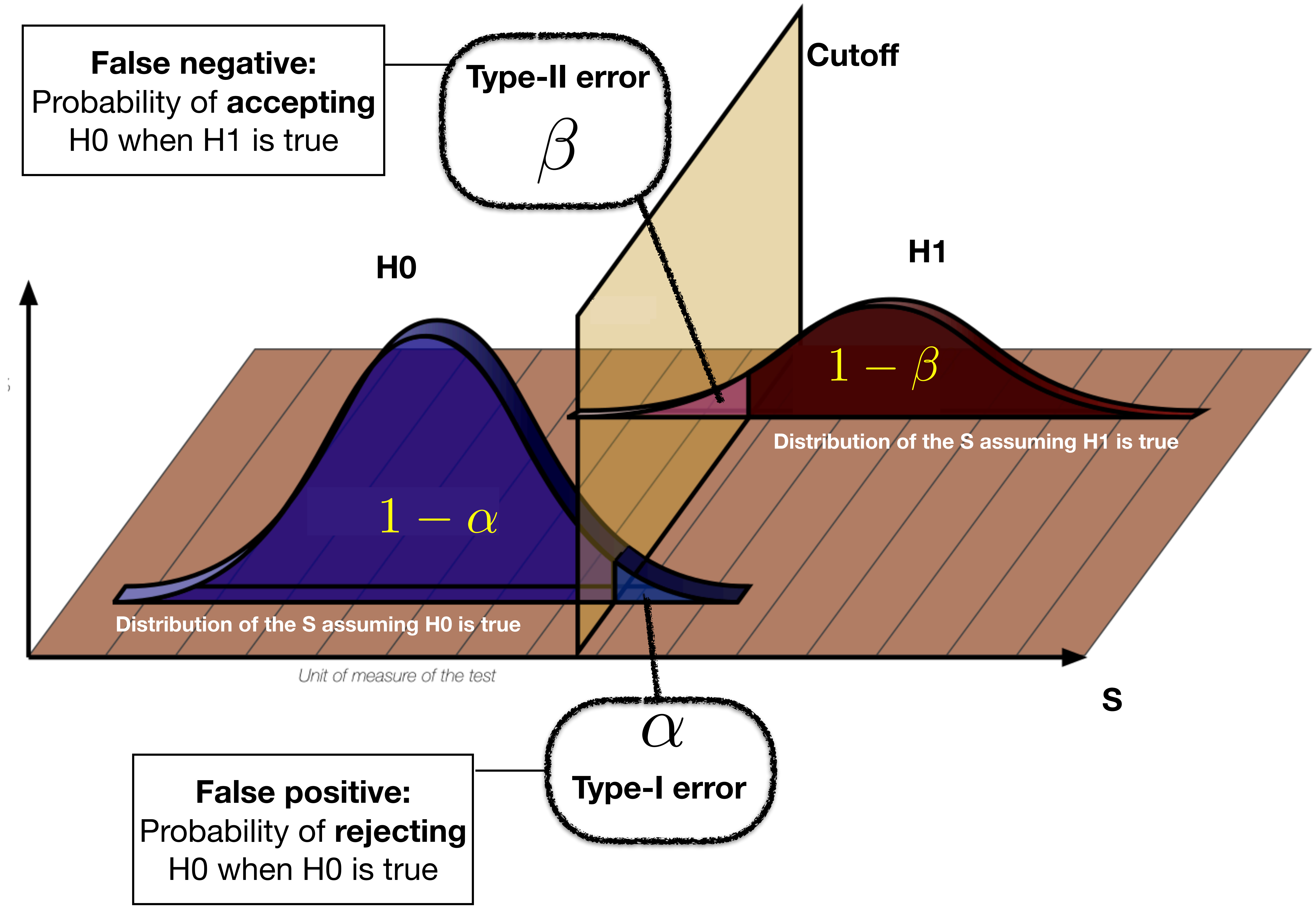
It does not take into account the **alternative hypothesis** that might explain the outcome of an event



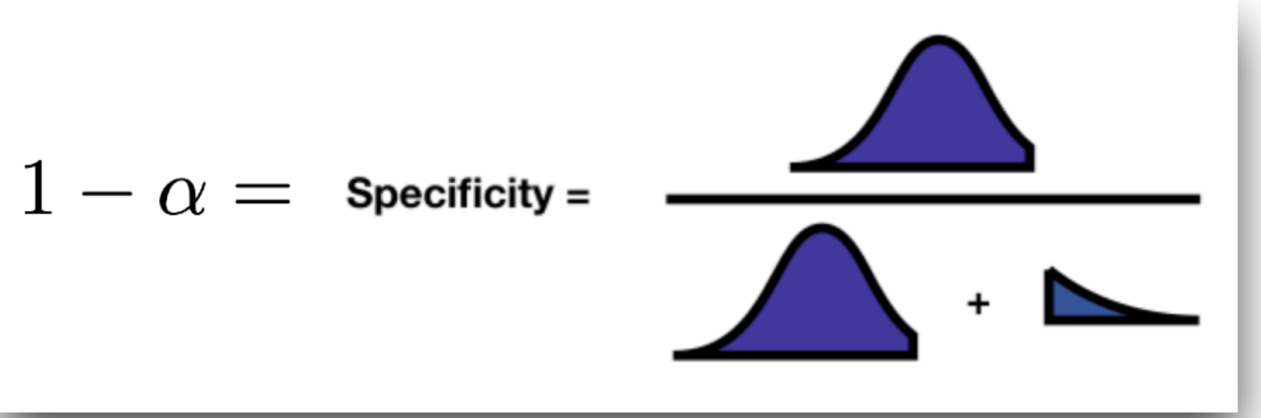
The observed value of 2 is actually more plausible being the outcome of the null hypothesis

By rejecting the null hypothesis we would have done the so-called ***type I error***

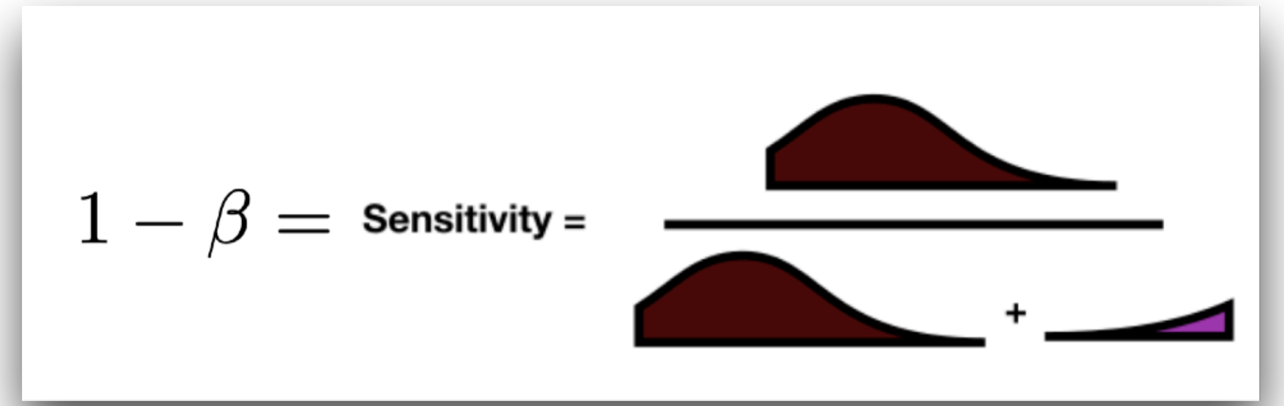
This is why a value of sigma bigger than 3 or even 5 is required for making a claim!



Confidence level

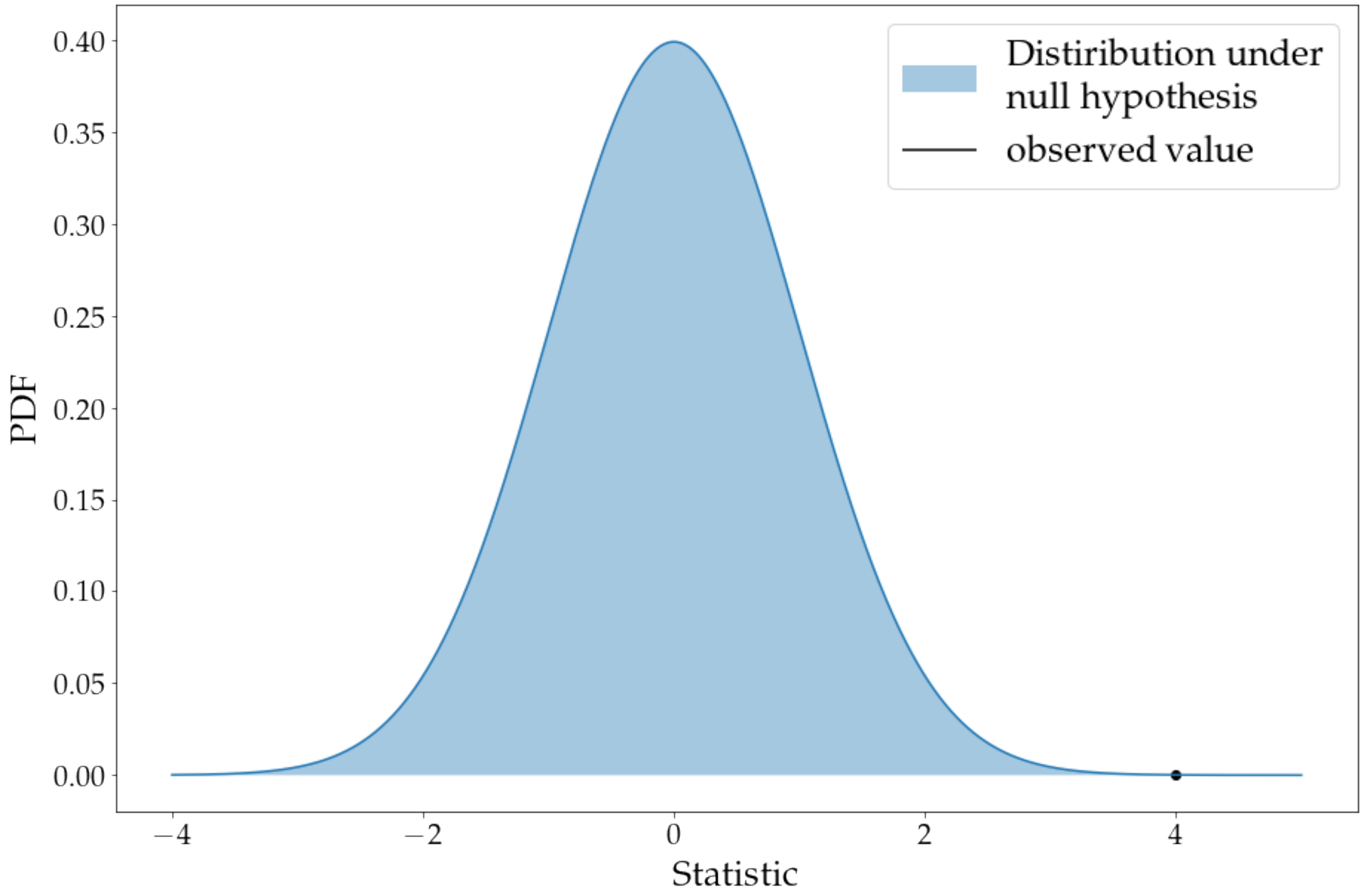


Power of the test



Issues of the frequentist approach:

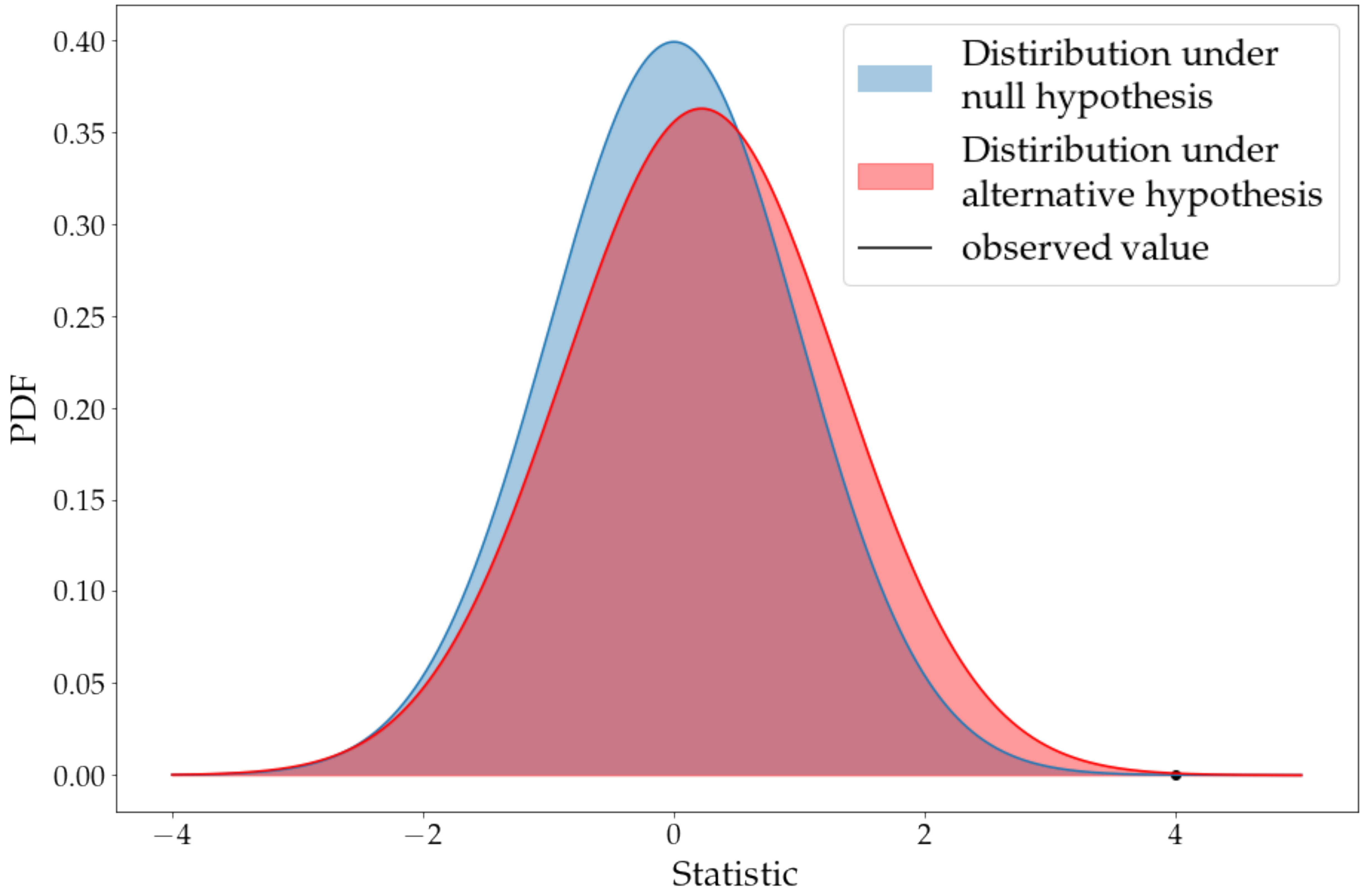
It does not take into account the **alternative hypothesis** that might explain the outcome of an event



So… with a significance of 4 we should be safe?

Issues of the frequentist approach:

It does not take into account the **alternative hypothesis** that might explain the outcome of an event

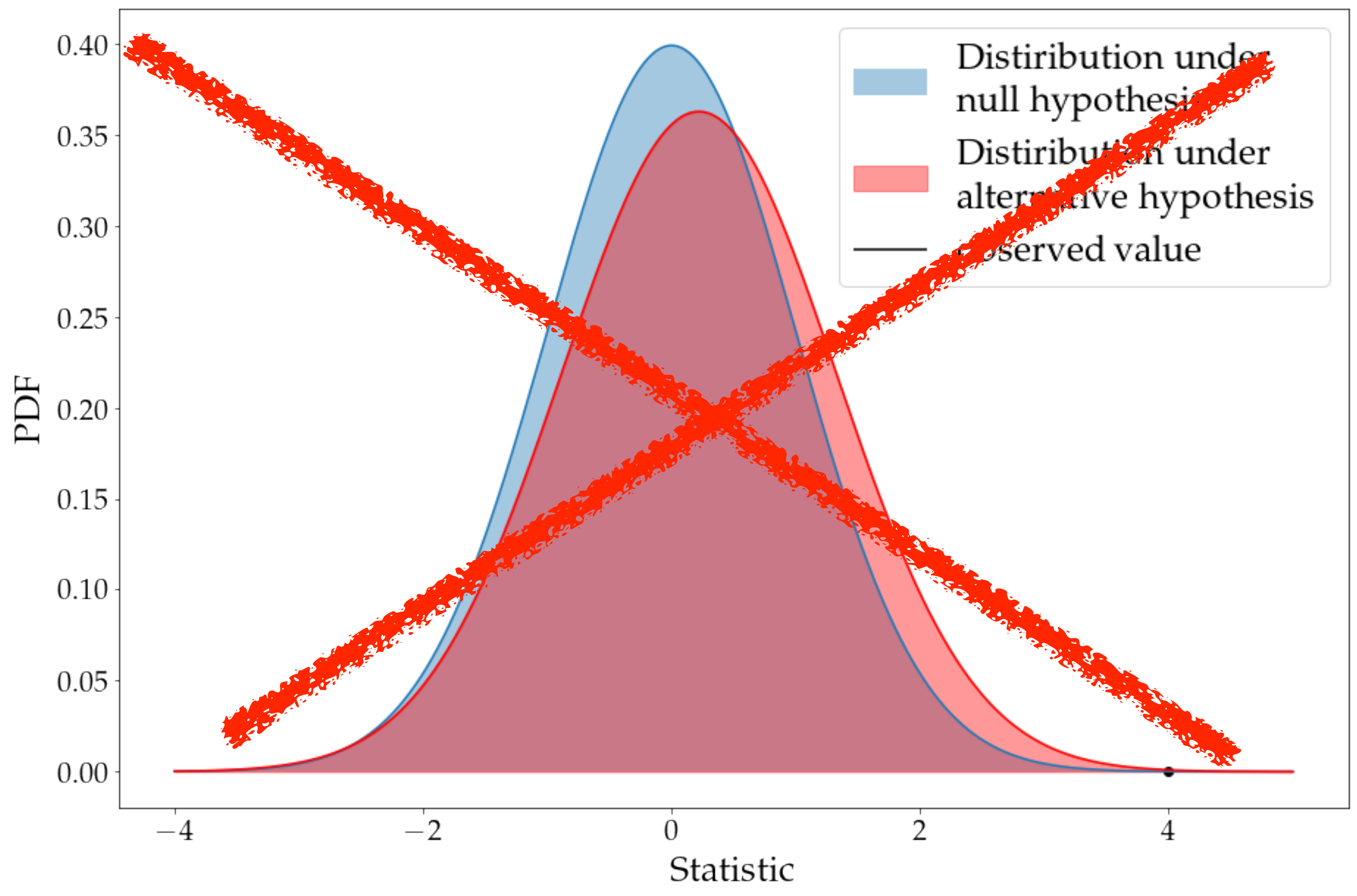


So… with a significance of 4 we should be safe?

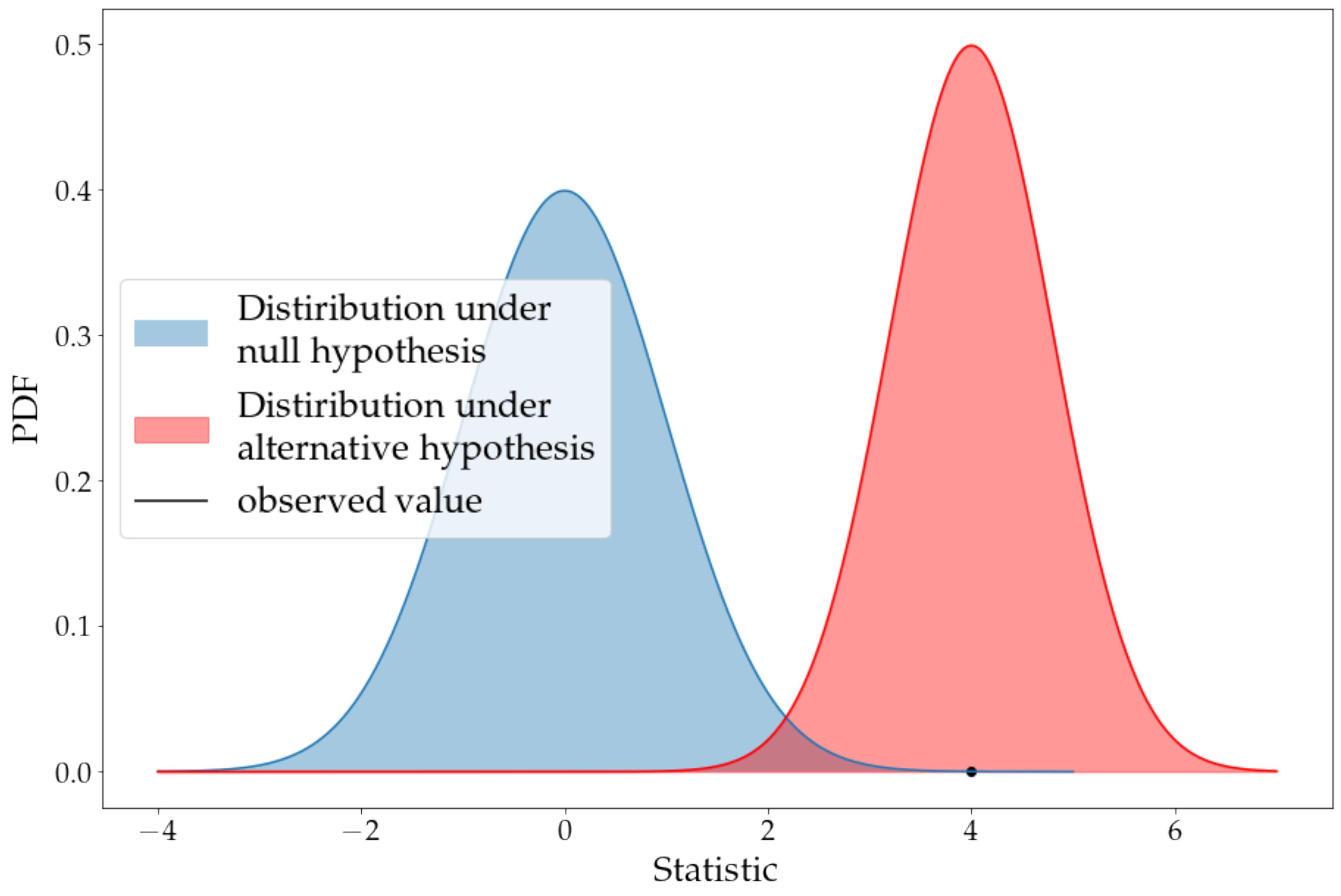
The value of 4 is unlikely to be the outcome also of the alternative hypothesis, thus again we could be doing a ***type I error***

Issues of the frequentist approach:

It does not take into account the **alternative hypothesis** that might explain the outcome of an event



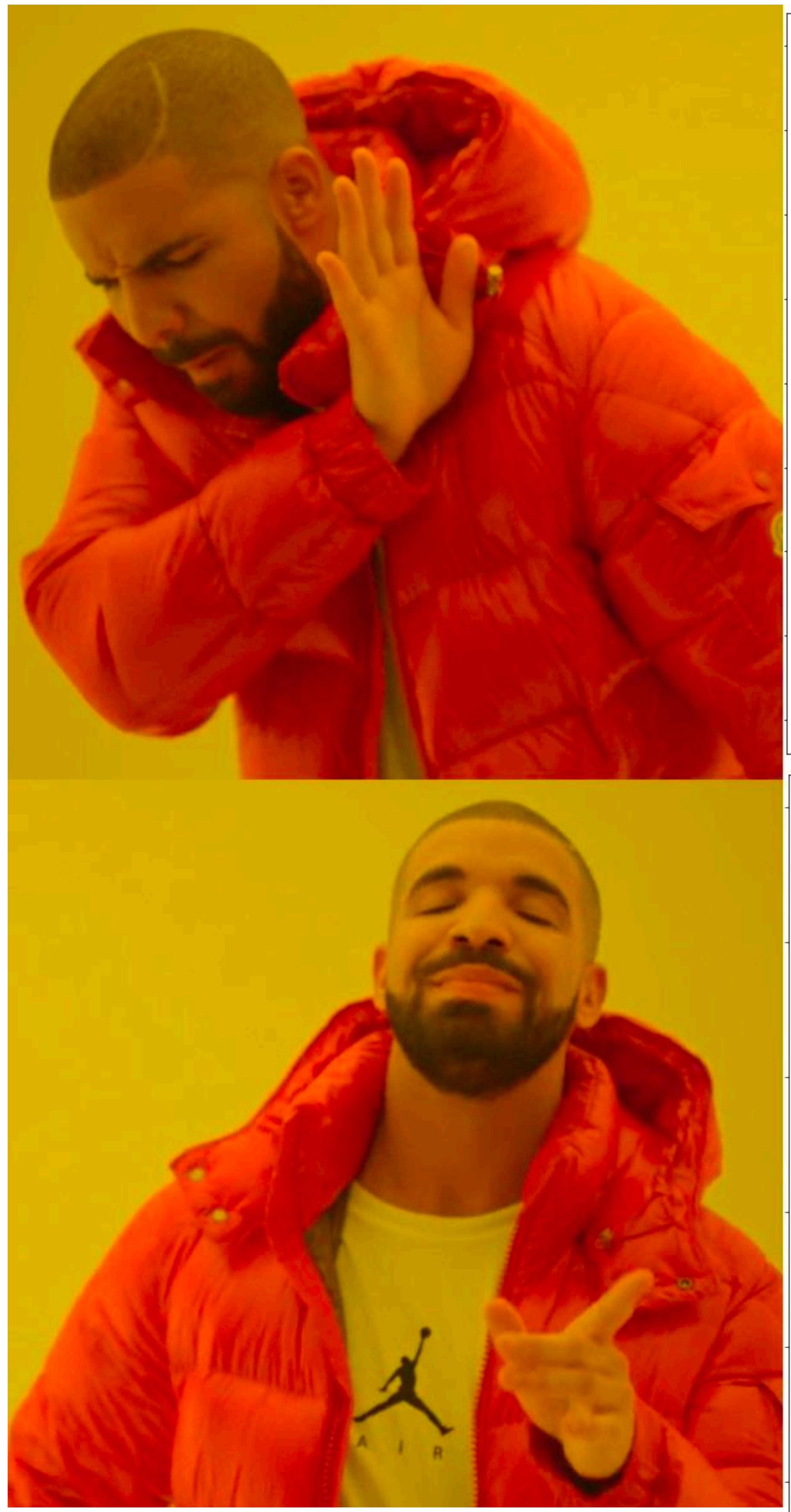
The ideal statistic is the one that makes you **reject** a hypothesis that is false!

