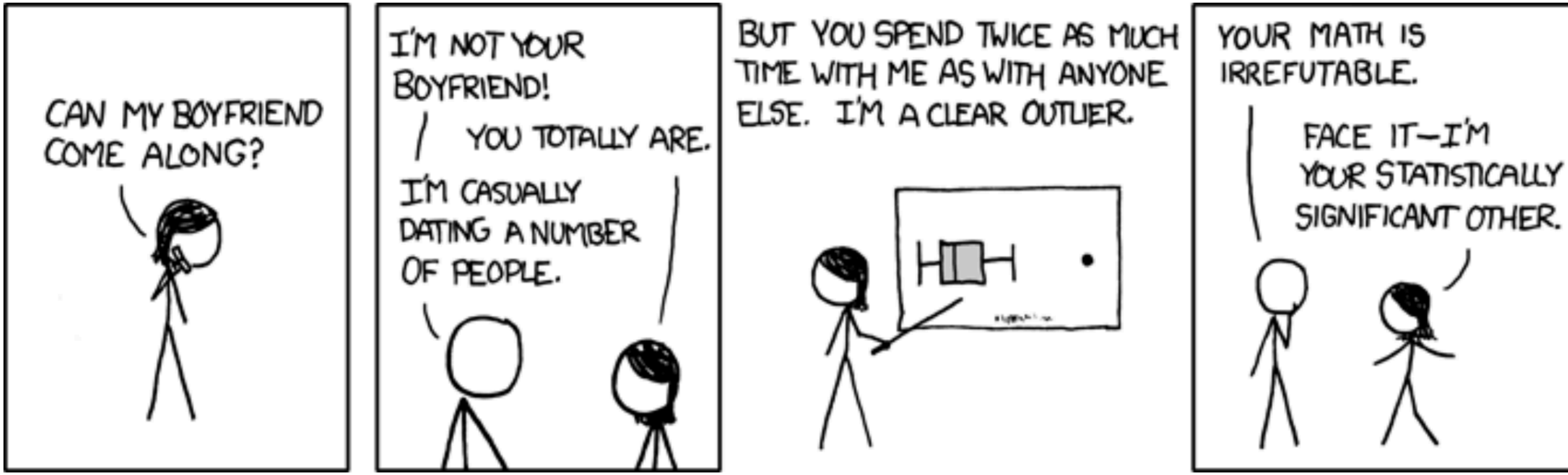
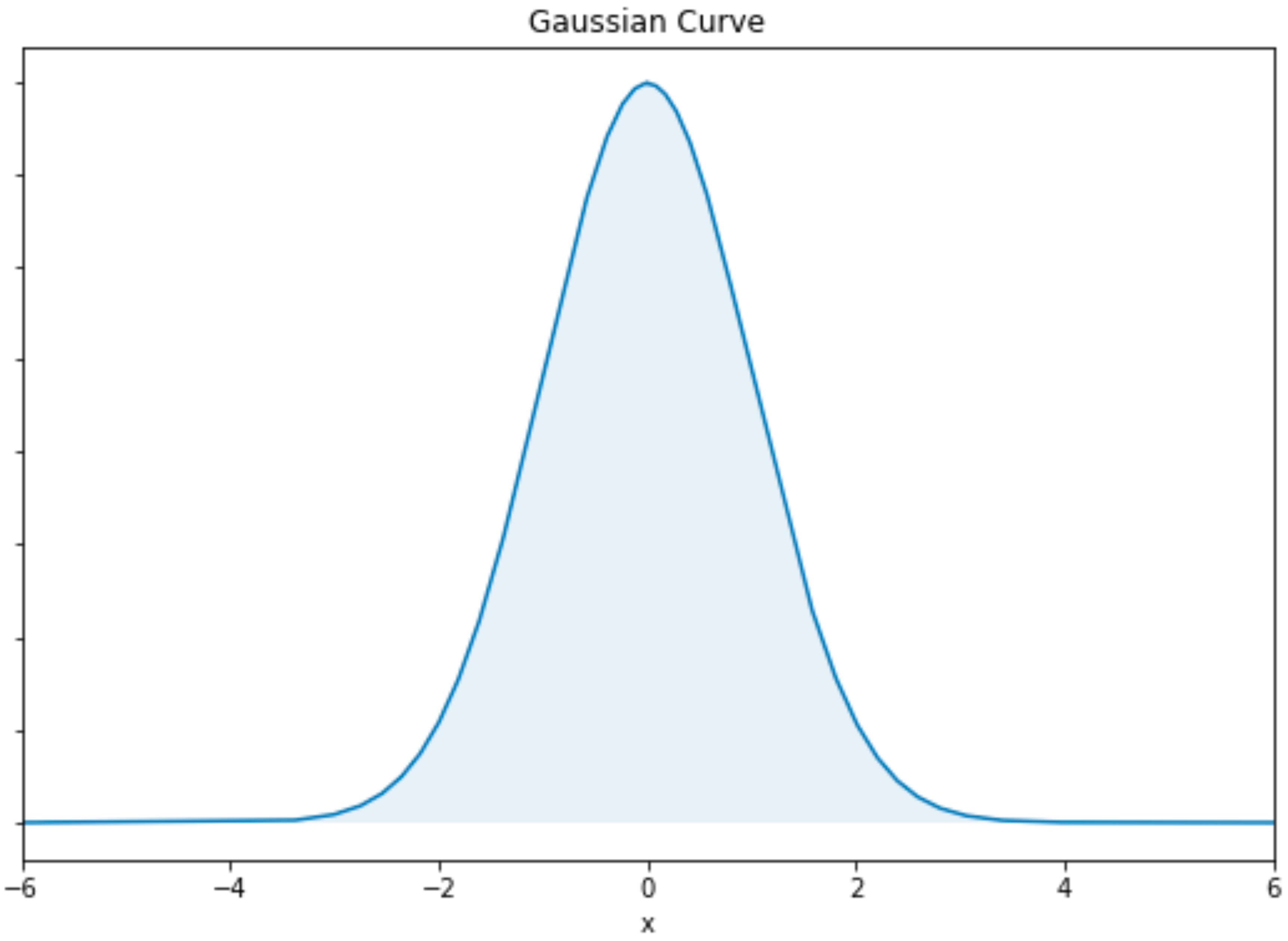


Lecture 10: Hypothesis Testing

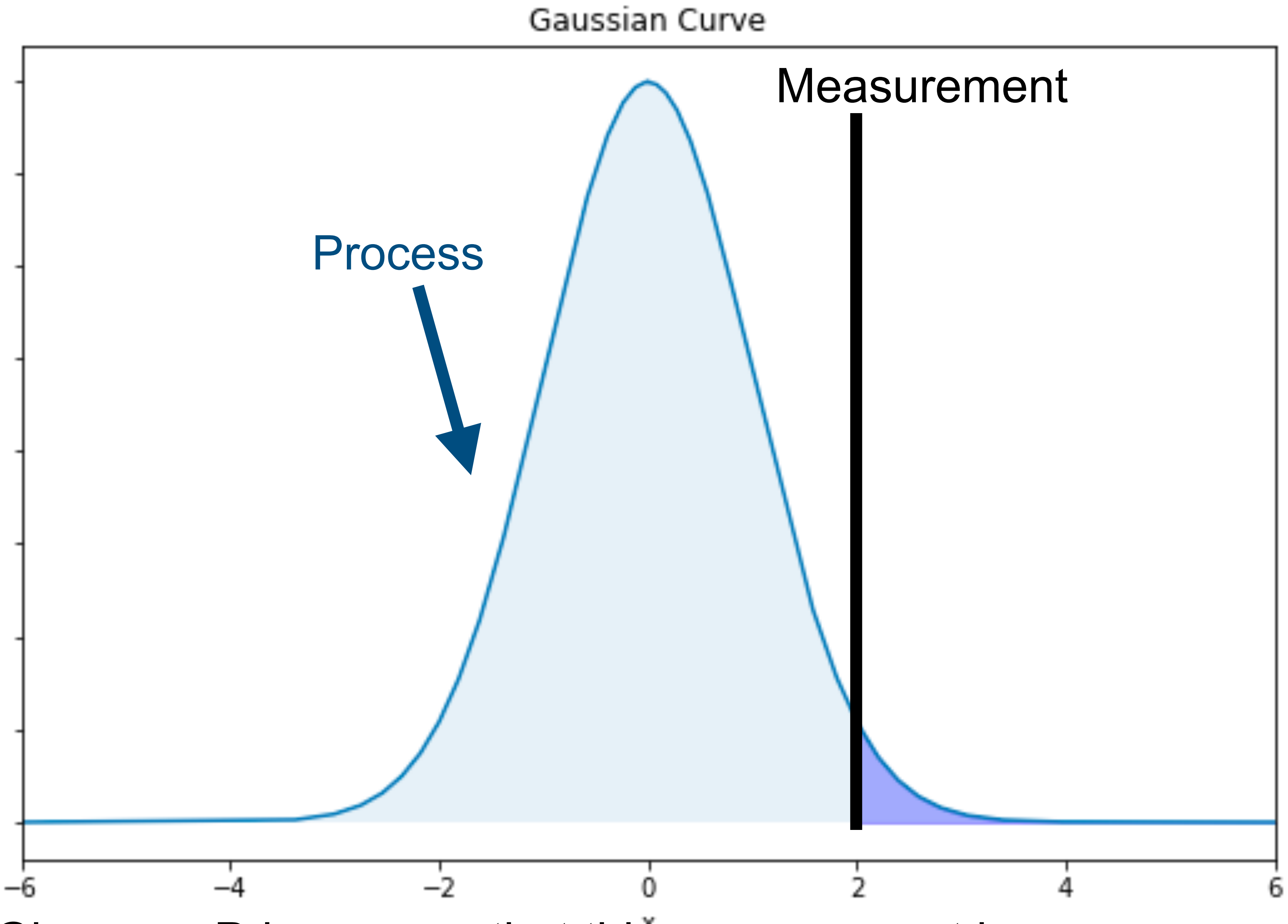
What is a measurement?

