Introduction to Statistical Inference

Giacomo D'Amico

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 /ɪnˈfɜː(r)/; NAmE /ɪnˈfɜːr/

+ 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

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

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

- 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

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

$$\underbrace{f(x|y, I)}_{\textbf{Posterior}} = \frac{\overbrace{f(y|x, I)}^{\textbf{Likelihood}}\overbrace{f(x|I)}^{\textbf{Prior}}}{\underbrace{\int f(y|x, I)f(x|I)dx}_{\textbf{Normalisation}}}$$

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

The Monty Hall problem

1 2 3

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

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

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

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

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

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

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

The Monty Hall problem

1

2

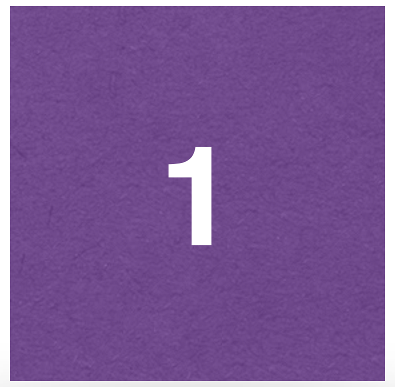
$$f(H_1|E,I) = \frac{f(E|H_1,I)f(H_1|I)}{f(E|I)} = \frac{\cdot 1/3}{1/2}$$

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

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

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

The Monty Hall problem



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

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

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

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

The Monty Hall problem

1

2

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

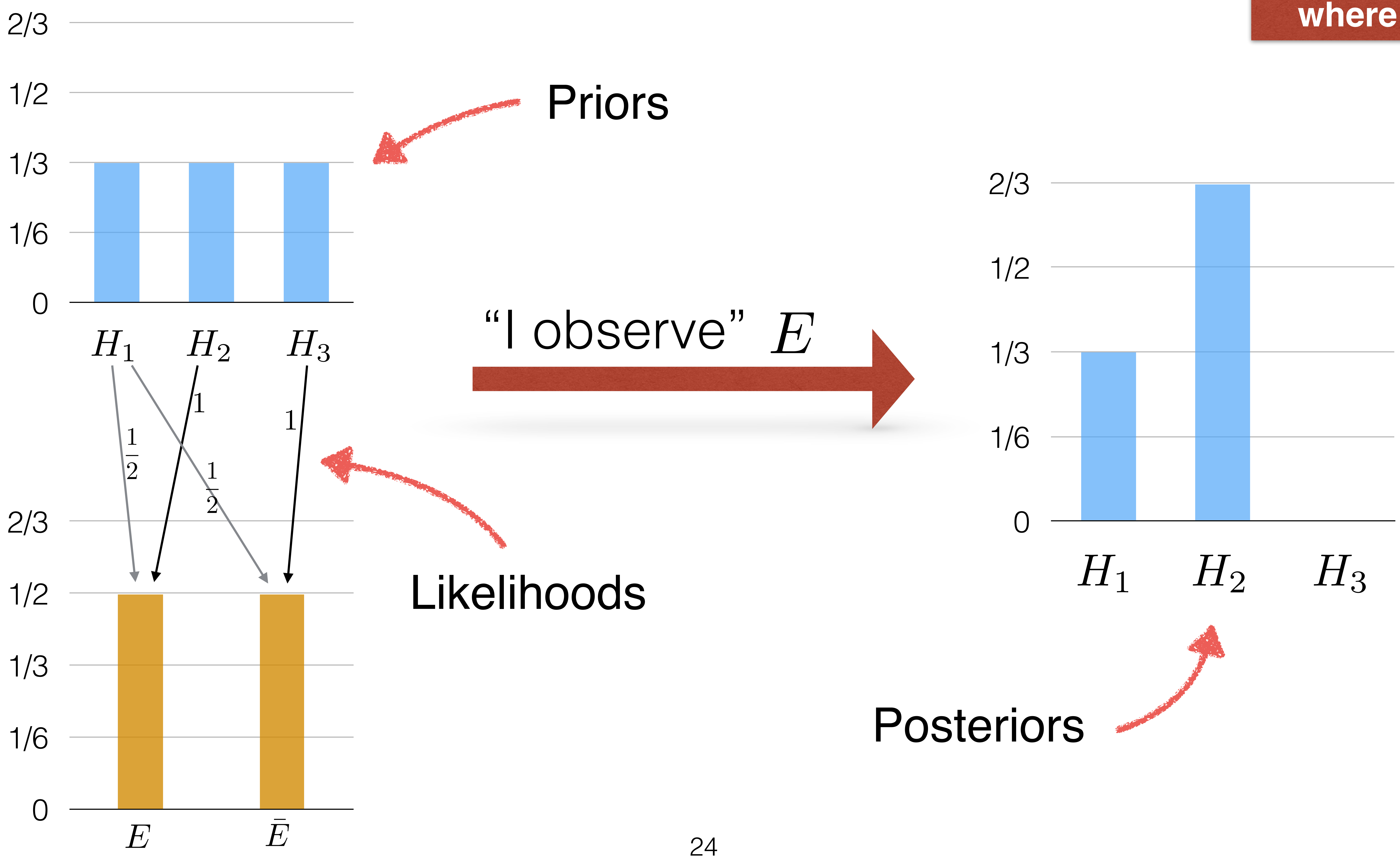
$$f(H_2|E,I) = \frac{f(E|H_2,I)f(H_2|I)}{f(E|I)} = \frac{1 \ \cdot 1/3}{1/2} = \frac{2}{3}$$

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

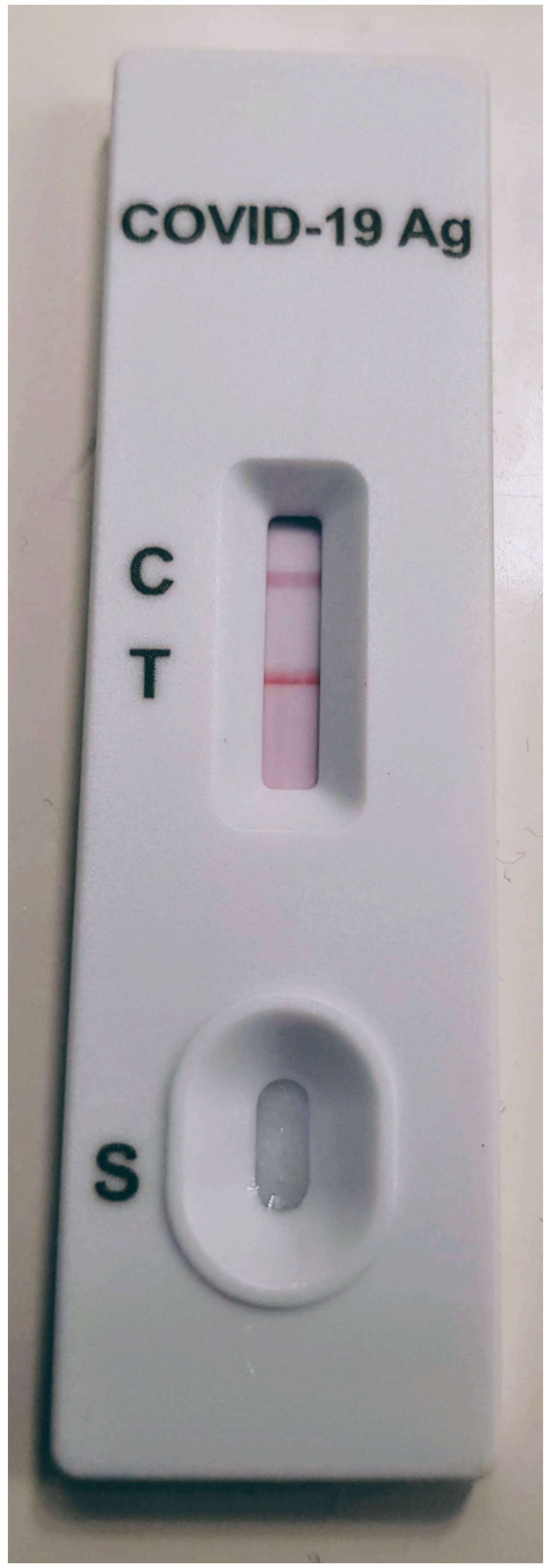
If we want to win the car, we should change the box!

The Monty Hall problem

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



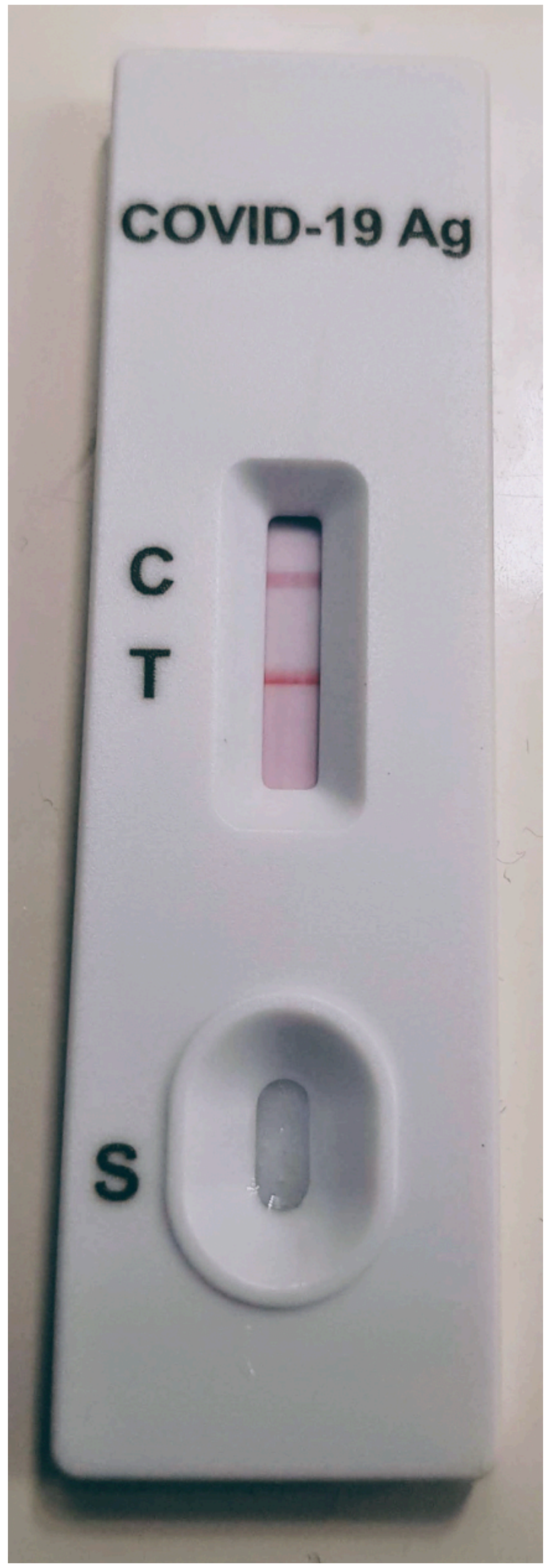
Another example: Covid-19 test



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) ?

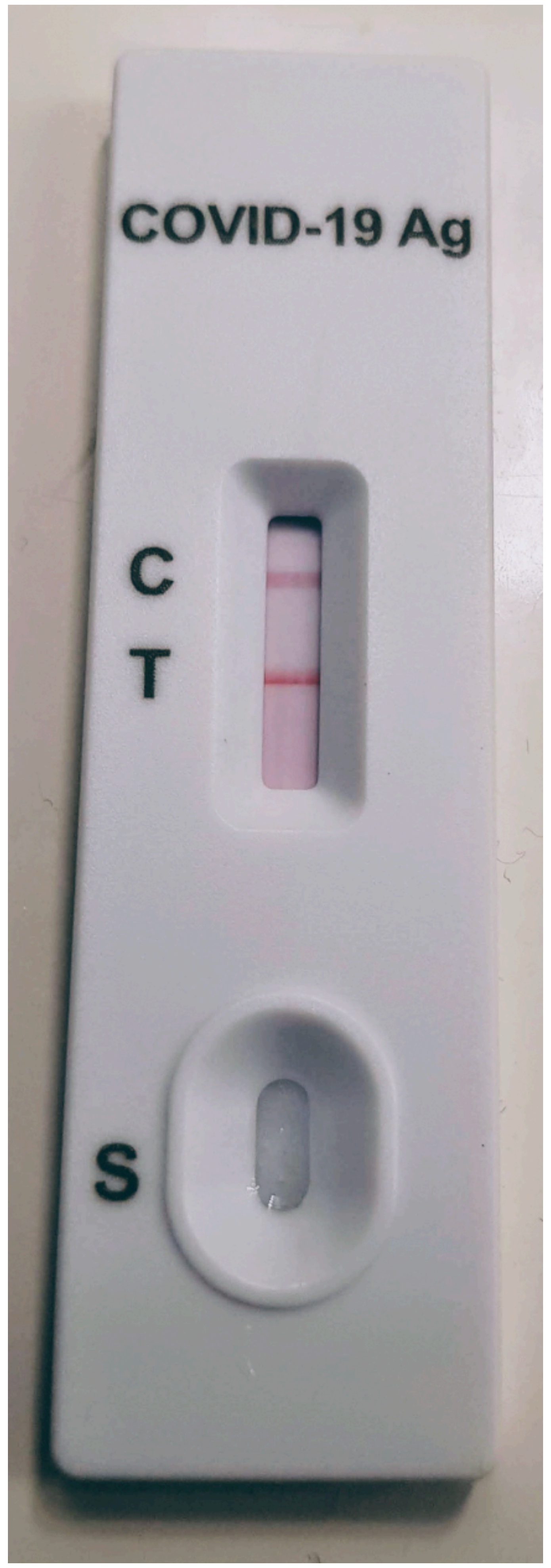
$$p(S\,|\,+) = \frac{P(+\,|\,S)P(S)}{\underbrace{P(+\,|\,S)}_{\text{Probability of True positive}}P(S) + \underbrace{P(+\,|\,\bar{S})}_{\text{Probability of False positive}}P(\bar{S})}$$