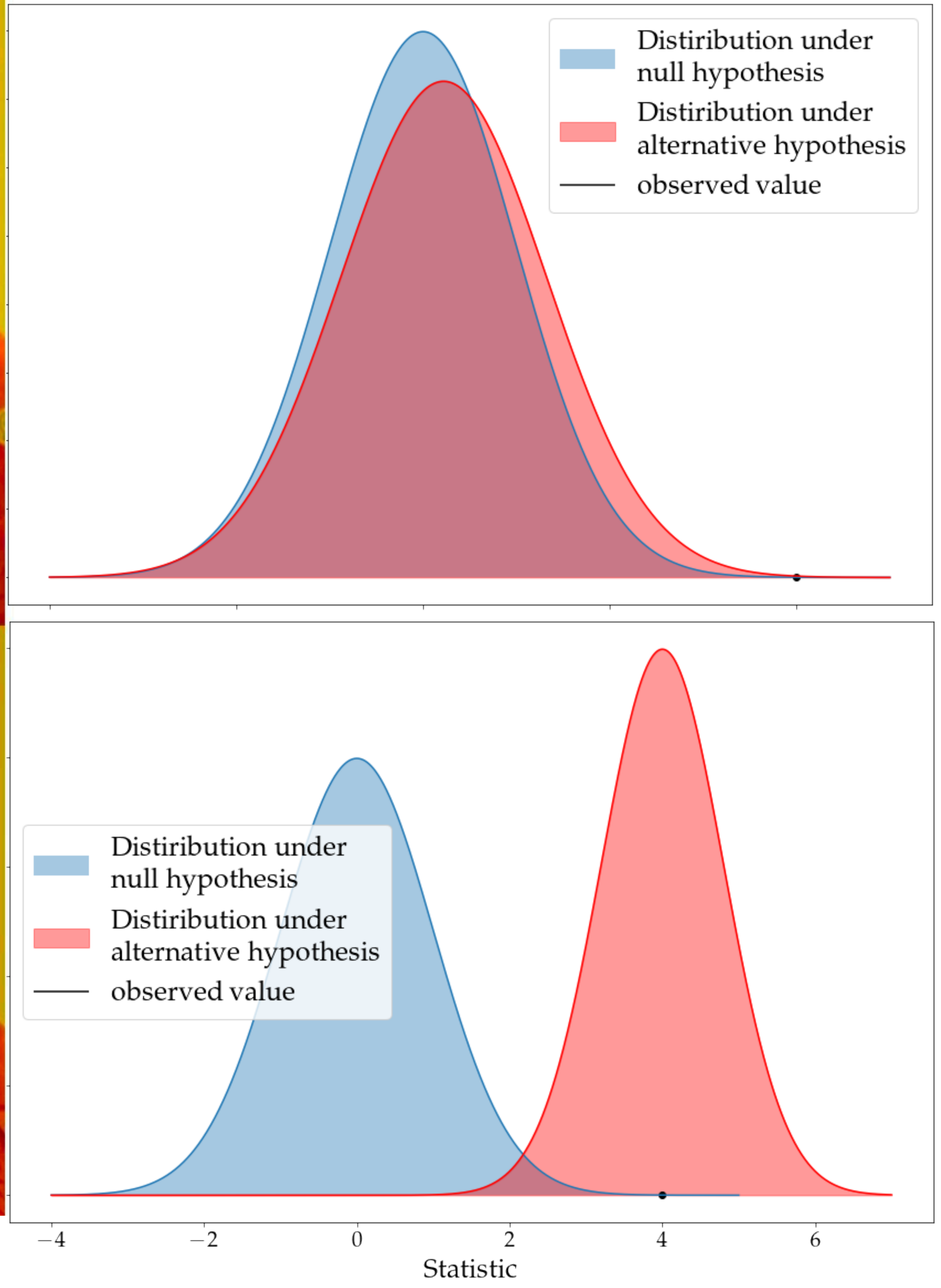


Issues of the frequentist approach:

The probability of rejecting a hypothesis that is false is called the “**power**” of the statistic

Your statistic must be ***POWERFUL***!





Issues of the frequentist approach:

Arbitrariness in the choice of the statistic

https://en.wikipedia.org/wiki/Category:Statistical_tests

Pages in category "Statistical tests"

The following 104 pages are in this category, out of 104 total. This list may not reflect recent changes (learn more).

A
- ABX test
- Analysis of similarities
- Analysis of variance
- Anderson–Darling test

B
- Bartlett's test
- Binomial test
- Breusch–Godfrey test
- Breusch–Pagan test
- Brown–Forsythe test

C
- Chauvenet's criterion
- Checking whether a coin is fair
- Closed testing procedure
- Cochran's C test
- Cochran's Q test
- Continuity correction
- Cramér–von Mises criterion
- Cuzick–Edwards test

D
- Dixon's Q test
- Duncan's new multiple range test
- Dunnett's test
- Durbin test

E
- Exact test

F
- F-test
- F-test of equality of variances
- False positive rate
- Fay and Wu's H
- Fisher's method
- Friedman test

G
- Goodman and Kruskal's gamma
- Glejser test
- Goldfeld–Quandt test
- GRIM test
- Grubbs's test

H
- Hartley's test
- Hoeffding's independence test
- Holm–Bonferroni method
- Hosmer–Lemeshow test

I
- Information matrix test
- Item-total correlation

J
- Jonckheere's trend test

K
- Kaiser–Meyer–Olkin test
- Kendall rank correlation coefficient
- Kolmogorov–Smirnov test
- Kruskal–Wallis one-way analysis of variance
- Kuiper's test

L
- Lepage test
- Levene's test
- Lexis ratio
- Likelihood-ratio test
- Wilks' theorem
- Location test
- Location testing for Gaussian scale mixture distributions
- Logrank test

M
- Mann–Whitney U test
- Mantel test
- Mauchly's sphericity test
- McNemar's test
- Median test
- Multinomial test

N
- Nemenyi test
- Neyman–Pearson lemma
- Normality test

O
- Omnibus test
- One- and two-tailed tests
- One-way analysis of variance

P
- P-rep
- Page's trend test
- Paired data
- Park test
- Permutation test
- Phillips–Perron test

Q
- Q-statistic
- QST (genetics)

R
- Ramsey RESET test
- Randomness test

S
- Sargan–Hansen test
- Scheirer–Ray–Hare test
- Score test
- Separation test
- Sequential probability ratio test
- Shapiro–Francia test
- Shapiro–Wilk test
- Siegel–Tukey test
- Sign test
- Sobel test
- Spearman's rank correlation coefficient
- Squared ranks test
- *Structural break test*
- Student's t-test
- Surrogate data testing

T
- Tajima's D
- Test statistic
- Tukey–Duckworth test
- Tukey's range test
- Tukey's test of additivity

V
- Van der Waerden test
- Vuong's closeness test

W
- Wald test
- Wald–Wolfowitz runs test
- Welch's t-test
- White test
- Wilcoxon signed-rank test
- Durbin–Wu–Hausman test

Z
- Z-test

Thankfully the Neyman-Pearson Lemma tells us that the most “powerful” statistic is the **likelihood ratio**:

Parameter of the null hypothesis

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

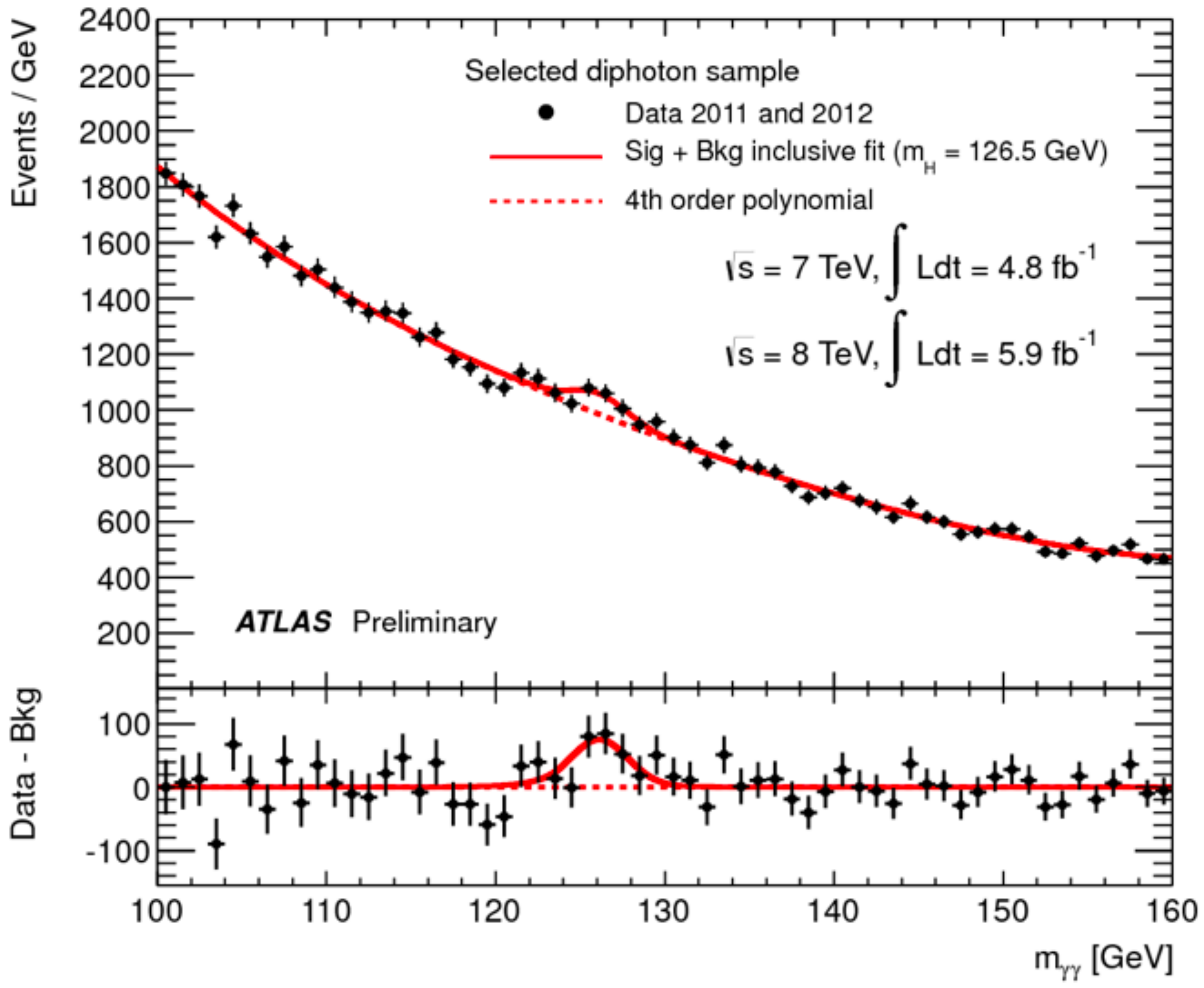
Best fit or value that maximises the likelihood

Observed data

Likelihood

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

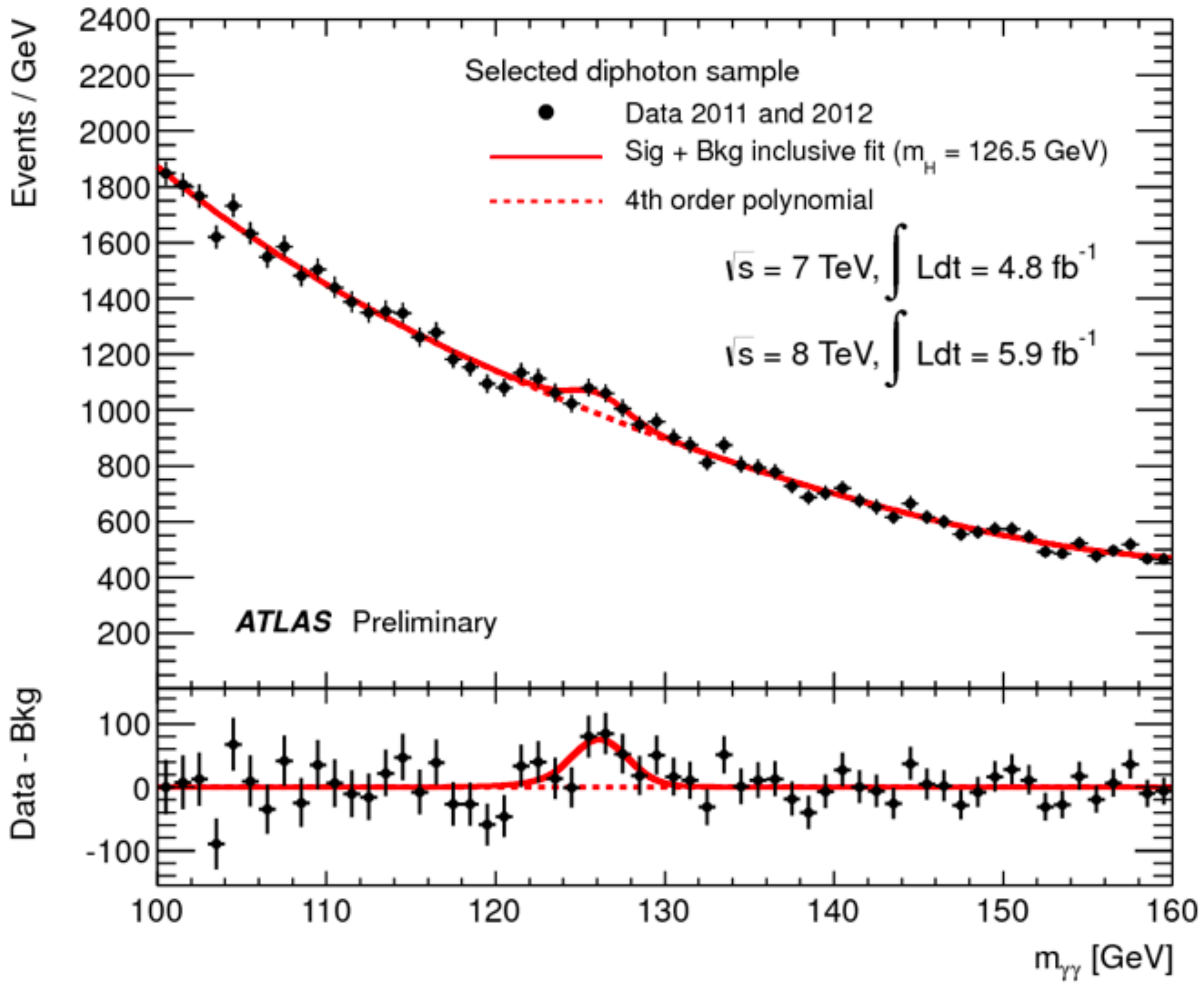
Example:



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

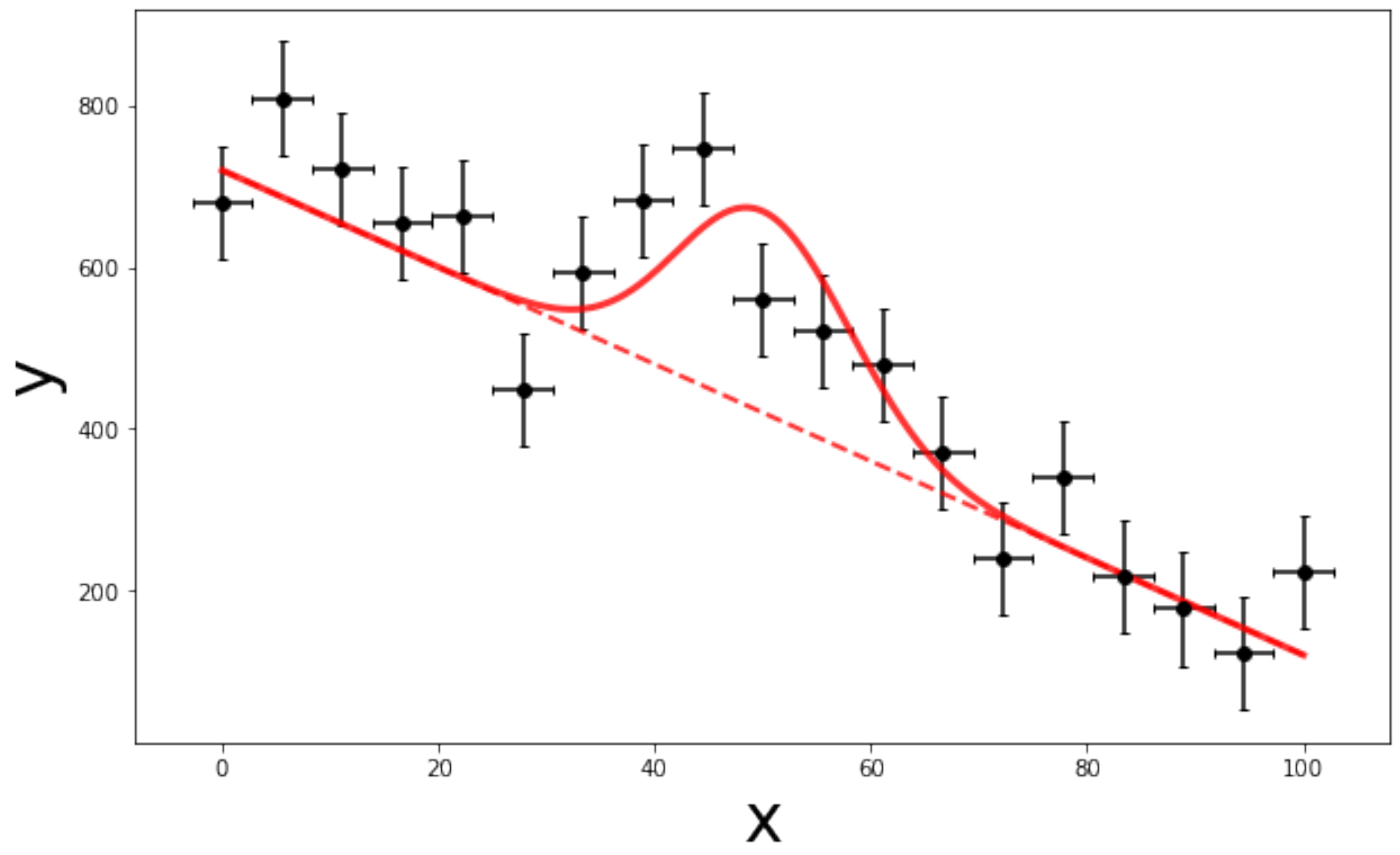
Example:



Toy Model

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

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



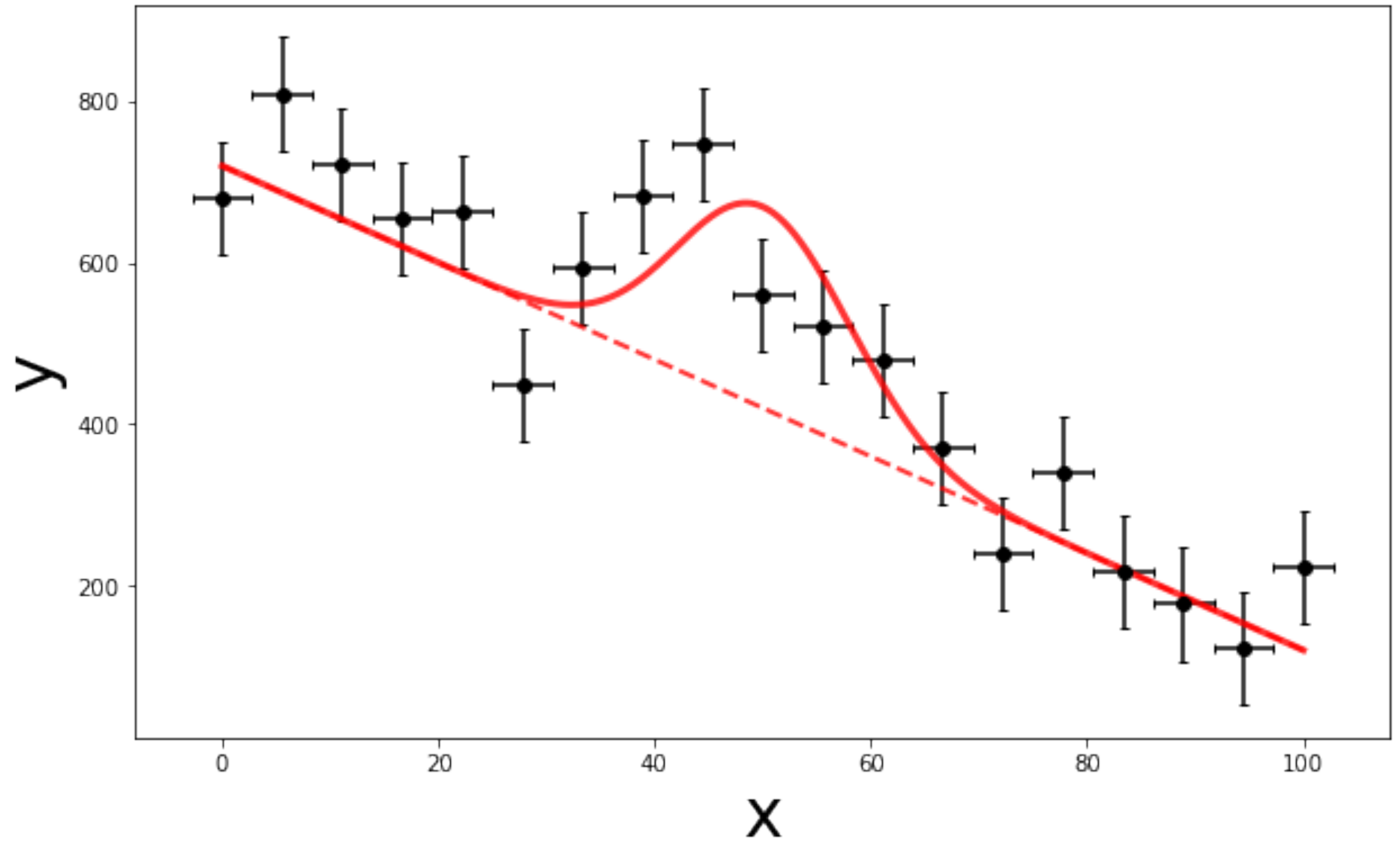
Null hypothesis H0
a = 0

Alternative hyp. H1
a = 5

Example:

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

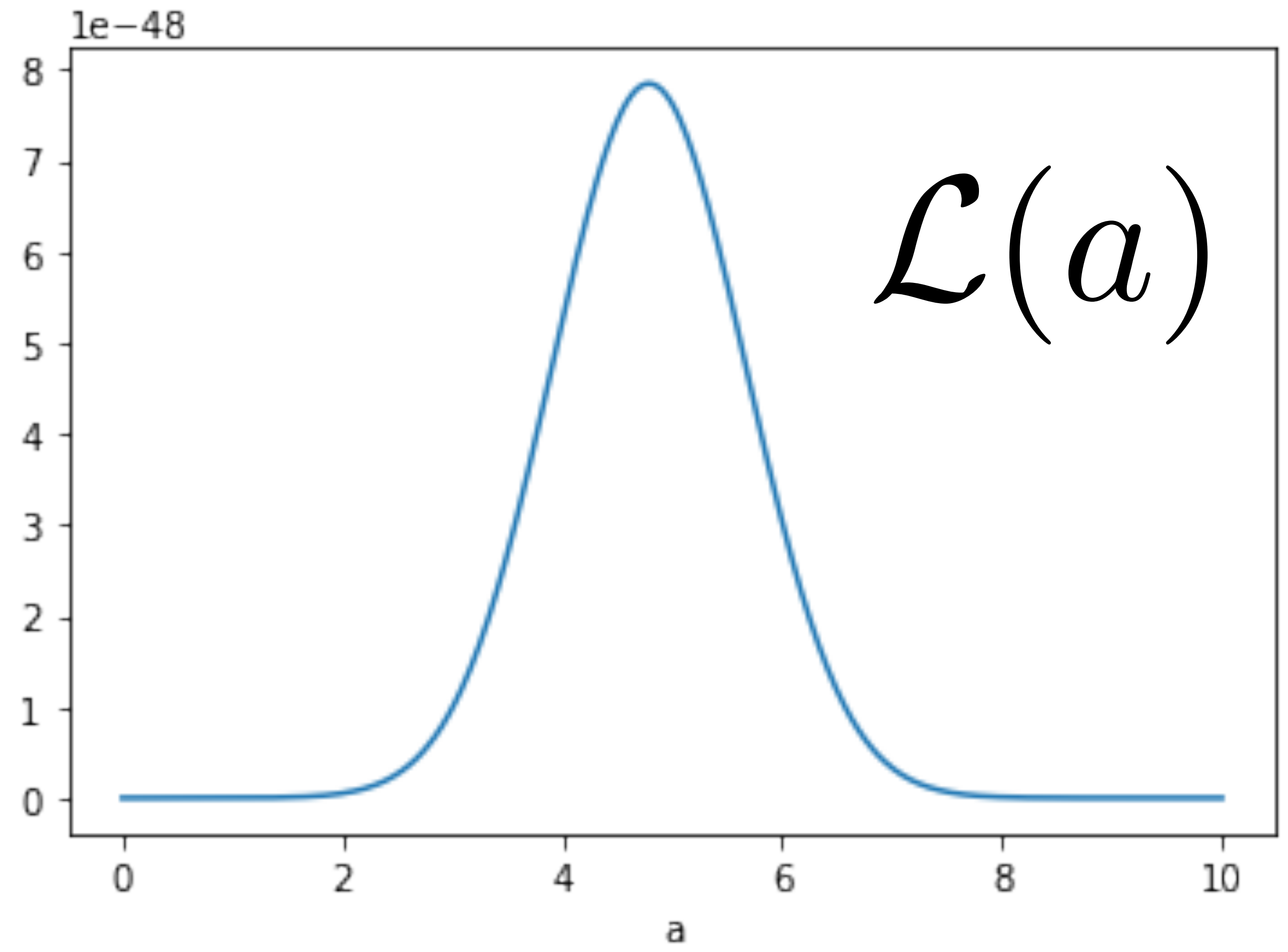
$$y \sim \mathcal{N}(\mu = y', \sigma = 70)$$



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

Example:

Likelihood

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

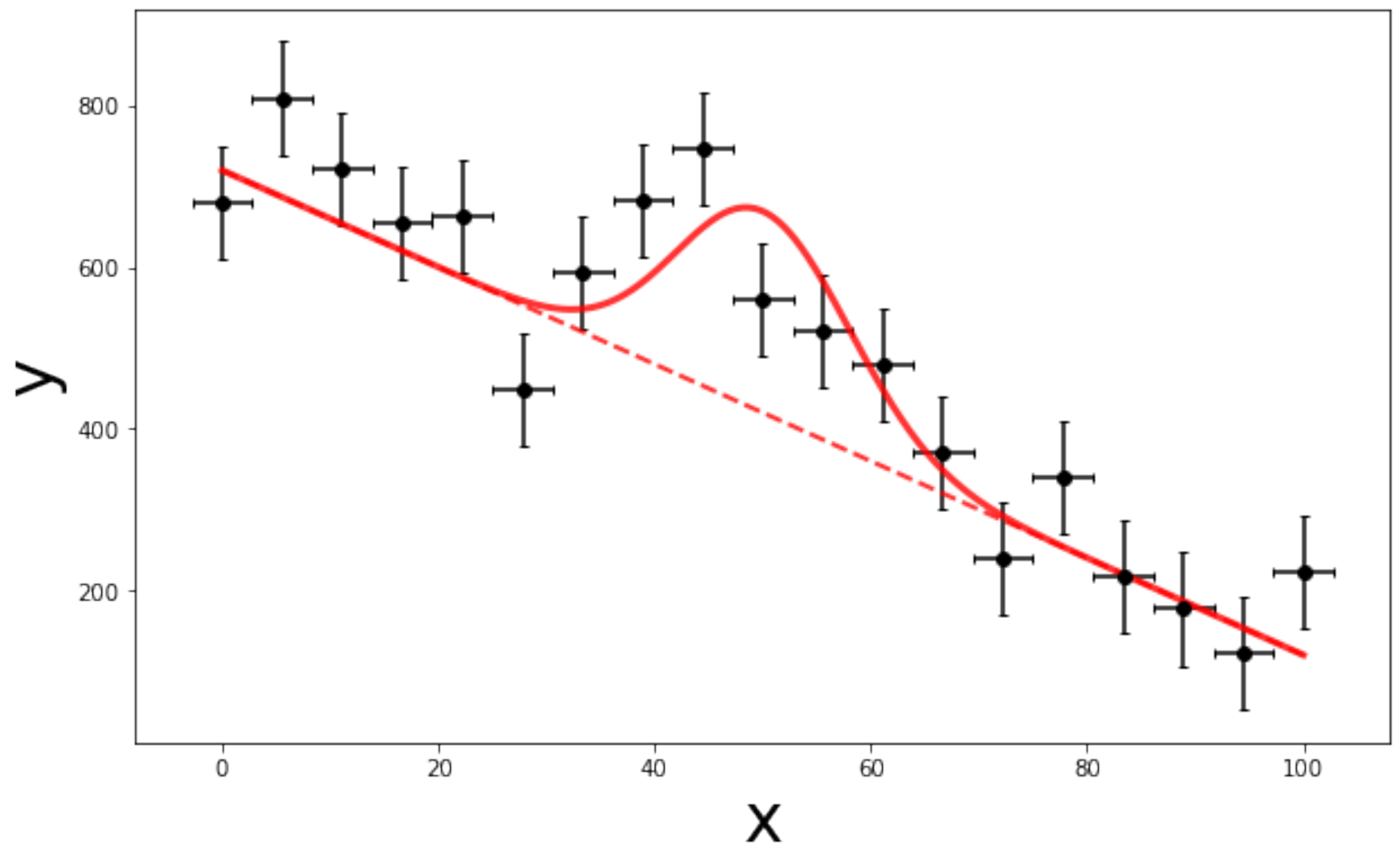
$$p(x_i, y_i|a) \propto e^{-\frac{1}{2}\left(\frac{y_i'(a)-y_i}{\sigma}\right)^2}$$

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

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

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

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



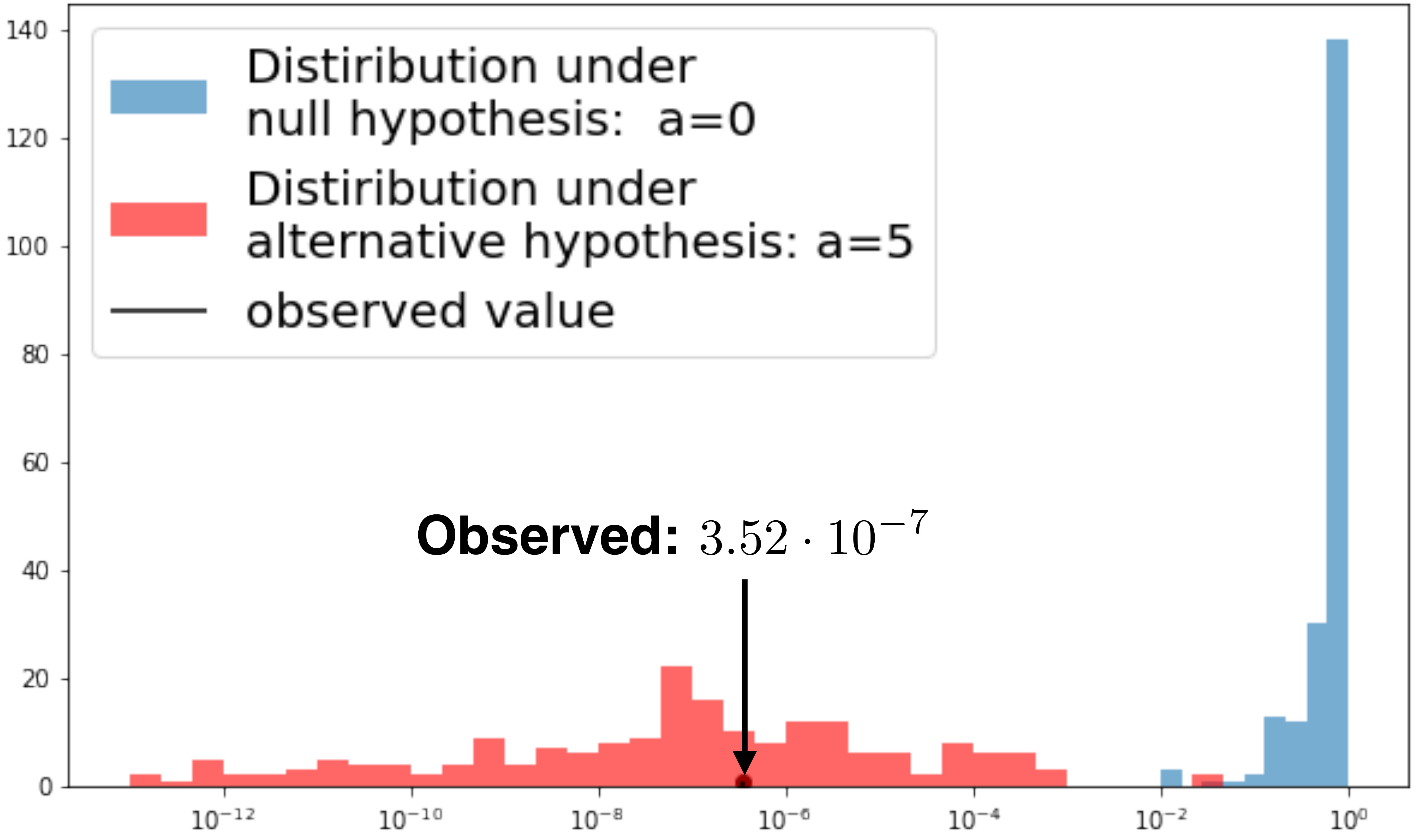
Null hypothesis H0
a = 0

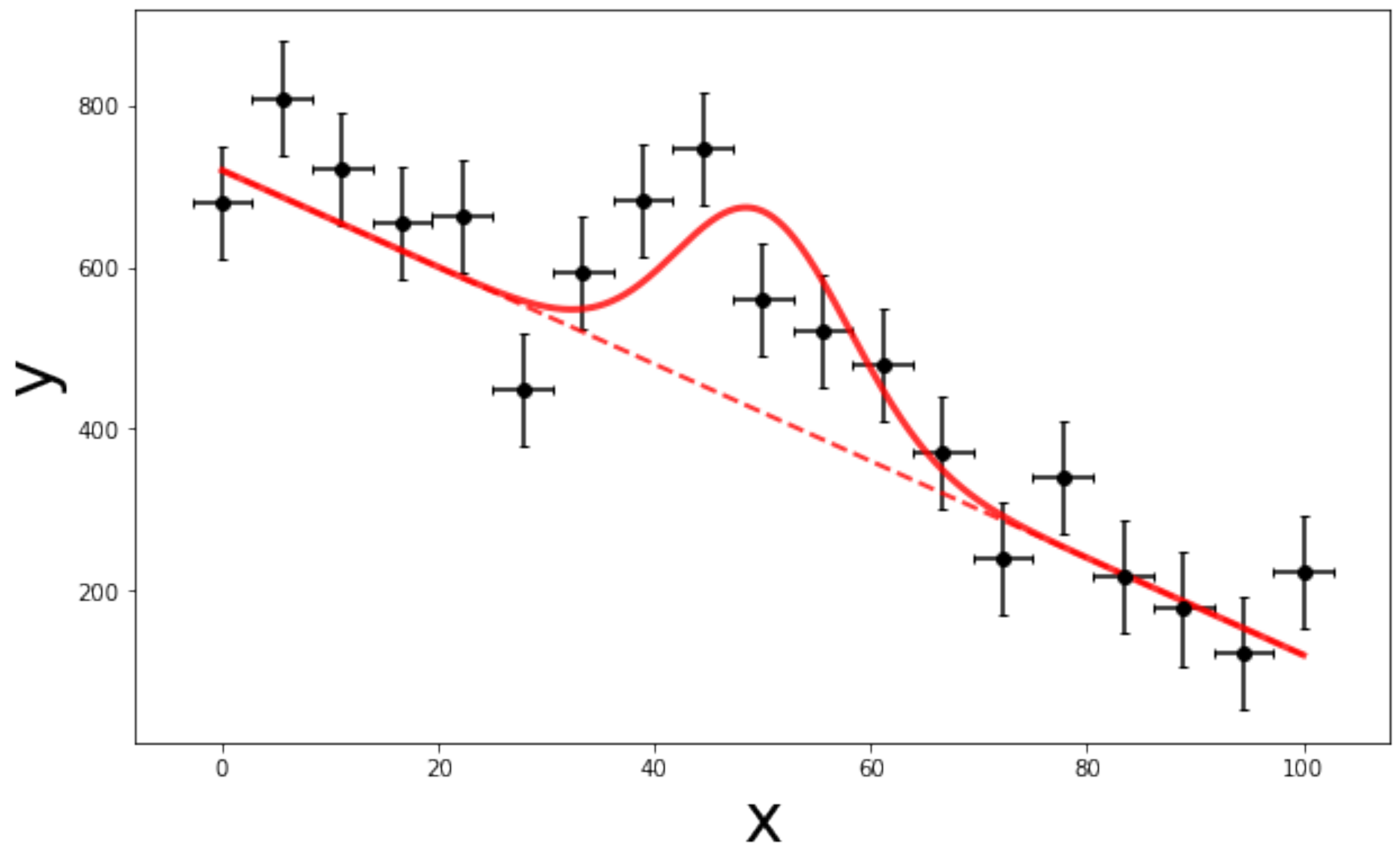
Alternative hyp. H1
a = 5

Example:

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

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



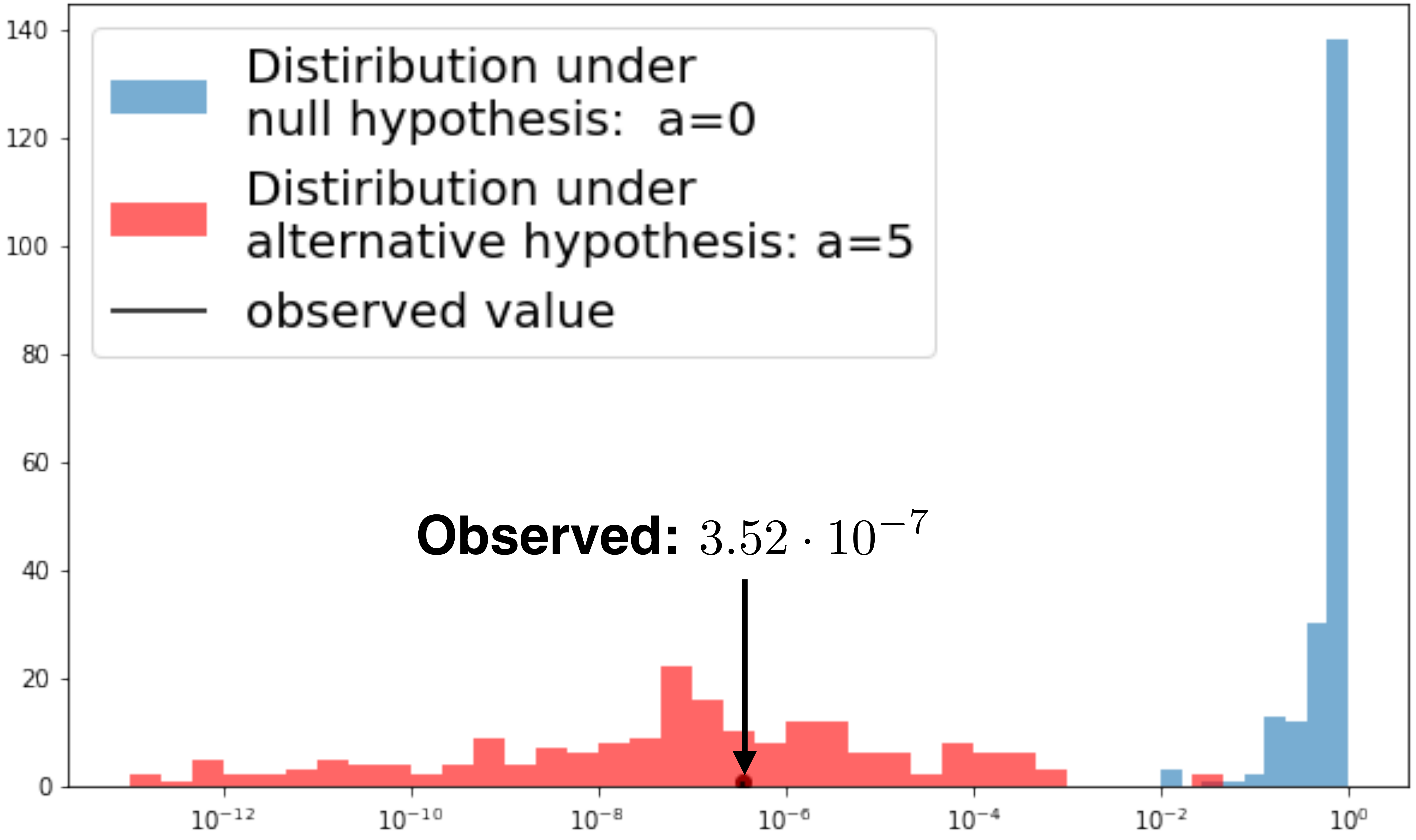


$$\mathcal{S} = \frac{\mathcal{L}(a=0)}{\mathcal{L}(a=\hat{a})}$$

Null hypothesis H0
a = 0

Alternative hyp. H1
a = 5

Example:



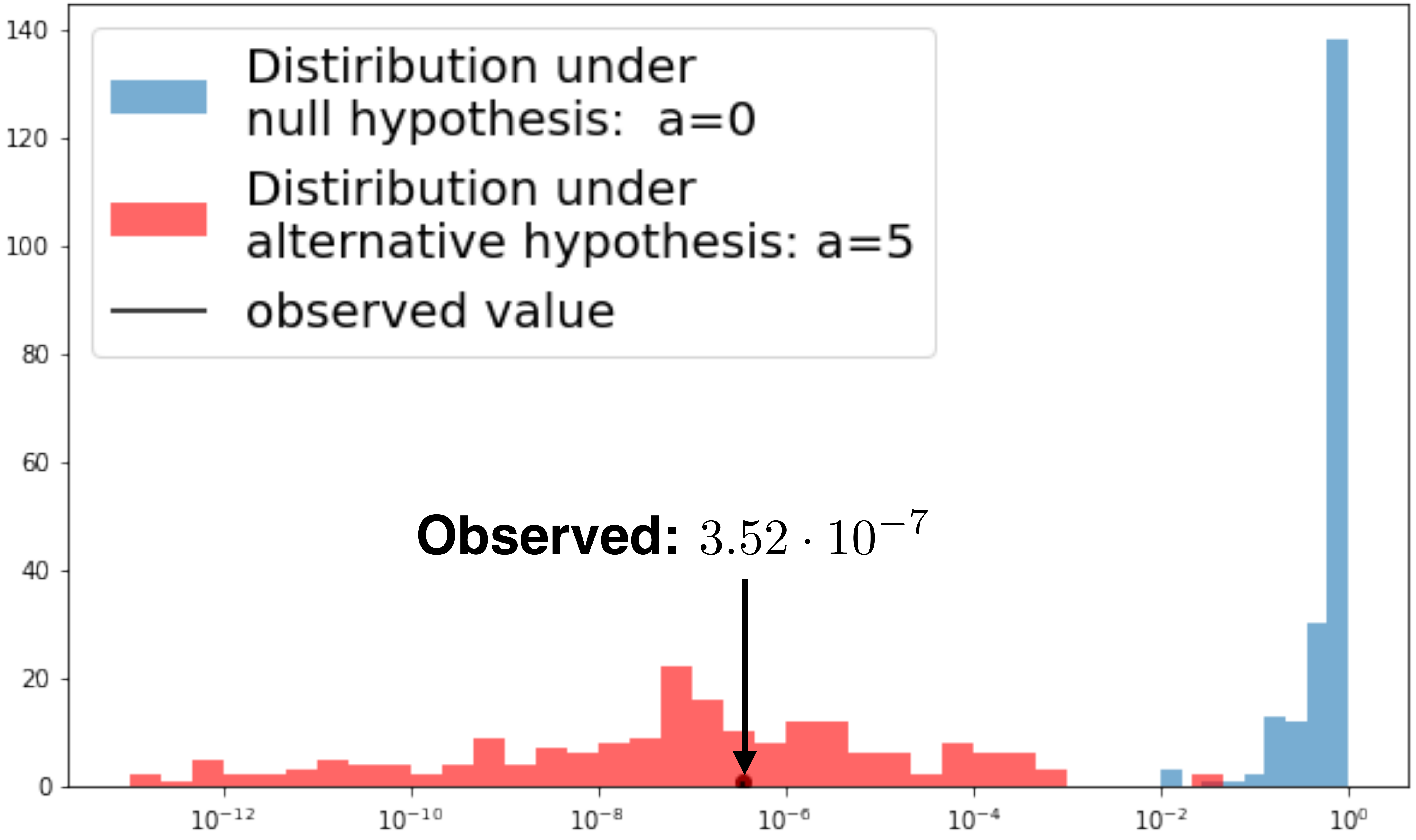
$$\mathcal{S} = \frac{\mathcal{L}(a=0)}{\mathcal{L}(a=\hat{a})}$$

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?

Example:

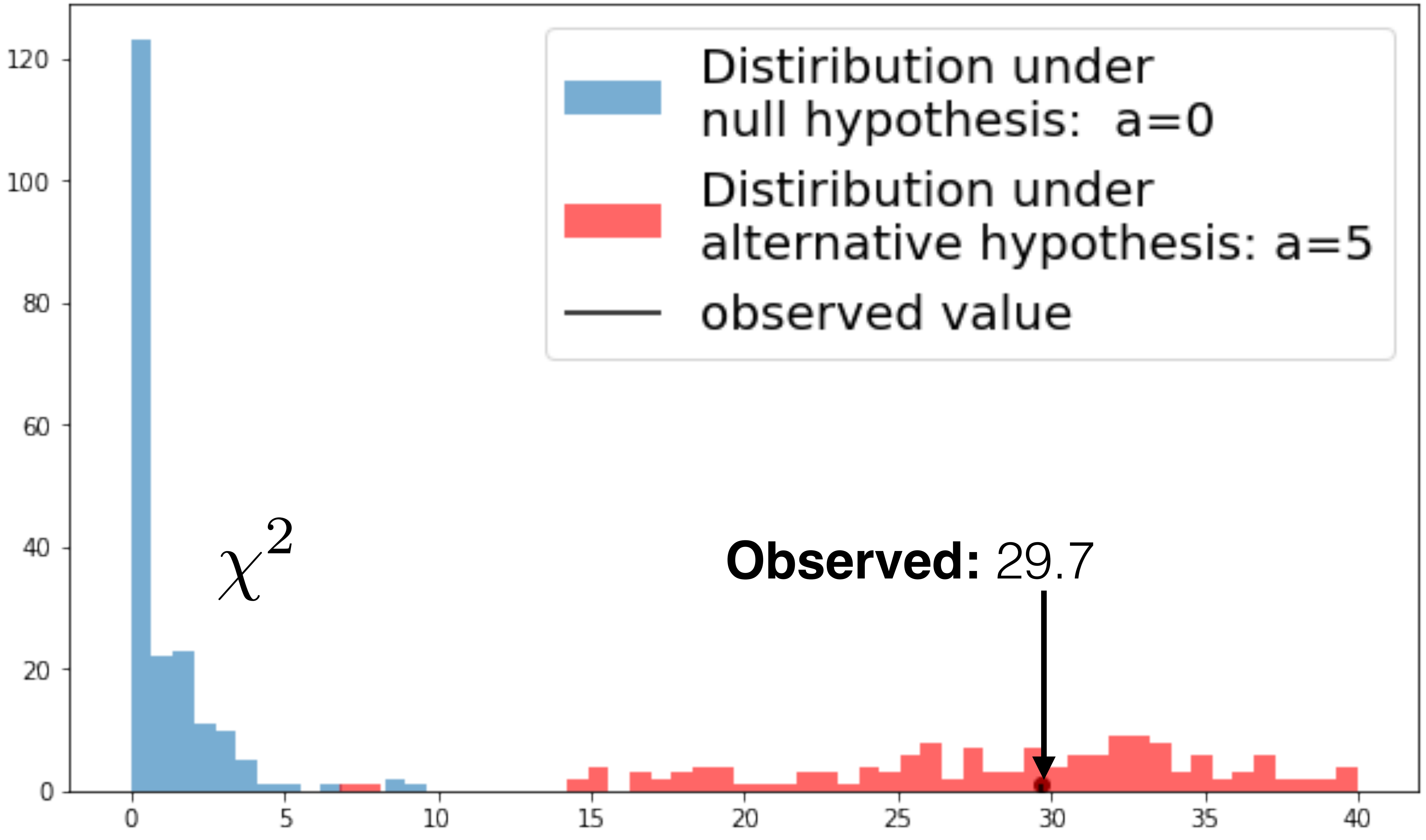


$$\mathcal{S} = \frac{\mathcal{L}(a=0)}{\mathcal{L}(a=\hat{a})}$$

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

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

Example:

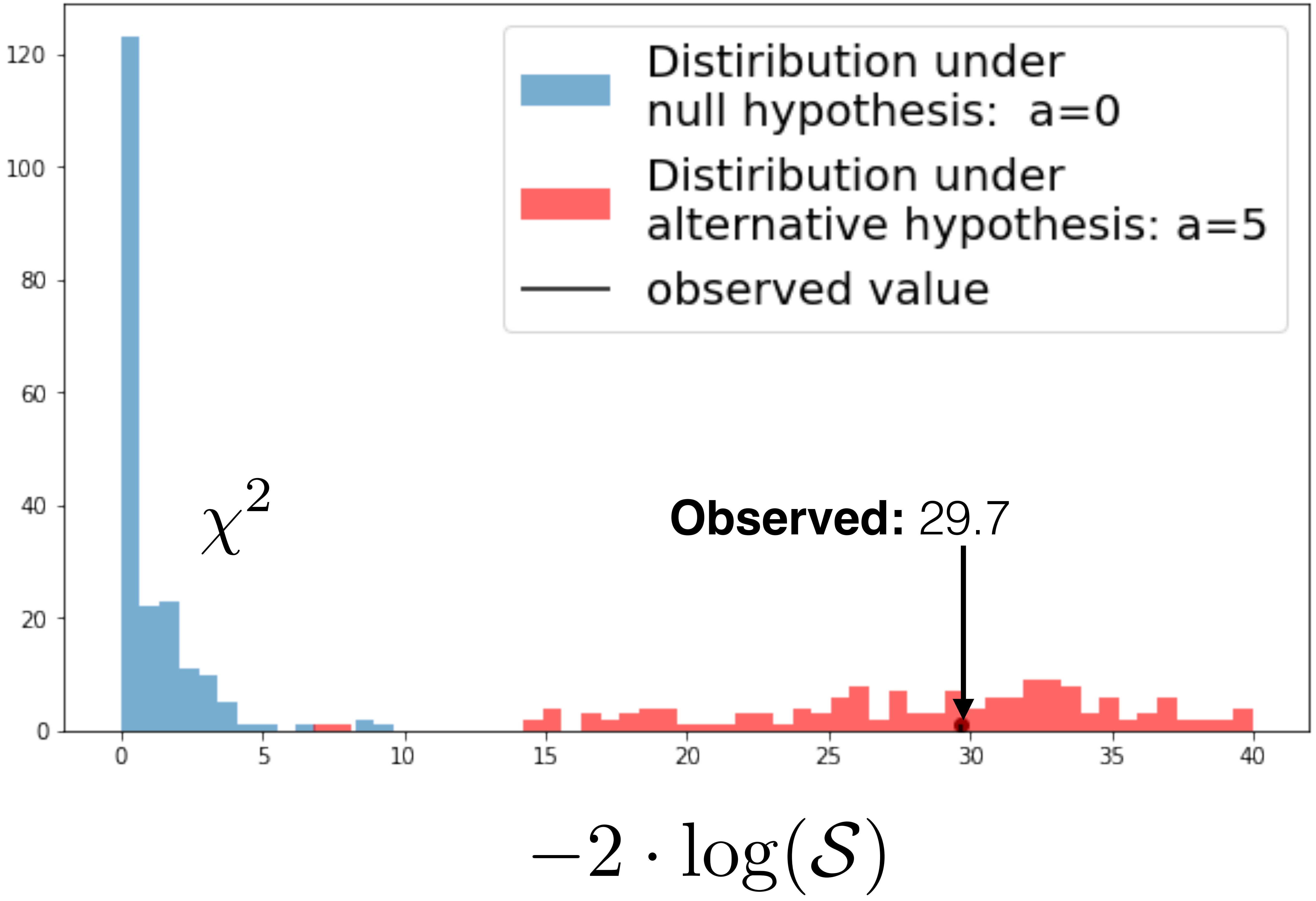


$$-2 \cdot \log(\mathcal{S})$$

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

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

Example:



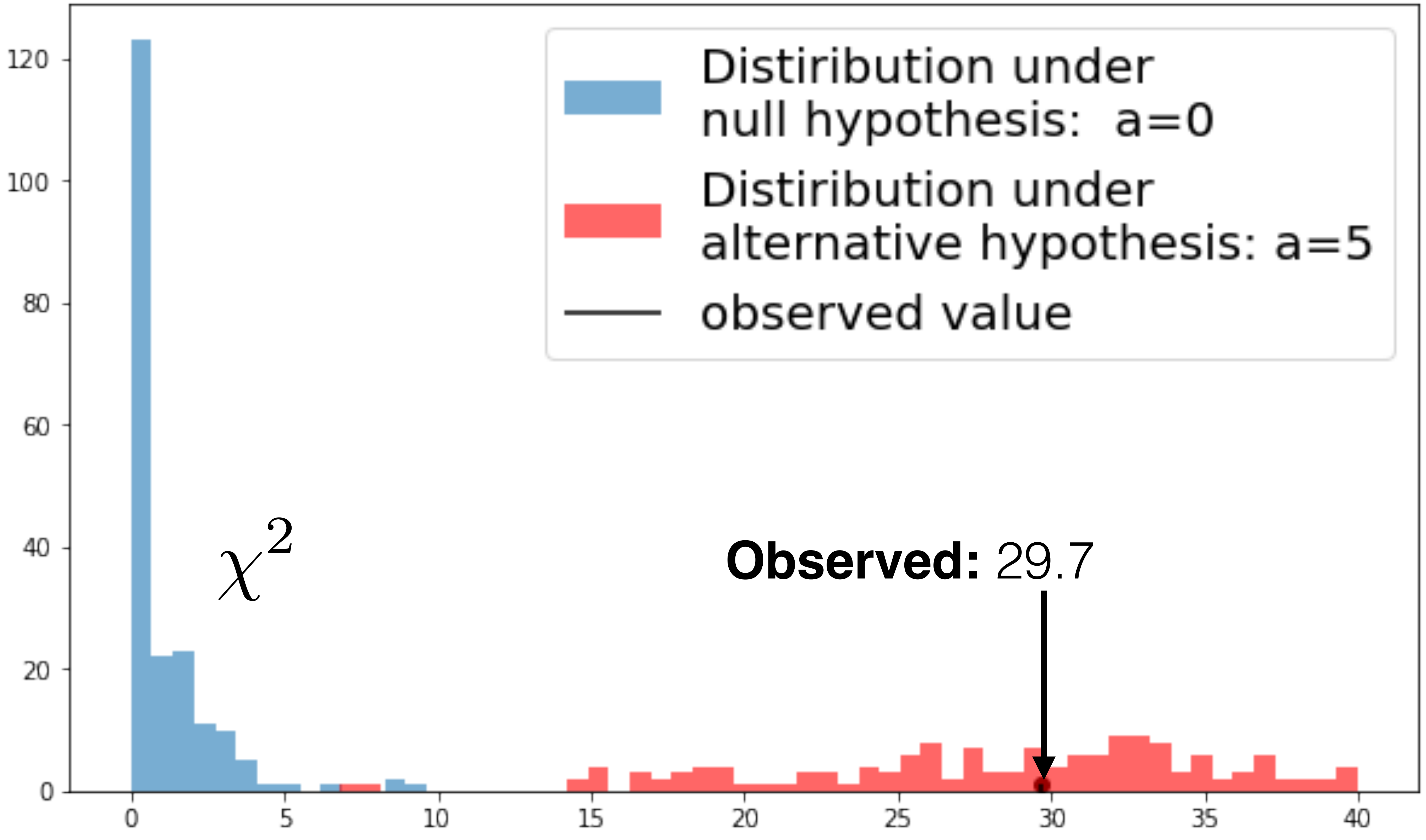
$$\text{p-value} = \int_{29.7}^{\infty} dx\ \chi^2(x) \simeq 5 \cdot 10^{-8}$$

Converting the p-value to a "sigma"

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

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

Example:



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

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

Converting the p-value to a "sigma"

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

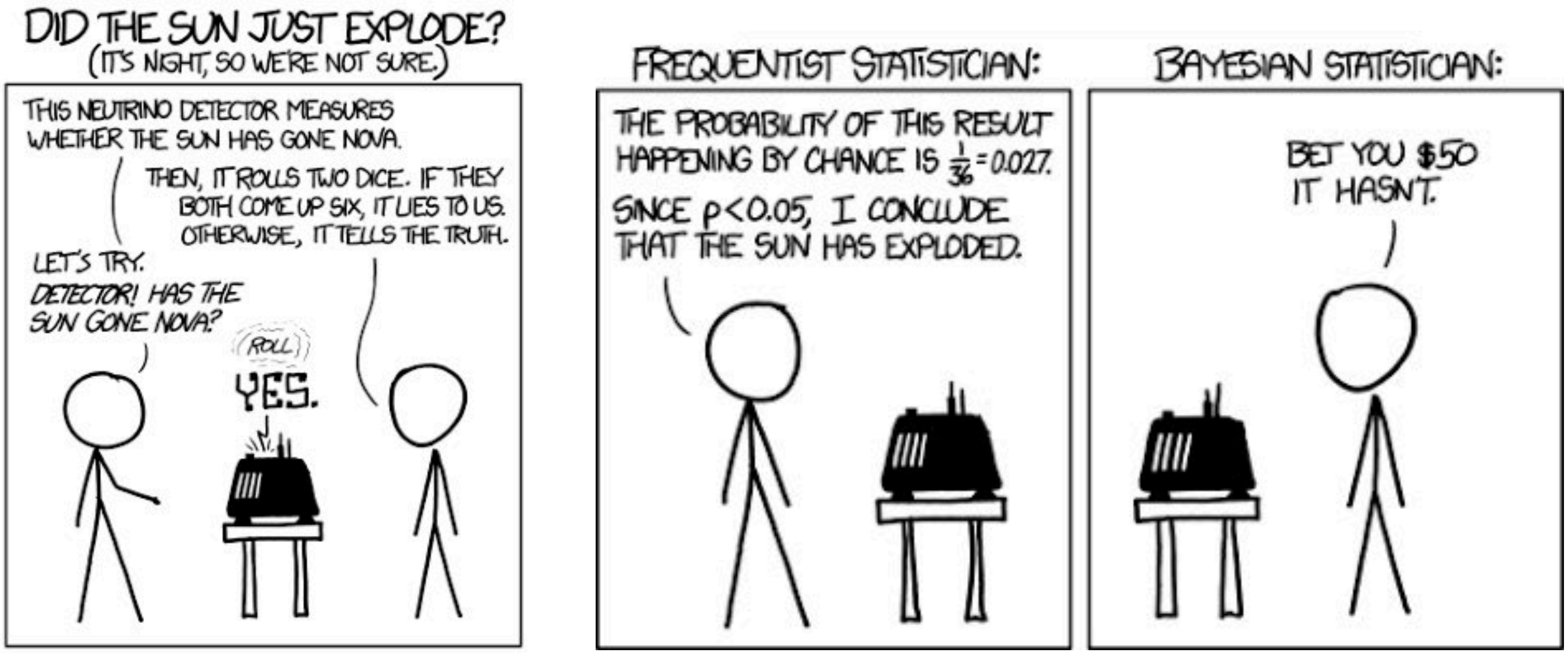
Notice that $\sqrt{29.7} \simeq 5.45$

Why?

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.


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS 1/36 = 0.027.
SINCE p<0.05, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.

Introduction to Statistical Inference

part II

Giacomo D'Amico

ArQus School 2022, Bergen, Norway

5-9 Sep. 2022

Statistical inference applied in gamma-ray astronomy

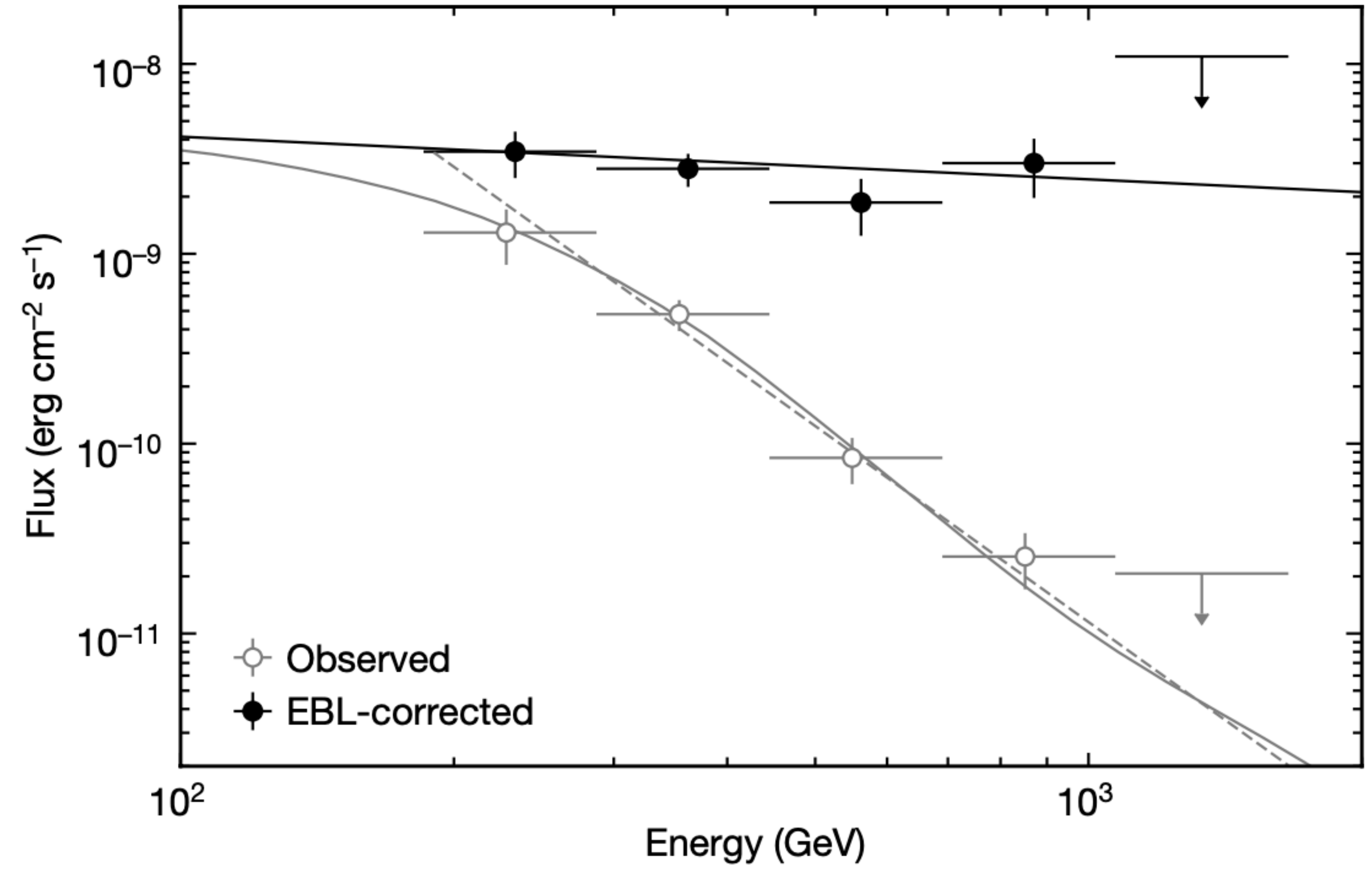
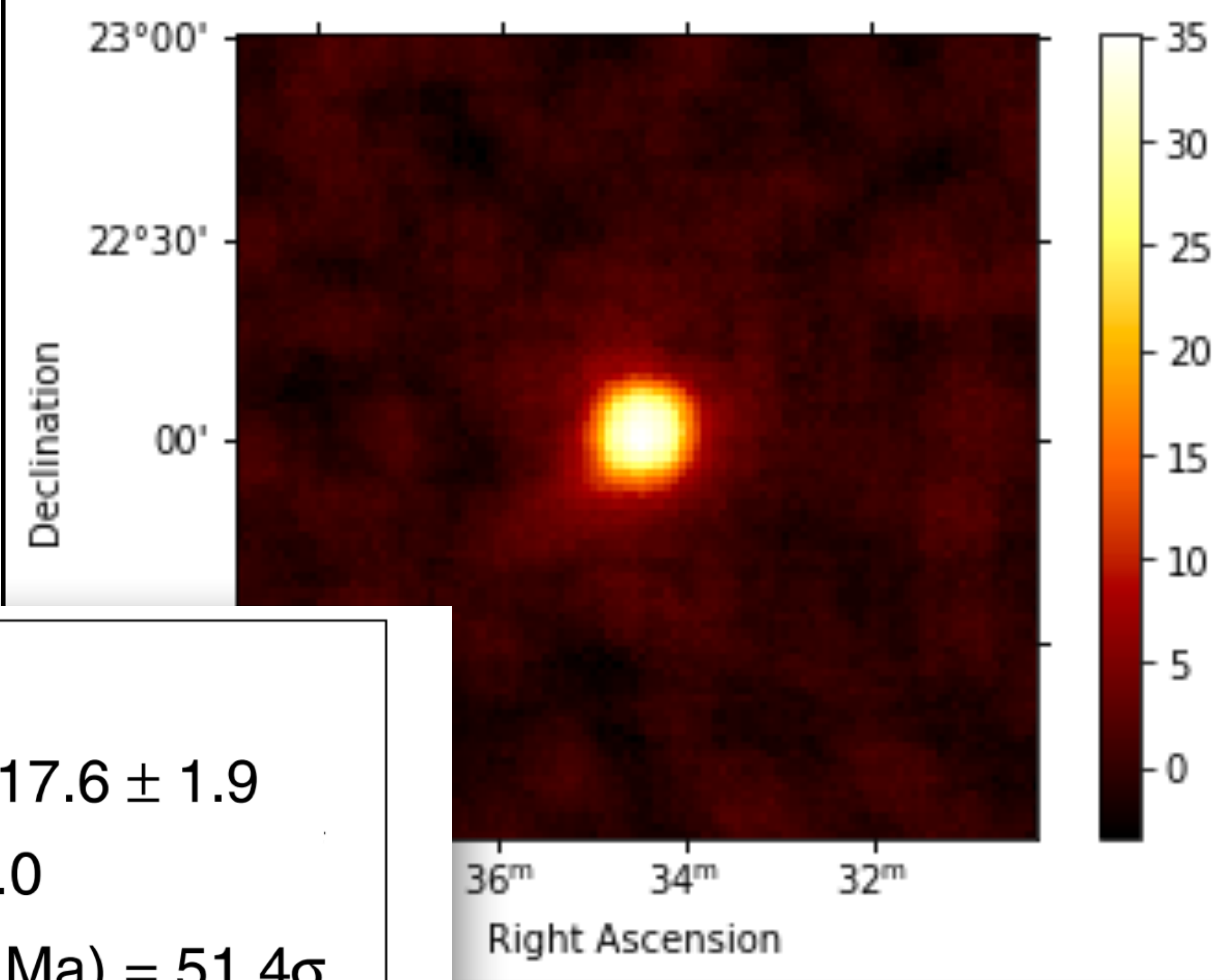
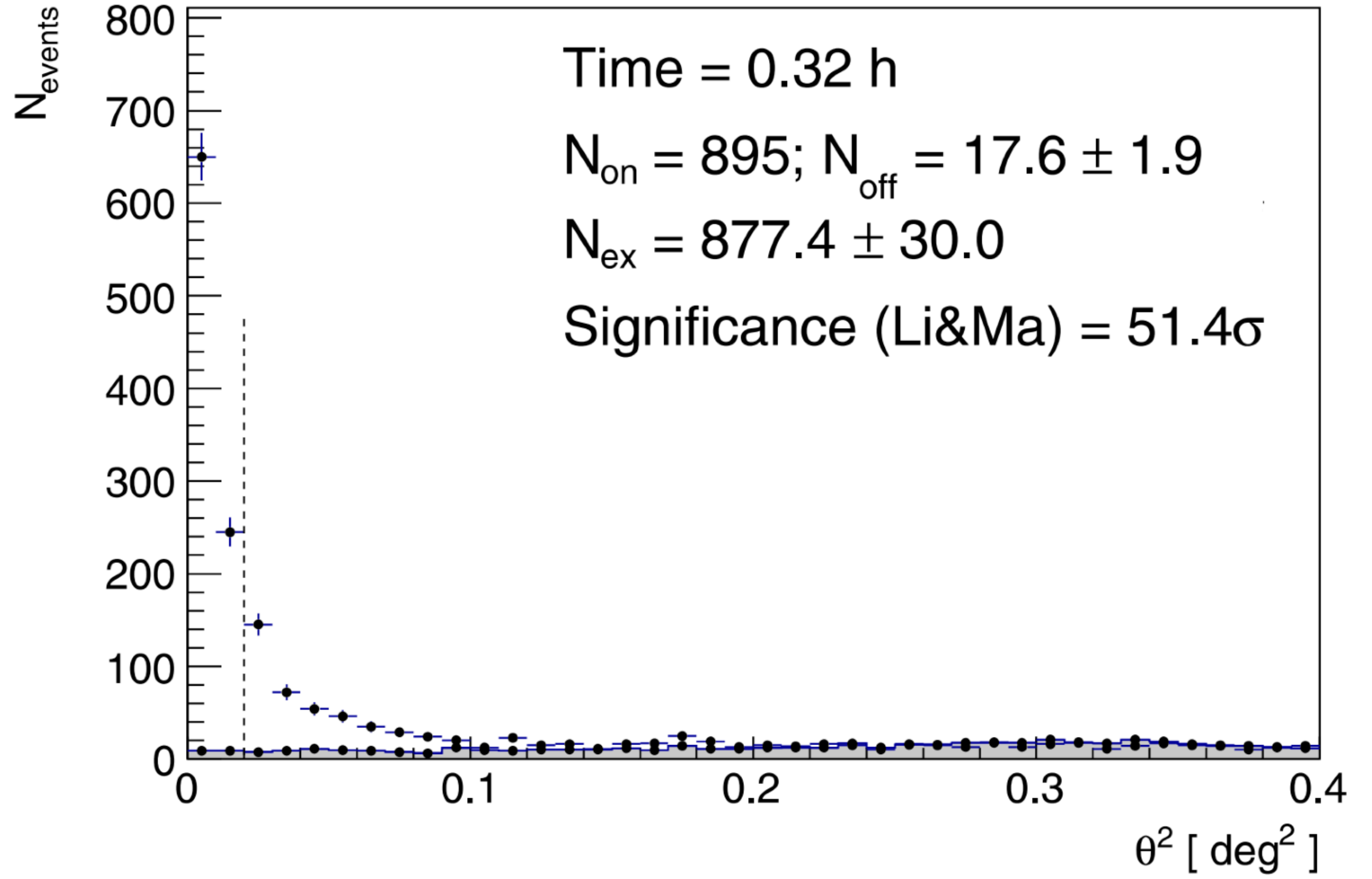


Fig. 2 | Spectrum above 0.2 TeV averaged over the period between T_0 + 62 s and T_0 + 2,454 s for GRB 190114C. Spectral-energy distributions for the spectrum observed by MAGIC (grey open circles) and the intrinsic spectrum corrected for EBL attenuation[25] (blue filled circles). The errors on the flux 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].

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"?

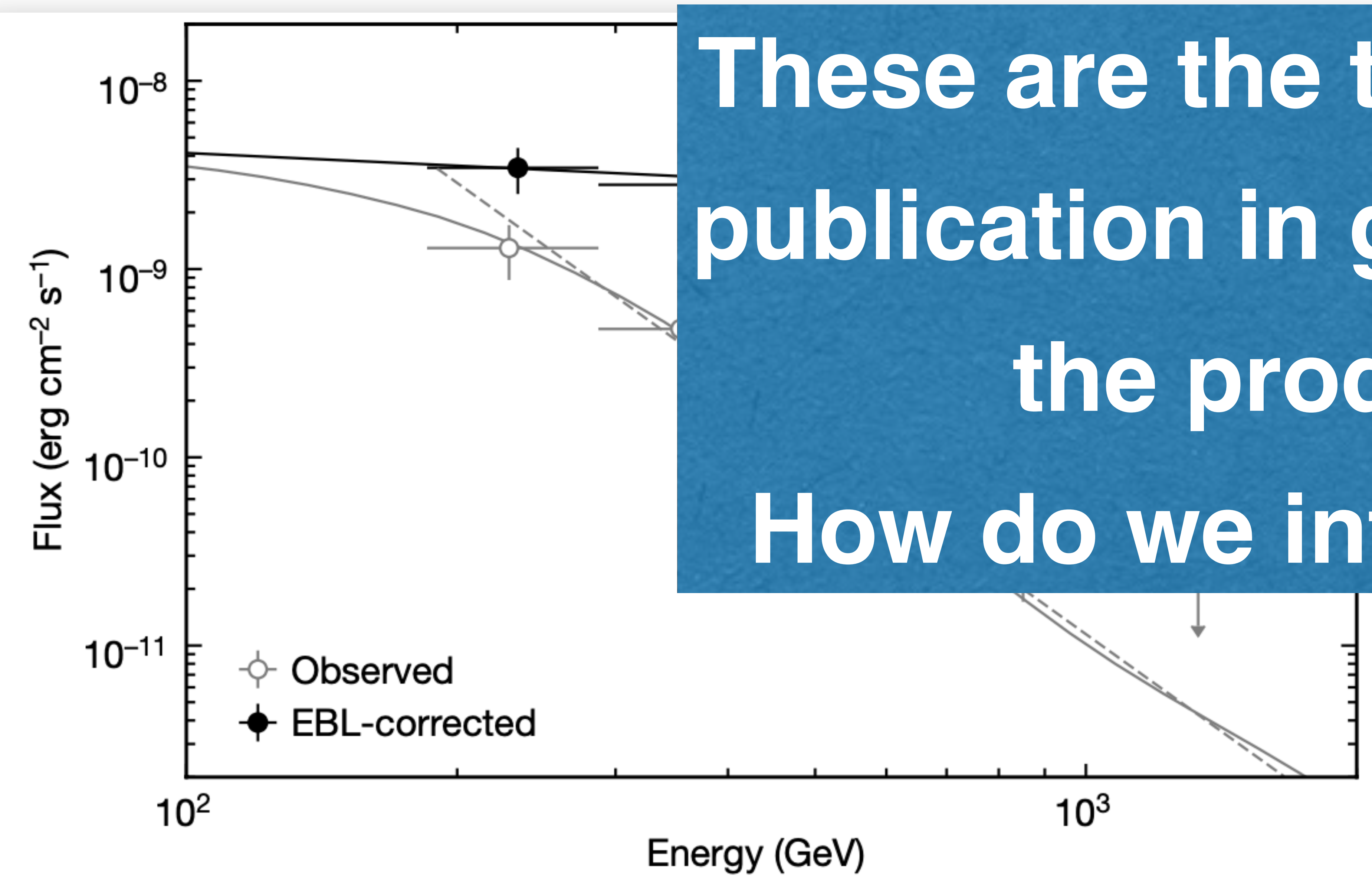
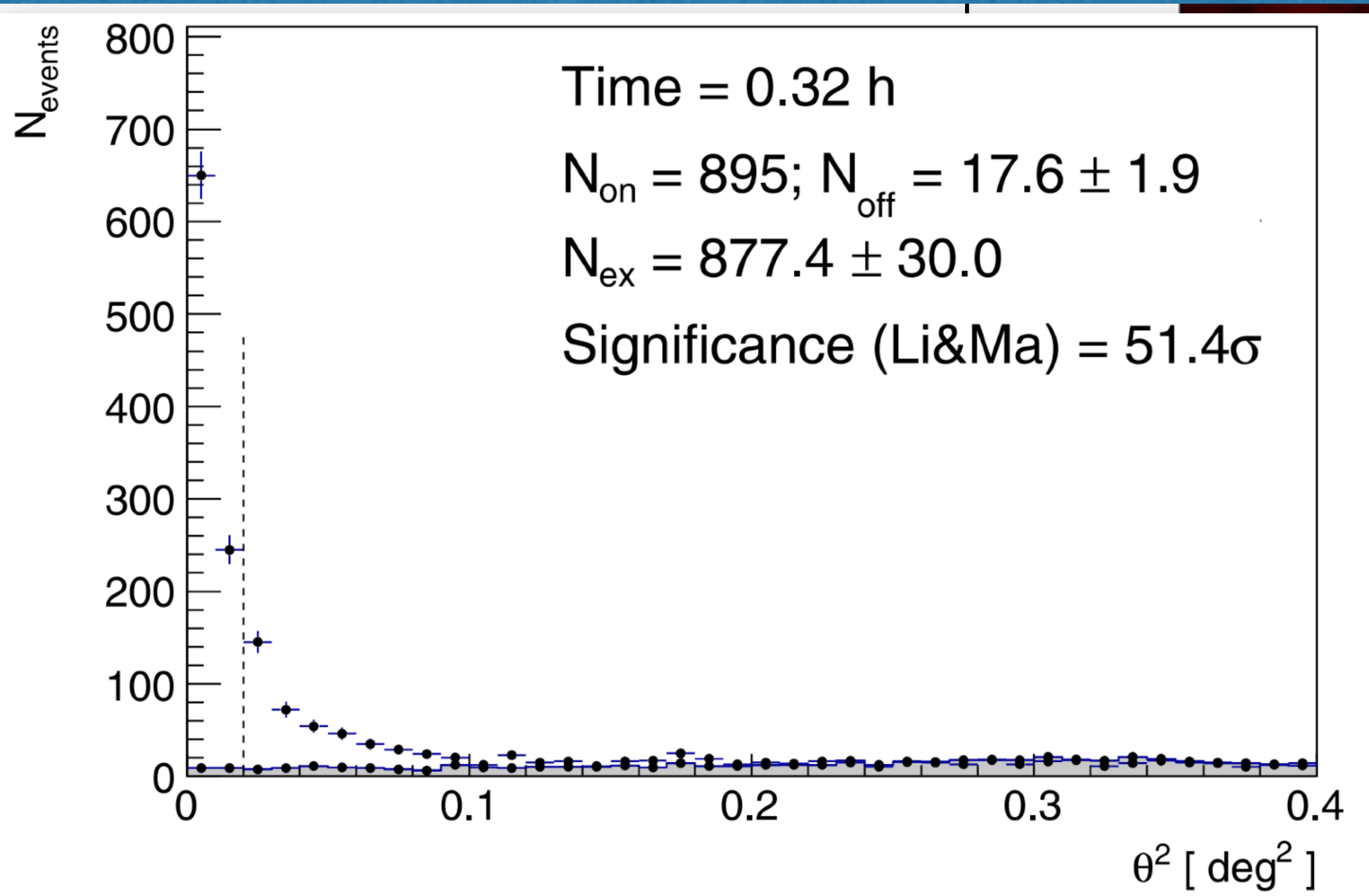
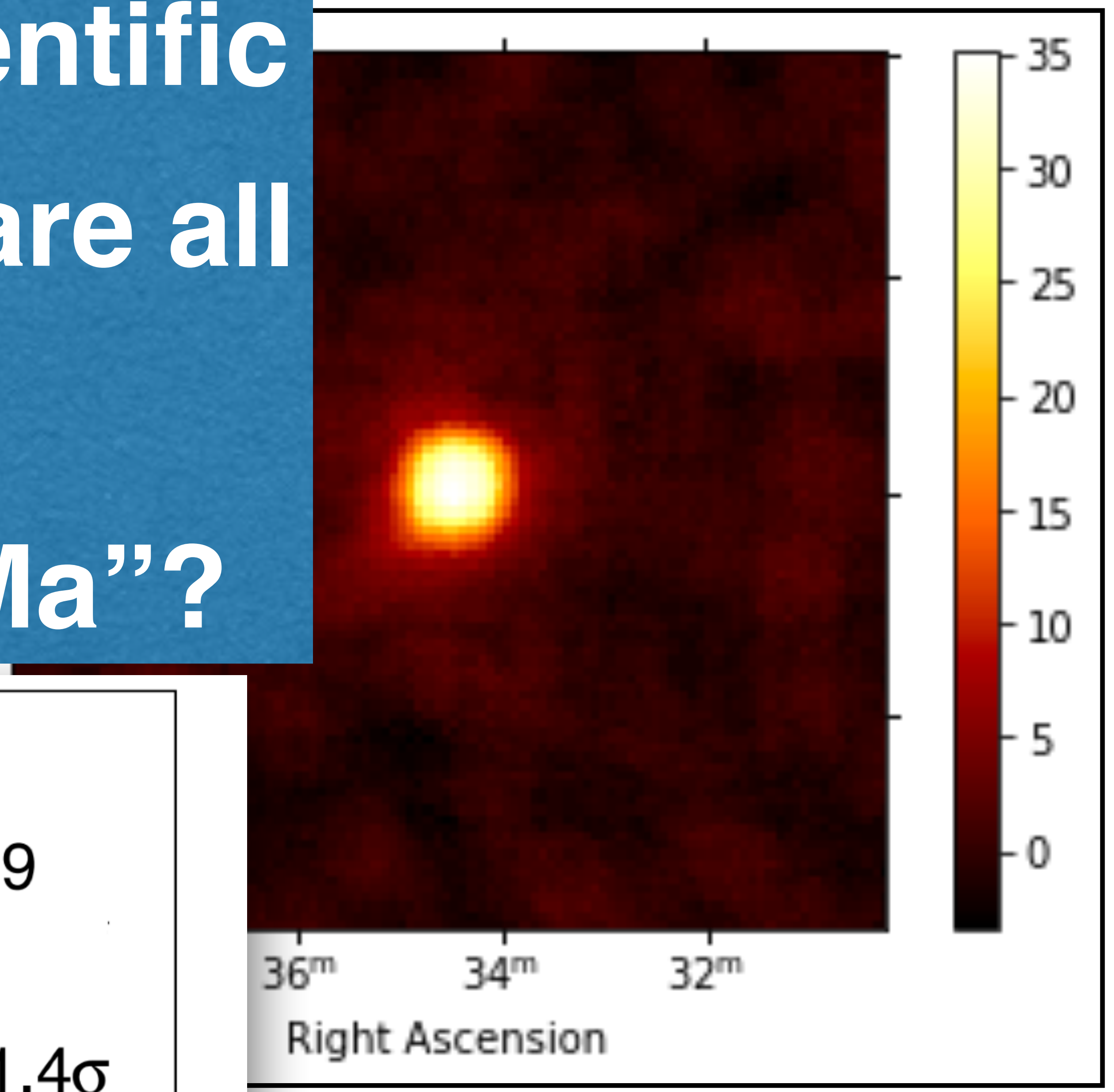
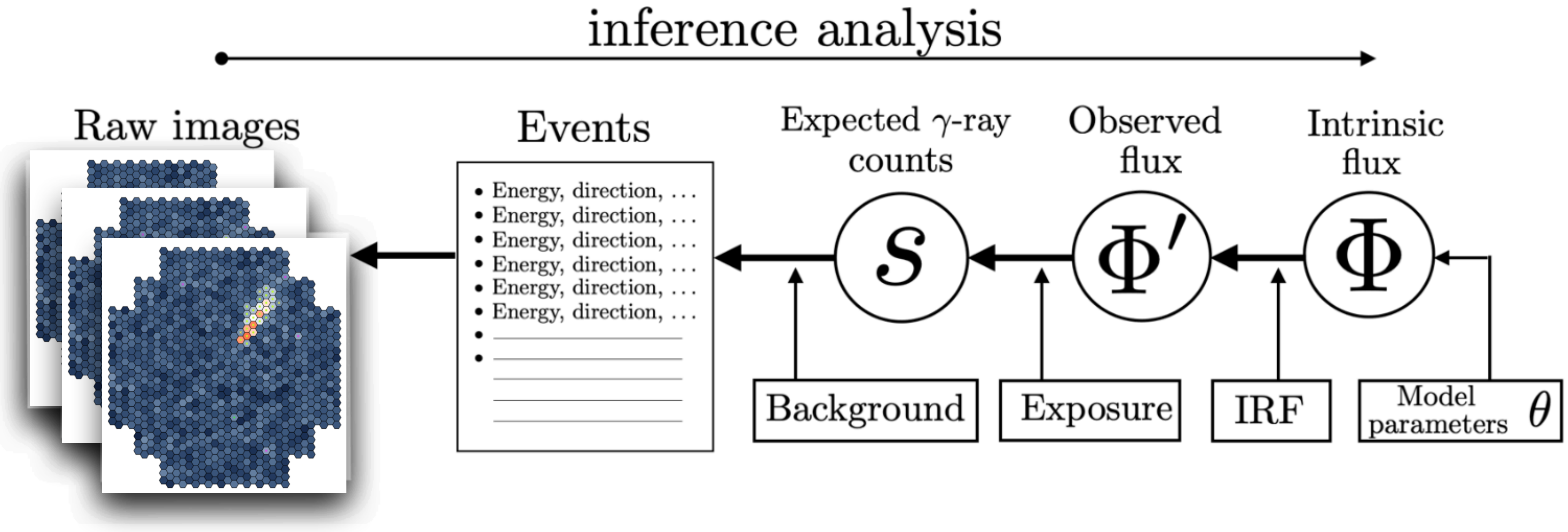