- Lets say we have a process that we believe behaves like this
  - This process is what we call our **Prior**



What is a measurement?

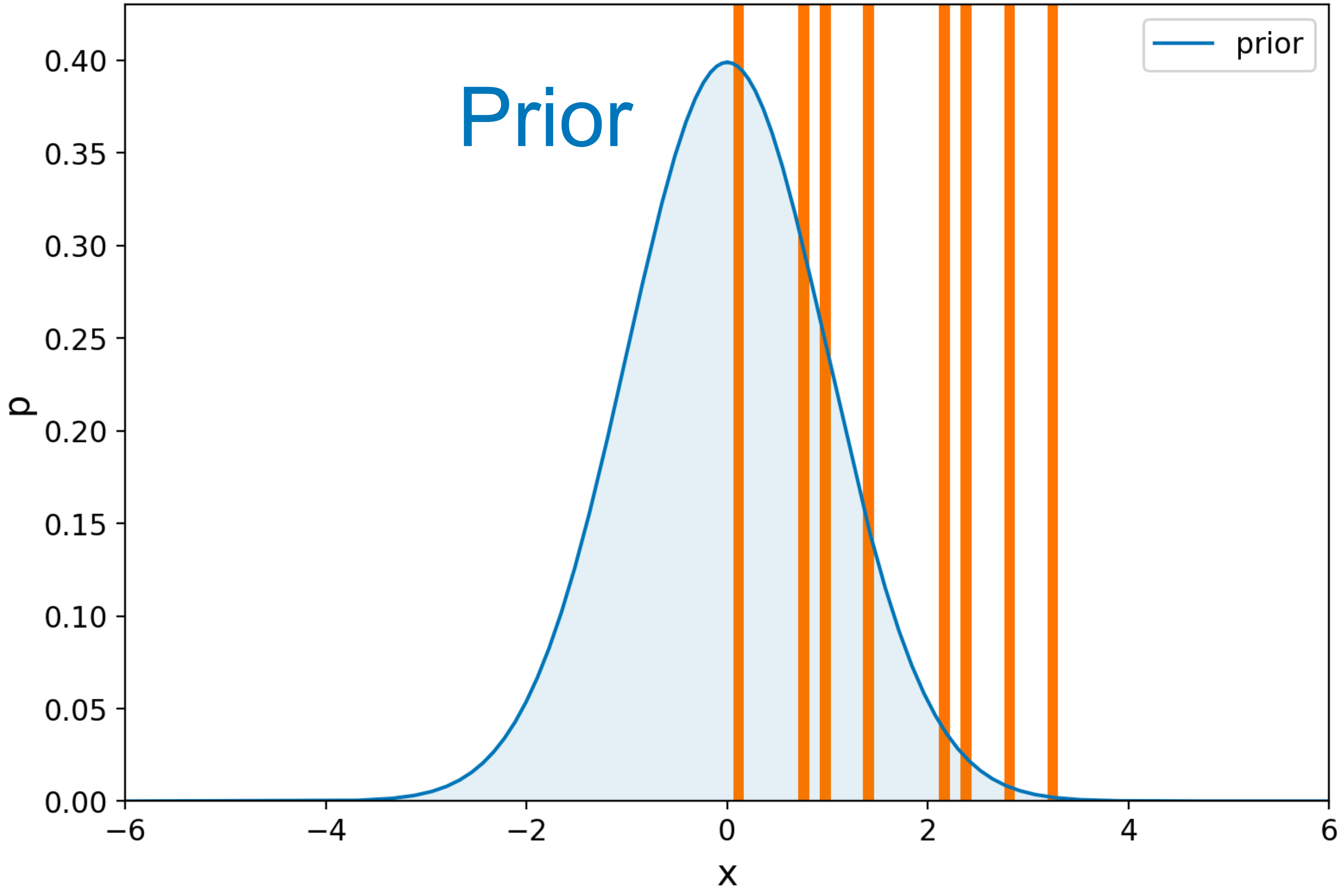
- We have assumed a model for this process
  - Example what is the direction of your eye reading slide?



Given our Prior we say that this measurement has a z-score of 2

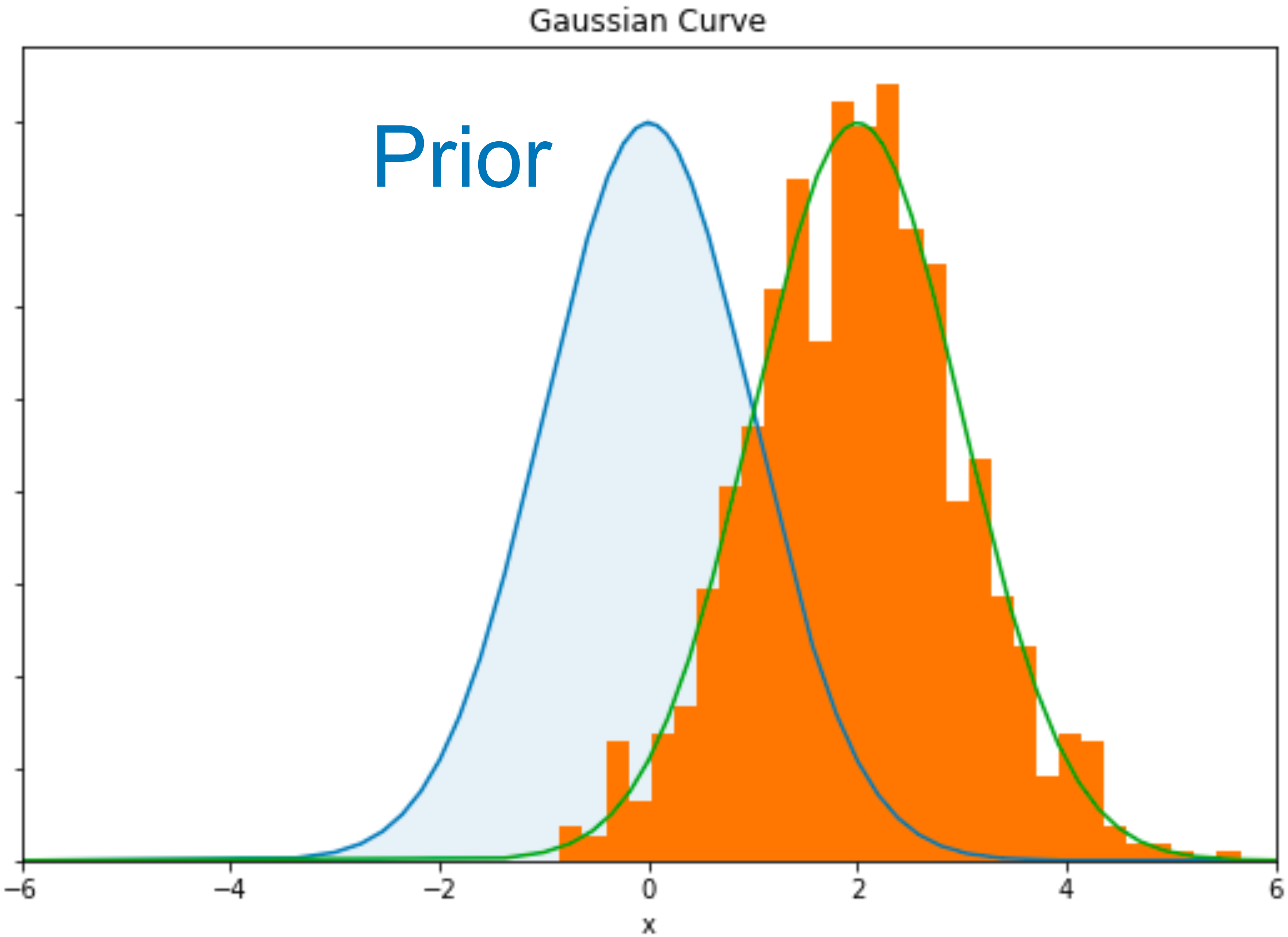
What happens with many measurements?

- What if the distribution we observe is different?