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

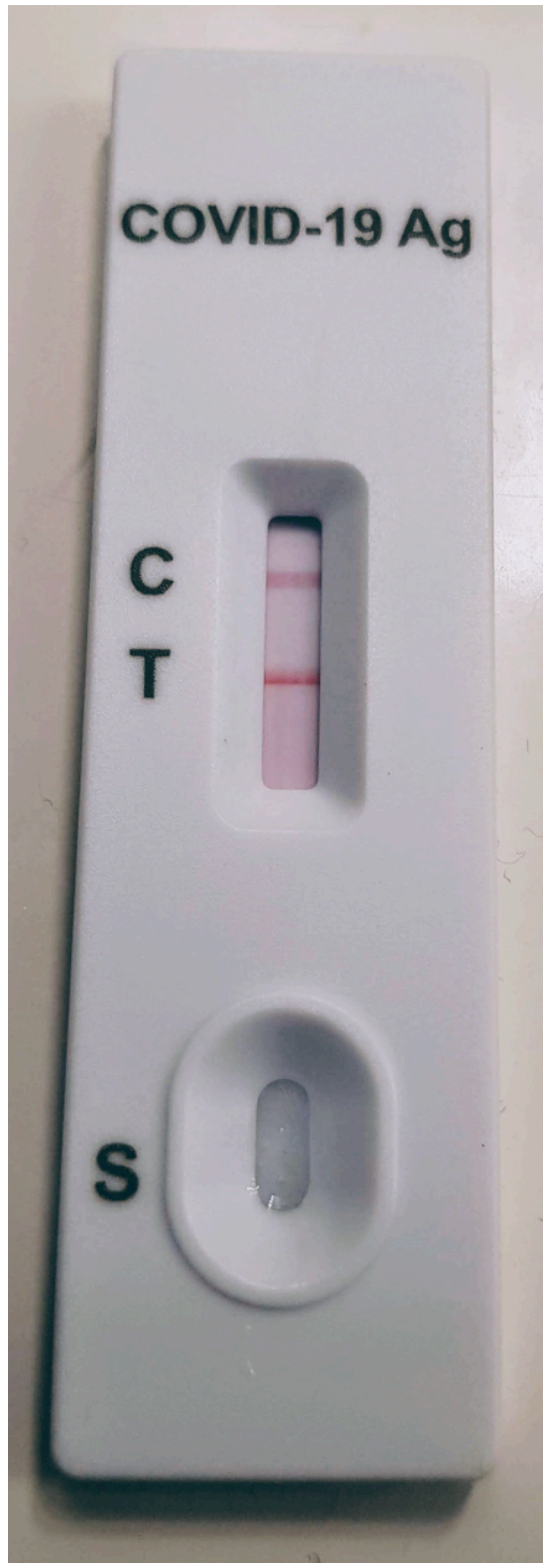
Probability of True positive

Probability of False positive

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

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

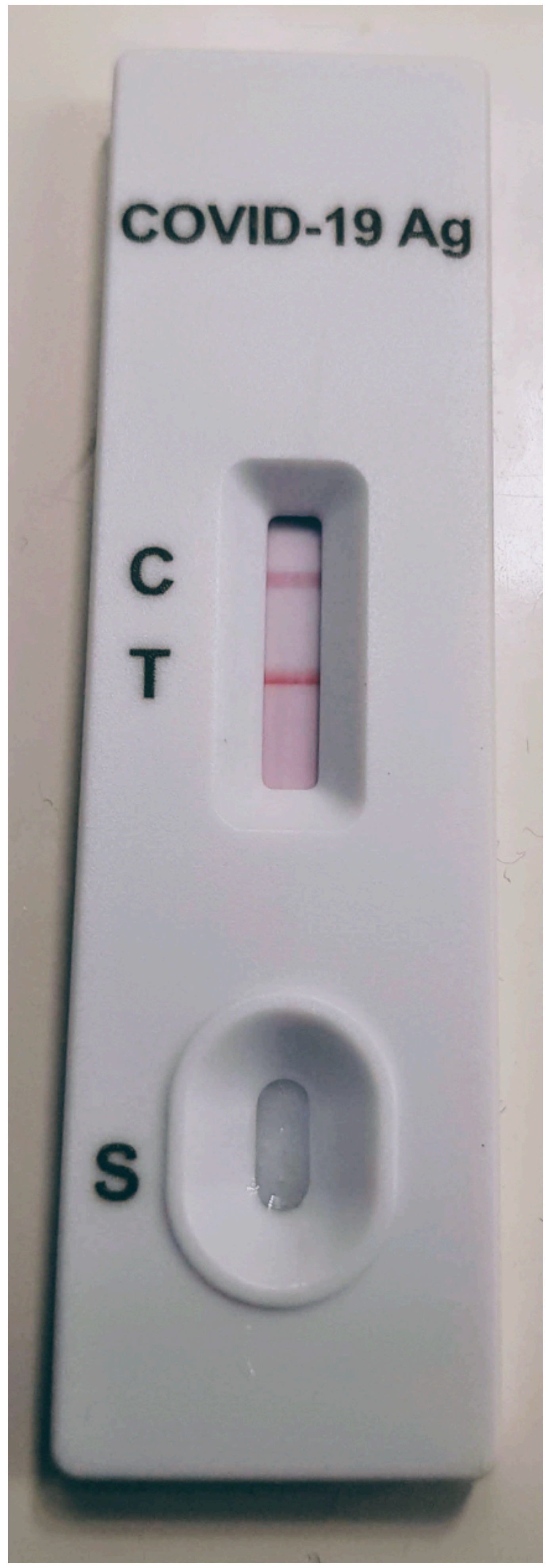
Another example: Covid-19 test



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

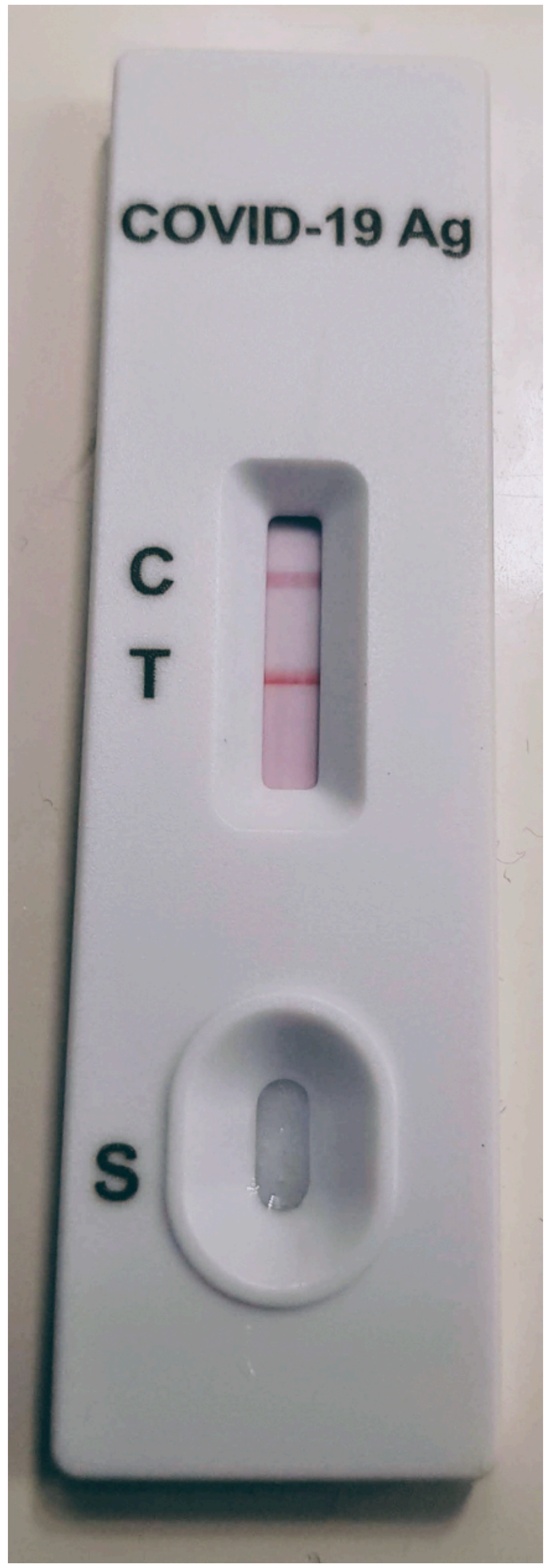
› Author Affiliations | Article Information

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

COVID-19 Resource Center

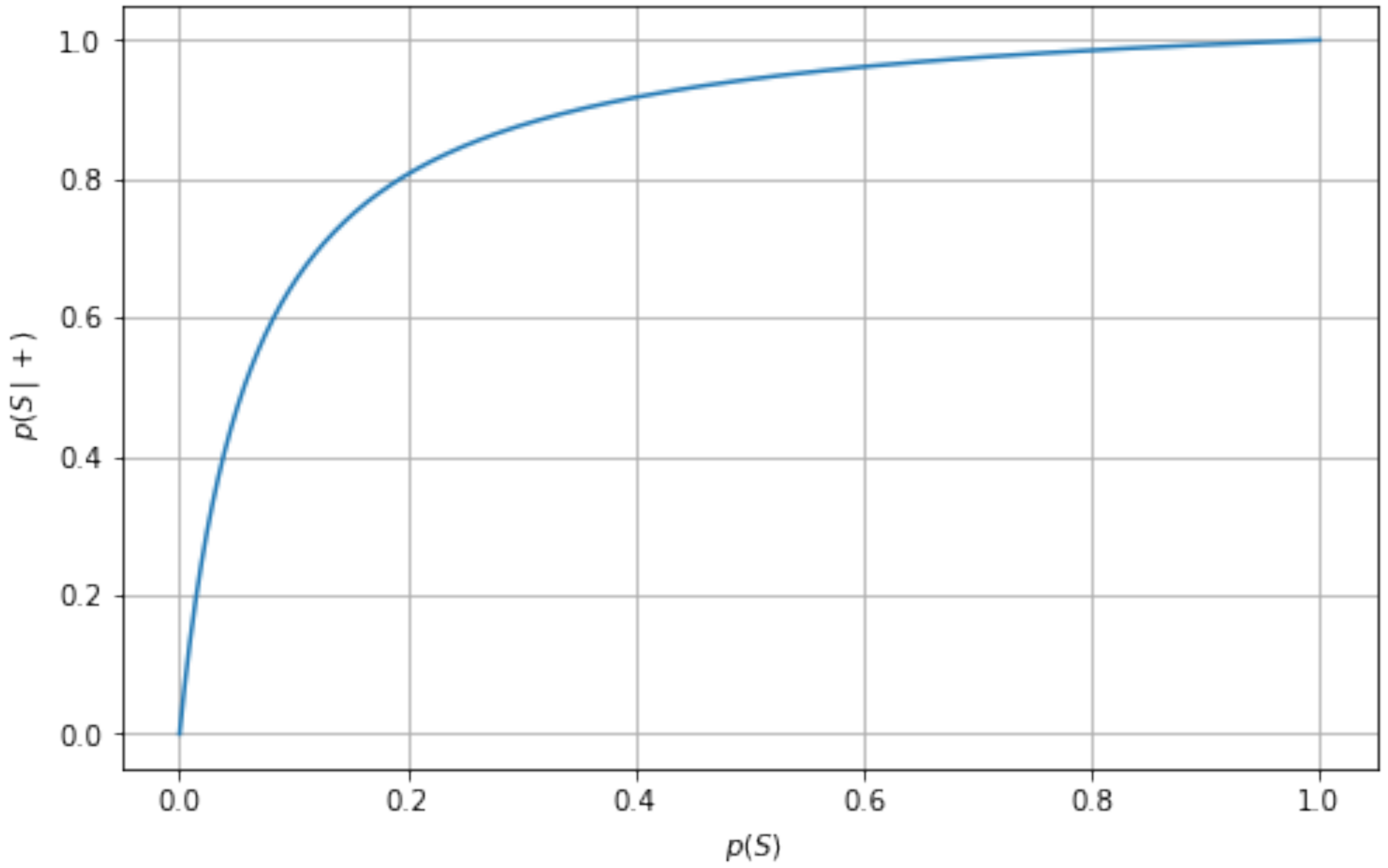
Overall sensitivity of home antigen tests for detecting cases was 50% (95% CI, 45%-55%) (Figure 3), whereas specificity was 97% (95% CI, 95%-98%). Sensitivity was higher for symptomatic cases (53%; 95% CI, 48%-57%) compared with asymptomatic cases (20%; 95% CI,

Another example: Covid-19 test



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

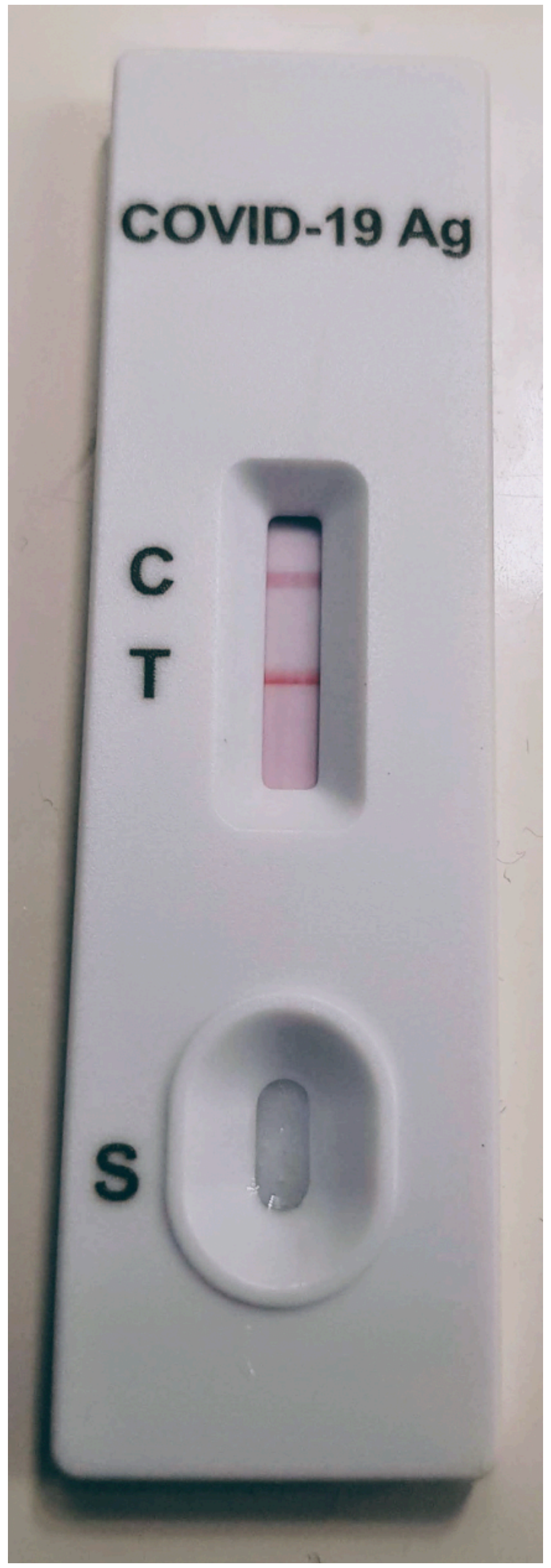
$$p(S\,|\,+) = \left(1 + \frac{1-\mathrm{Sp.}}{\mathrm{Se.}} \cdot \frac{p(\bar{S})}{p(S)}\right)^{-1}$$



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

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

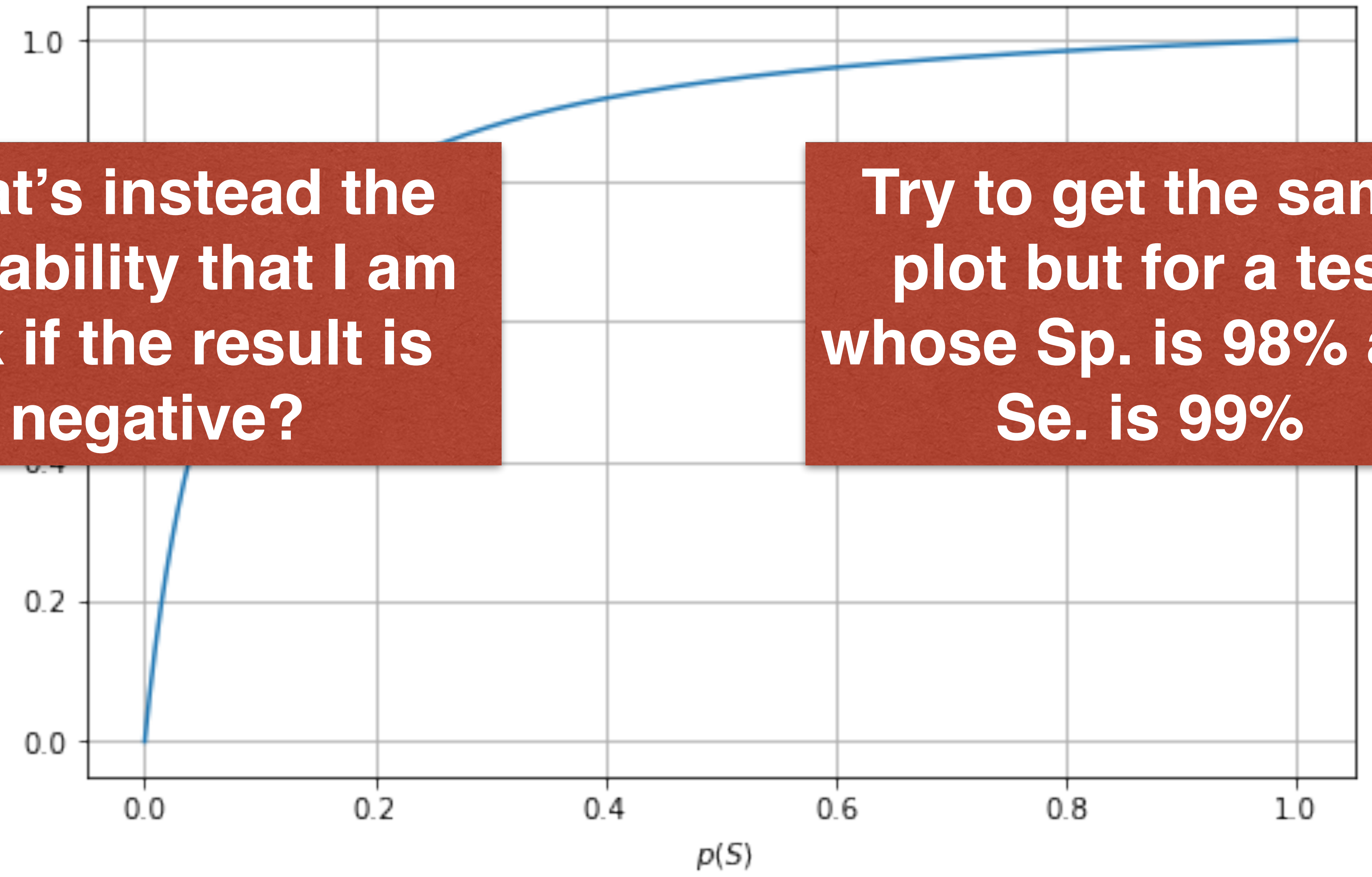
Another example: Covid-19 test



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**

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

= 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

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

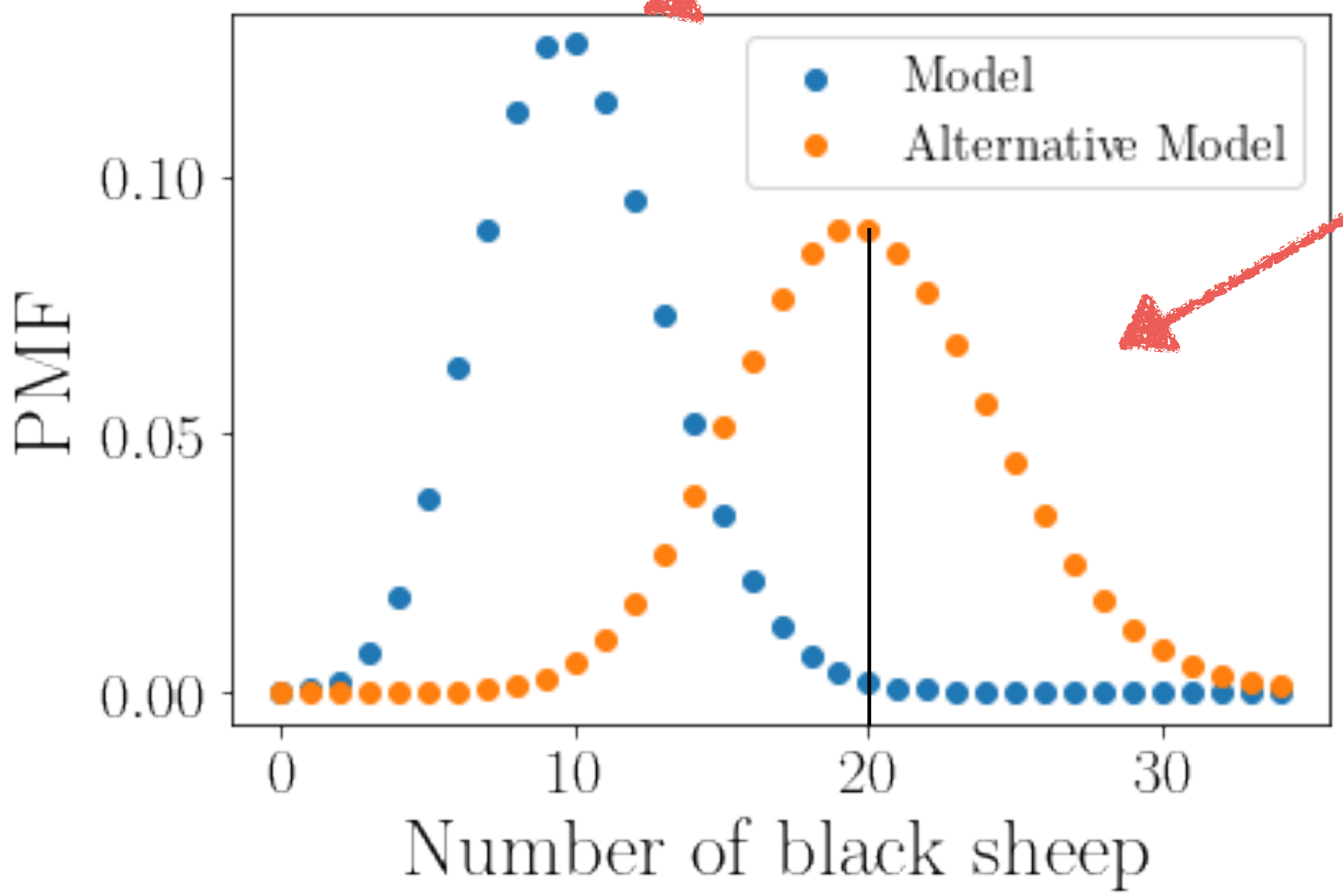
Our prior knowledge:

- 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

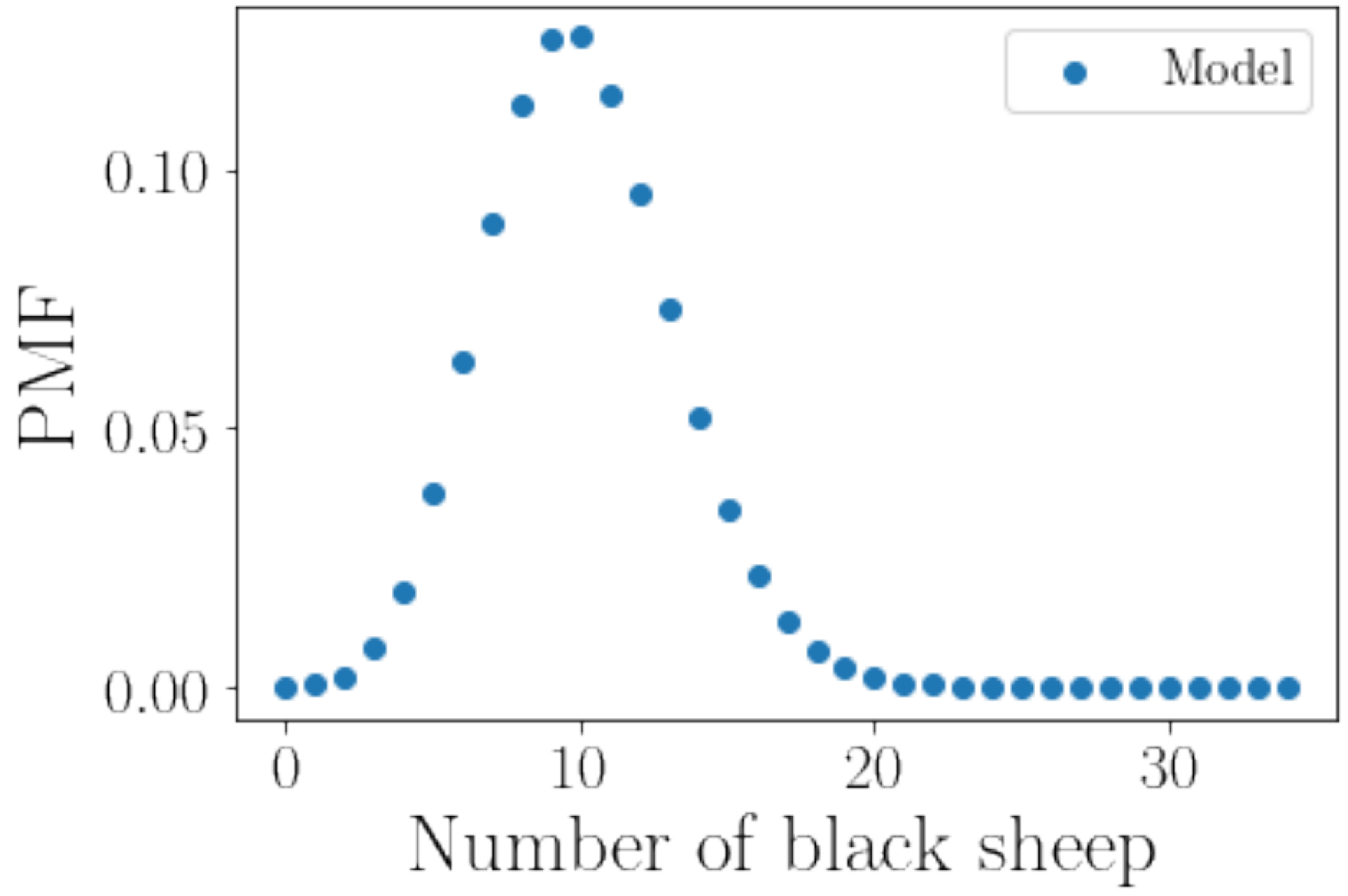
We can use the number of sheep observed as statistics and ask ourselves:

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.

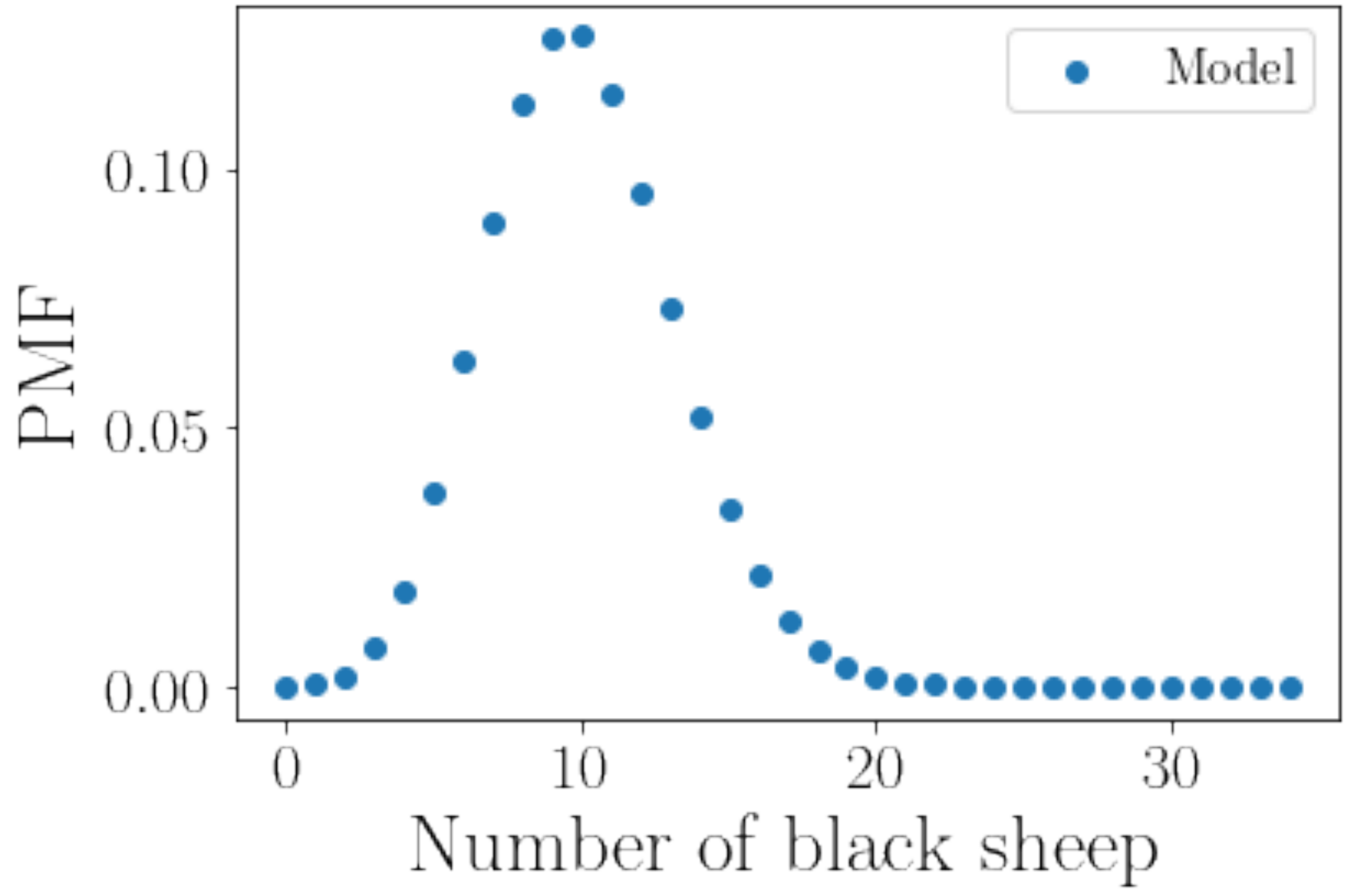
Let's go back to the "sheep" example

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.

[ro-ph.HE] 19 Jun 2022

PKS 1413+135: Bright

Ying-Ying Gan,[1] Jin

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

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 **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

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

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

Let's go back to the "sheep" example

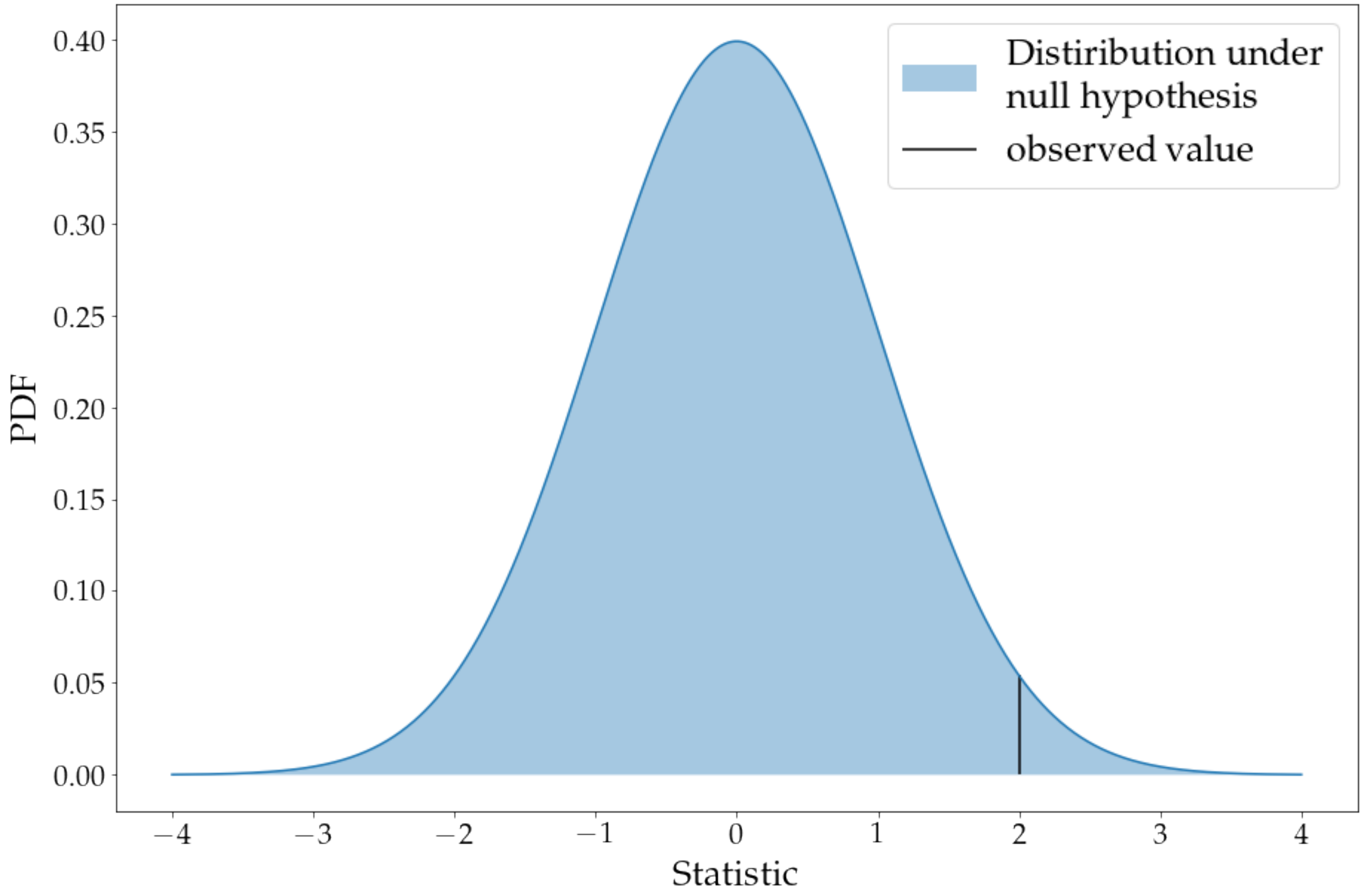
The Model *1% of the sheep are black*

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:

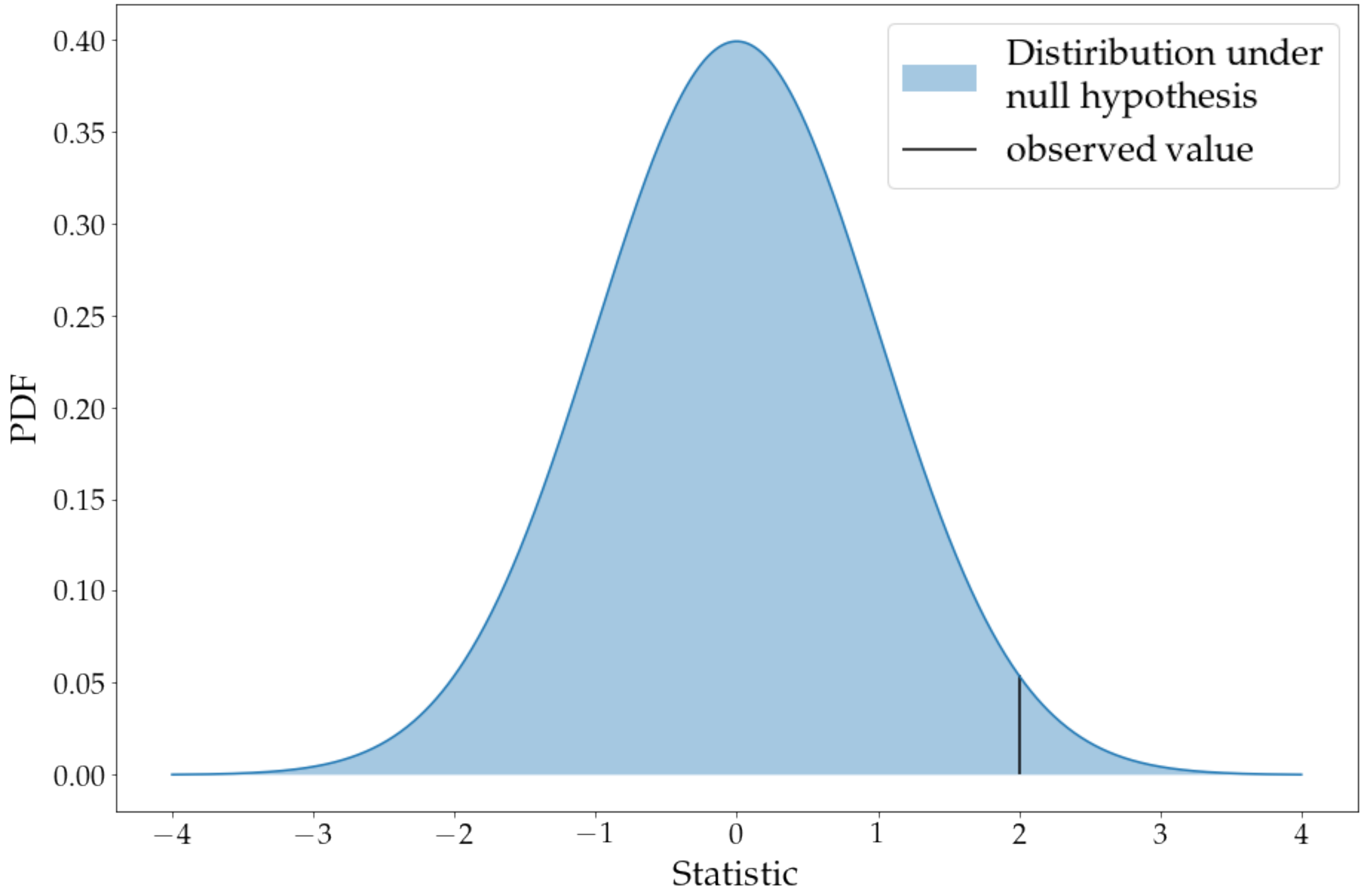
It does not take into account the **alternative hypothesis** that might explain the outcome of an event