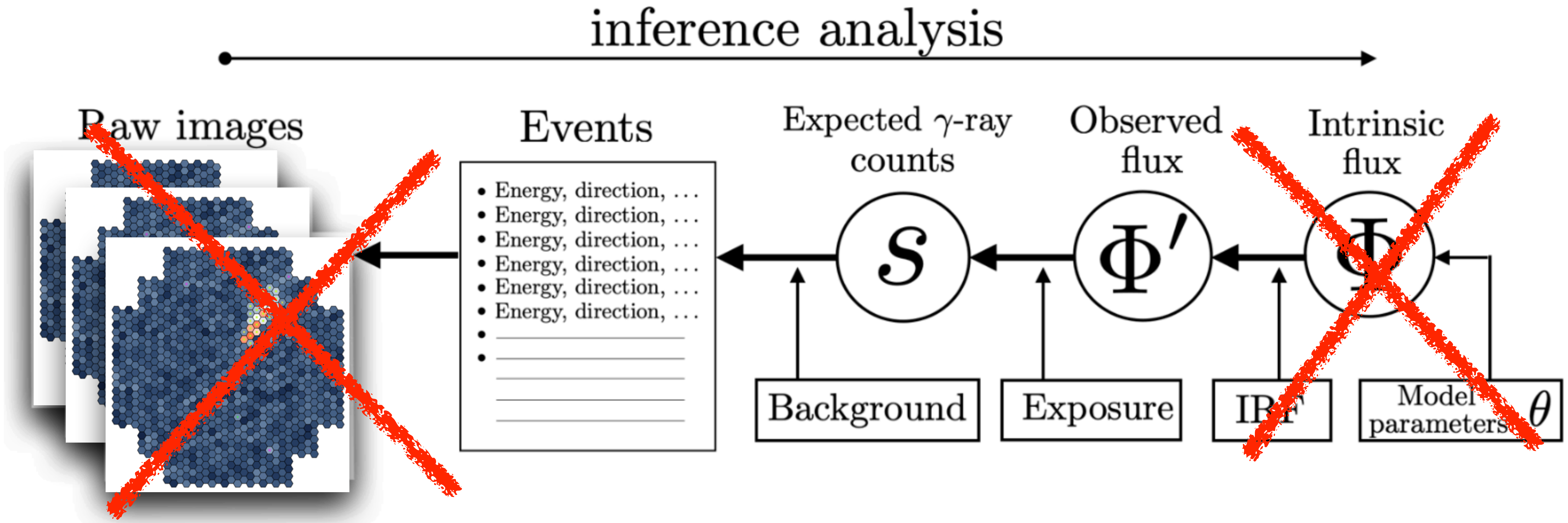
Fig. 2 | Spectrum above 0.2 TeV averaged over the period between $T_0 + 62$ s and $T_0 + 2{,}454$ s for GRB 190114C. Spectral-energy distributions for the spectrum observed by MAGIC (grey open circles) and the intrinsic spectrum corrected for EBL attenuation[25] (blue filled circles). The errors on the 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].

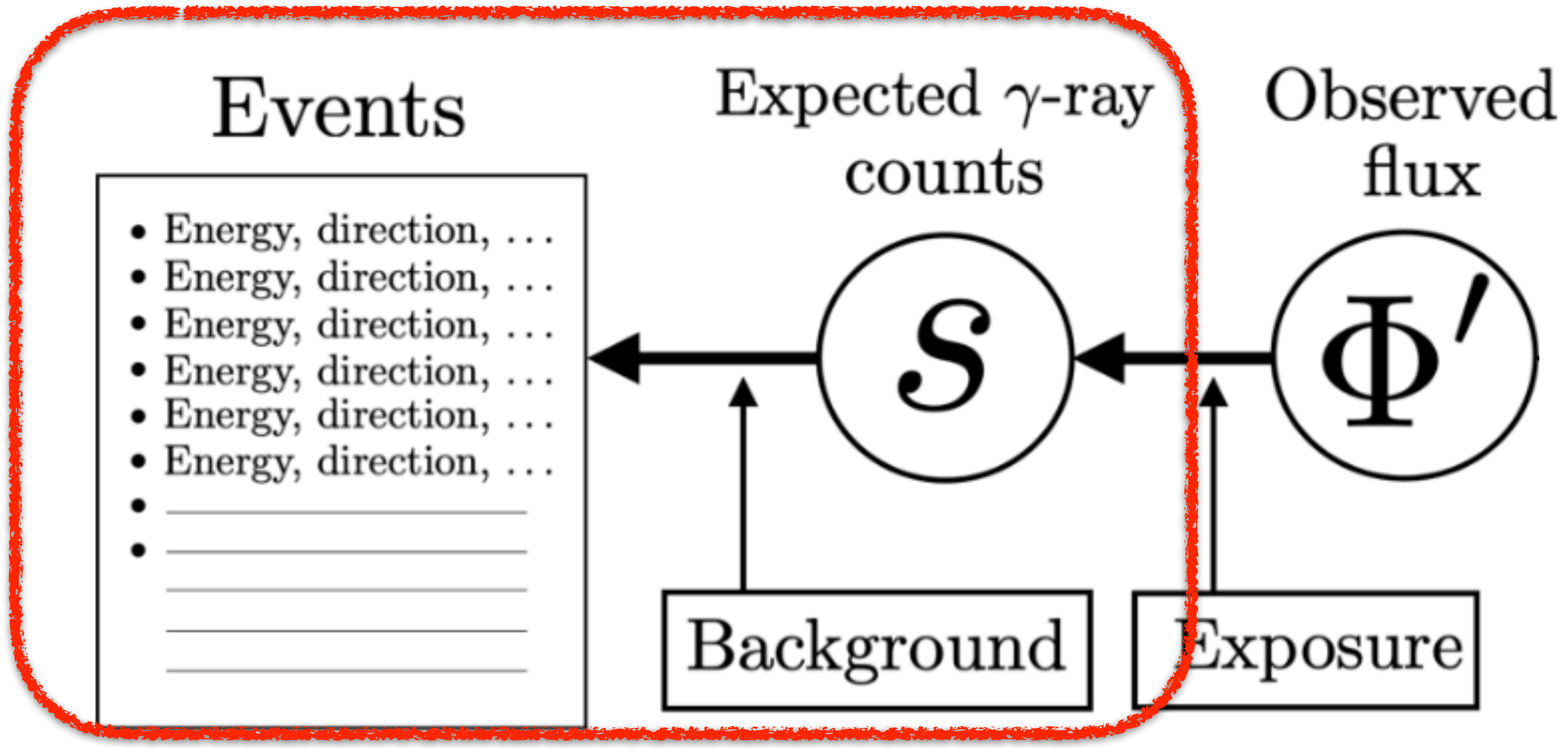



inference analysis
Raw images
Events
Energy, direction, ...
Energy, direction, ...
Energy, direction, ...
Energy, direction, ...
Energy, direction, ...
Energy, direction, ...
Expected γ-ray counts
Observed flux
Intrinsic flux
S
Φ′
Φ
Background
Exposure
IRF
Model parameters θ



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?



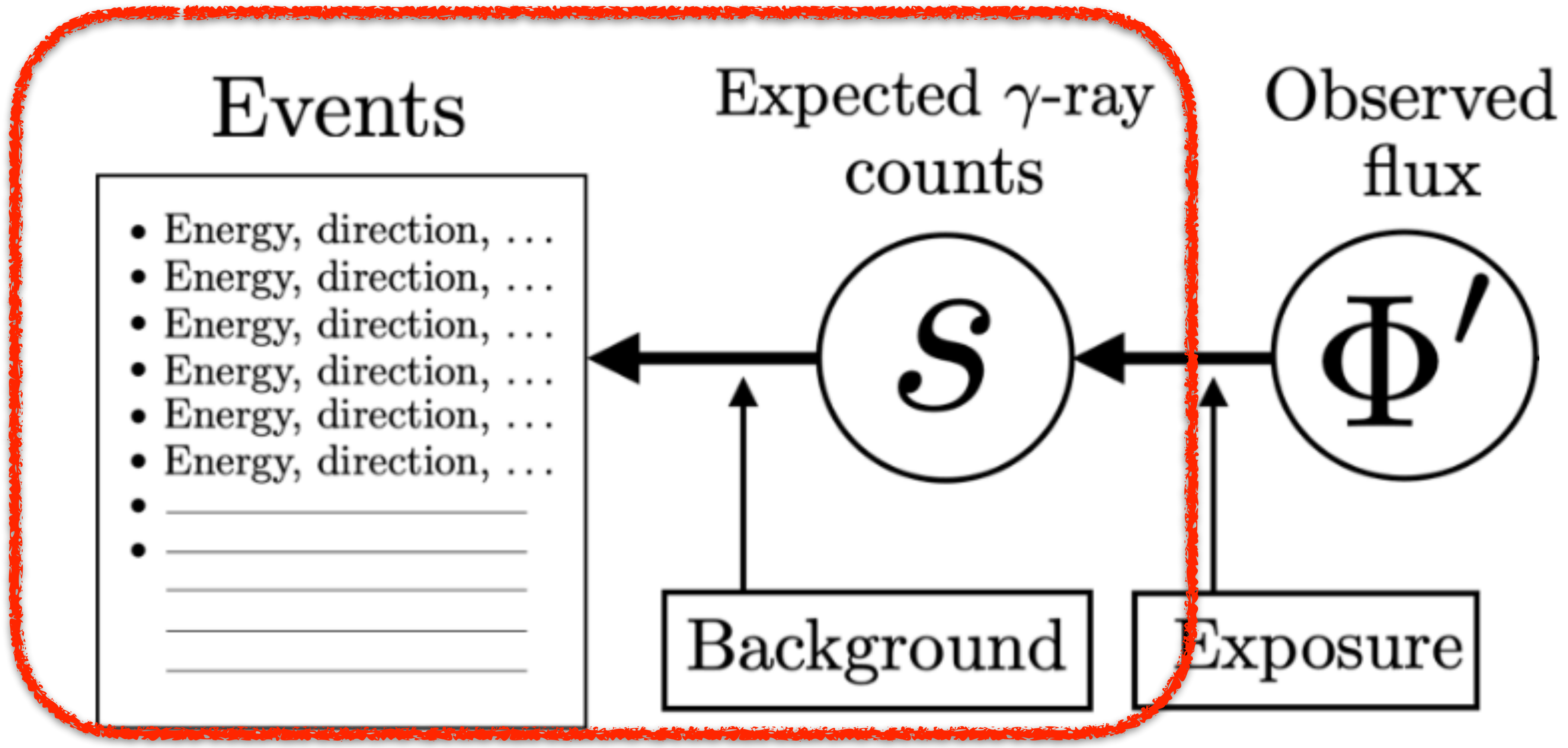
Given your event list what's the expected number of gamma-ray?

In [4]:
```
from astropy.io import fits
from astropy.table import Table
```

In [5]:
```
name_file = "run_05029747_DL3.fits"
Table.read(name_file  ,hdu=1)
```

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



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

In [4]:
```
from astropy.io import fits
from astropy.table import Table
```

In [5]:
```
name_file = "run_05029747_DL3.fits"
Table.read(name_file  ,hdu=1)
```

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…)?

The “ingredients”

The flux

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

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

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

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

Expected background events “b”

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

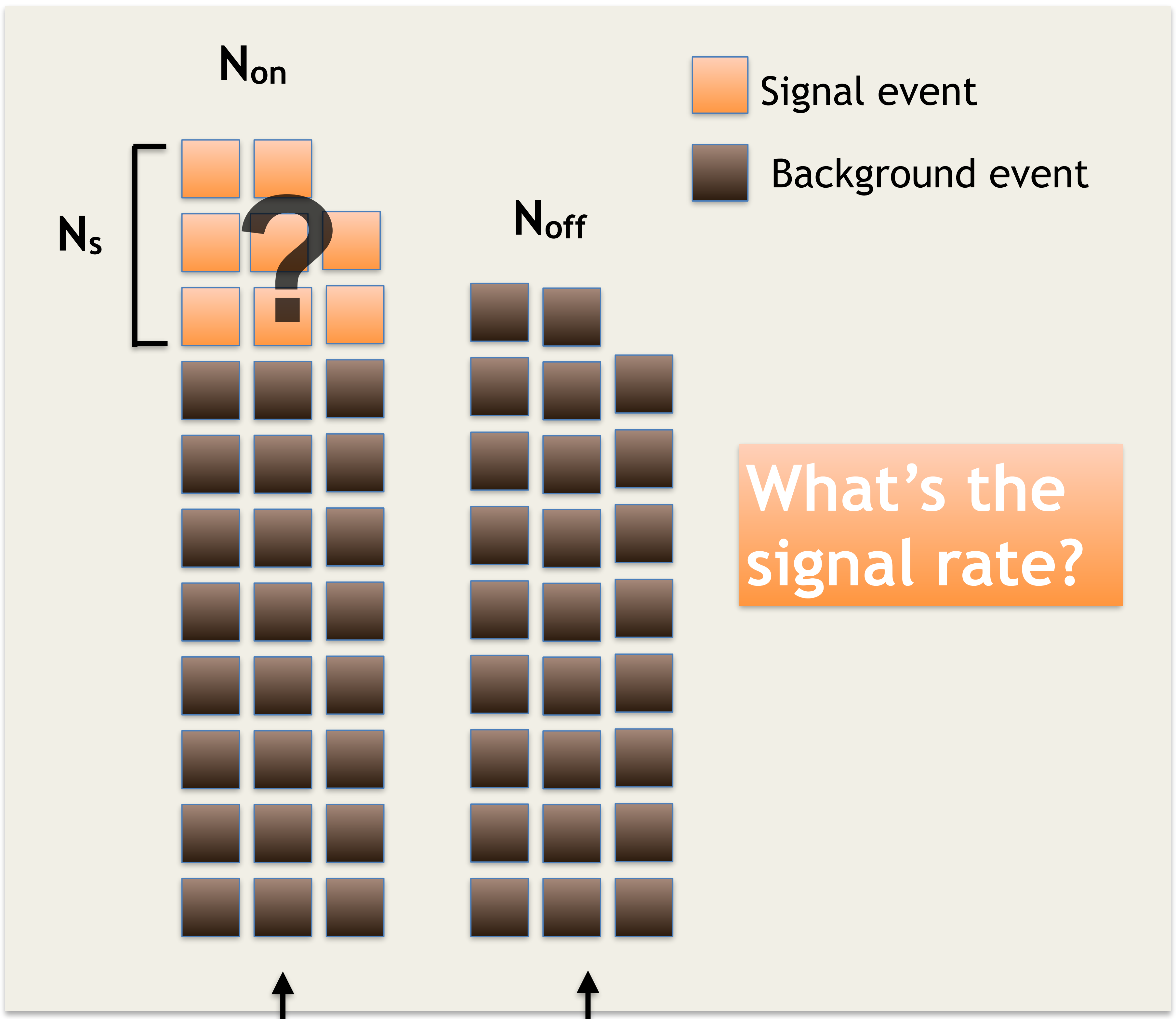
Total number of observed events “on source”

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

Total number of observed events “off source”

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

On/Off measurement



In an **On/Off experiment**:

- a background-control (**Off**) region, which is supposedly void of any signal, is defined to estimate the **background rate** (b)
- the **On** source measurement instead provides an estimate of the **signal rate** (s) plus b, with the latter term supposed to be equal to that in the Off region.

The following **variables** are therefore introduced:

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

Probability mass function of observing N_{on} and N_{off} or
Likelihood function of the signal (s) and background (b) rate

On measurement Off measurement

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

On/Off measurement

Signal estimation in the **frequentist approach:**

Likelihood function:

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

On/Off measurement

Signal estimation in the **frequentist approach:**

Likelihood function:

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

$$\downarrow$$

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

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

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

On/Off measurement

Signal estimation in the **frequentist approach:**

Likelihood function:

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

$\downarrow$

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

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

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

$\downarrow$

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

On/Off measurement

Signal estimation in the **frequentist approach:**

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

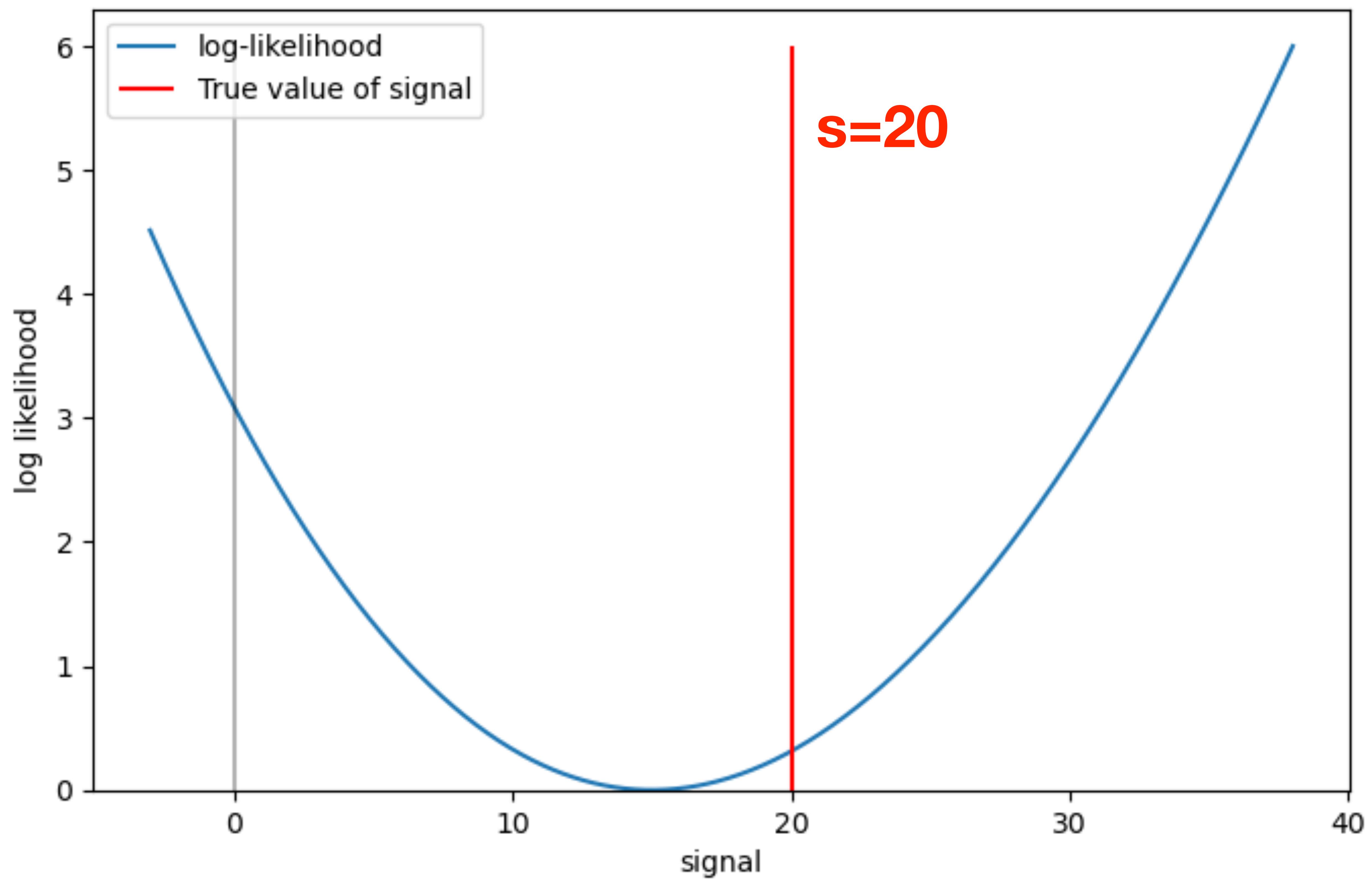
Example with:

Non = 57
Noff = 85
α = 0.5

which have been produced from a Poissonian sampling with **s=20** and b=90:

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

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



On/Off measurement

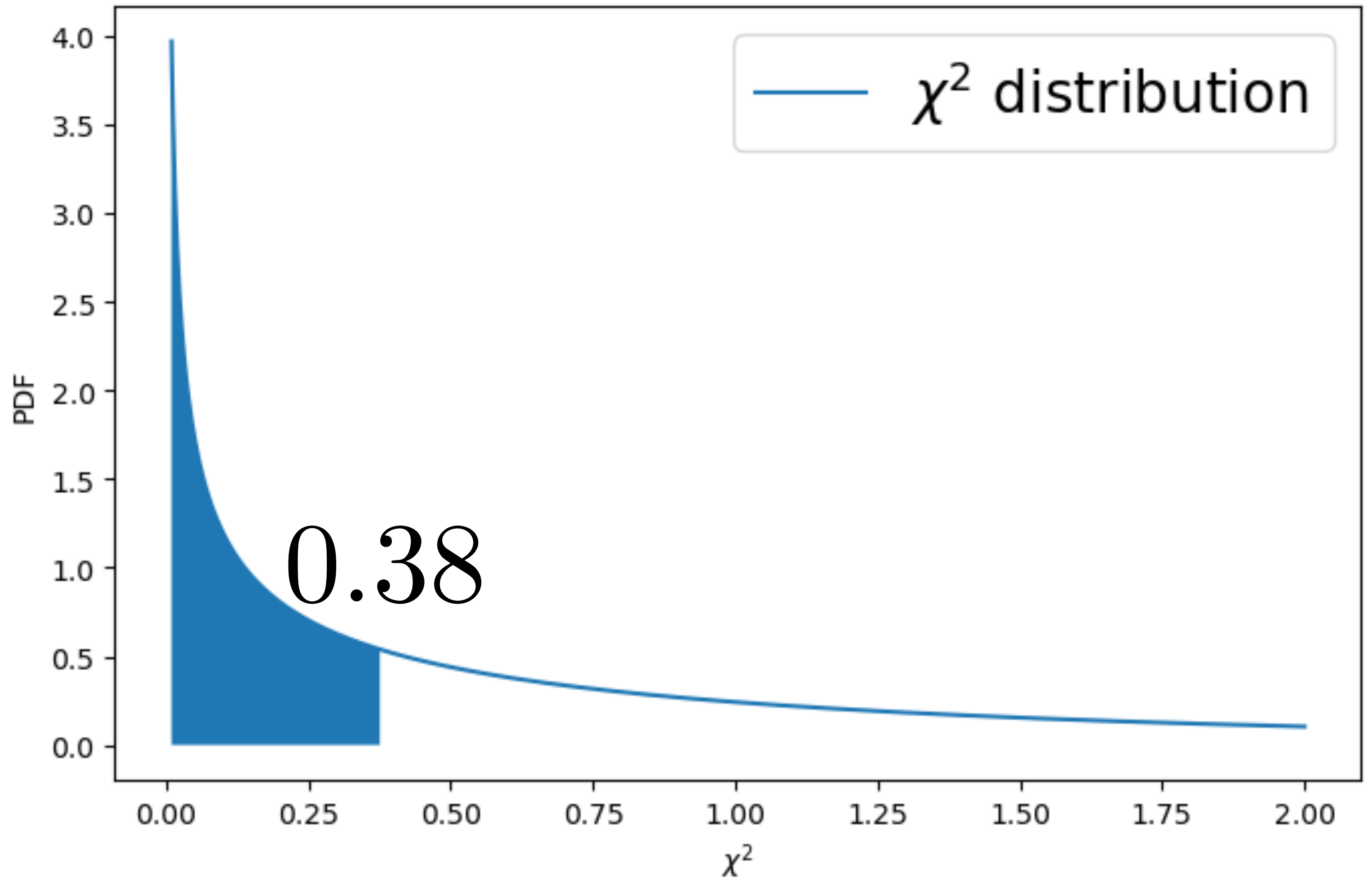
Signal estimation in the **frequentist approach:**