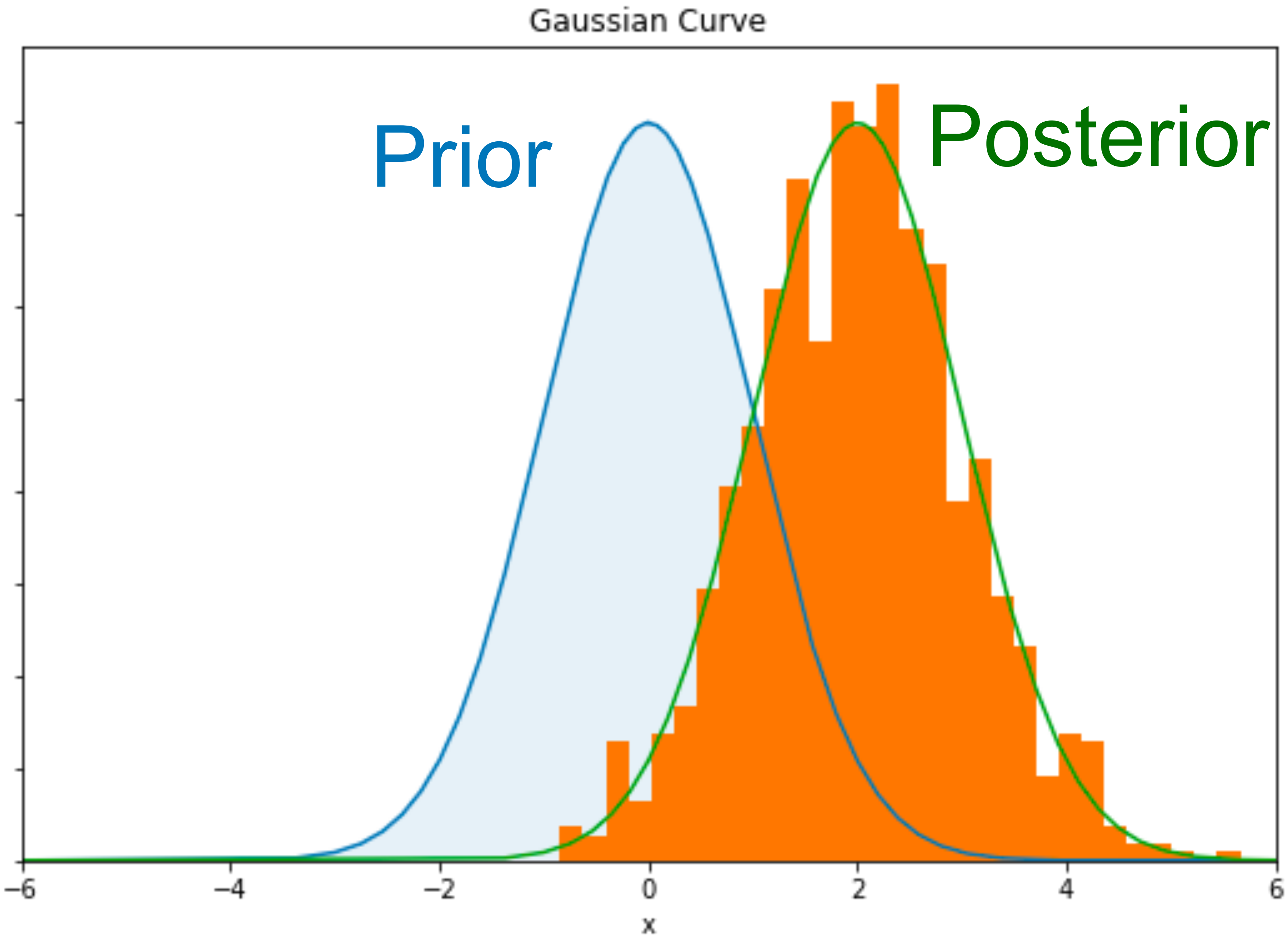
What happens with many measurements?

- What if the distribution we observe is different?

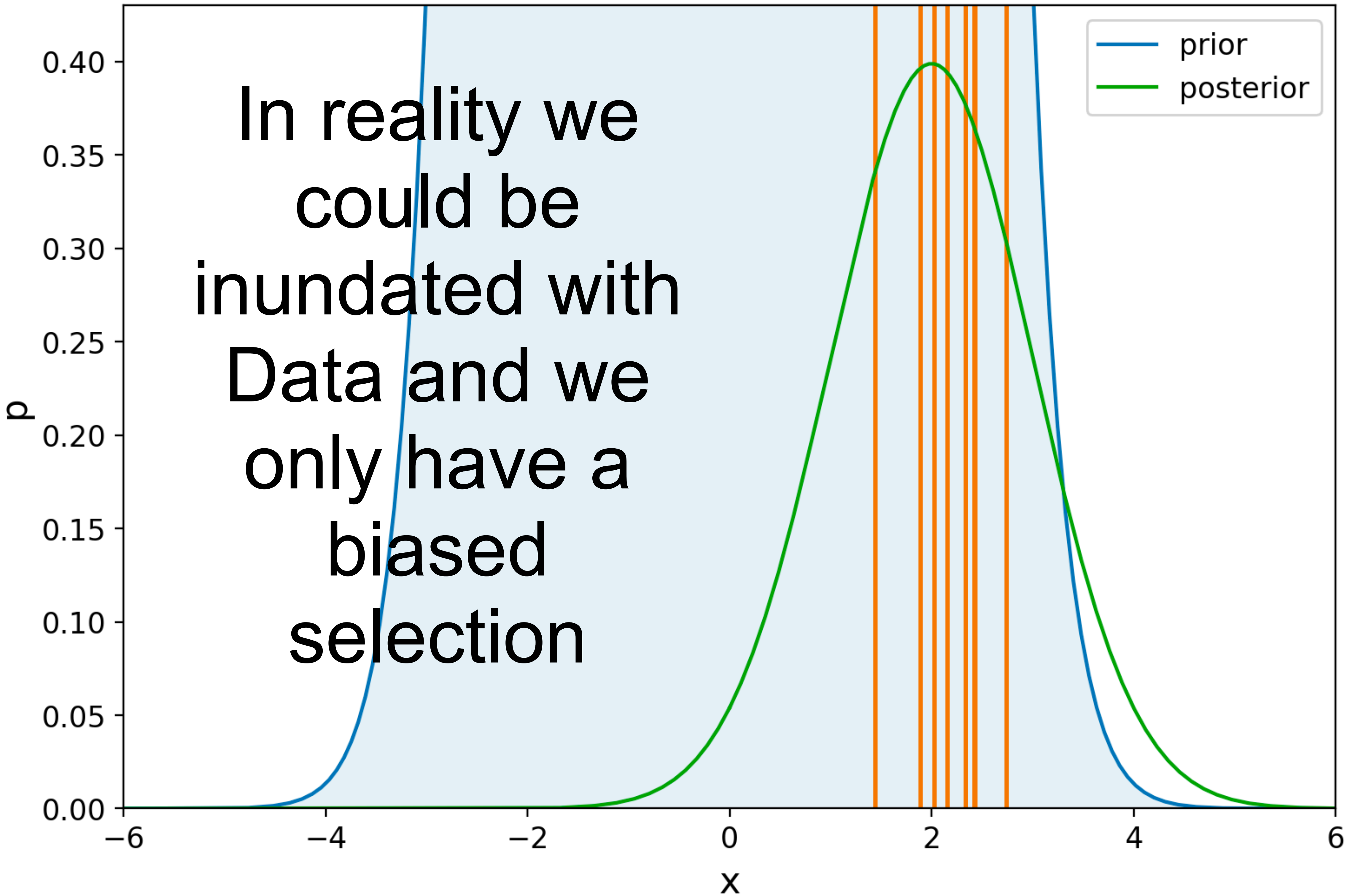


What happens with many measurements?

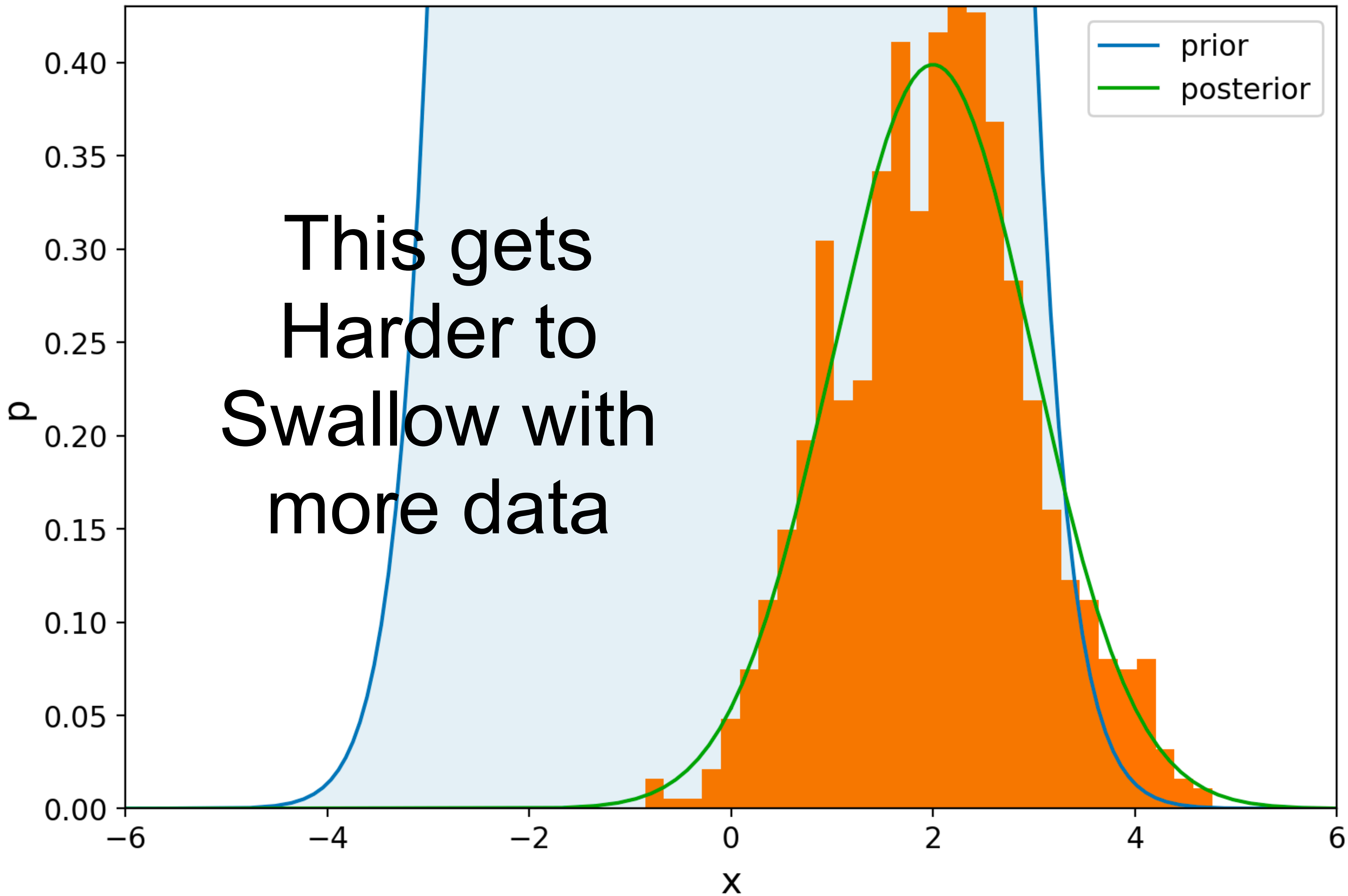
- What if the distribution we observe is different?