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

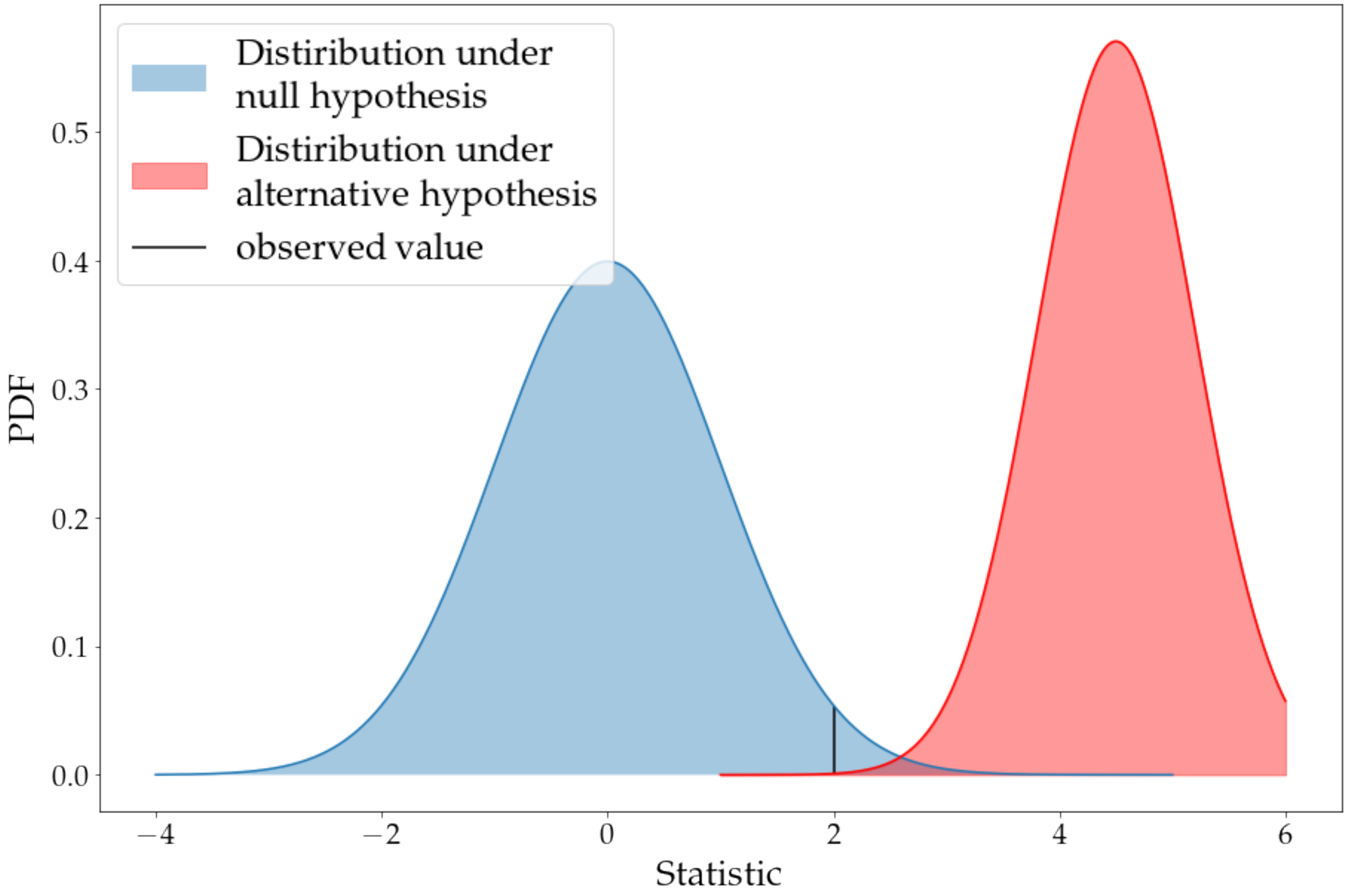


Conclusion:
The null hypothesis is rejected with a **2 sigma** significance

But what about the alternative hypothesis?

Issues of the frequentist approach:

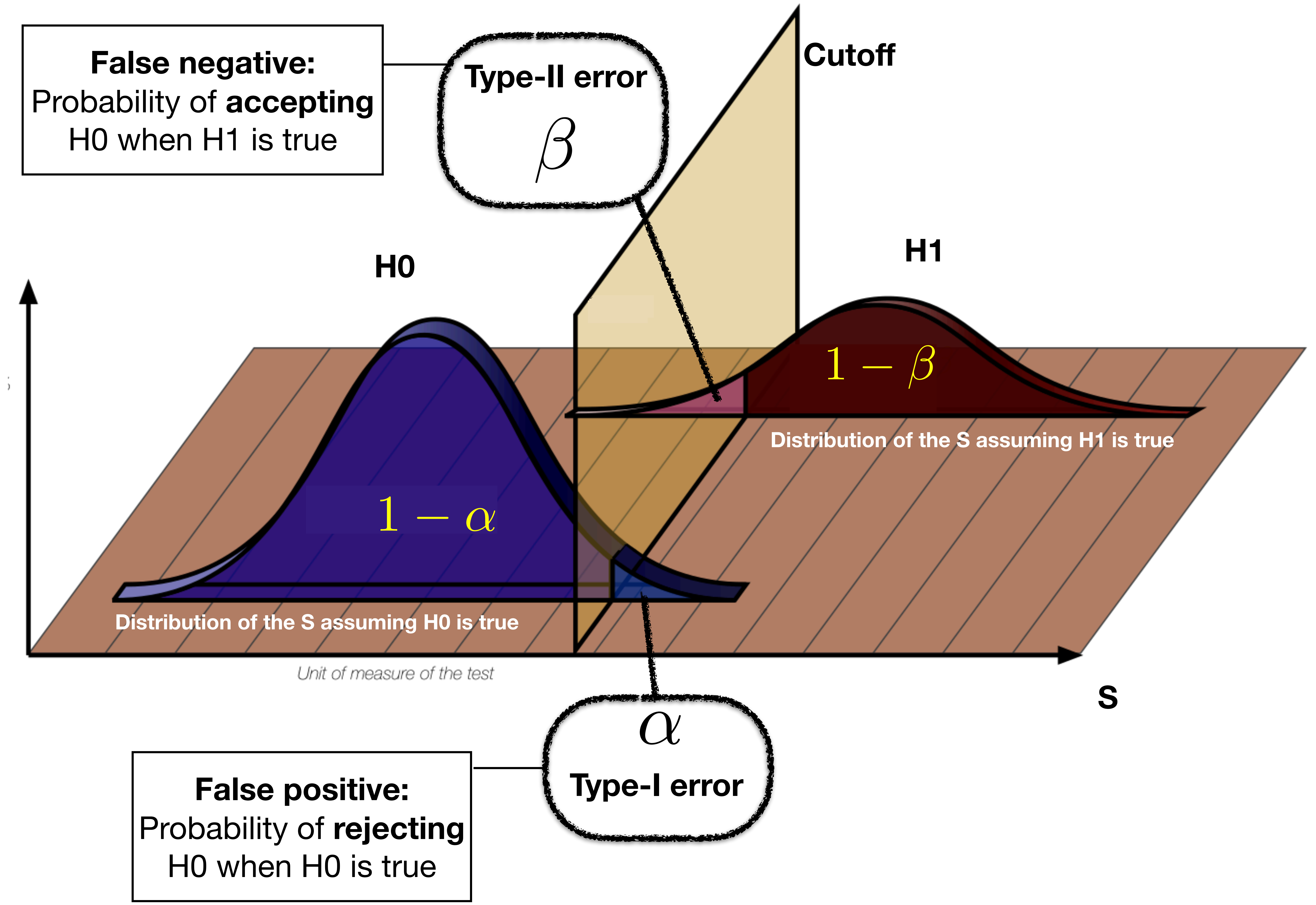
It does not take into account the **alternative hypothesis** that might explain the outcome of an event



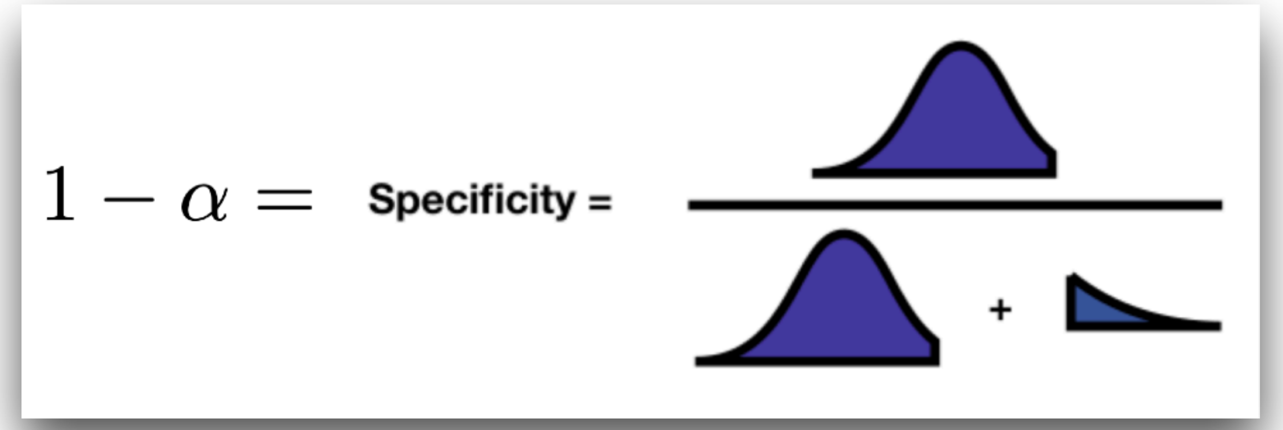
The observed value of 2 is actually more plausible being the outcome of the null hypothesis

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

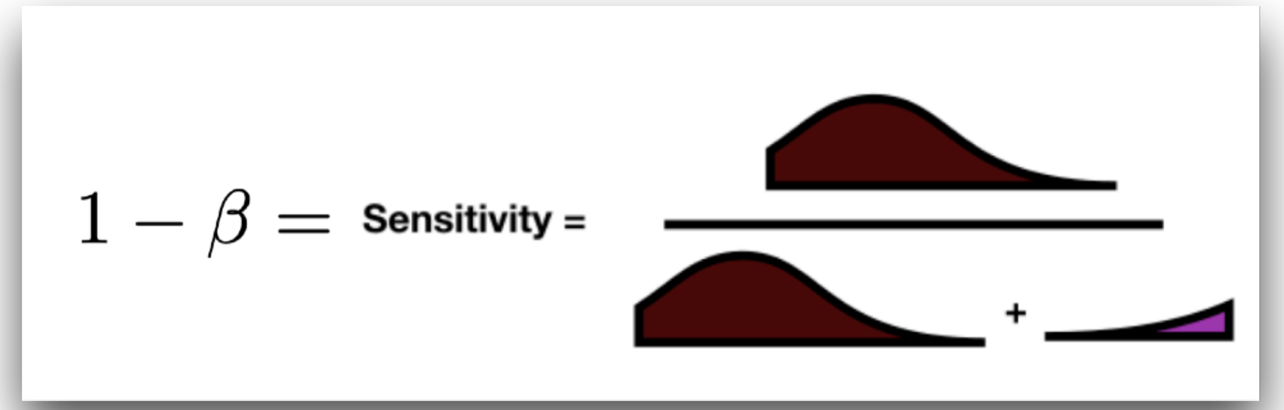
This is why a value of sigma bigger than 3 or even 5 is required for making a claim!



Confidence level

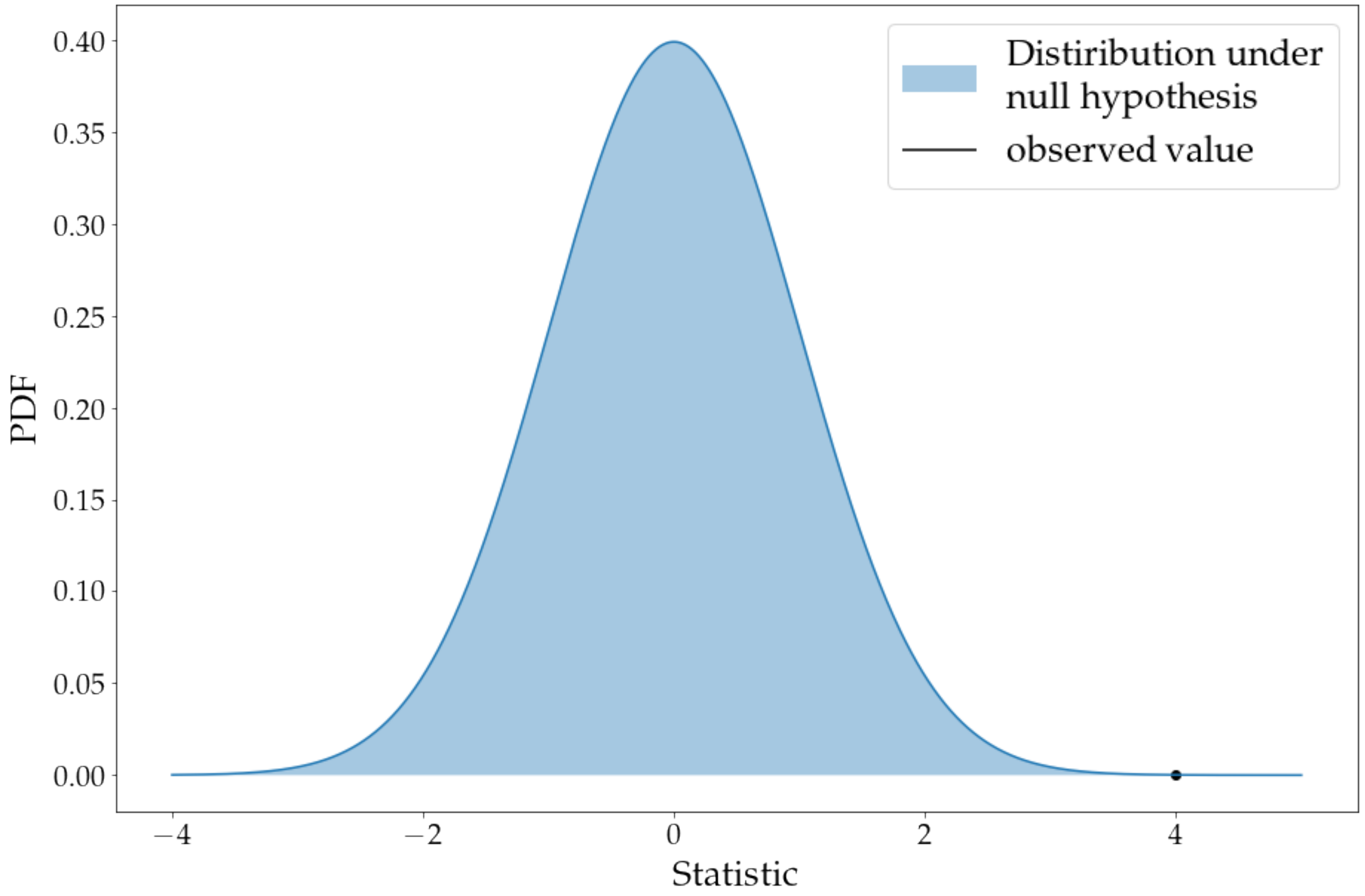


Power of the test



Issues of the frequentist approach:

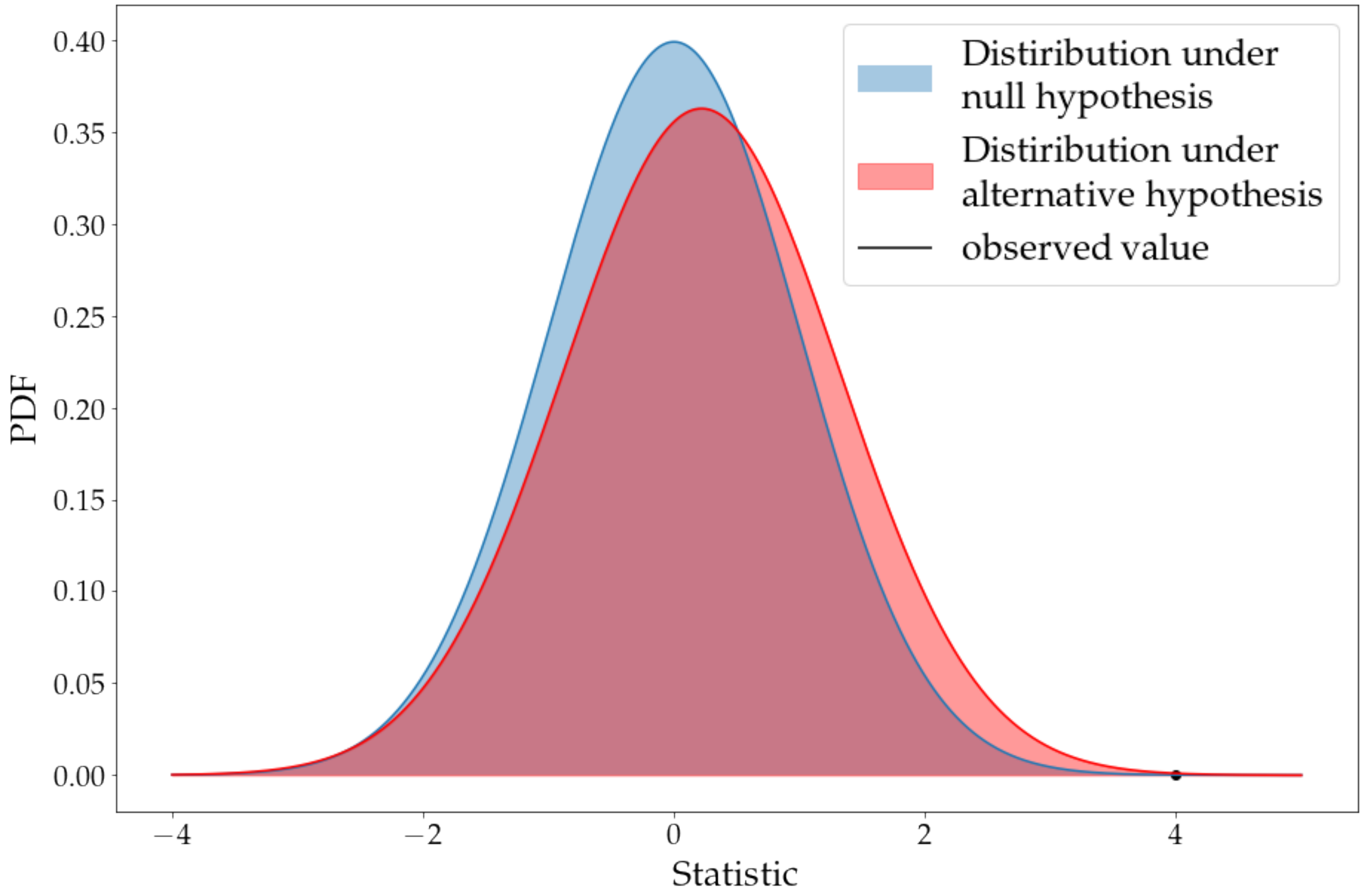
It does not take into account the **alternative hypothesis** that might explain the outcome of an event