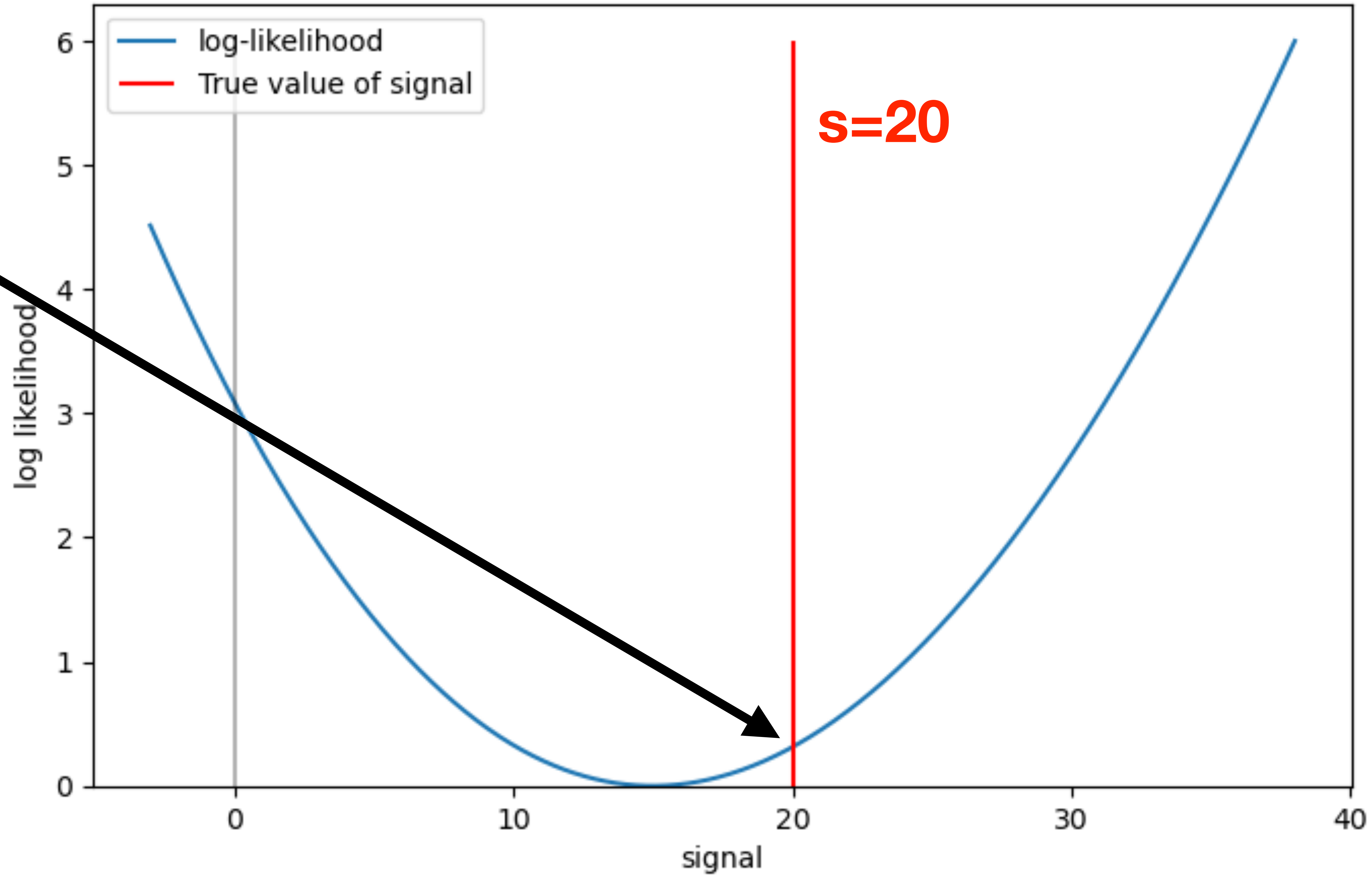
$$-2\log\lambda(s) = 2\left[N_{on}\log\left(\frac{N_{on}}{s+\alpha\hat{b}}\right) + N_{off}\log\left(\frac{N_{off}}{\hat{b}}\right) + s + (1+\alpha)\hat{b} - N_{on} - N_{off}\right]$$

Our statistic is

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

Which is an expected value for a chi-squared variable





On/Off measurement

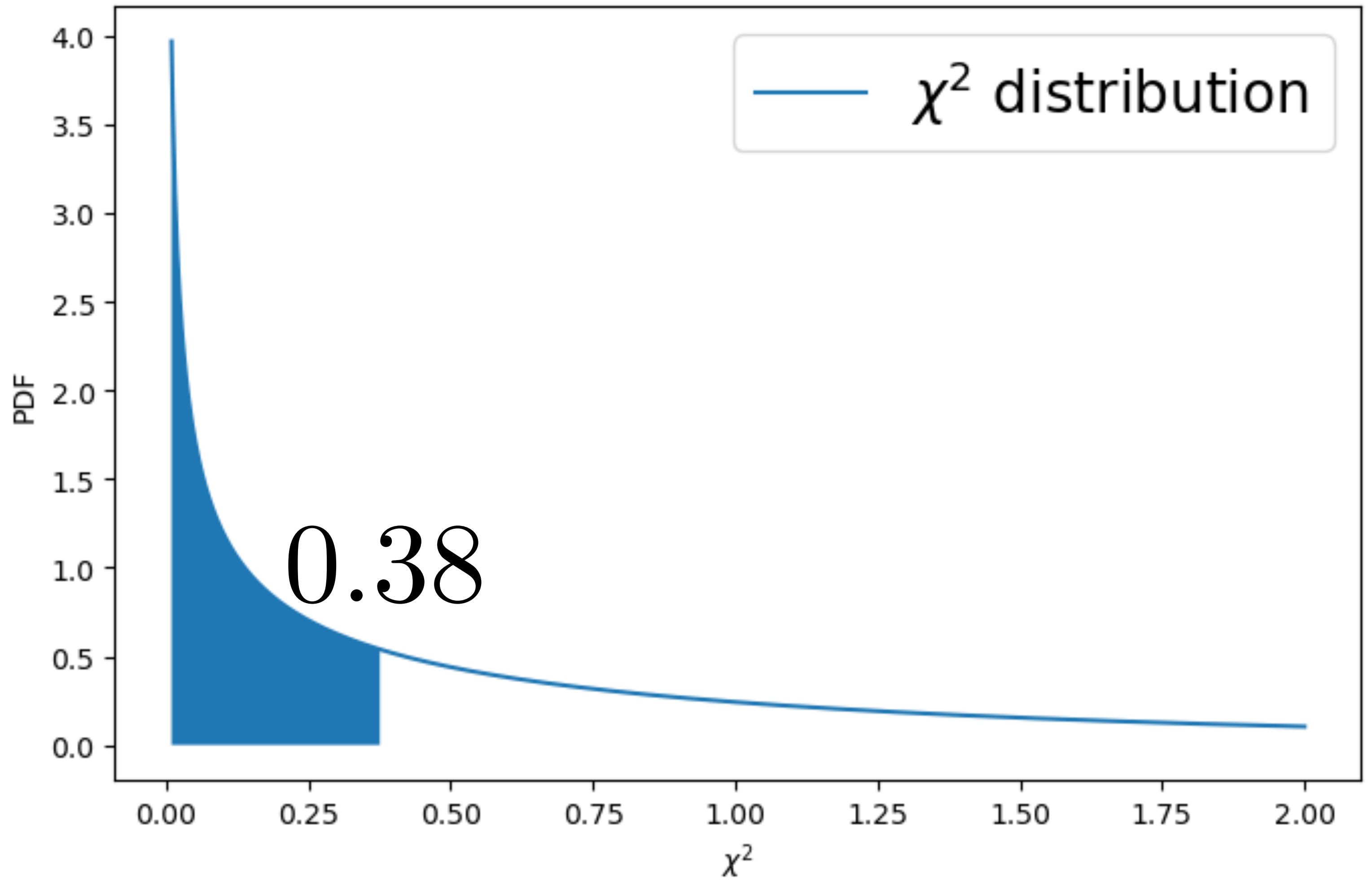
Signal estimation in the **frequentist approach:**

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

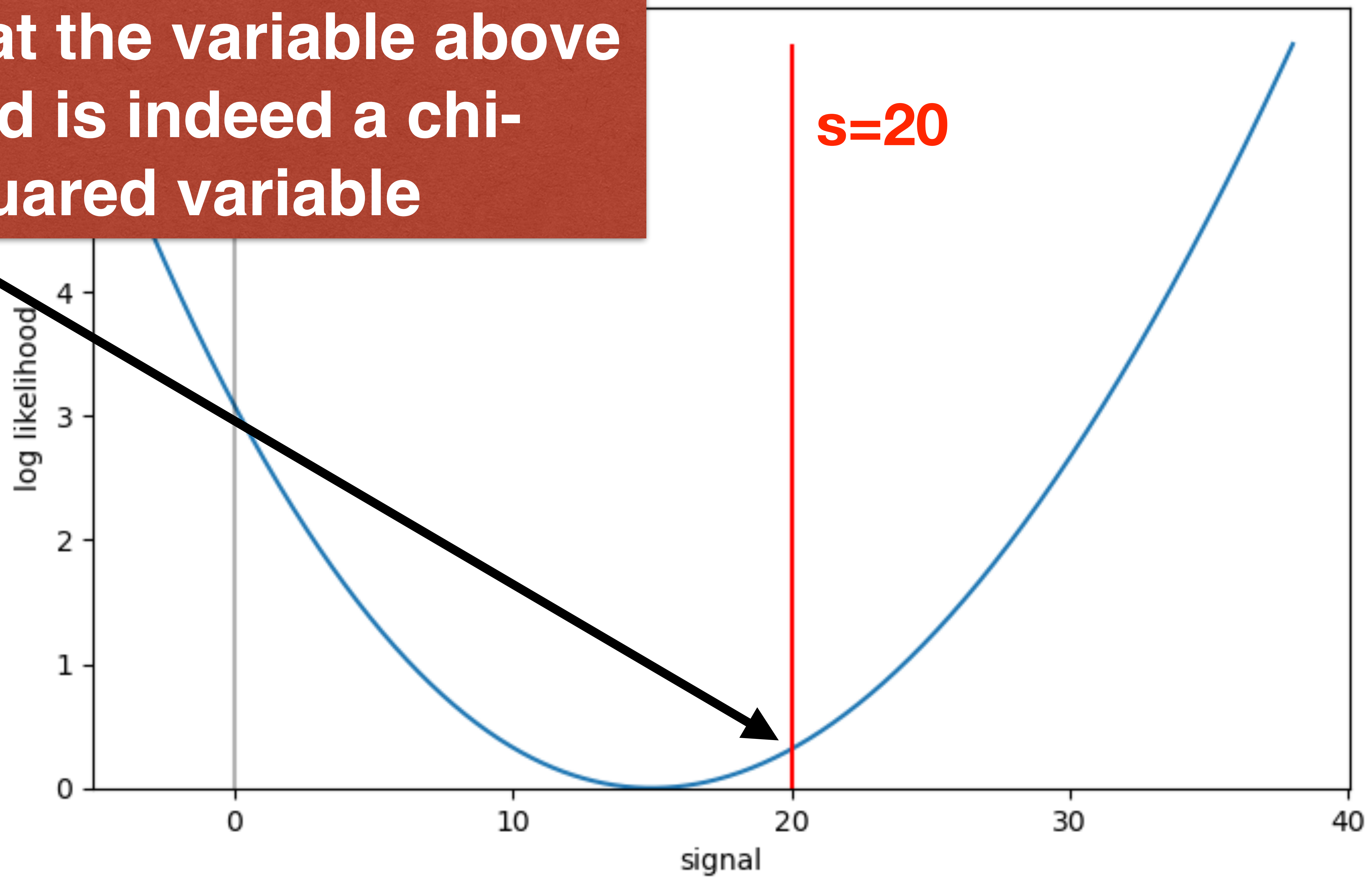
Our statistic is

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

Which is an expected value for a chi-squared variable



Check that the variable above defined is indeed a chi-squared variable

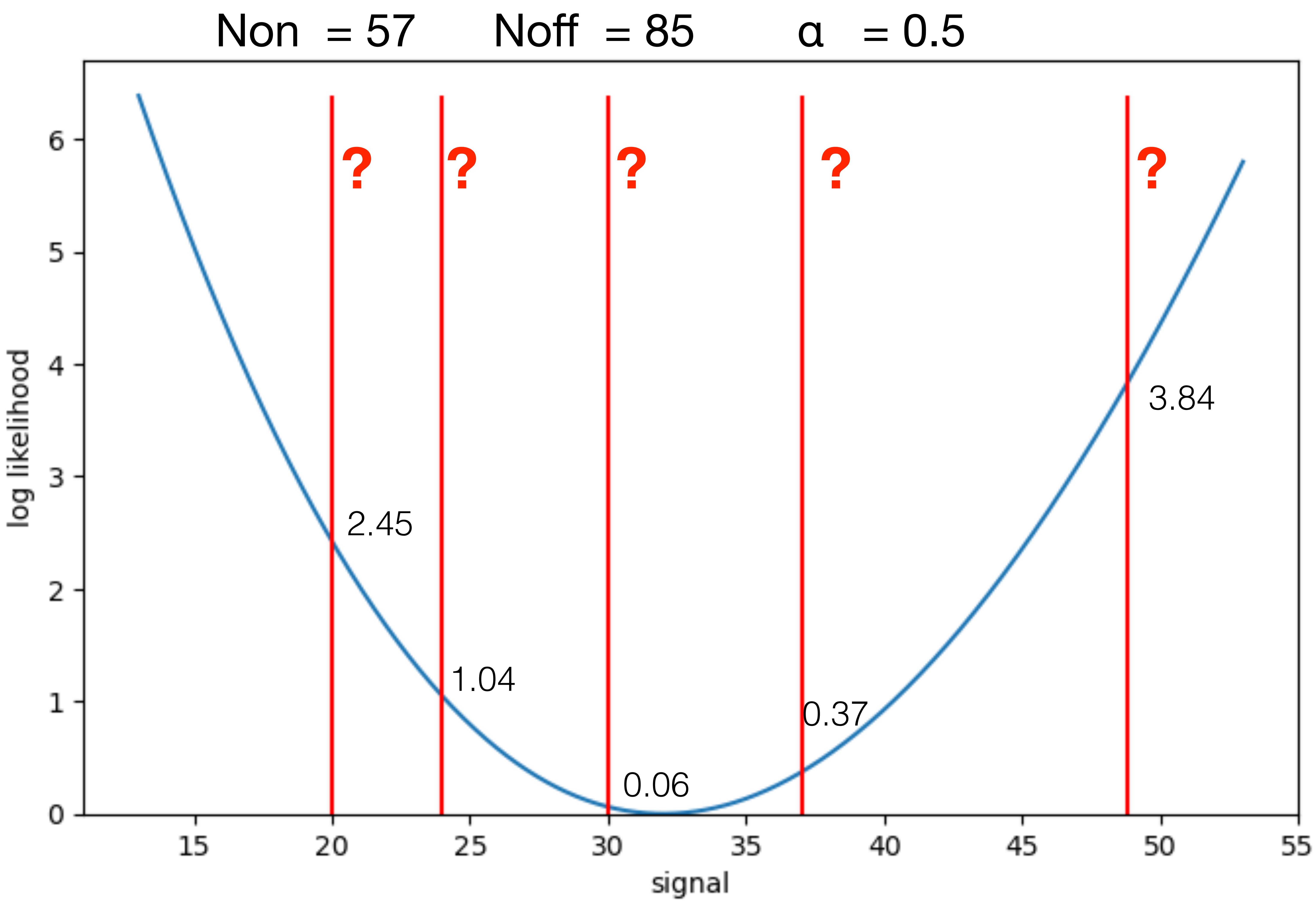


On/Off measurement

Signal estimation in the **frequentist approach:**

In a real data analysis, we do not know the true signal (this is what we want to estimate) but we only know the counts in the OFF and ON regions.

So what can we do?



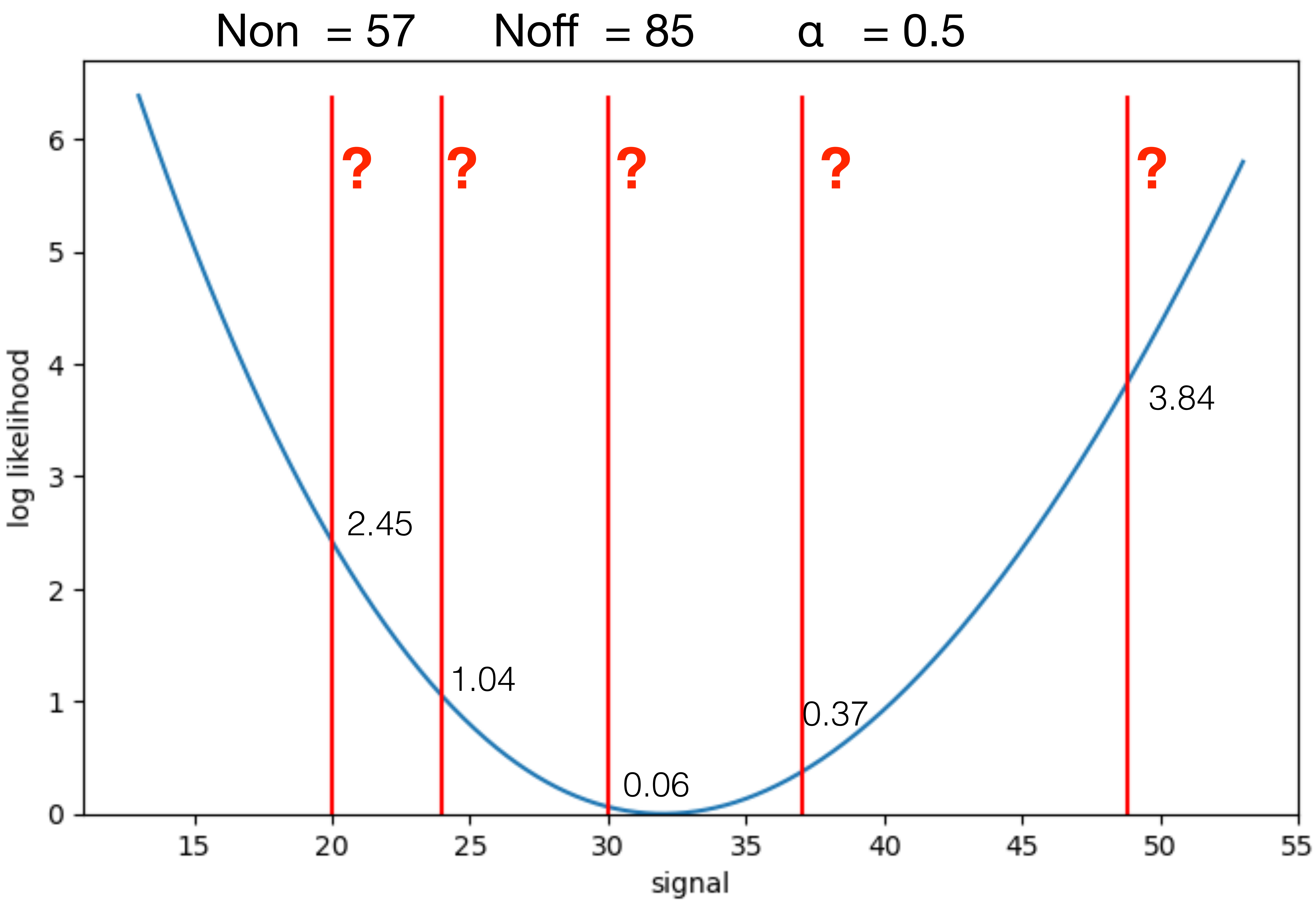
On/Off measurement

Signal estimation in the **frequentist approach:**

All these values should follow a chi-squared distribution assuming a given signal to be true.

For example, if the true signal was 20, we would have observed a value smaller than 2.45 88% of the time.

We can therefore claim that a value of 20 is **excluded with a 88% confidence level.**