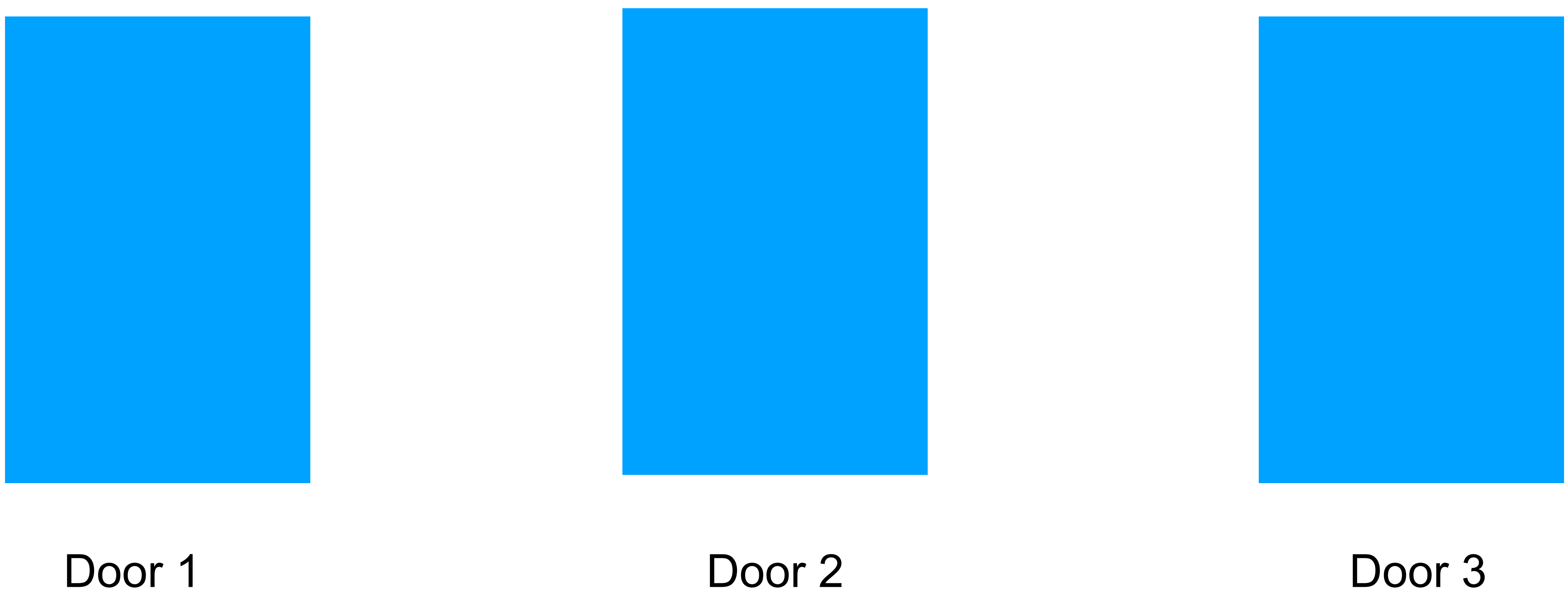
Are we missing something?



Are we missing something?