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

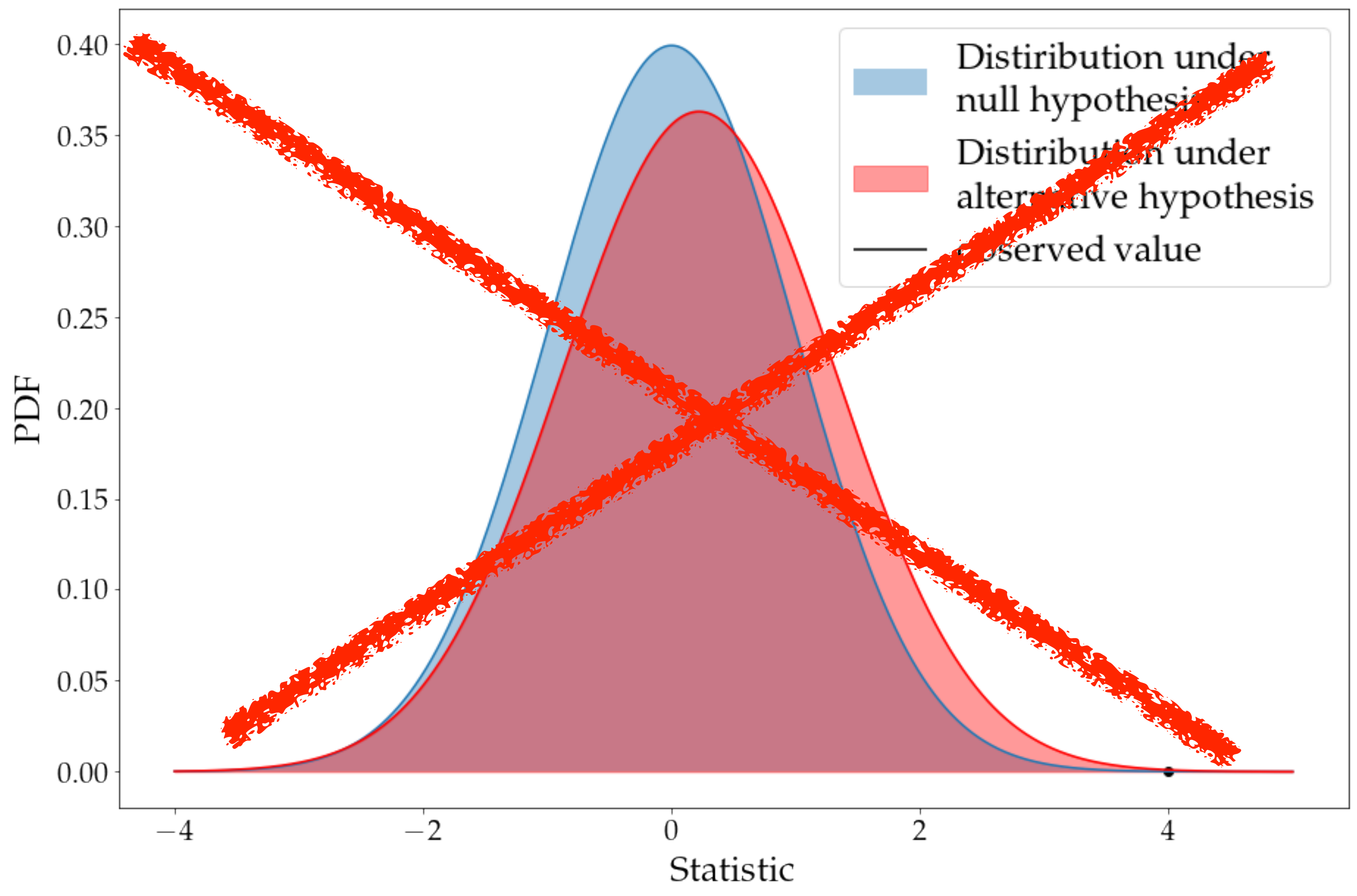


So… with a significance of 4 we should be safe?

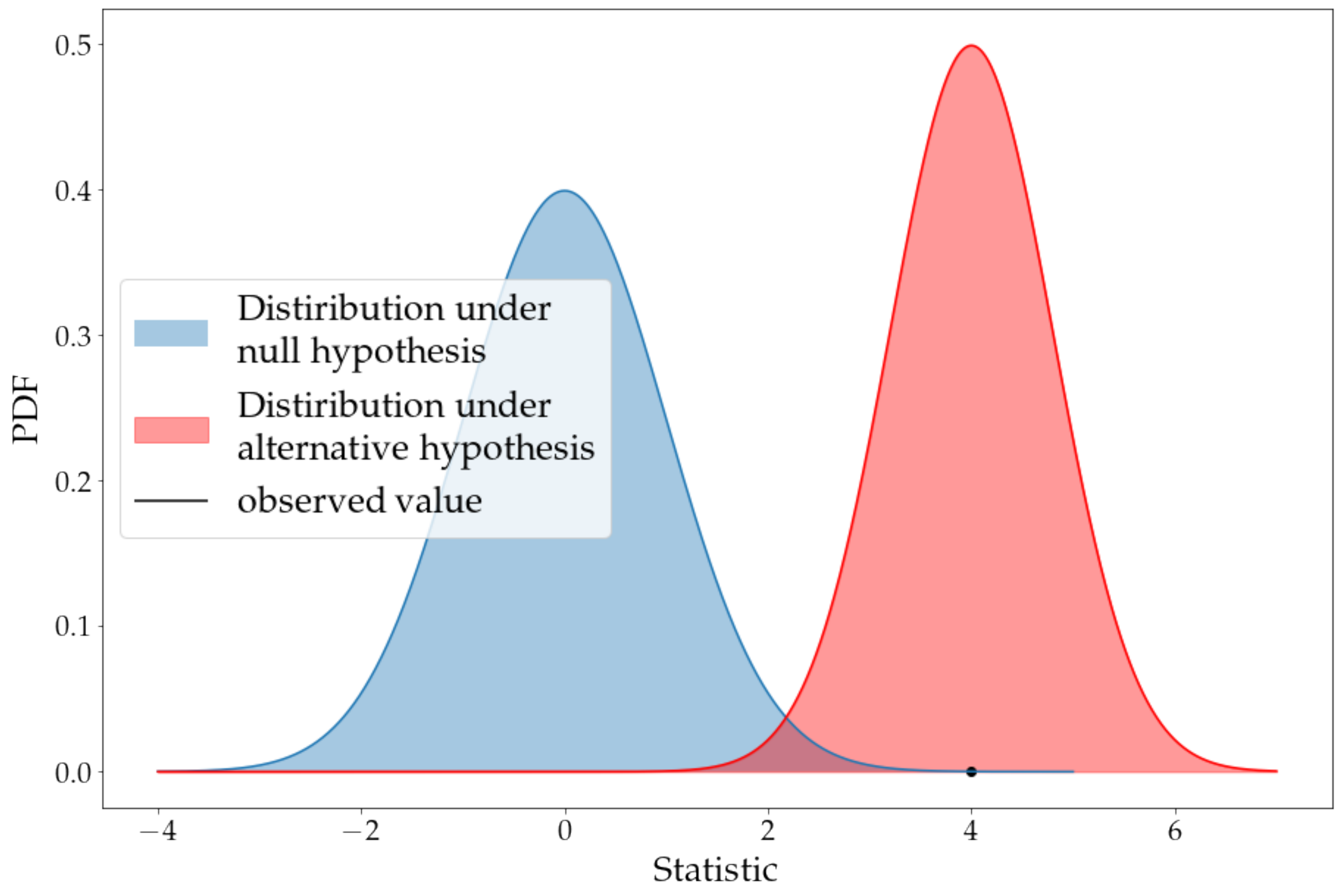
The value of 4 is unlikely to be the outcome also of the alternative hypothesis, thus again we could be doing a ***type I error***

Issues of the frequentist approach:

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



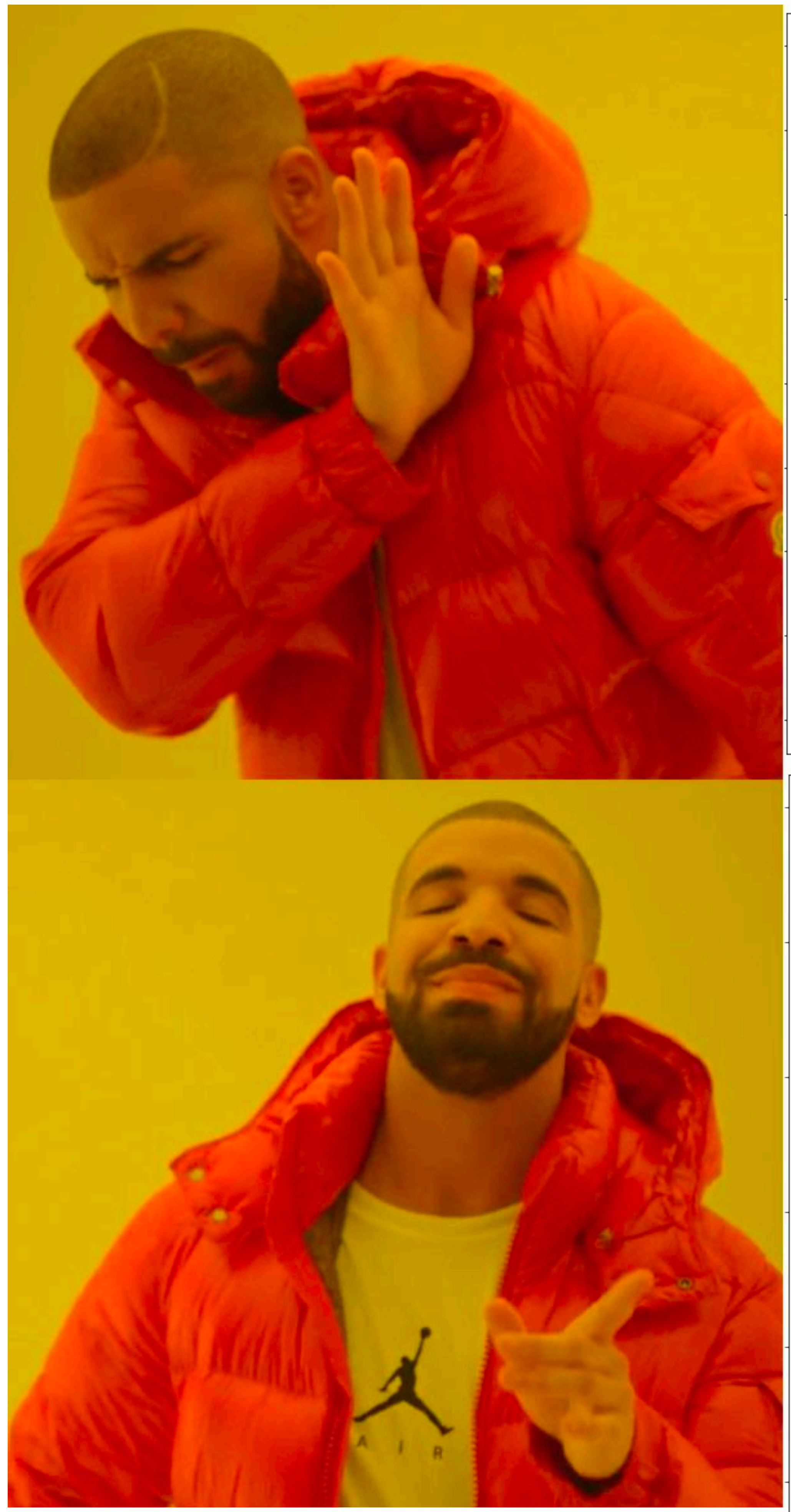
The ideal statistic is the one that makes you **reject** a hypothesis that is false!

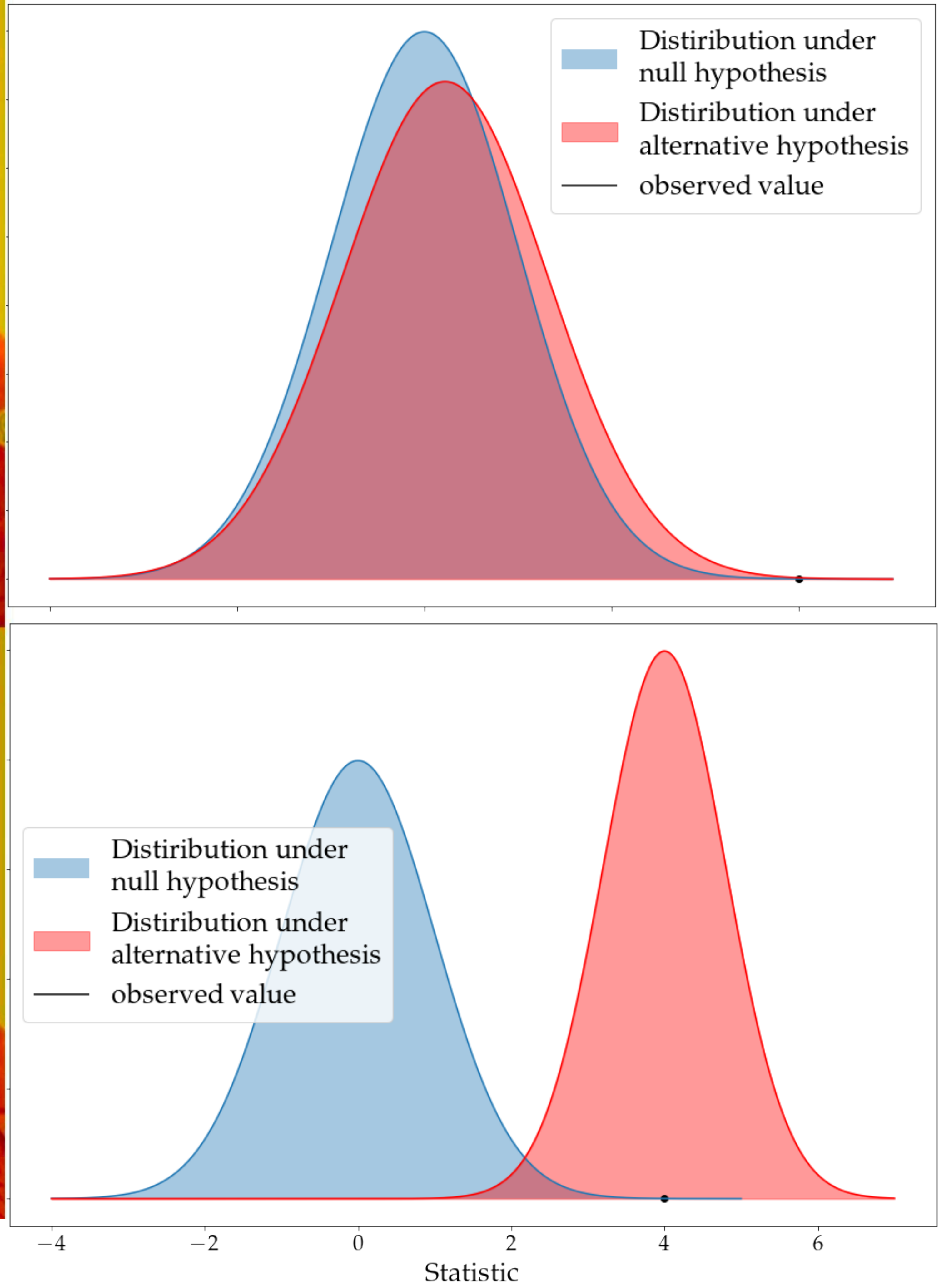


Issues of the frequentist approach:

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**:

Likelihood

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

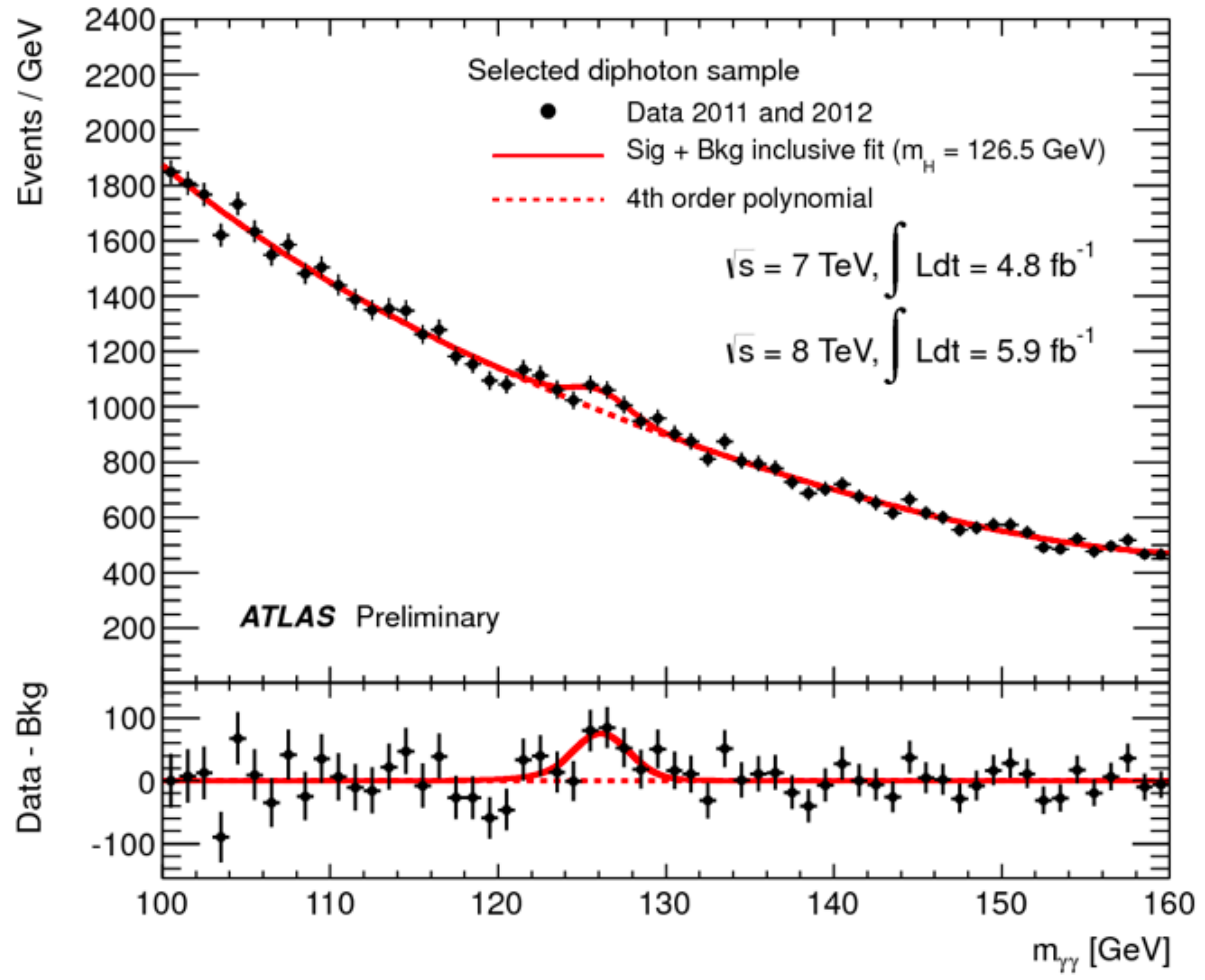
Parameter of the null hypothesis

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

Best fit or value that maximises the likelihood

Observed data

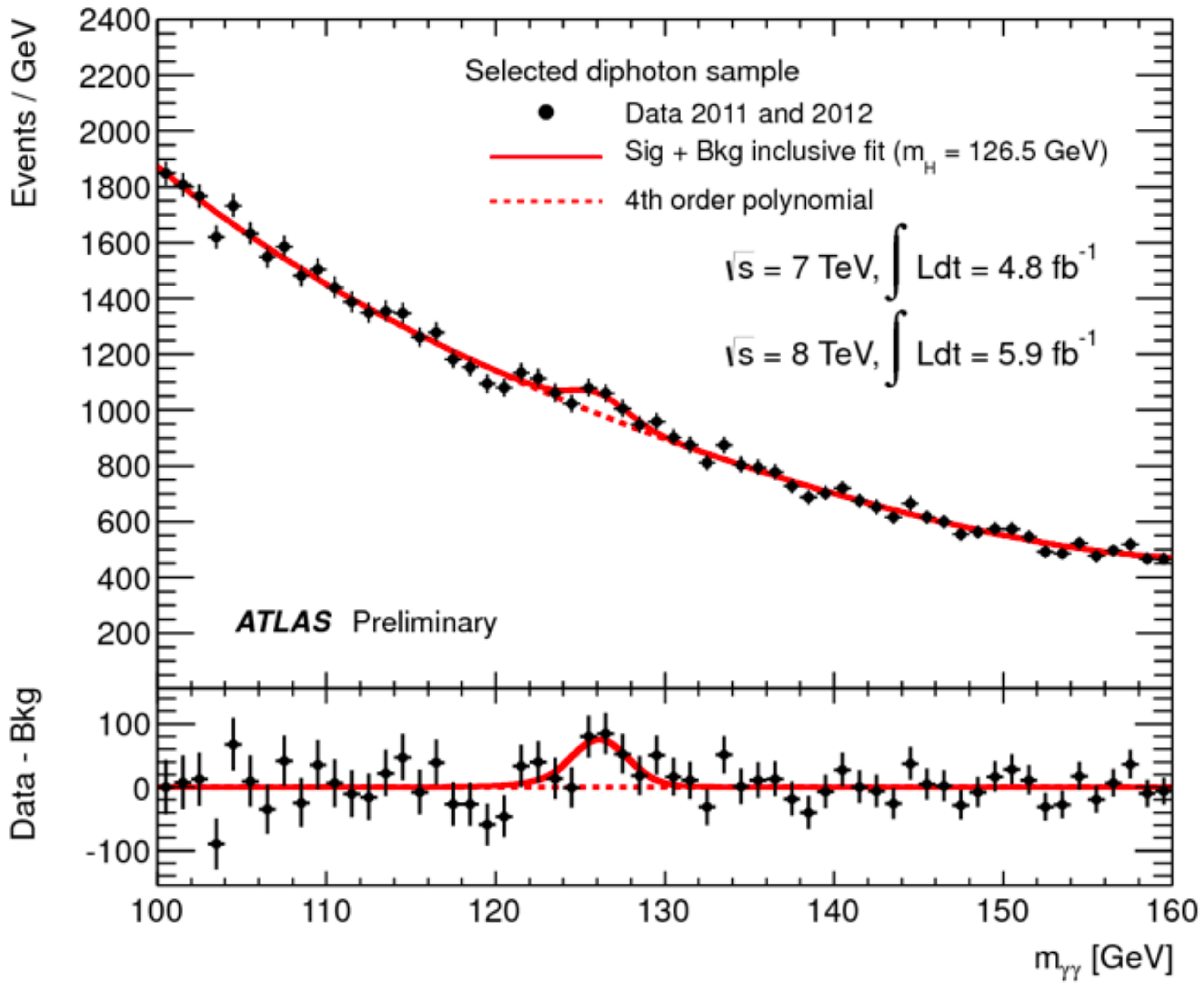
Example:



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

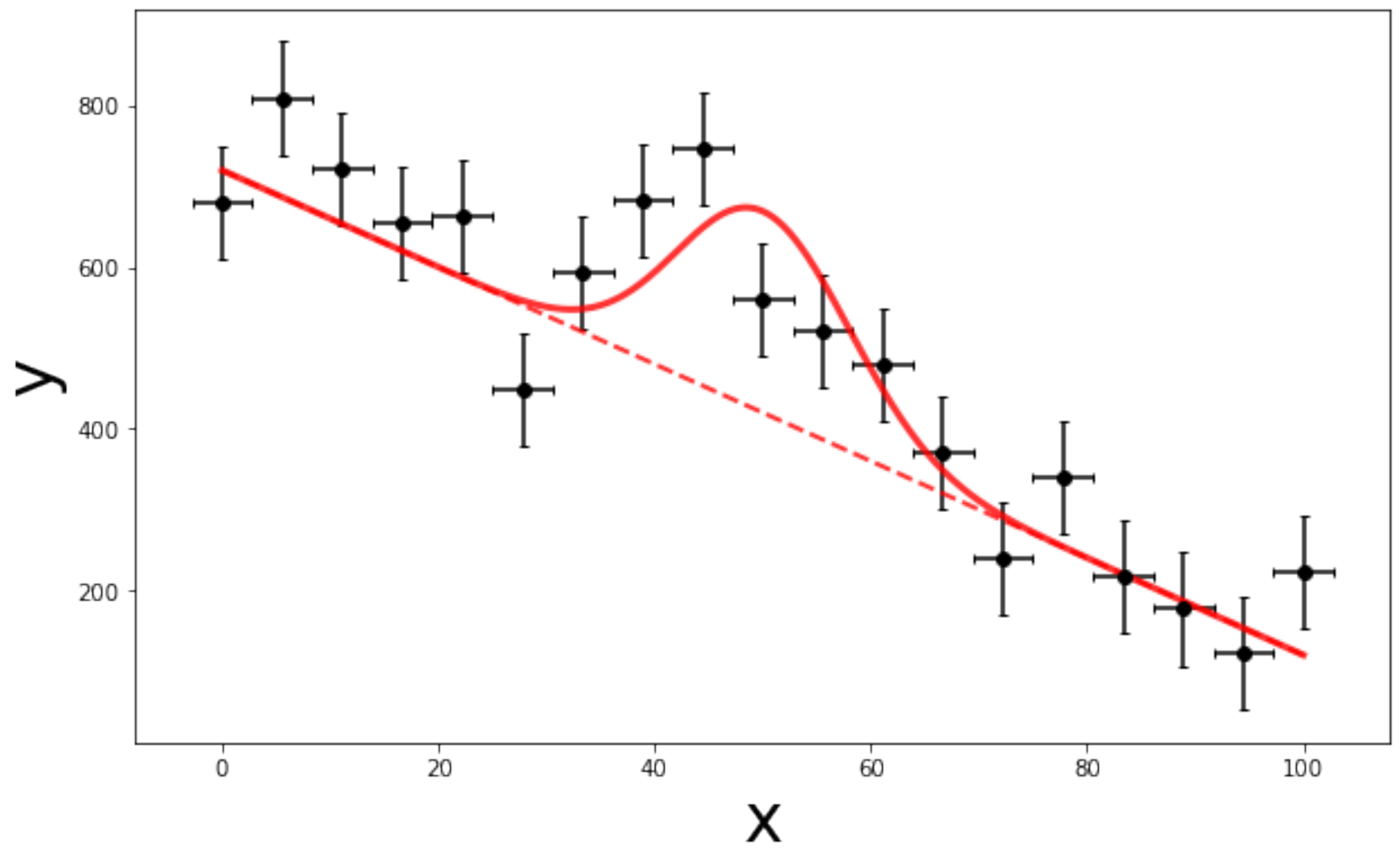
Example:



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:

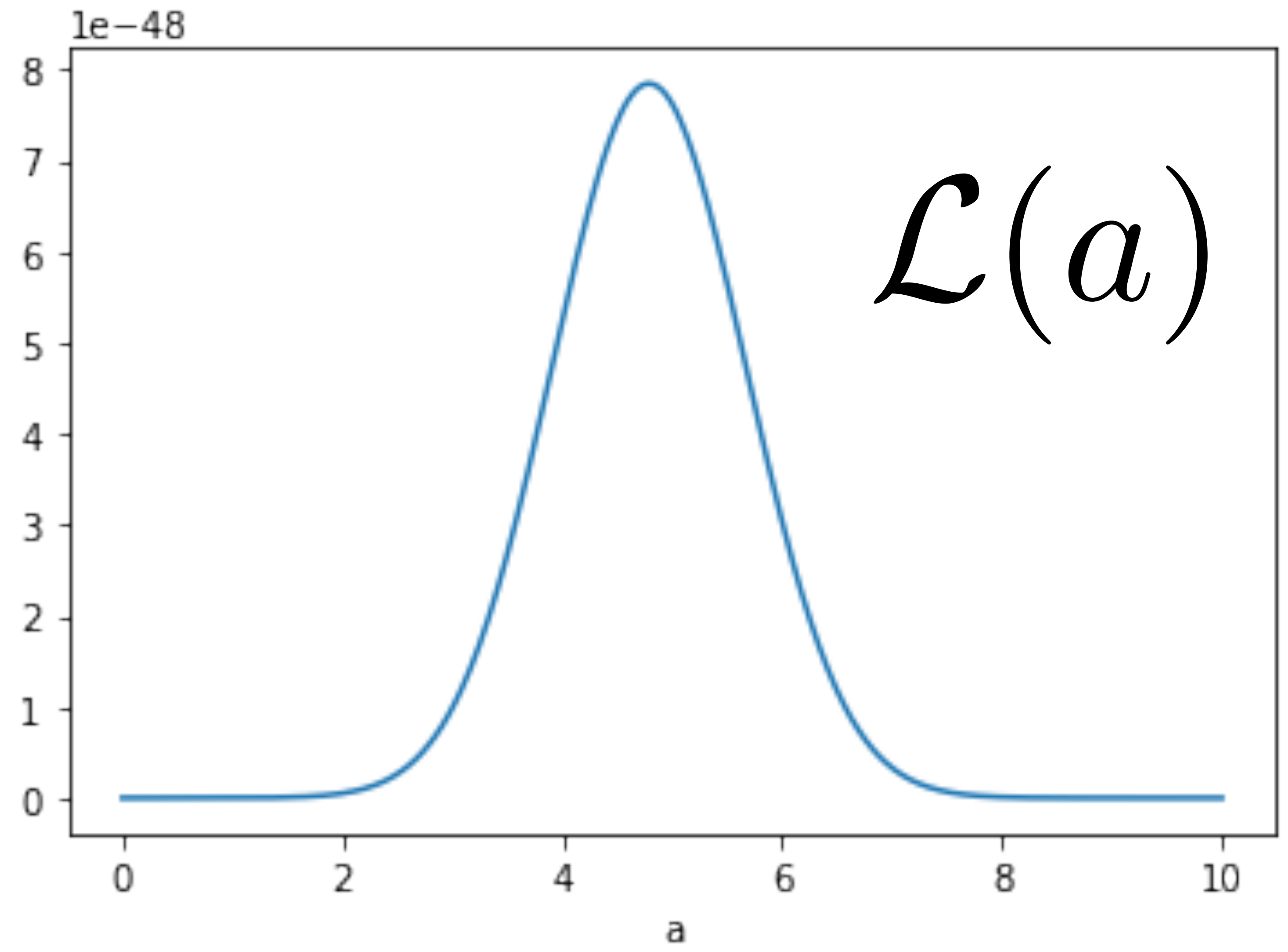
$$y' = mx + q + a \cdot \mathcal{G}(x; \mu = 50, \sigma = 8)$$

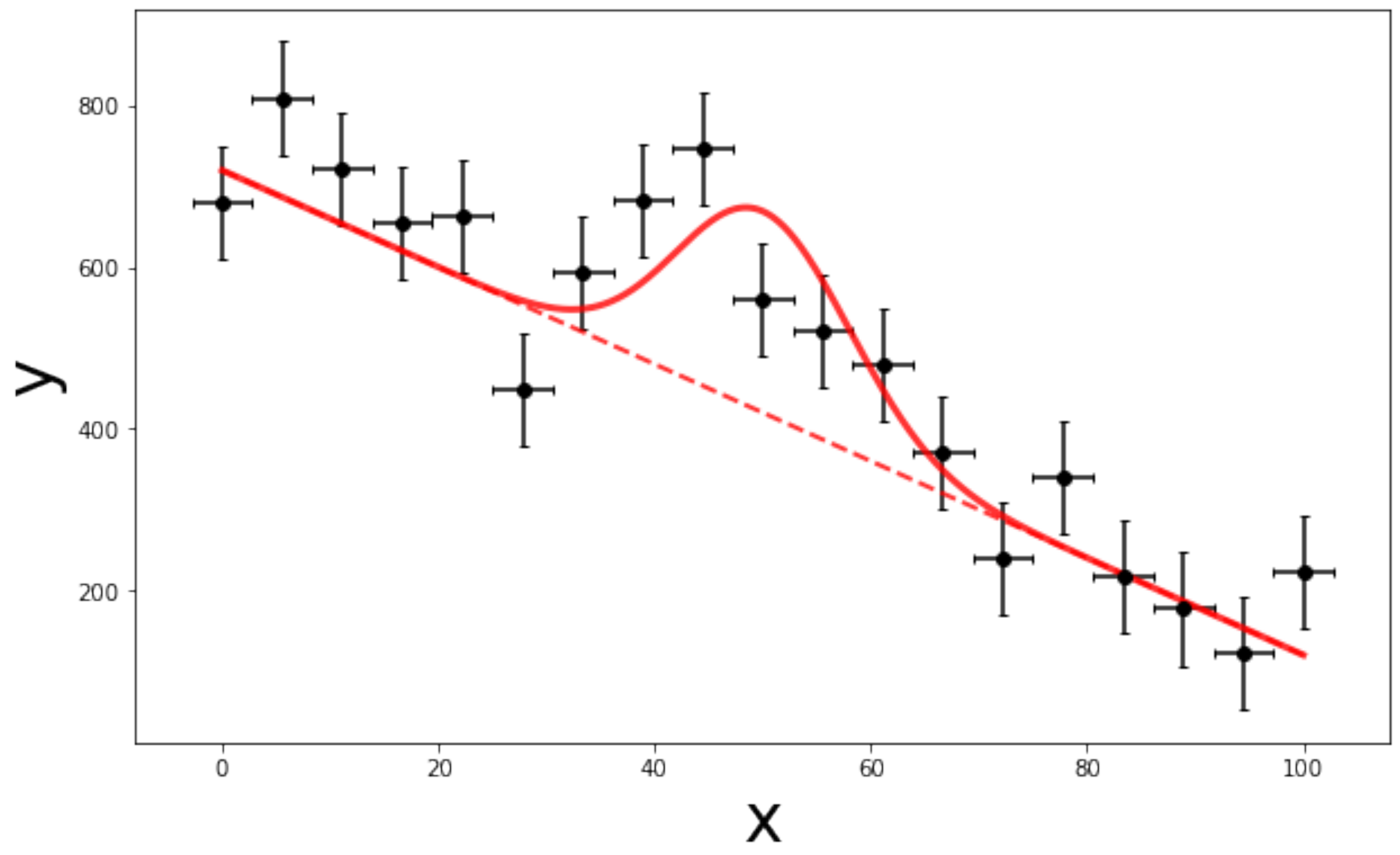
$$y \sim \mathcal{N}(\mu = y', \sigma = 70)$$