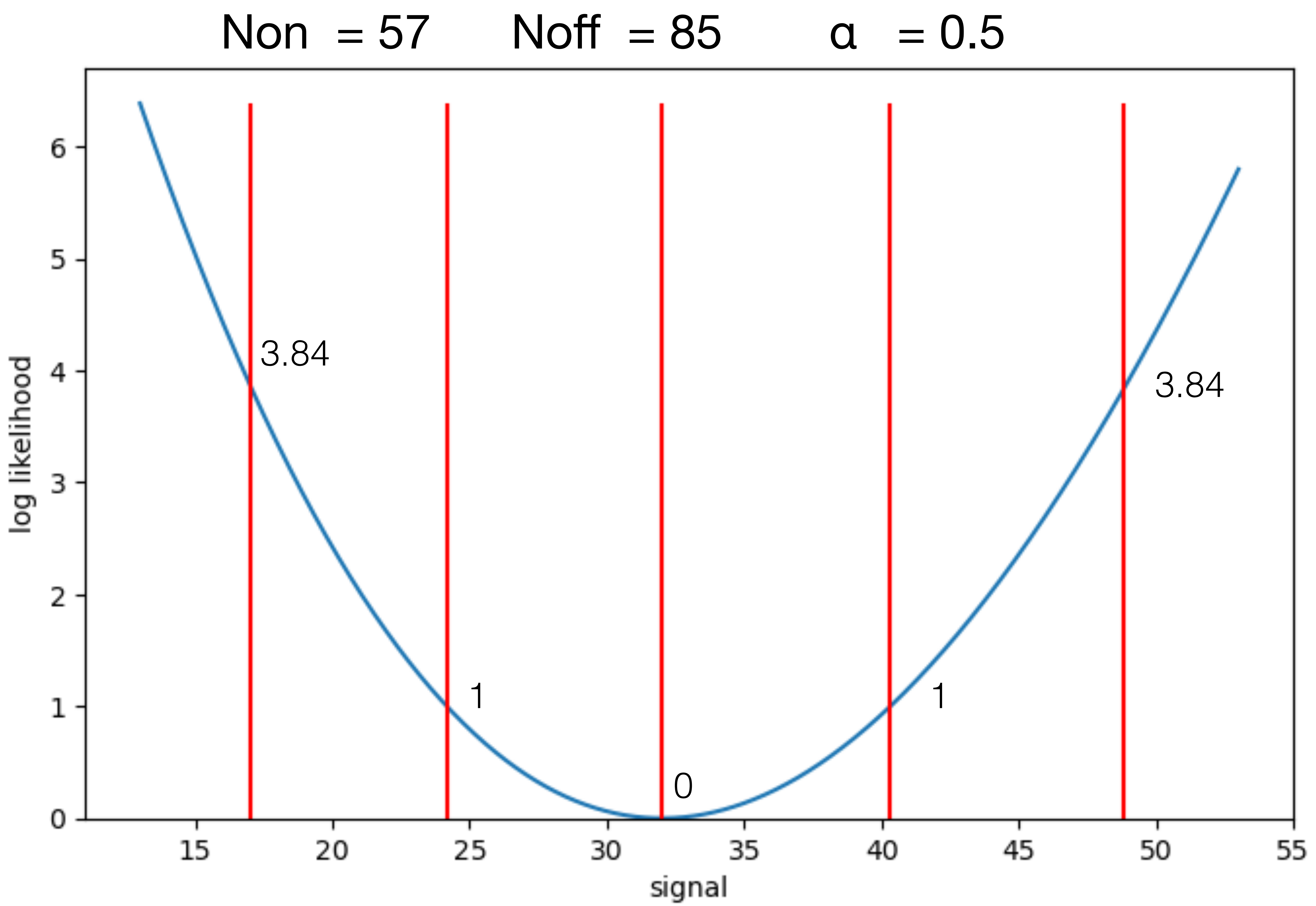


On/Off measurement

Signal estimation in the **frequentist approach:**

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**



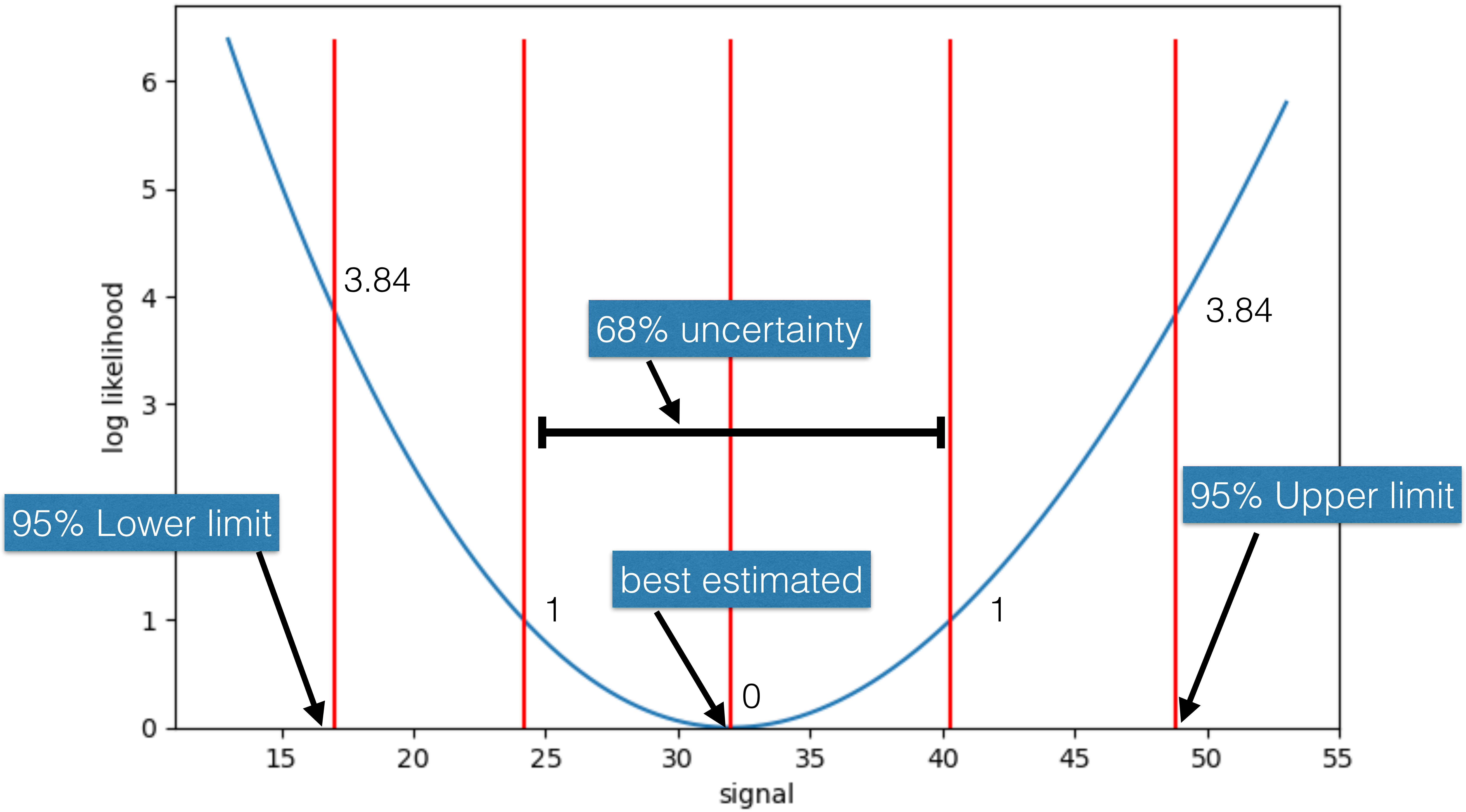
On/Off measurement

Signal estimation in the **frequentist approach:**

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**

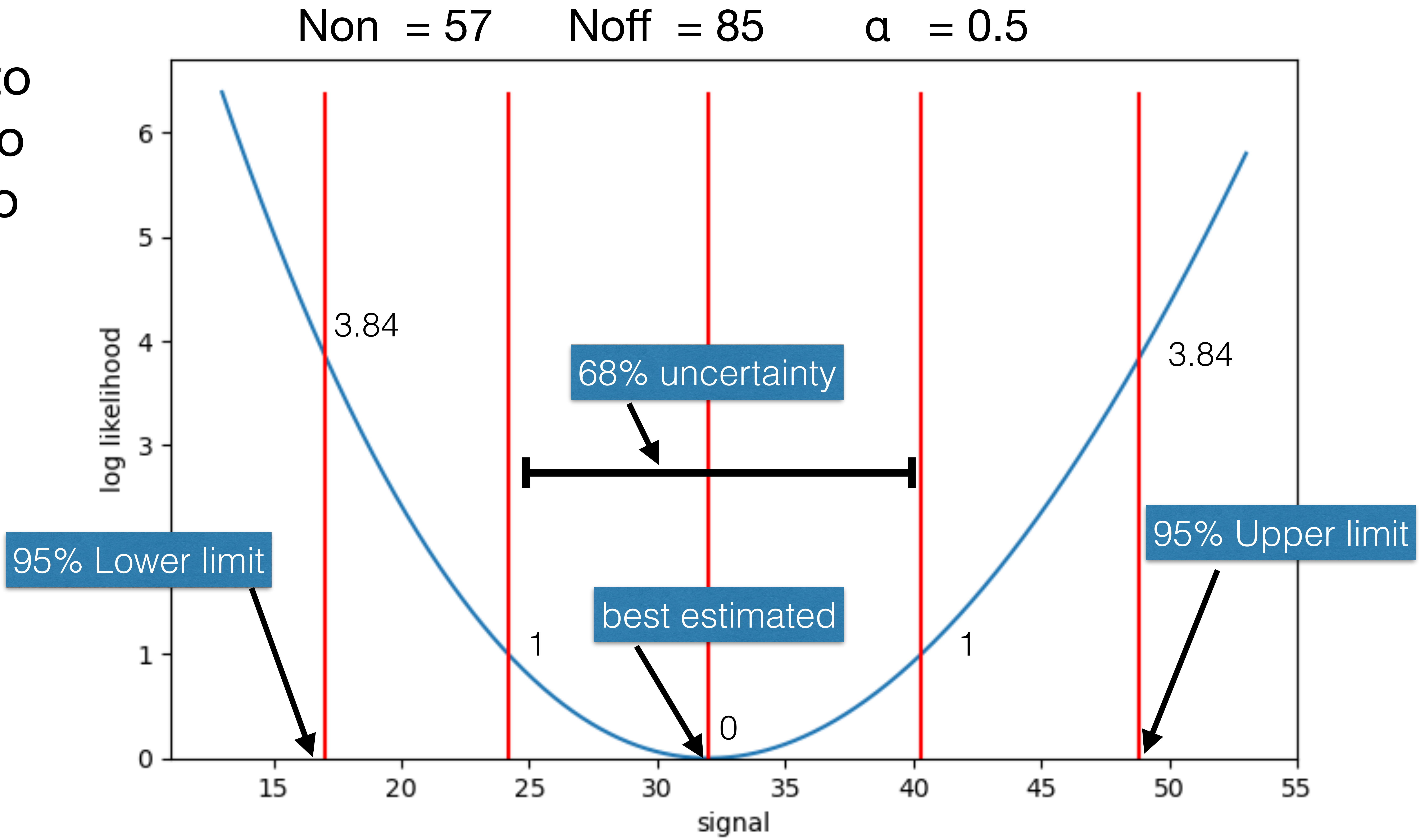
Non = 57 Noff = 85 α = 0.5



On/Off measurement

Signal estimation in the **frequentist approach:**

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

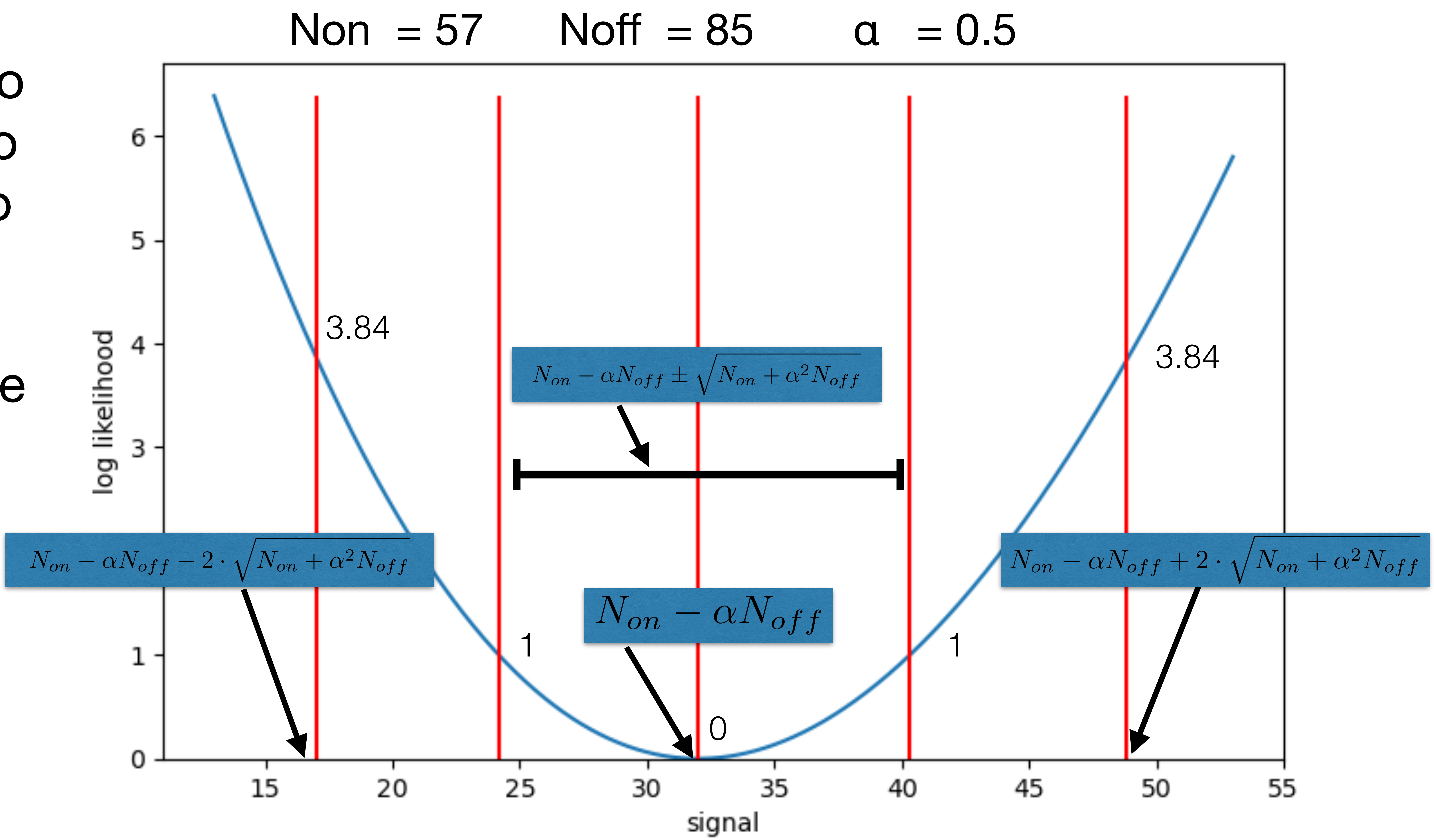


On/Off measurement

Signal estimation in the **frequentist approach:**

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



On/Off measurement

Signal estimation in the **frequentist approach:**

Non = 57 Noff = 85 α = 0.5

Non - α Noff = 32

sqrt(Non + α² Noff) = 8.06

The signal estimation is:

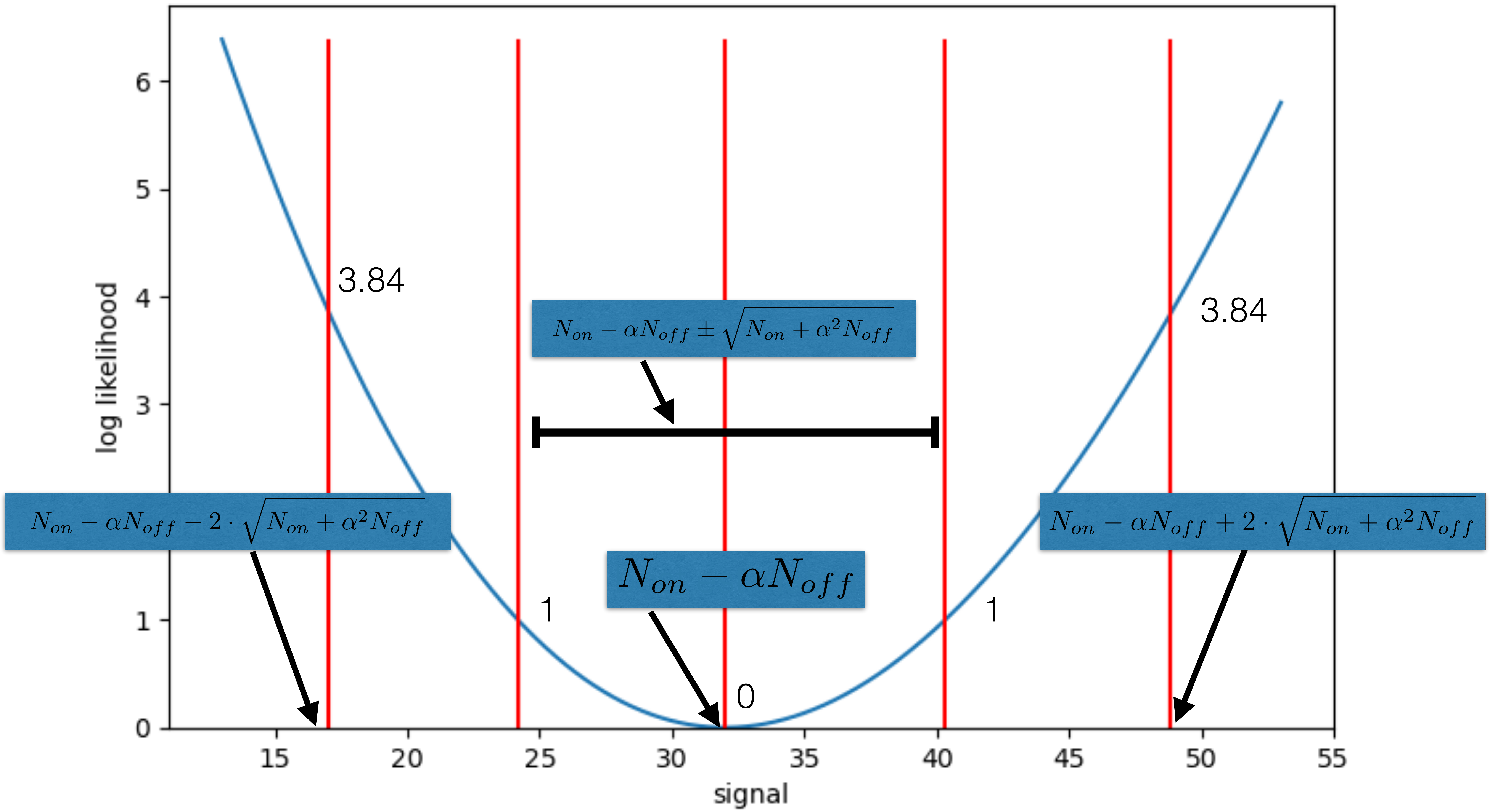
32 ± 8

with upper limit

48.1

and lower limit

15.9



On/Off measurement

Signal estimation in the **frequentist approach:**

Non = 57 Noff = 85 α = 0.5

Non - α Noff = 32

sqrt(Non + α² Noff) =

Check that these are indeed good approximations

↓

The signal estimation is:

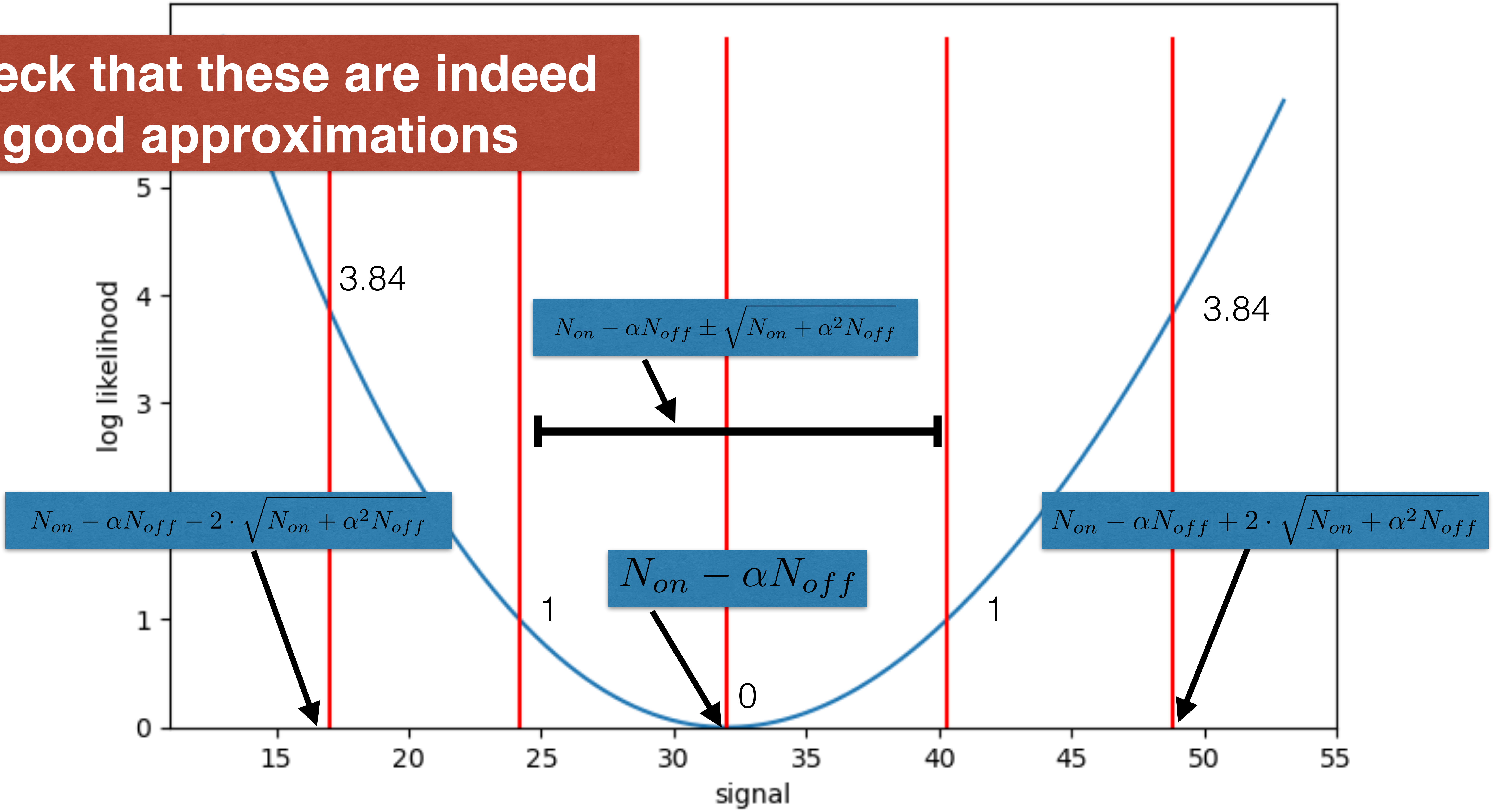
32 ± 8

with upper limit

48.1

and lower limit

15.9

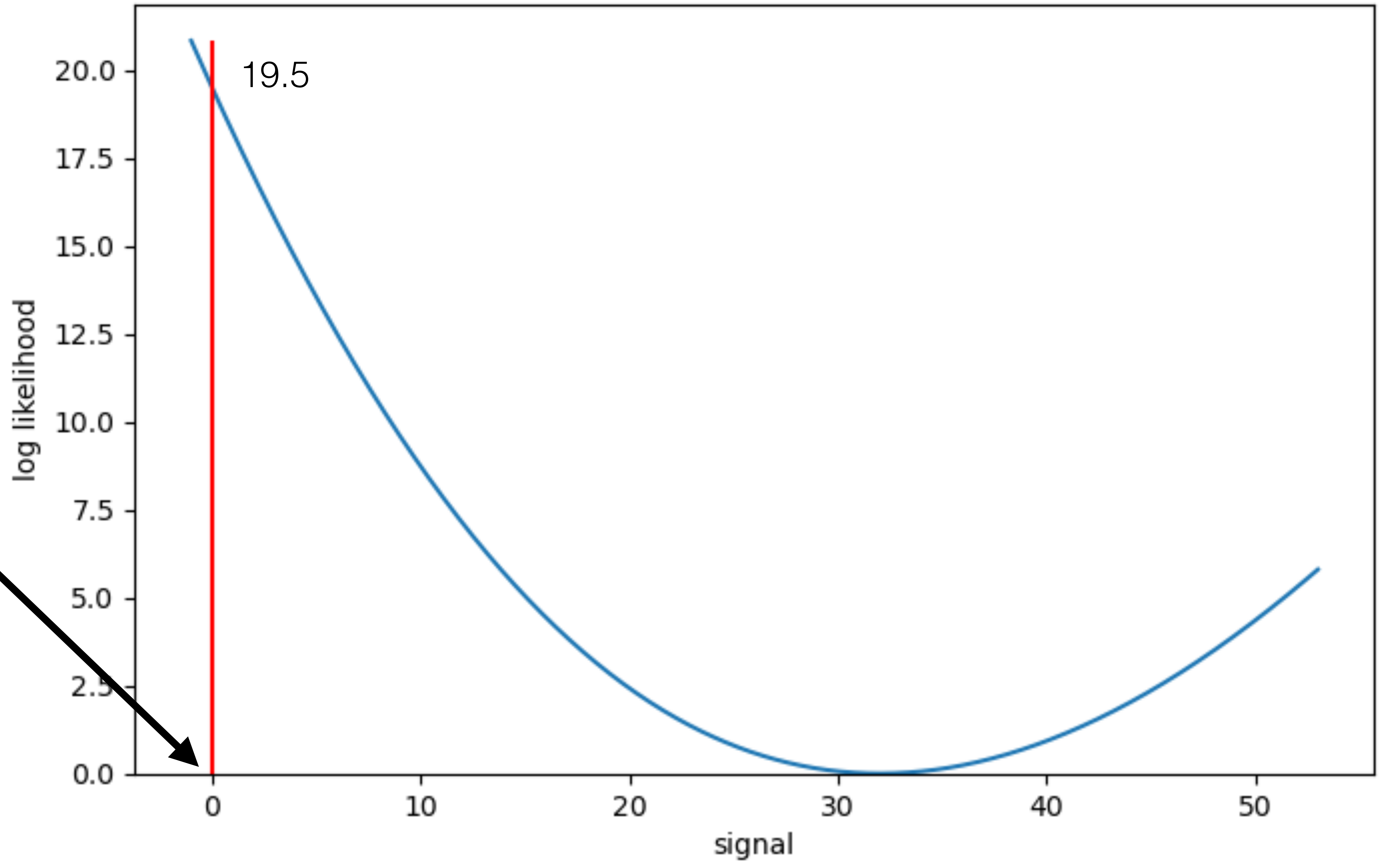


On/Off measurement

Signal estimation in the **frequentist approach:**

Non = 57 Noff = 85 α = 0.5

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**



On/Off measurement

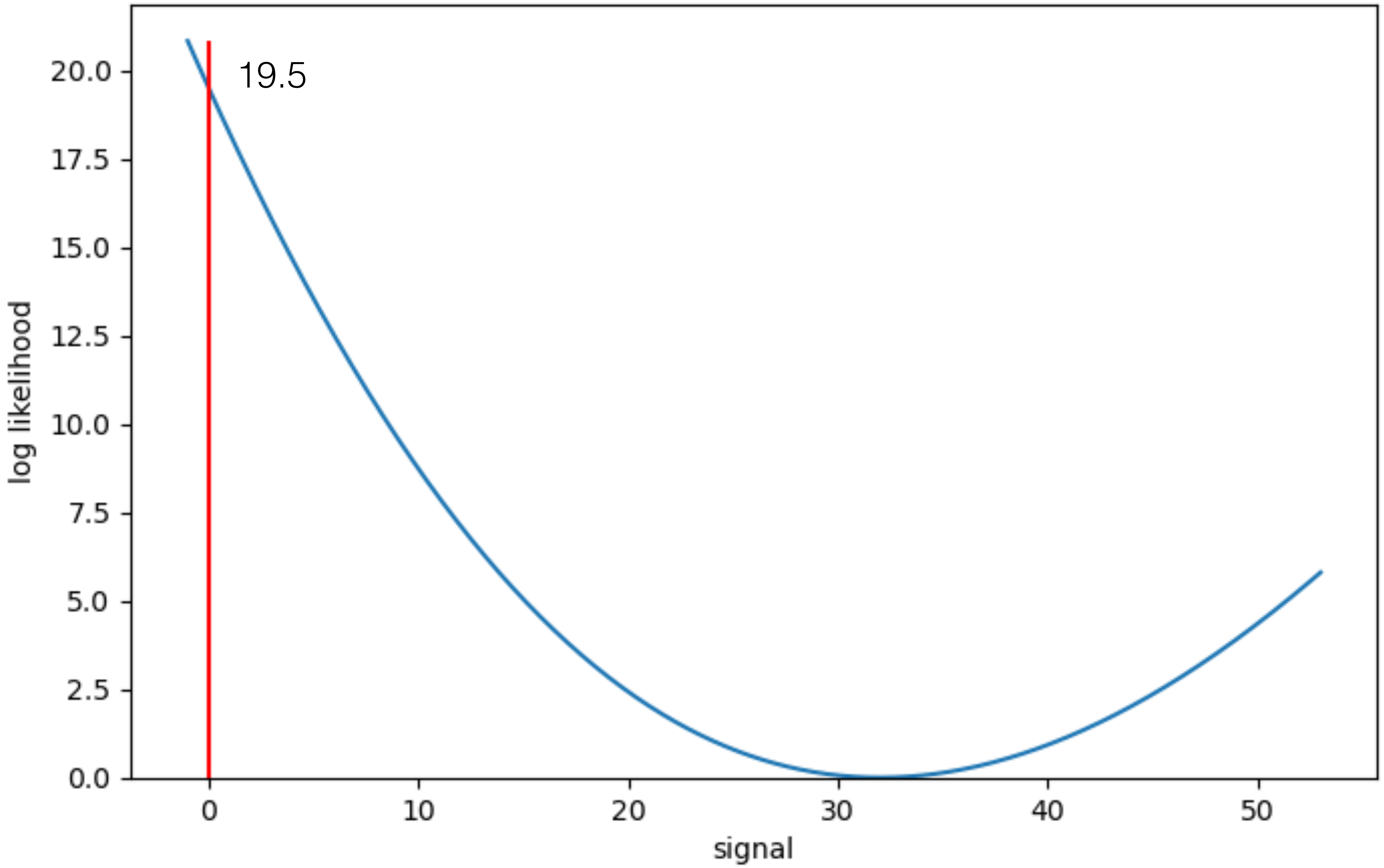
Signal estimation in the **frequentist approach:**

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**

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 … (?)

Non = 57 Noff = 85 α = 0.5



On/Off measurement

Signal estimation in the **frequentist approach:**

If you take the **log-likelihood** expression

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

put **s=0** and take the **square root** (in order to get a normal variable from a chi-squared one), you get the famous **“Li&Ma” Significance**

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

where the sign + or - is arbitrary chosen to be positive when the excess is positive

On/Off measurement

Signal estimation in the **frequentist approach:**

If you take the **log-lik**

$$-2\log\lambda(s) = 2\left[N_{on}\log\left(\right. \right. \quad \left. N_{off}\right]$$

Perform a simulation of On/Off counts with fixed 's'=0 and 'b' and get each time the Li&Ma significance.
Which distribution do you get from it?
What happens if 's' is not fixed to zero?