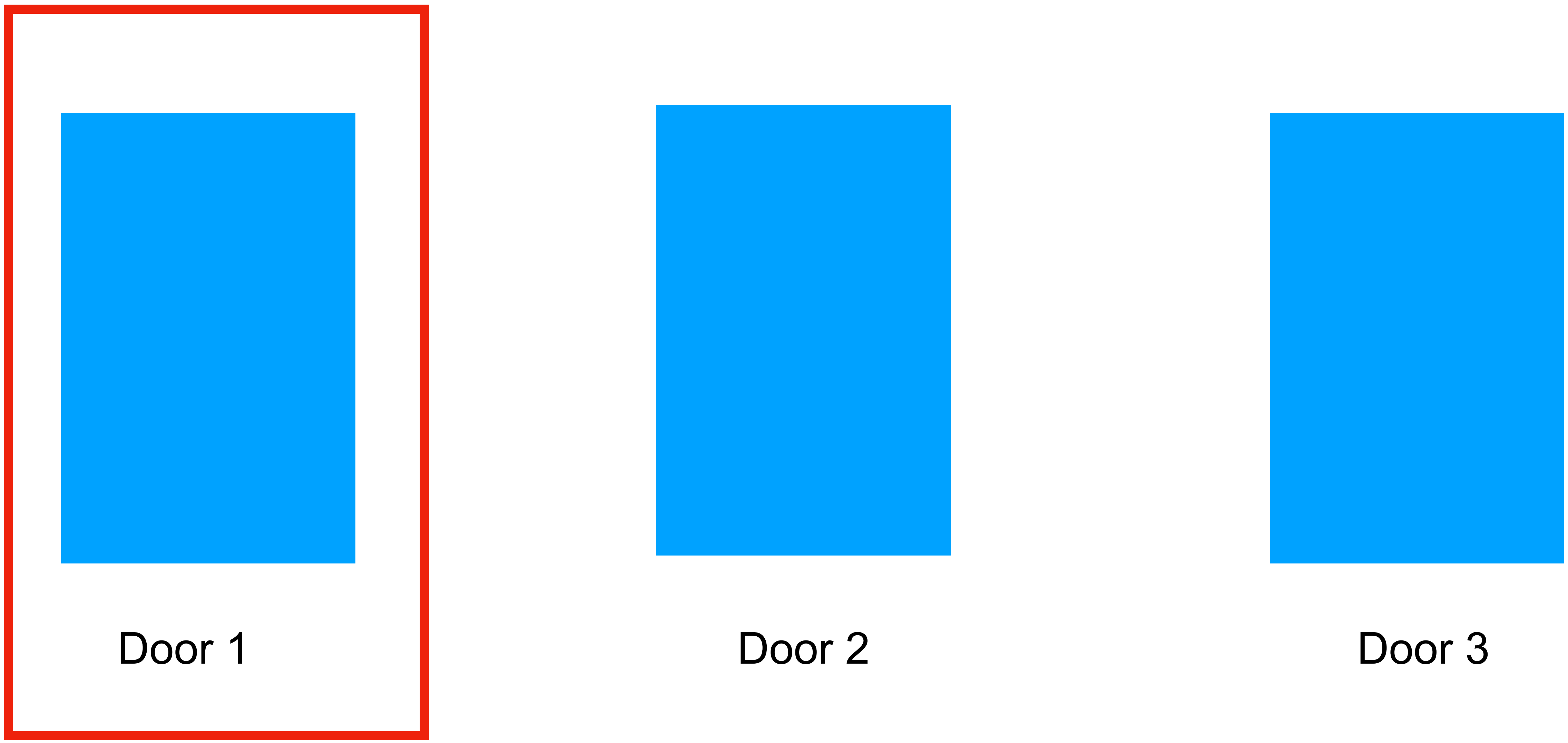


Monty Hall Problem



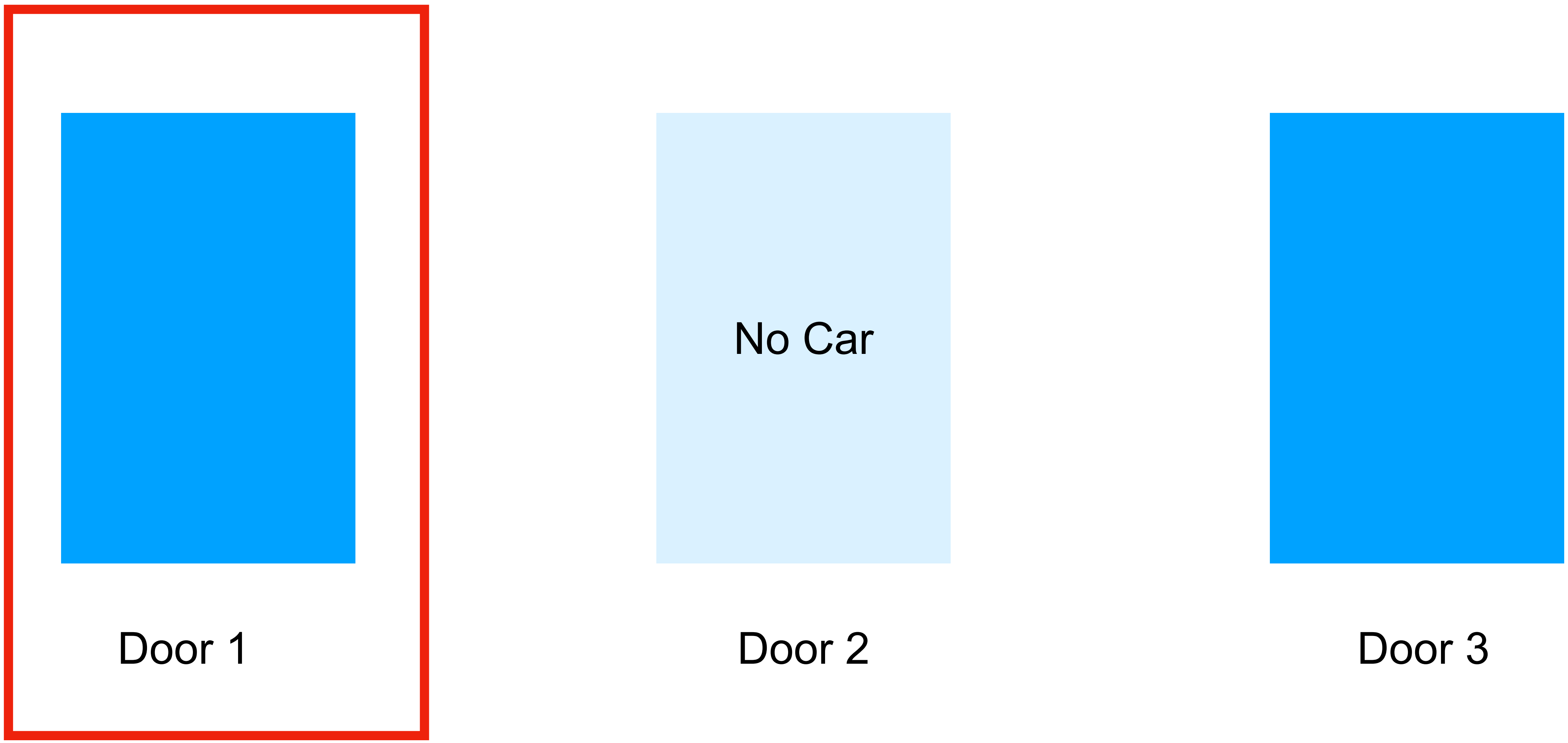
There is a car behind one of these doors?
Which one do you choose?

Monty Hall Problem



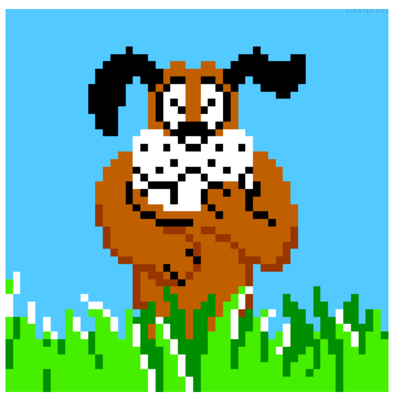
There is a car behind one of these doors?
Which one do you choose?

Monty Hall Problem



Now we open one door and it doesn’t have a car
Do you stay on Door 1 or do you go to Door 3

Monty Hall Problem



No Car

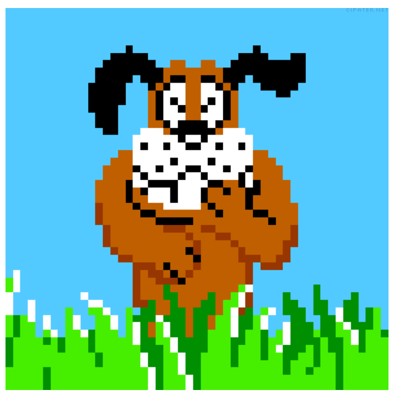
Door 1

Door 2

Door 3

Now we open one door and it doesn't have a car
Do you stay on Door 1 or do you go to Door 2

Monty Hall Problem



No Car

Door 1

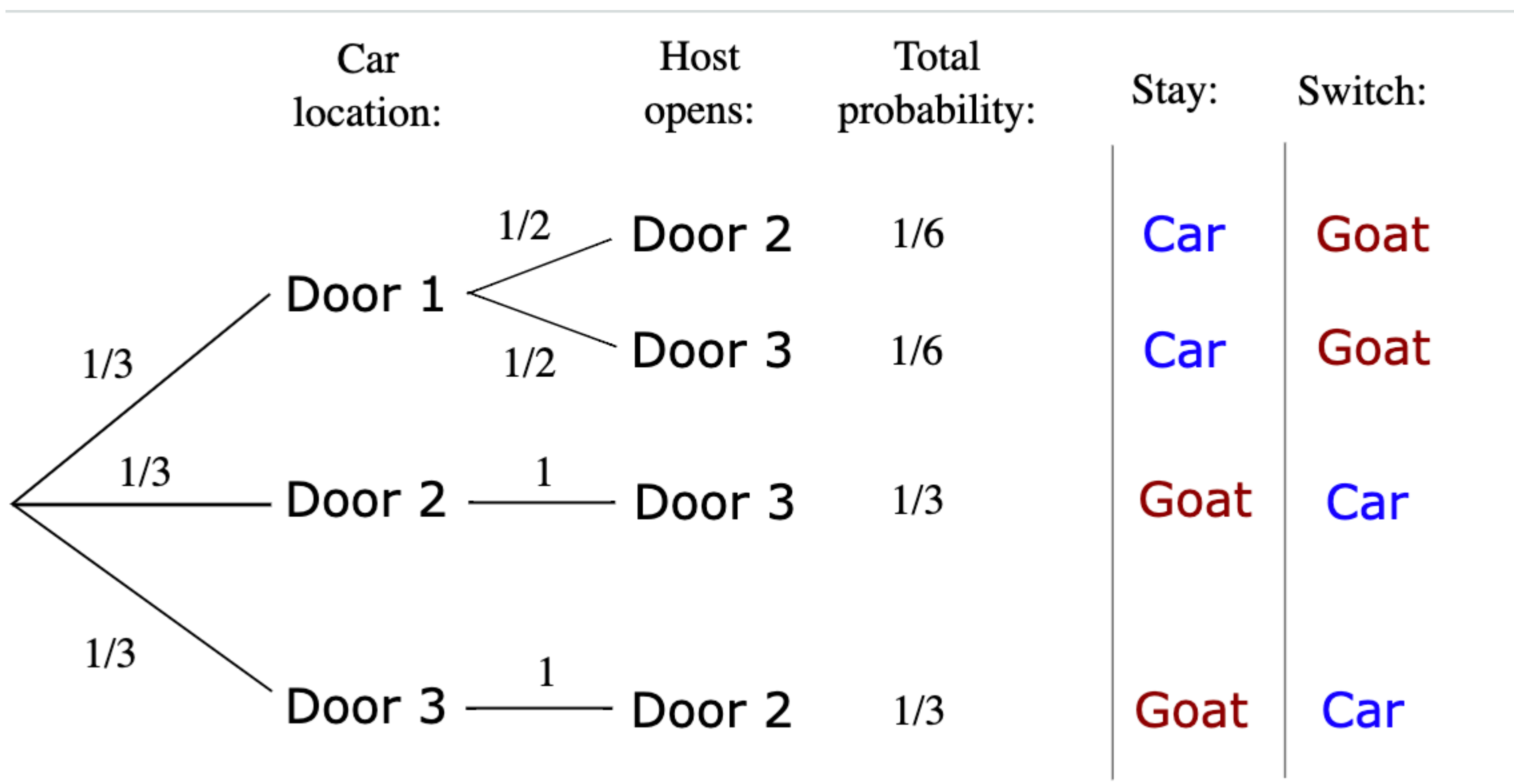
Door 2

Door 3

Now we open one door and it doesn’t have a car
Do you stay on Door 1 or do you go to Door 2

Monty Hall Problem

- Let's make a deal
- https://www.youtube.com/watch?v=iBdjqtR2iK4
- https://medium.com/@NickDoesData/applying-bayes-theorem-simulating-the-monty-hall-problem-with-python-5054976d1fb5



What happens with many measurements?

- How did we go from Prior to Posterior?

- For many cases this is an application of bayes theorem

$$\boxed{P(\mathcal{H}|\mathcal{D})} = \frac{\boxed{P(\mathcal{D}|\mathcal{H})}\,\boxed{P(\mathcal{H})}}{P(\mathcal{D})}$$

$$\boxed{P(\text{hypothesis}|\text{data})} = \frac{\boxed{P(\text{data}|\text{hypothesis})}\,\boxed{P(\text{hypothesis})}}{P(\text{data})}$$

Prior

Likelihood

Posterior

-

What happens with many measurements?