Likelihood

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

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





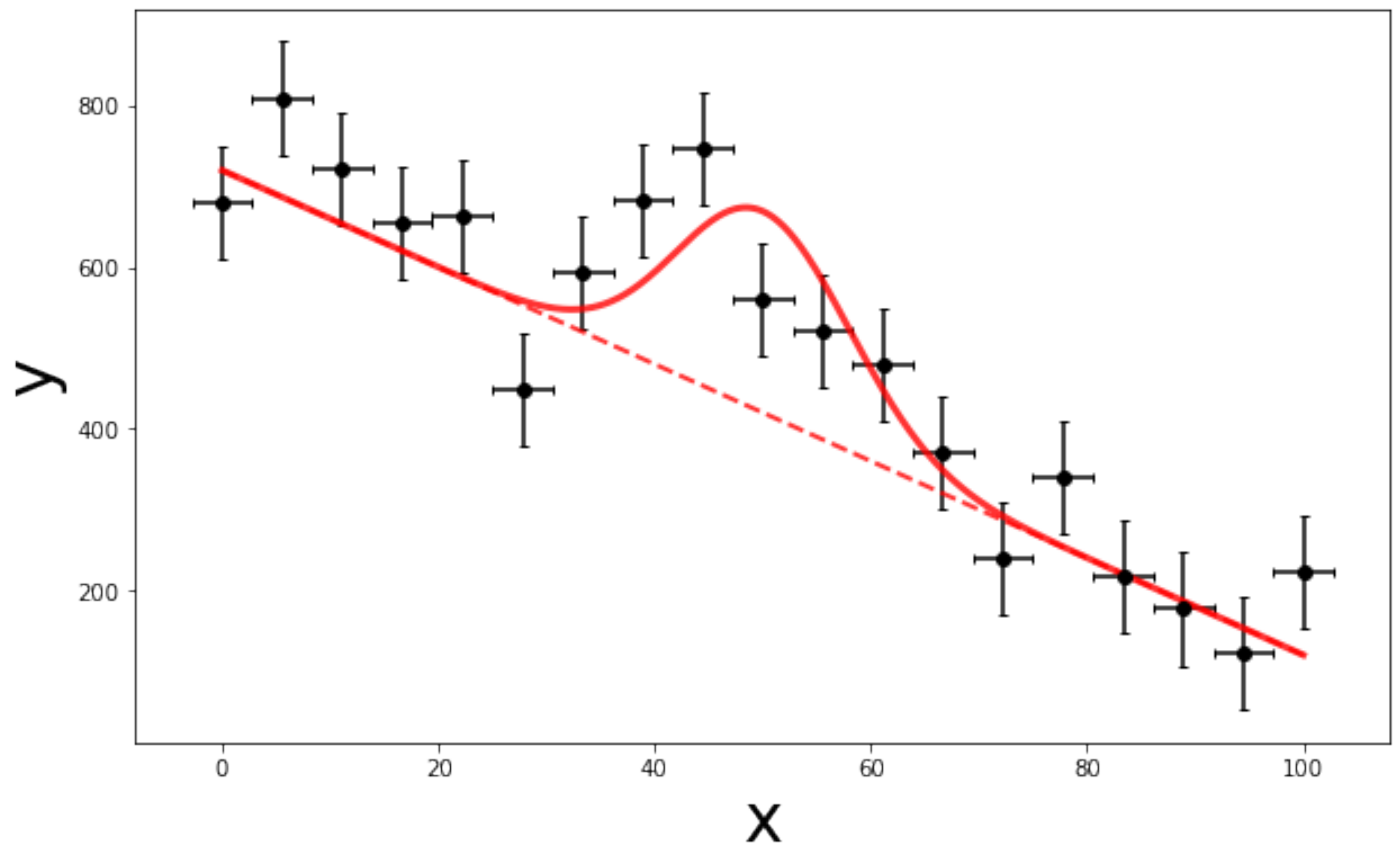
Null hypothesis H0
a = 0

Alternative hyp. H1
a = 5

Example:

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

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



Likelihood

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

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

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

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

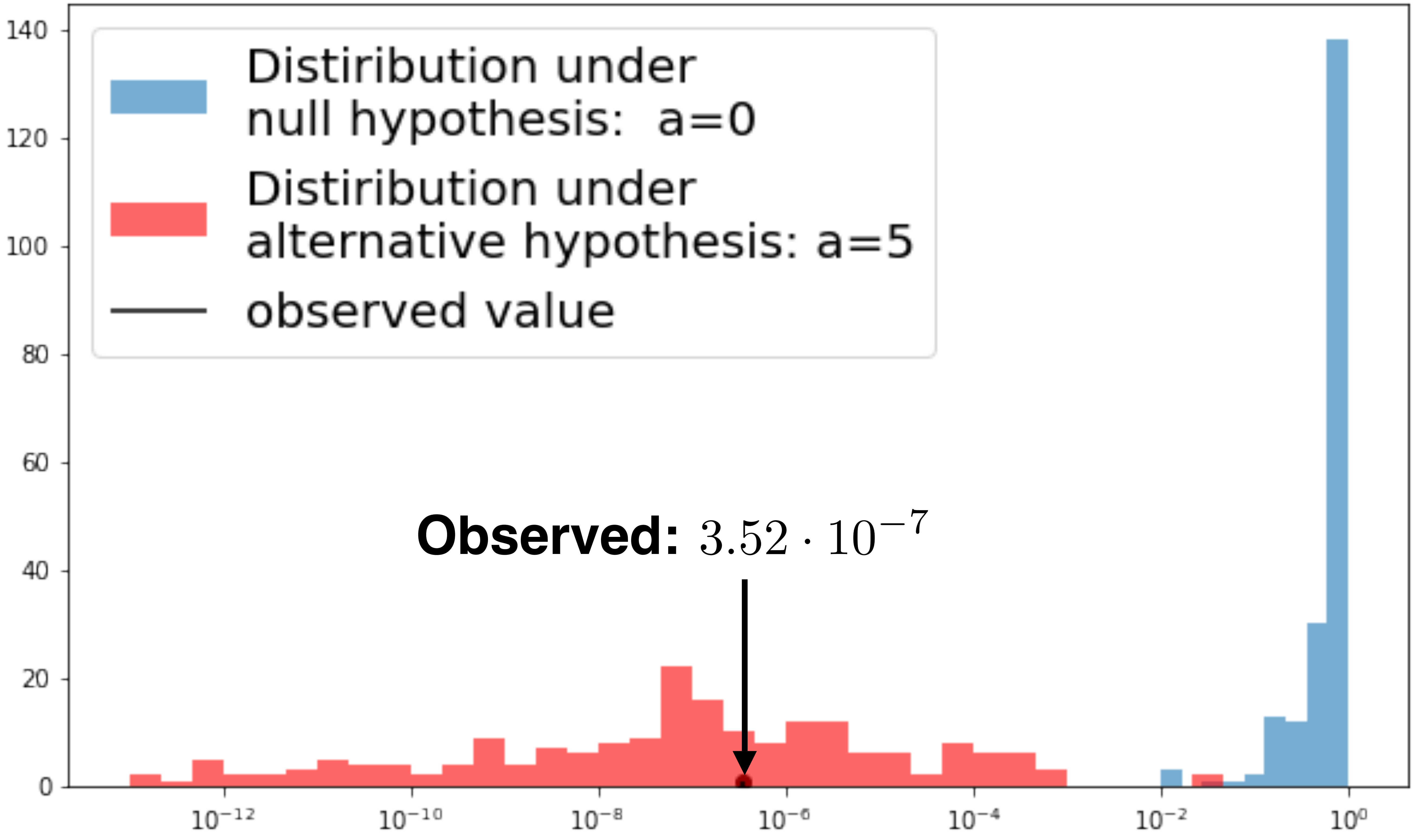
Null hypothesis H0
a = 0

Alternative hyp. H1
a = 5

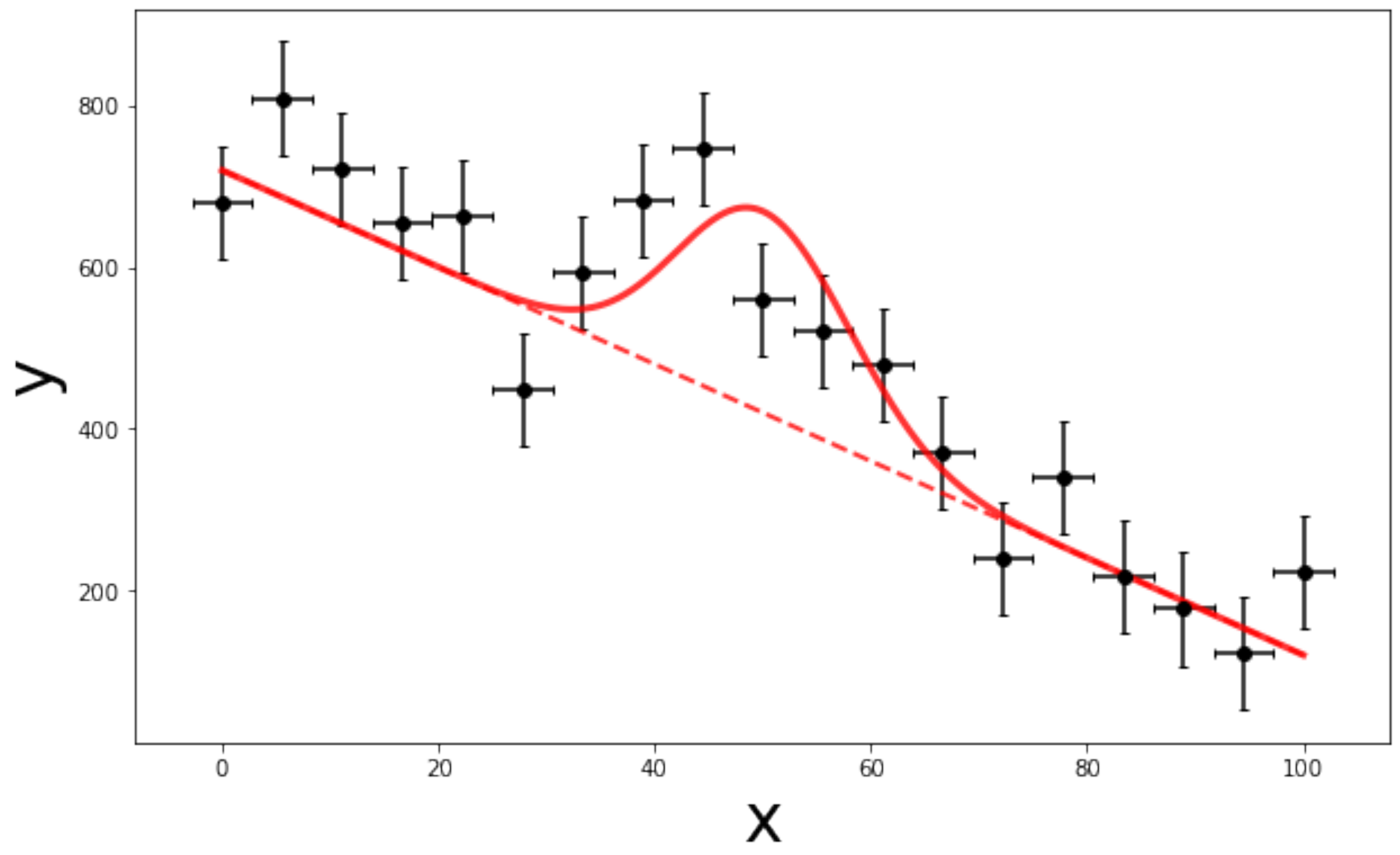
Example:

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

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



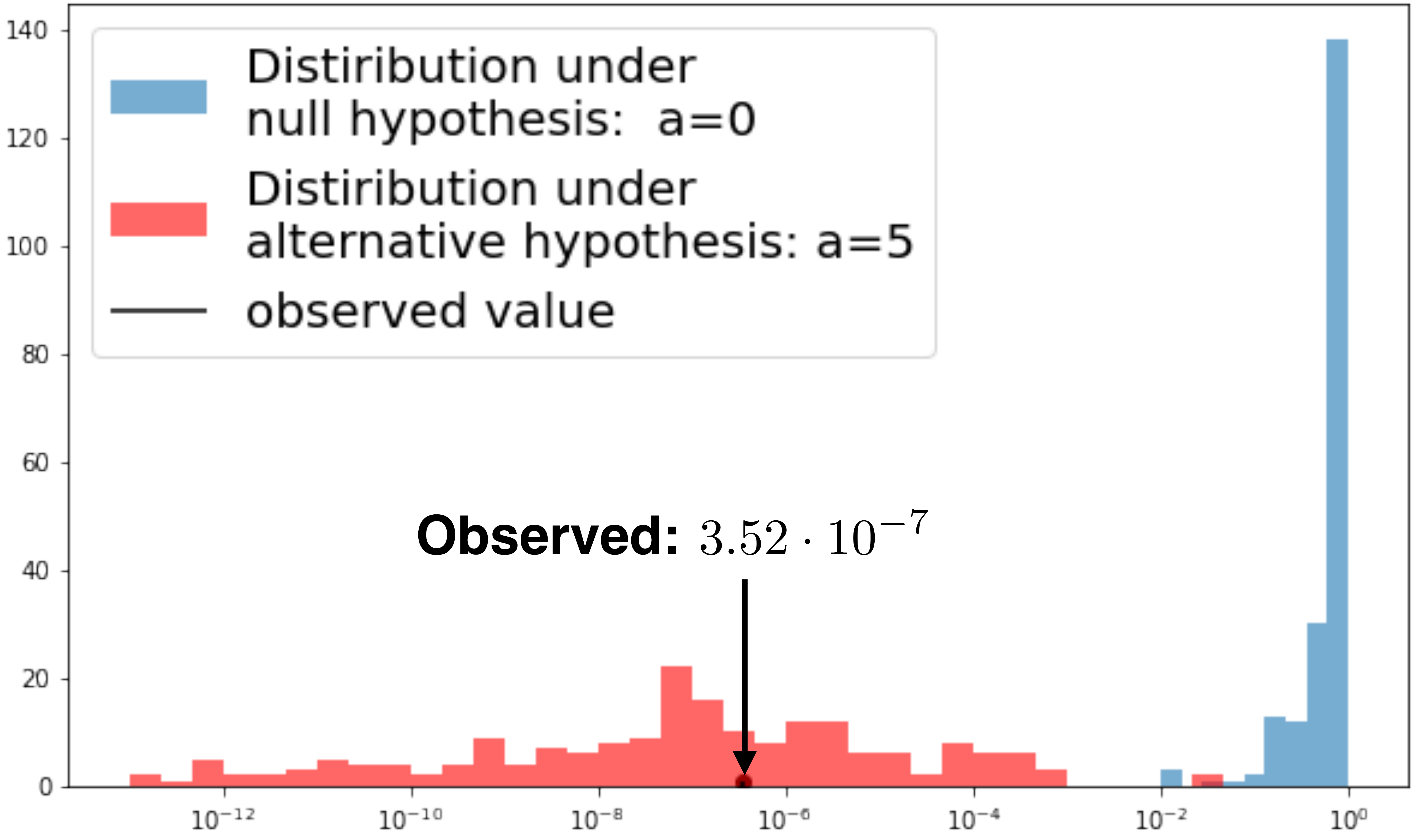
$$\mathcal{S} = \frac{\mathcal{L}(a=0)}{\mathcal{L}(a=\hat{a})}$$



Null hypothesis H0
a = 0

Alternative hyp. H1
a = 5

Example:



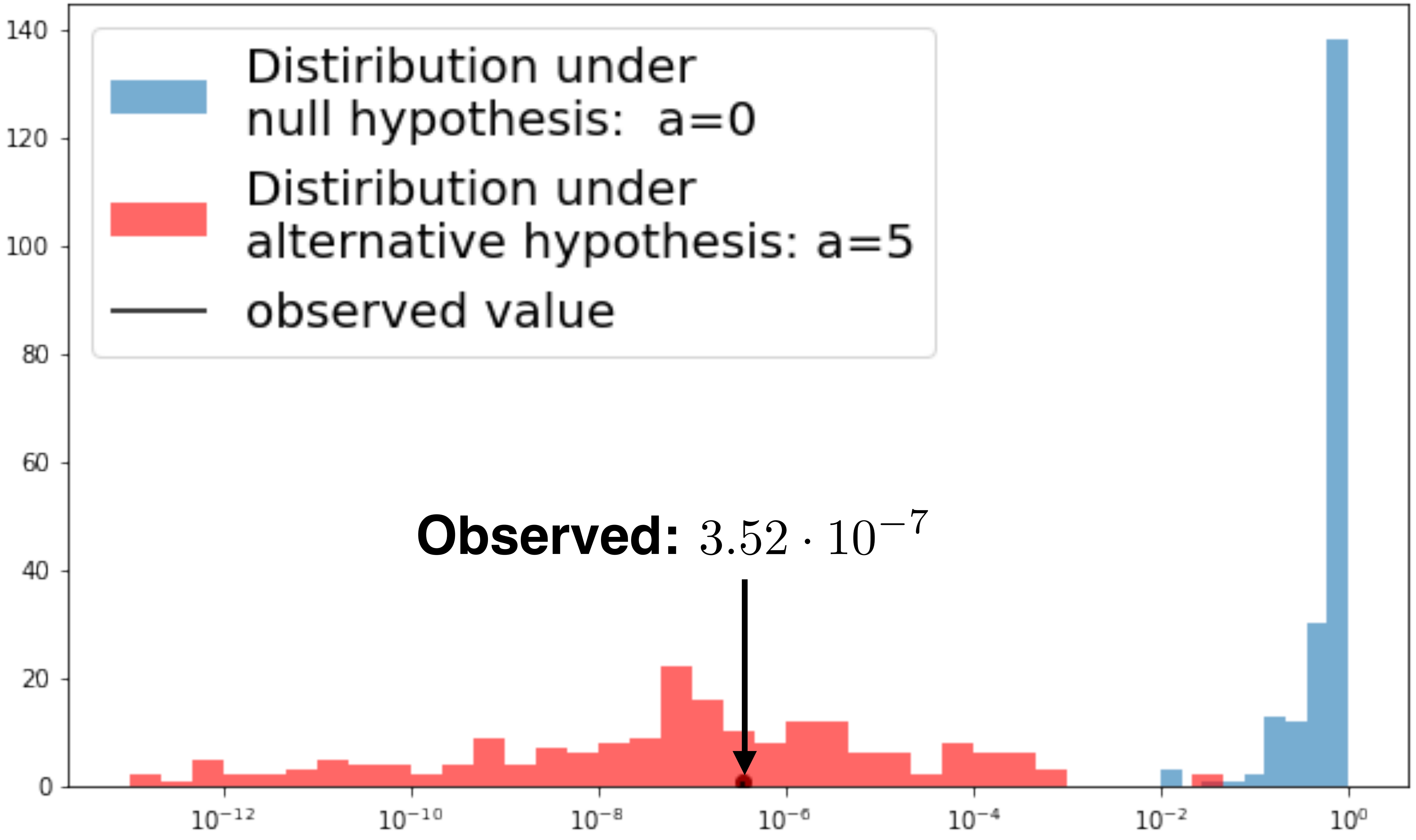
$$\mathcal{S} = \frac{\mathcal{L}(a=0)}{\mathcal{L}(a=\hat{a})}$$

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:

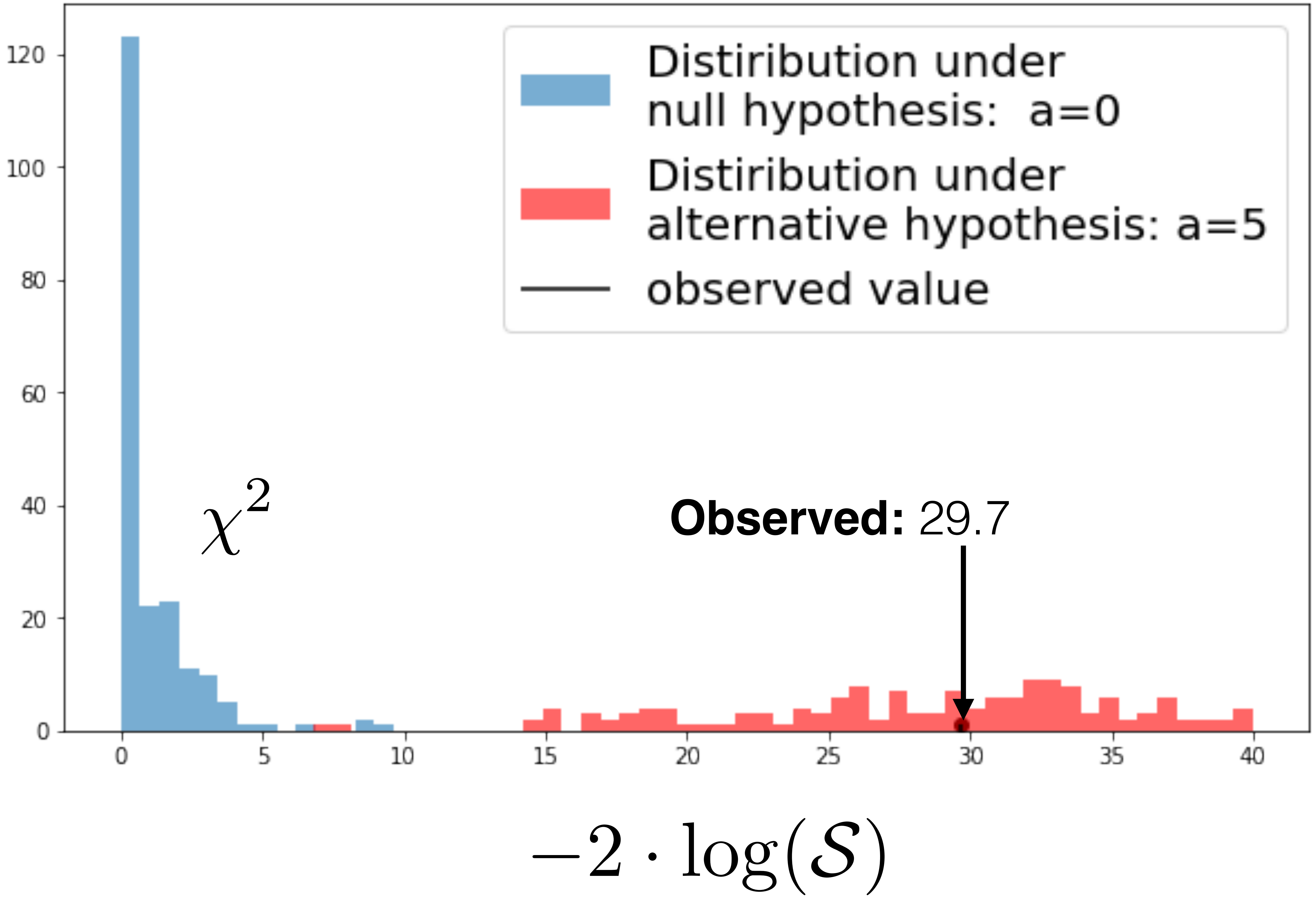


$$\mathcal{S} = \frac{\mathcal{L}(a=0)}{\mathcal{L}(a=\hat{a})}$$

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

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

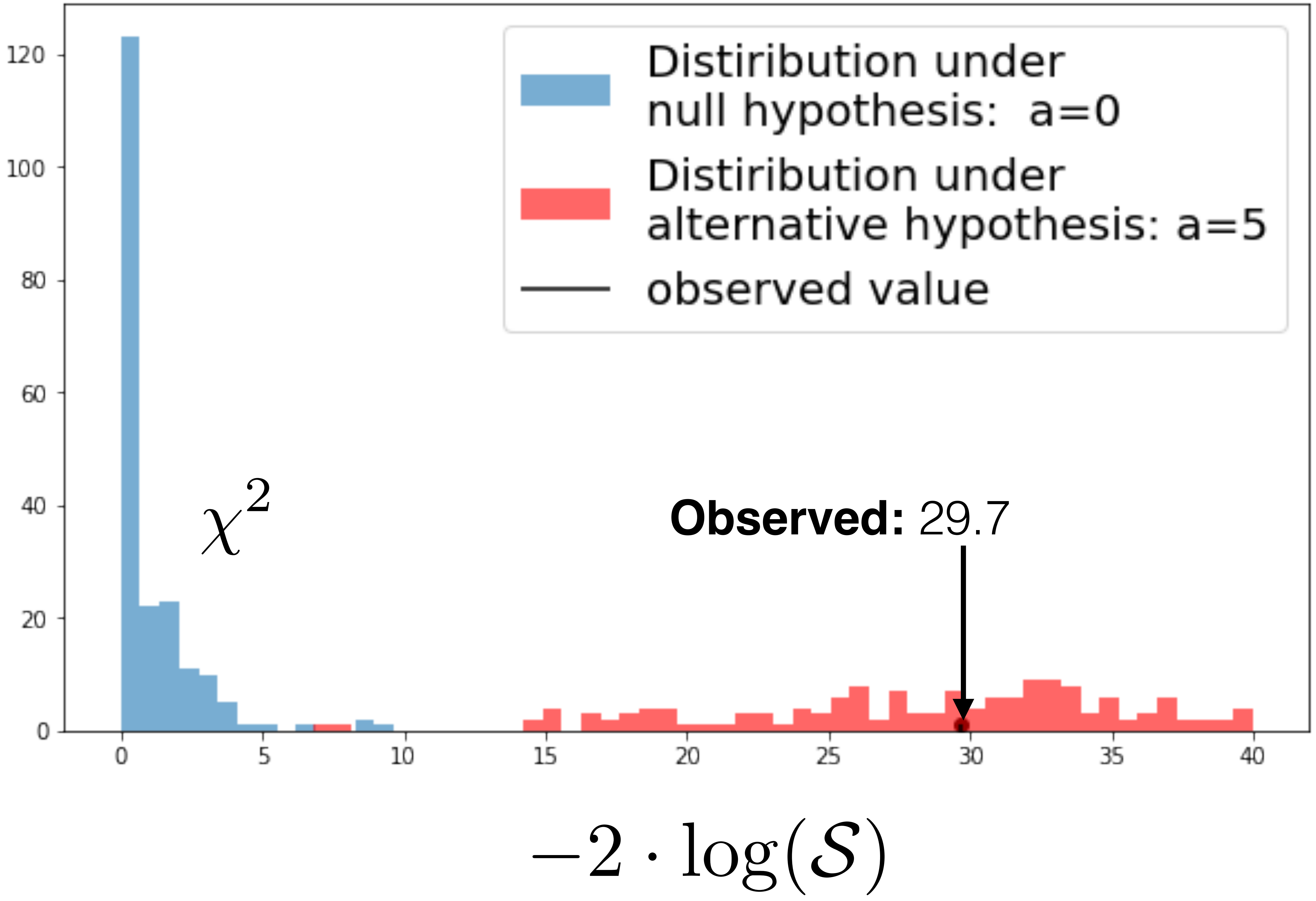
Example:



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

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

Example:



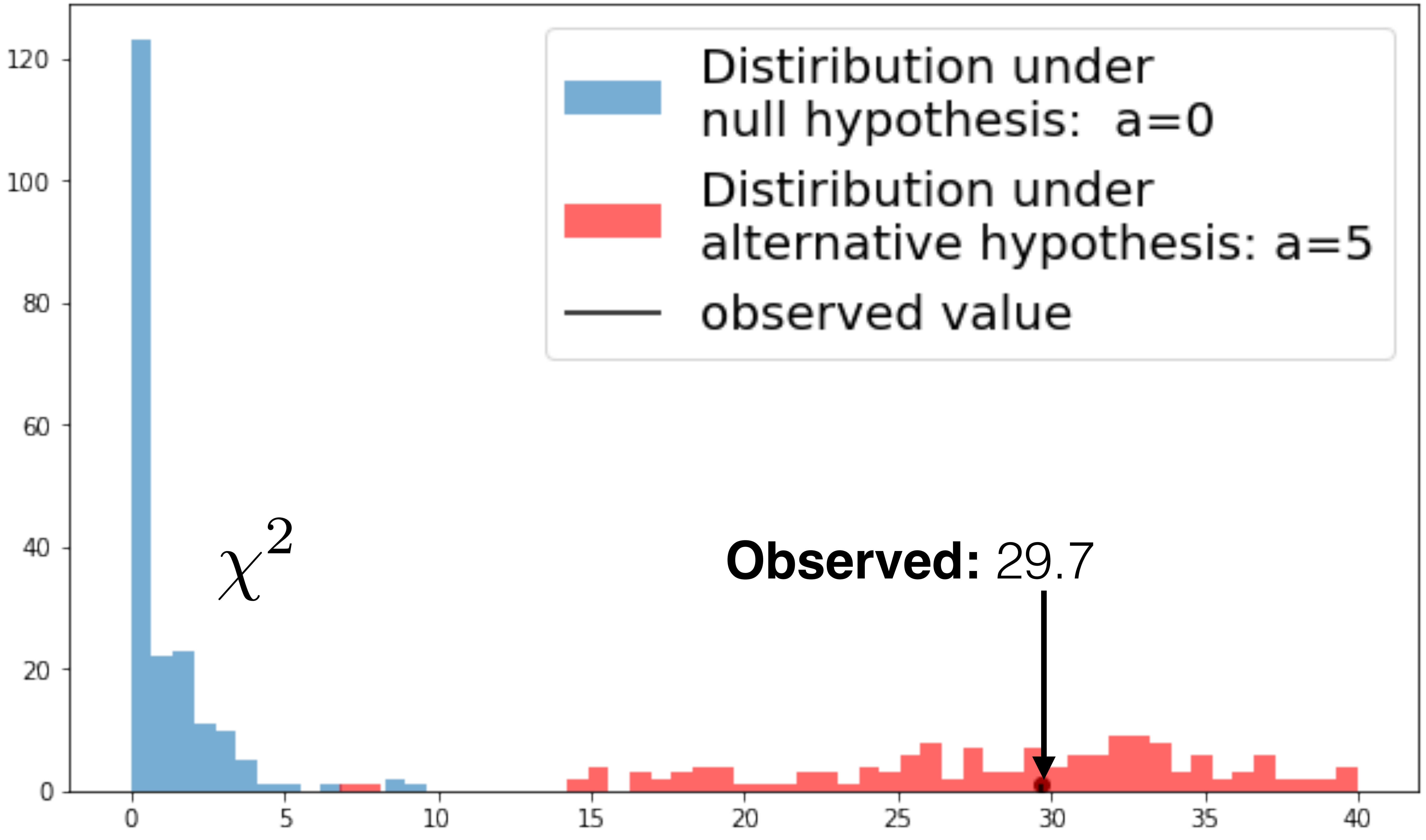
$$\text{p-value} = \int_{29.7}^{\infty} dx\ \chi^2(x) \simeq 5 \cdot 10^{-8}$$

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"

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

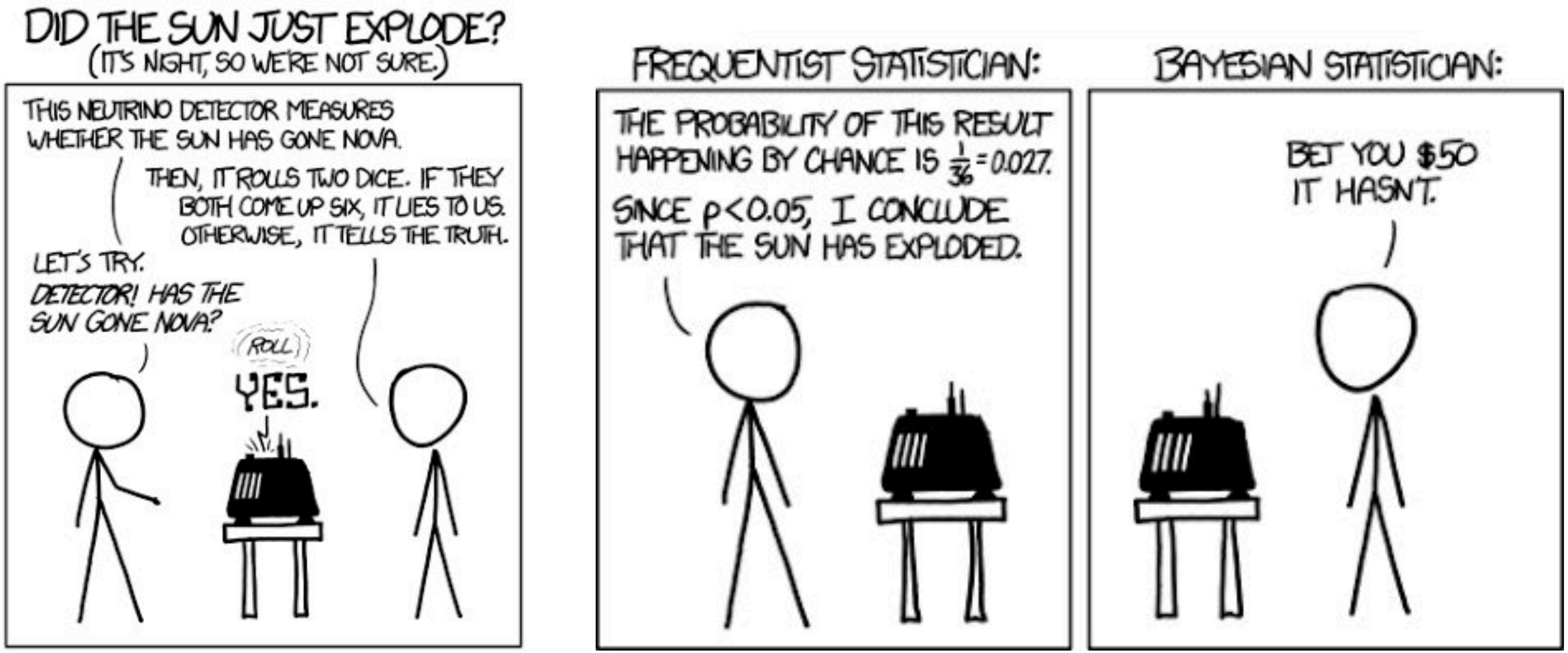
Notice that $\sqrt{29.7} \simeq 5.45$

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.