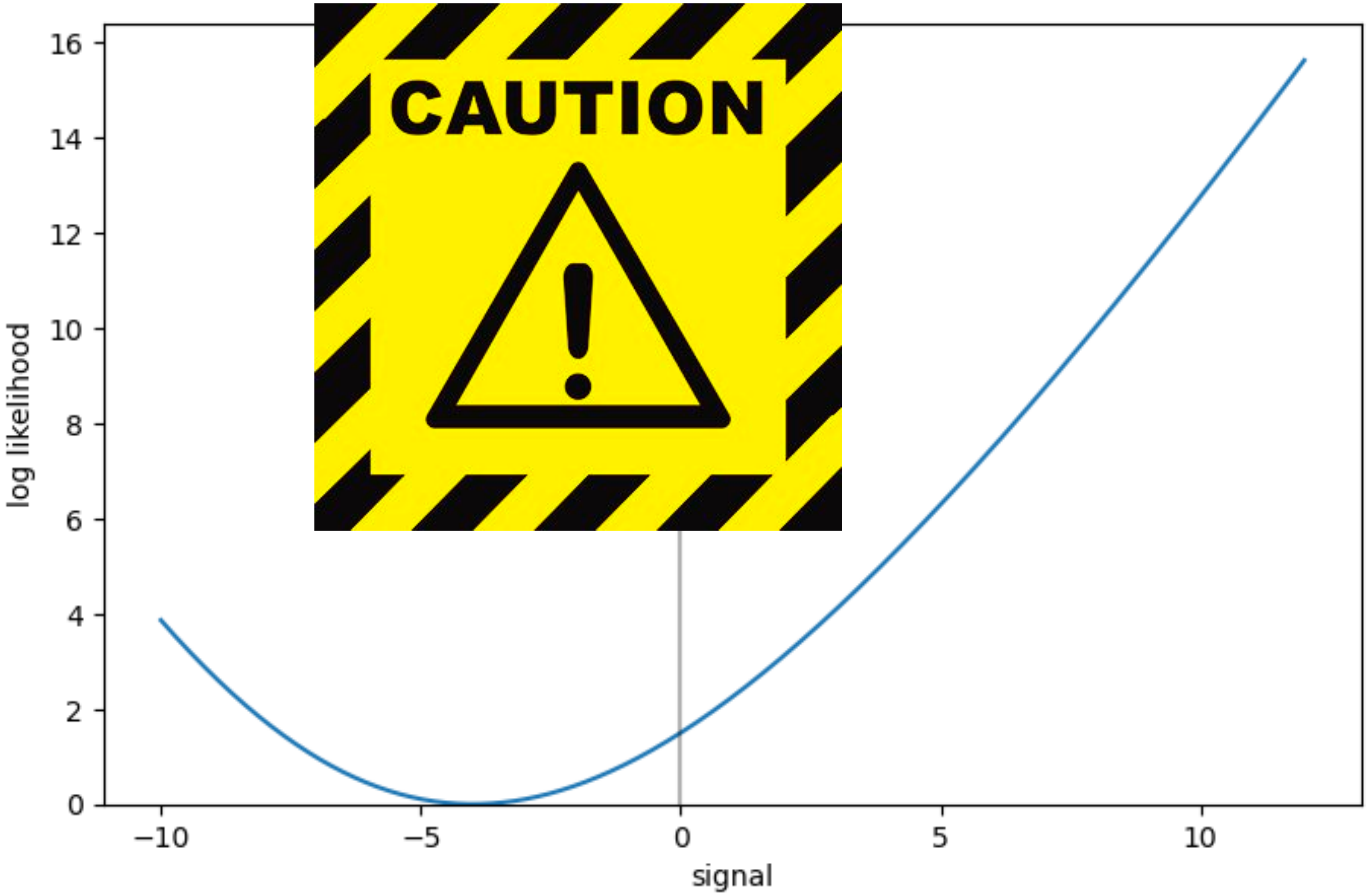
put **s=0** and take the **square root** (in order to get a normal variable from a chi-squared one), you get the famous **"Li&Ma" Significance**

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

where the sign + or - is arbitrary chosen to be positive when the excess is positive

On/Off measurement

Signal estimation in the **frequentist approach:**



For small or negative excess Wilks' theorem cannot be applied anymore.
This means that the value of 3.84 should not be used for putting 95% upper limit on the signal

What can we do then?

In these cases 'ad-hoc' **adjustments** are required, the most famous one being the **Rolke et al. method**

On/Off measurement

Signal estimation in the **frequentist approach:**

ELSEVIER

Nuclear Instruments and Methods in Physics Research A 458 (2001) 745–758

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A

www.elsevier.nl/locate/nima

Confidence intervals and upper bounds for small signals in the presence of background noise☆

Wolfgang A. Rolke[a],*, Angel M. López[b]

[a]*Department of Mathematics, University of Puerto Rico-Mayagüez, Mayagüez, P.O. Box 5959, PR 00681, USA*
[b]*Department of Physics, University of Puerto Rico-Mayagüez, Mayagüez, PR 00681, USA*

Received 6 April 2000; received in revised form 17 July 2000; accepted 22 July 2000

Abstract

We discuss a new method for setting limits on small signals in the presence of background noise. The method is based on a combination of a two-dimensional confidence region and the large sample approximation to the likelihood ratio test statistic. It automatically quotes upper limits for small signals and two-sided confidence intervals for larger samples. We show that this method gives the correct coverage and also has good power.

Keywords: Maximum likelihood; Profile likelihood; Confidence regions; Coverage; Monte Carlo; Sensitivity

Table 1
90% CI for Poisson signal μ with $\tau = 1$[a]

x	y = 0	1	2	3	4
0	0,2.21	0,1.71	0,0.86	0,NA	0,NA
1	0,3.65	0,3.62	0,2.71	0,1.98	0,1.3
2	0.03,5.3	0,4.52	0,4.19	0,3.41	0,2.74
3	0.73,6.81	0,6.02	0,5.22	0,4.81	0,4.17
4	1.42,8.25	0,7.44	0,6.63	0,5.82	0,5.57
5	2.11,9.63	0.16,8.81	0,7.99	0,7.17	0,6.35
6	2.81,10.98	0.93,10.15	0,9.32	0,8.49	0,7.66
7	3.5,12.3	1.7,11.46	0.22,10.63	0,9.79	0,8.96
8	4.2,13.59	2.49,12.76	1.02,11.92	0,11.07	0,10.23
9	4.92,14.87	3.27,14.03	1.82,13.19	0.45,12.34	0,11.49
10	5.66,16.14	4.06,15.29	2.63,14.44	1.27,13.59	0,12.74
11	6.41,17.39	4.85,16.54	3.44,15.69	2.09,14.83	0.78,13.97
12	7.17,18.63	5.65,17.78	4.26,16.92	2.92,16.06	1.61,15.2
13	7.94,19.86	6.45,19	5.08,18.14	3.75,17.28	2.45,16.42
14	8.71,21.09	7.26,20.23	5.9,19.36	4.58,18.5	3.29,17.63
15	9.5,22.3	8.07,21.44	6.72,20.57	5.41,19.7	4.13,18.83
16	10.29,23.51	8.89,22.64	7.55,21.77	6.25,20.9	4.97,20.03
17	11.09,24.71	9.71,23.84	8.38,22.97	7.09,22.1	5.82,21.22
18	11.89,25.91	10.53,25.03	9.21,24.16	7.93,23.29	6.67,22.41
19	12.7,27.1	11.35,26.22	10.05,25.35	8.77,24.47	7.52,23.59
20	13.52,28.28	12.18,27.41	10.89,26.53	9.62,25.65	8.37,24.77

x	y = 5	6	7	8	9
0	0,NA	0,NA	0,NA	0,NA	0,NA
1	0,0.59	0,NA	0,NA	0,NA	0,NA
2	0,2.12	0,1.36	0,0.41	0,NA	0,NA
3	0,3.52	0,2.66	0,1.9	0,1.05	0,0.16
4	0,4.86	0,4.02	0,3.26	0,2.39	0,1.5
5	0,6.32	0,5.39	0,4.67	0,3.72	0,2.86
6	0,6.83	0,6.8	0,6.02	0,4.97	0,4.17
7	0,8.12	0,7.27	0,7.27	0,6.32	0,5.54
8	0,9.39	0,8.54	0,7.69	0,7.68	0,6.83
9	0,10.64	0,9.79	0,8.94	0,8.08	0,8.05
10	0,11.88	0,11.03	0,10.17	0,9.31	0,8.45
11	0,13.12	0,12.26	0,11.4	0,10.53	0,9.67
12	0.33,14.34	0,13.48	0,12.61	0,11.75	0,10.88
13	1.17,15.55	0,14.69	0,13.82	0,12.95	0,12.08
14	2.02,16.76	0.76,15.89	0,15.02	0,14.15	0,13.28
15	2.86,17.96	1.62,17.09	0.38,16.22	0,15.35	0,14.47
16	3.71,19.16	2.47,18.28	1.24,17.41	0.02,16.53	0,15.66
17	4.57,20.35	3.33,19.47	2.1,18.59	0.88,17.72	0,16.84
18	5.42,21.53	4.18,20.65	2.96,19.77	1.75,18.9	0.54,18.01
19	6.27,22.71	5.04,21.83	3.83,20.95	2.62,20.07	1.41,19.19
20	7.13,23.89	5.91,23.0	4.69,22.12	3.48,21.24	2.28,20.35

[a] y is the number of events observed in the background region and x is the number of events observed in the signal region. In this case, the estimated background rate would be $b = y$.

In these cases 'ad-hoc' **adjustments** are required, the most famous one being the **Rolke et al. method**

On/Off measurement

Signal estimation in the **bayesian approach:**

Likelihood:

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

↓

Bayes theorem with uniform priors:

$$p(s \mid N_{on}, N_{off}; \alpha) = \frac{\int db\, p(N_{on}, N_{off} \mid s, b; \alpha) p(b)\, p(s)}{\int ds\, db\, p(N_{on}, N_{off}, s, b; \alpha)} \propto \int db\, p(N_{on}, N_{off} \mid s, b; \alpha)$$

↓

PDF of the signal rate

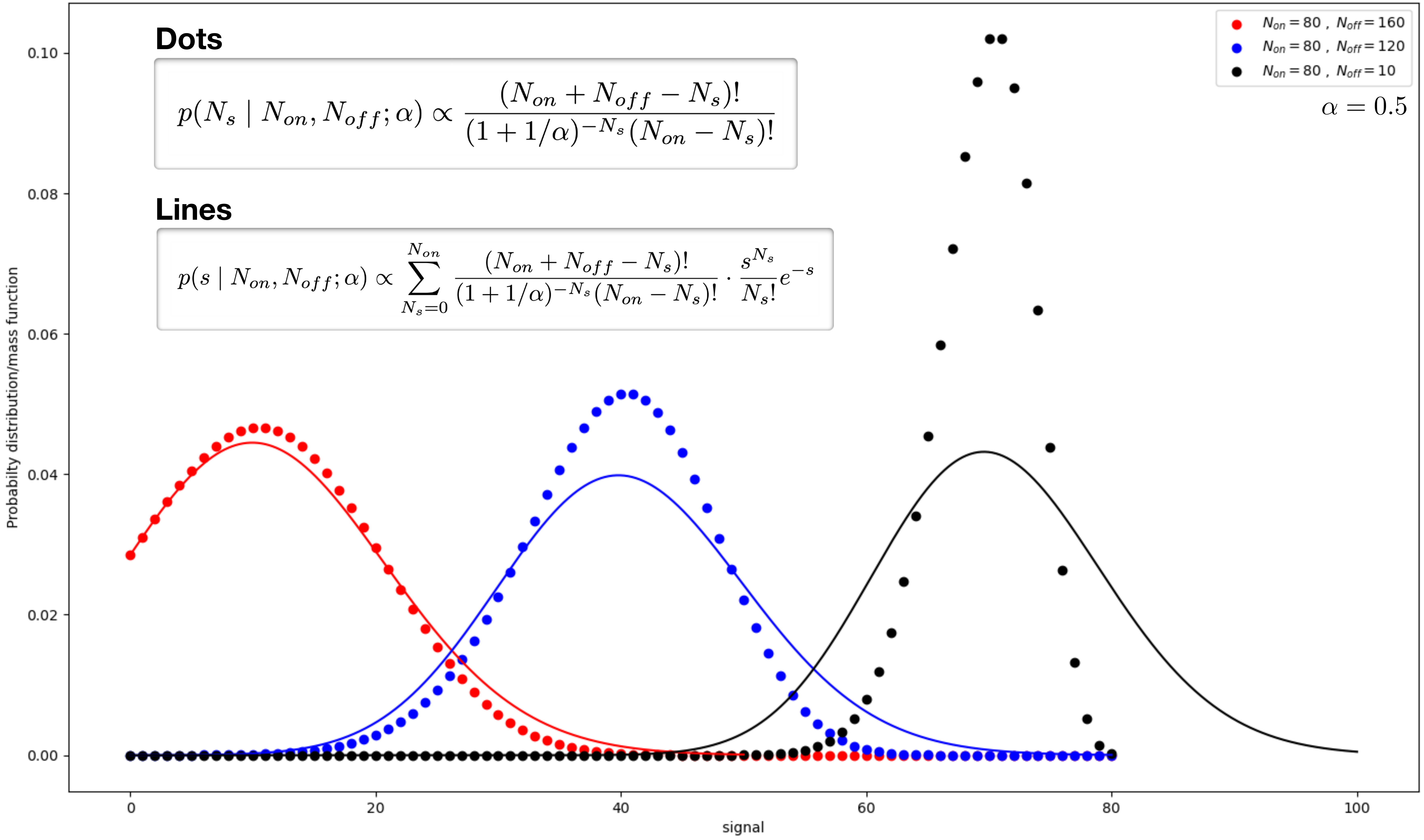
$$p(s \mid N_{on}, N_{off}; \alpha) \propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on} + N_{off} - N_s)!}{(1+1/\alpha)^{-N_s} (N_{on} - N_s)!} \cdot \frac{s^{N_s}}{N_s!} e^{-s}$$

PMS of the number of signal event

$$p(N_s \mid N_{on}, N_{off}; \alpha) \propto \frac{(N_{on} + N_{off} - N_s)!}{(1+1/\alpha)^{-N_s} (N_{on} - N_s)!}$$

On/Off measurement

Signal estimation in the **bayesian approach:**



On/Off measurement

Signal estimation in the **bayesian approach:**

Contrary to the frequentist approach, where one has to maximize the likelihood and apply the Wilks' theorem, in the Bayesian approach all the information on 's' (the signal rate) is included in its PDF.

The **best estimate** of 's' will be given by the most probable value, while the **68% credible interval** [s1, s2] is such that

$$\int_{s_1}^{s_2} p(s \mid N_{on}, N_{off}; \alpha)\, ds = 0.68$$

The **upper limit** is given straightforwardly by the value s_{UL} such that

$$\int_{s_{UL}}^{\infty} p(s \mid N_{on}, N_{off}; \alpha)\, ds = 0.05.$$

On/Off measurement

Signal estimation in the **bayesian approach:**

Contrary to the frequentist approach, where one has to maximize the likelihood and apply the Wilks' theorem, in the Bayesian approach all the information on 's' (the signal rate) is included in its PDF.

The **best estimate** of 's' will be given by the most probable value, while the **68% credible interval** [s1, s2] is such that

$$\int_{s_1}^{s_2} p(s \mid N_{on}, N_{off}; \alpha)\, ds = 0.68$$

Try to get the best estimate 's', with credible intervals and upper limits from the 3 examples shown in the previous slide

The **upper limit** is given straightforwardly by the value s_{UL} such that

$$\int_{s_{UL}}^{\infty} p(s \mid N_{on}, N_{off}; \alpha)\, ds = 0.05.$$

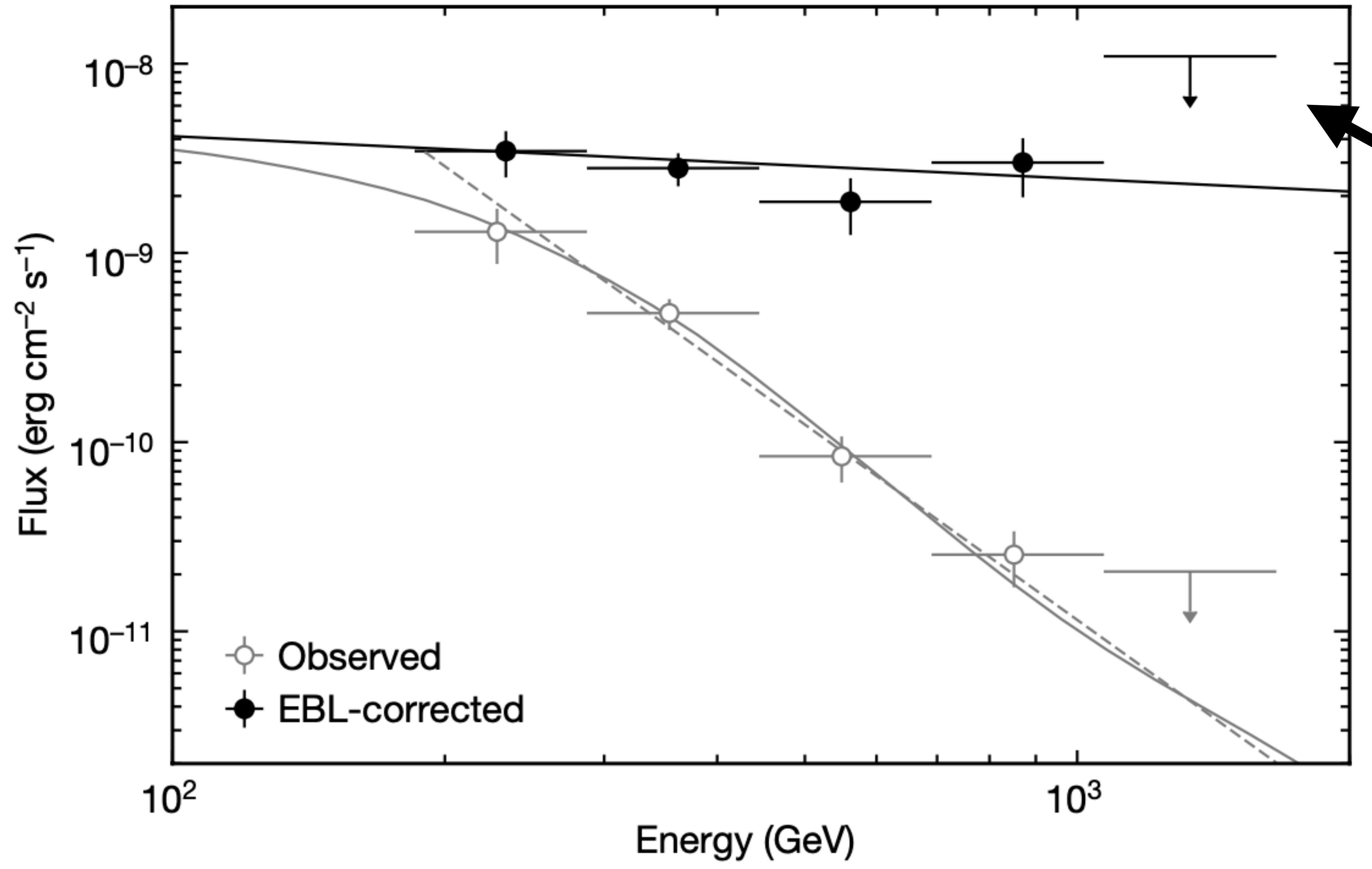
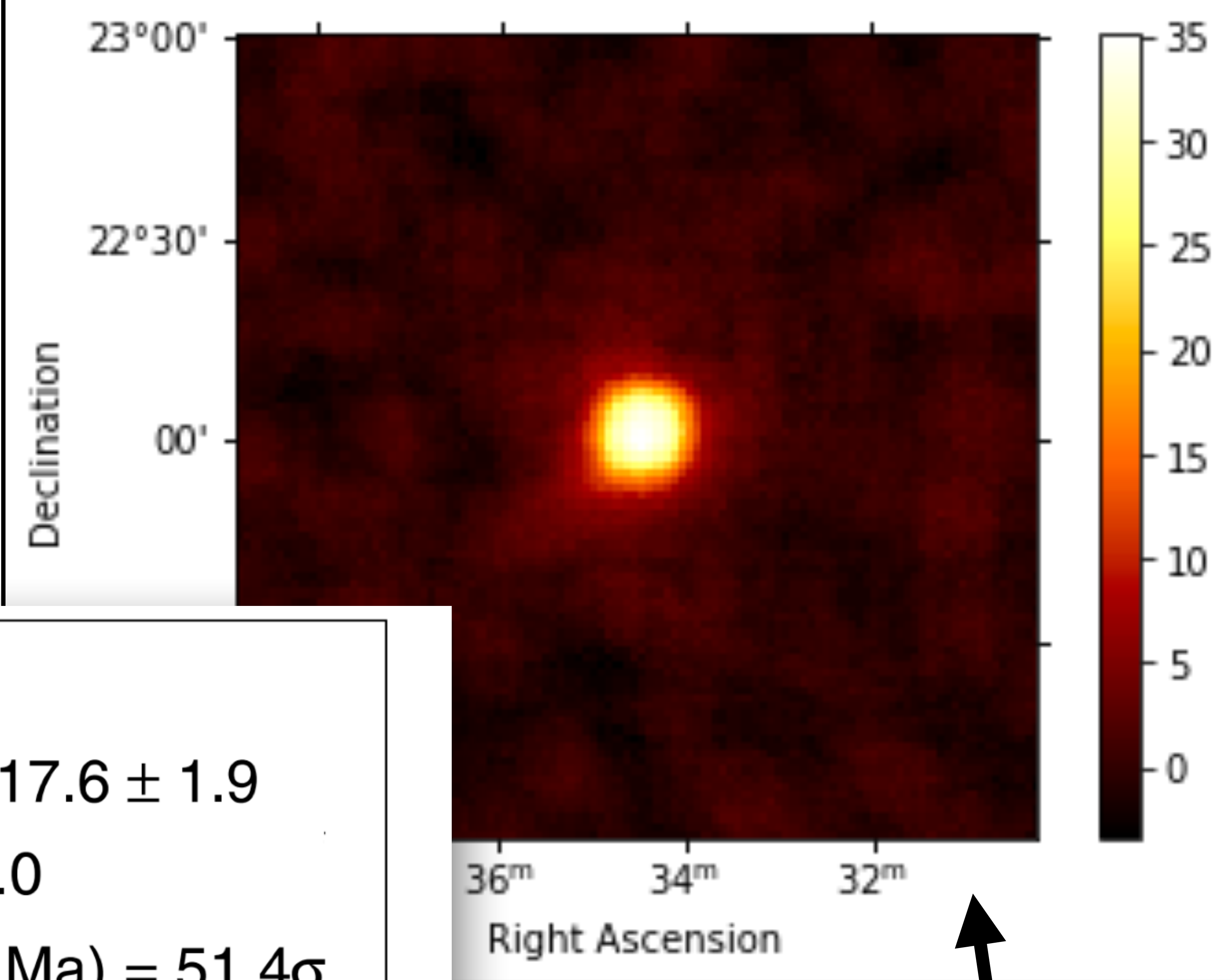
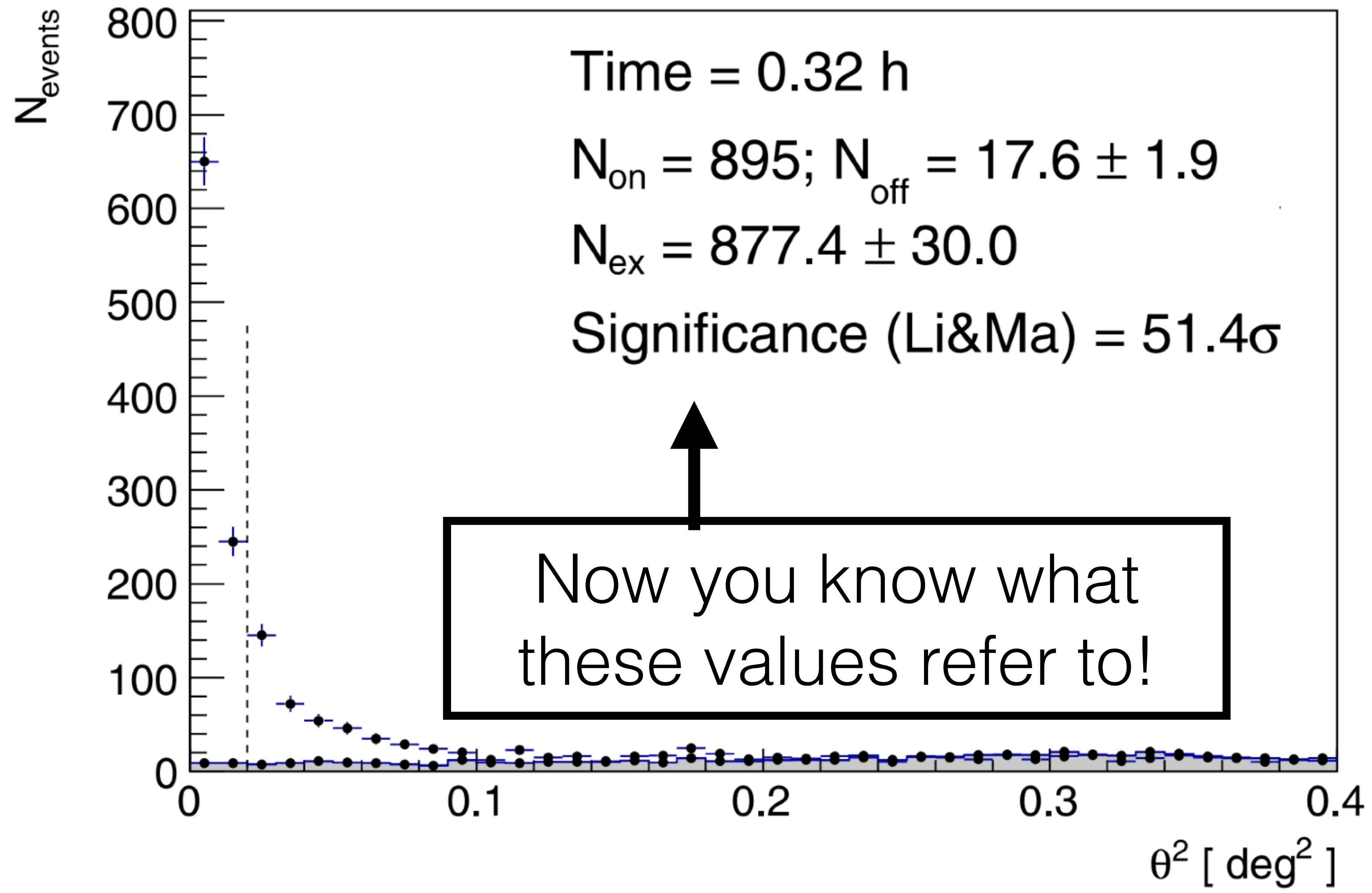


Fig. 2 | Spectrum above 0.2 TeV averaged over the period between $T_0 + 62$ s and $T_0 + 2{,}454$ s for GRB 190114C. Spectral-energy distributions for the spectrum observed by MAGIC (grey open circles) and the intrinsic spectrum corrected for EBL attenuation[25] (blue filled circles). The errors on the 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!





Now you know what these values refer to!

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].

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**
4. The **bayesian** gives us a **probability distribution** for the excess of gamma-ray

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