- How did we go from Prior to Posterior?
- For many cases this is an application of bayes theorem

$$P(\mathcal{H}|\mathcal{D}) = \frac{\boxed{P(\mathcal{D}|\mathcal{H})}\,\boxed{P(\mathcal{H})}}{\boxed{P(\mathcal{D})}}$$

Prior

$$P(\text{hypothesis}|\text{data}) = \frac{\boxed{P(\text{data}|\text{hypothesis})}\,\boxed{P(\text{hypothesis})}}{\boxed{P(\text{data})}}$$

Likelihood

Posterior

Normalizer

$$\boxed{\text{posterior} \;\propto\; \text{likelihood} \times \text{prior}}$$

What happens with many measurements?

- How did we go from Prior to Posterior?
- For many cases this is an application of bayes theorem

$$\boxed{P(\mathcal{H}|\mathcal{D})} = \frac{P(\mathcal{D}|\mathcal{H})\boxed{P(\mathcal{H})}}{P(\mathcal{D})}$$

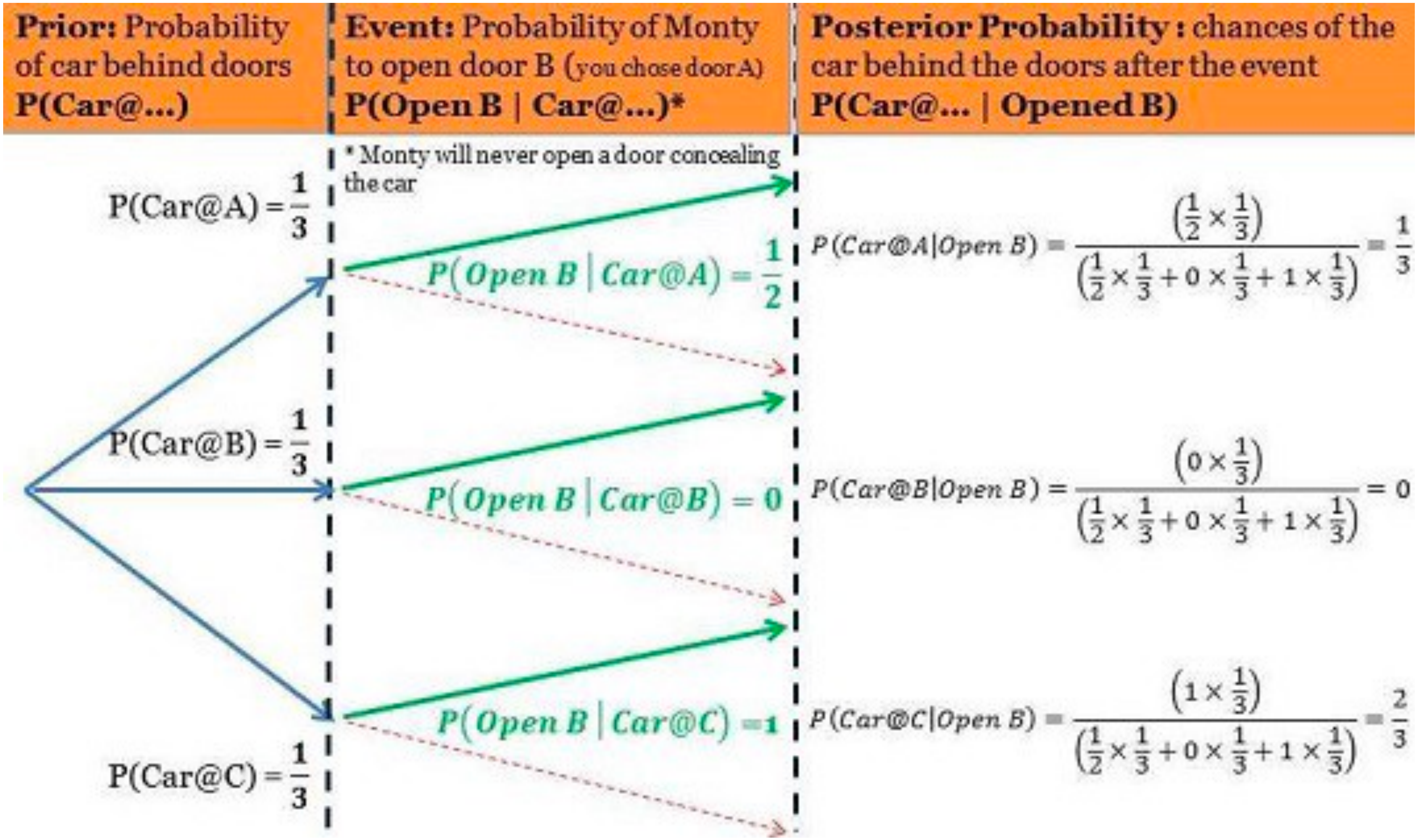
Prior

$$\boxed{P(\text{hypothesis}|\text{data})} = \frac{P(\text{data}|\text{hypothesis})\boxed{P(\text{hypothesis})}}{P(\text{data})}$$

Posterior

- Lets go back to Monty Hall
  - Re-evaluate

Monty Hall Problem



What is a good Prior?

Cornell University

arXiv > hep-ph > arXiv:2203.13180

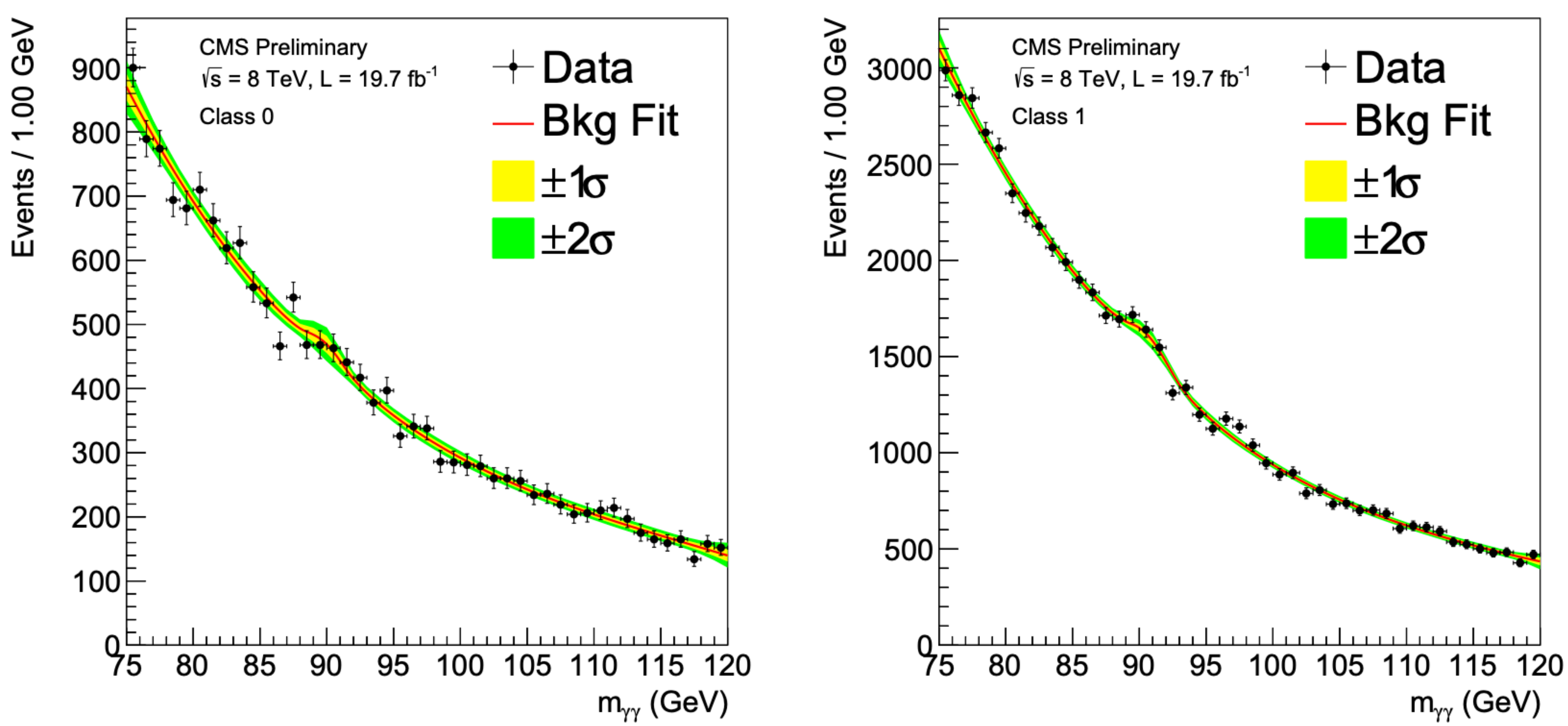
High Energy Physics – Phenomenology

[Submitted on 24 Mar 2022 (v1), last revised 21 Aug 2022 (this version, v2)]

Mounting evidence for a 95 GeV Higgs boson

Thomas Biekötter, Sven Heinemeyer, Georg Weiglein

In 2018 CMS reported an excess in the light Higgs-boson search in the diphoton decay mode at about 95GeV based on Run 1 and first year Run 2 data. The combined local significance of the excess was $2.8\,\sigma$. The excess is compatible with the limits obtained in the ATLAS searches from the diphoton search channel. Recently, CMS reported another local excess with a significance of $3.1\,\sigma$ in the light Higgs-boson search in the di-tau final state, which is compatible with the interpretation of a Higgs boson with a mass of about 95GeV. We show that the observed results can be interpreted as manifestations of a Higgs boson in the Two-Higgs Doublet Model with an additional real singlet (N2HDM). We find that the lightest Higgs boson of the N2HDM can fit both excesses simultaneously, while the second-lightest state is such that it satisfies the Higgs-boson measurements at 125GeV, and the full Higgs-boson sector is compatible with all Higgs exclusion bounds from the searches at LEP, the Tevatron and the LHC as well as with other theoretical and experimental constraints. Finally, we demonstrate that it is furthermore possible to accommodate the excesses observed by CMS in the two search channels together with a local $2.3\,\sigma$ excess in the $b\bar{b}$ final state observed at LEP in the same mass range.



(a) Class 0 (b) Class 1

A Strange LHC Excess

arXiv > hep-ph > arXiv:2203.13180

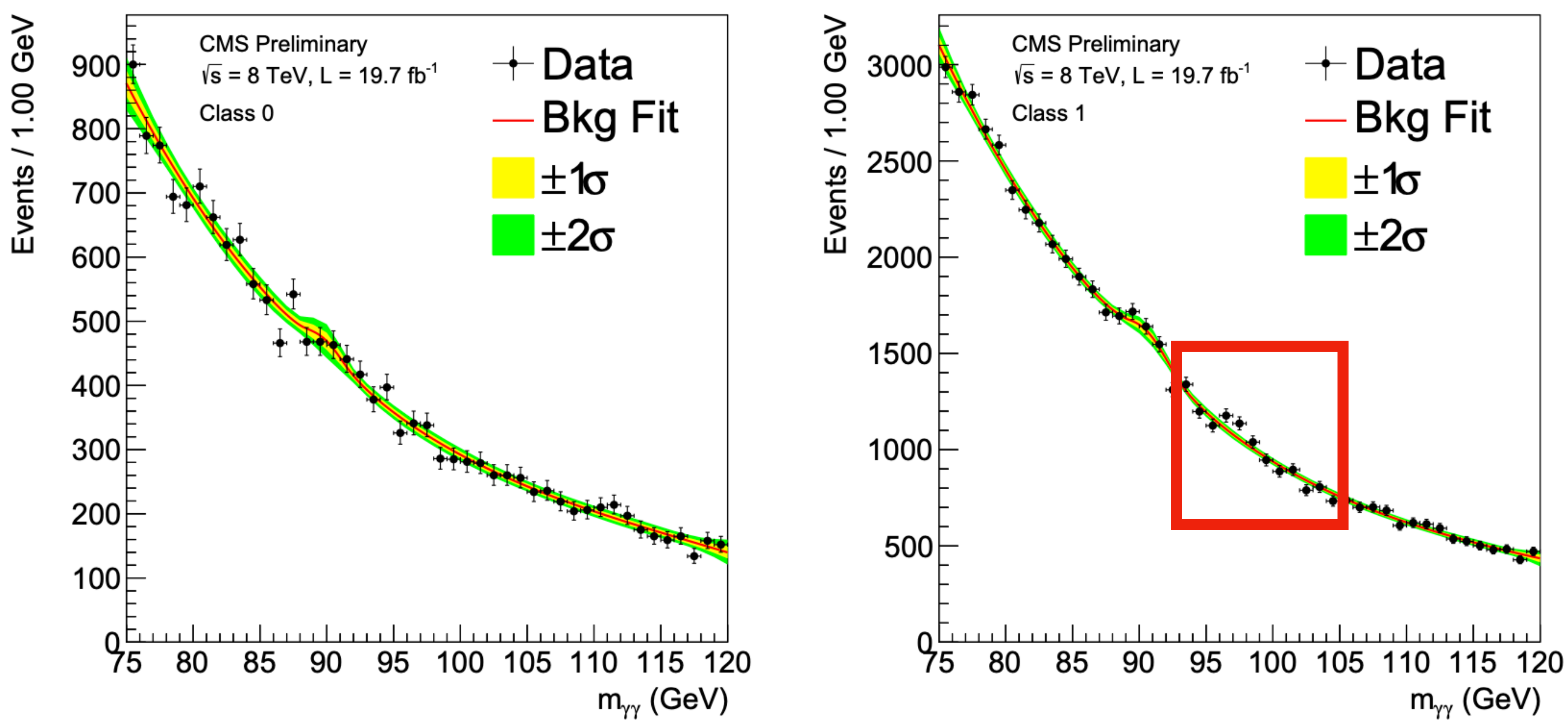
High Energy Physics – Phenomenology

[Submitted on 24 Mar 2022 (v1), last revised 21 Aug 2022 (this version, v2)]

Mounting evidence for a 95 GeV Higgs boson

Thomas Biekötter, Sven Heinemeyer, Georg Weiglein

In 2018 CMS reported an excess in the light Higgs-boson search in the diphoton decay mode at about 95GeV based on Run 1 and first year Run 2 data. The combined local significance of the excess was $2.8\,\sigma$. The excess is compatible with the limits obtained in the ATLAS searches from the diphoton search channel. Recently, CMS reported another local excess with a significance of $3.1\,\sigma$ in the light Higgs-boson search in the di-tau final state, which is compatible with the interpretation of a Higgs boson with a mass of about 95GeV. We show that the observed results can be interpreted as manifestations of a Higgs boson in the Two-Higgs Doublet Model with an additional real singlet (N2HDM). We find that the lightest Higgs boson of the N2HDM can fit both excesses simultaneously, while the second-lightest state is such that it satisfies the Higgs-boson measurements at 125GeV, and the full Higgs-boson sector is compatible with all Higgs exclusion bounds from the searches at LEP, the Tevatron and the LHC as well as with other theoretical and experimental constraints. Finally, we demonstrate that it is furthermore possible to accommodate the excesses observed by CMS in the two search channels together with a local $2.3\,\sigma$ excess in the $b\bar{b}$ final state observed at LEP in the same mass range.



(a) Class 0 (b) Class 1

Bayesian vs Frequentists

Probability (mathematics)

$$\boxed{P(H|D) = \frac{P(D|H)P(H)}{P(D)}}$$

Everyone uses Bayes' formula when the prior $P(H)$ is known.

Bayesian path

Frequentist path

Statistics (art)

$$\boxed{P_{\text{Posterior}}(H|D) = \frac{P(D|H)P_{\text{prior}}(H)}{P(D)}}$$

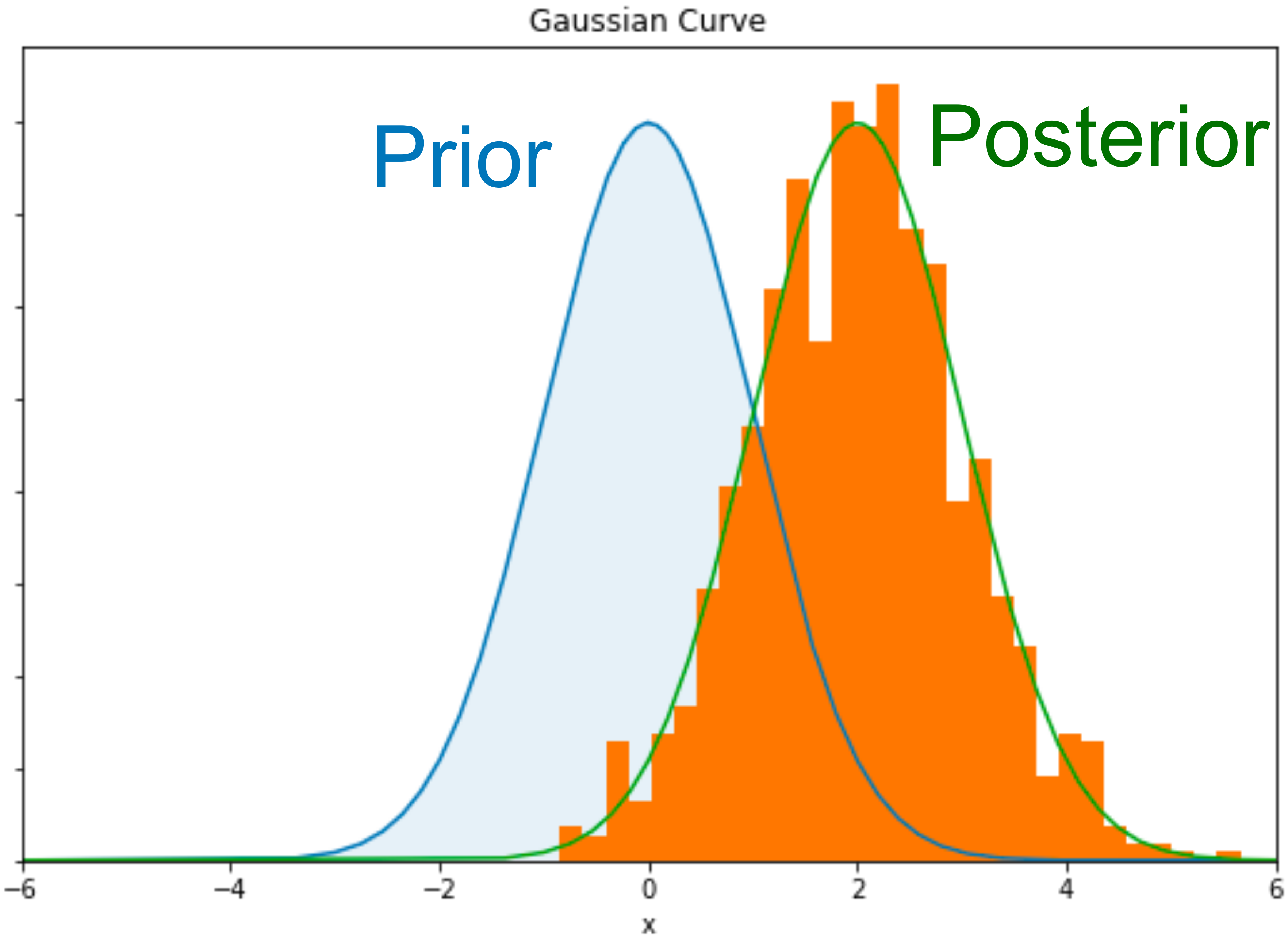
Bayesians require a prior, so they develop one from the best information they have.

$$\boxed{\text{Likelihood } L(H;D) = P(D|H)}$$

Without a known prior frequentists draw inferences from just the likelihood function.

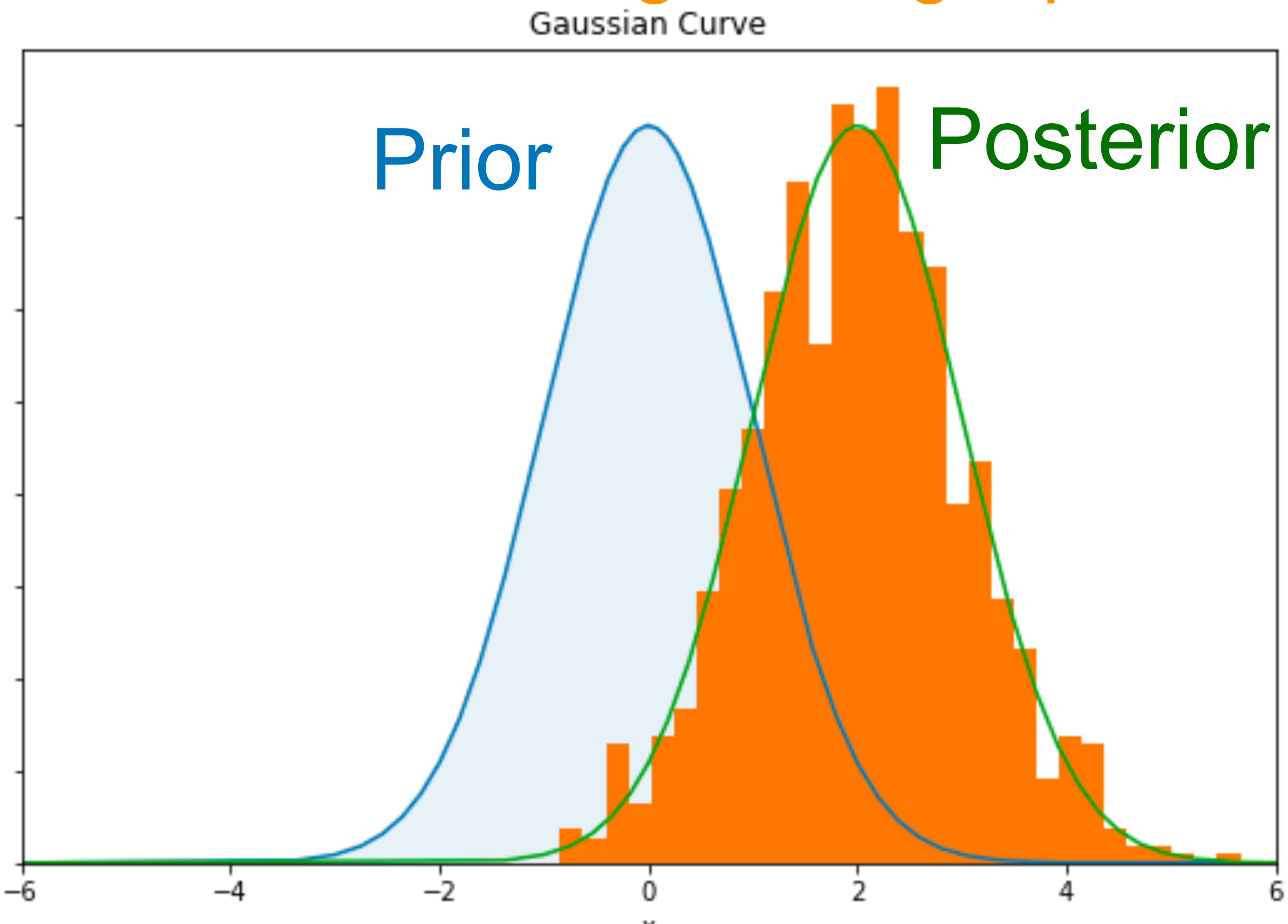
Bayesian View

My posterior is the result of some selection bias of a deeper model
I just need to fix my conditional probabilities



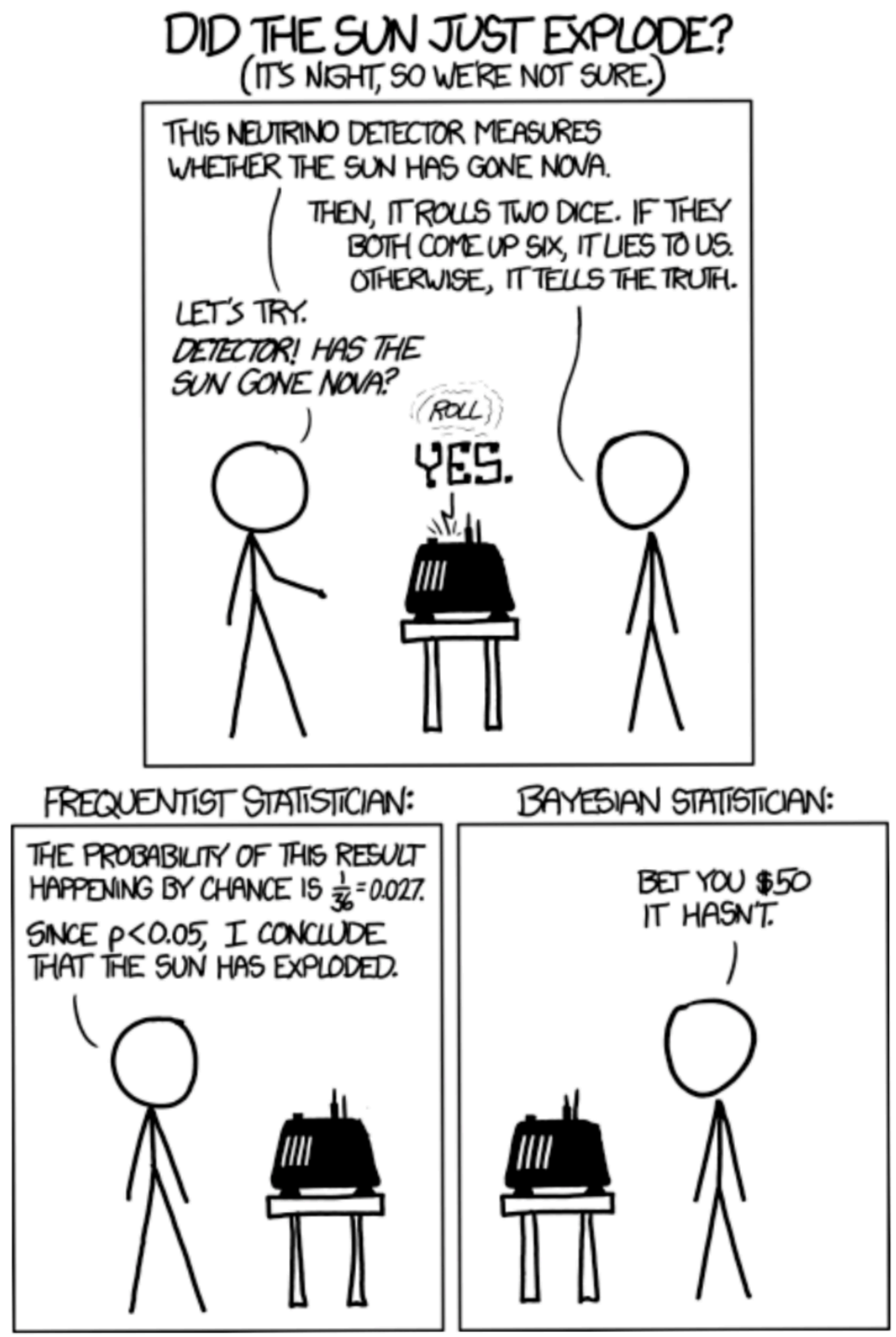
Frequentist View

All we know is our result gives us a gaussian at 2
Bayesian stats people are a bunch of
smart asses guessing a prior



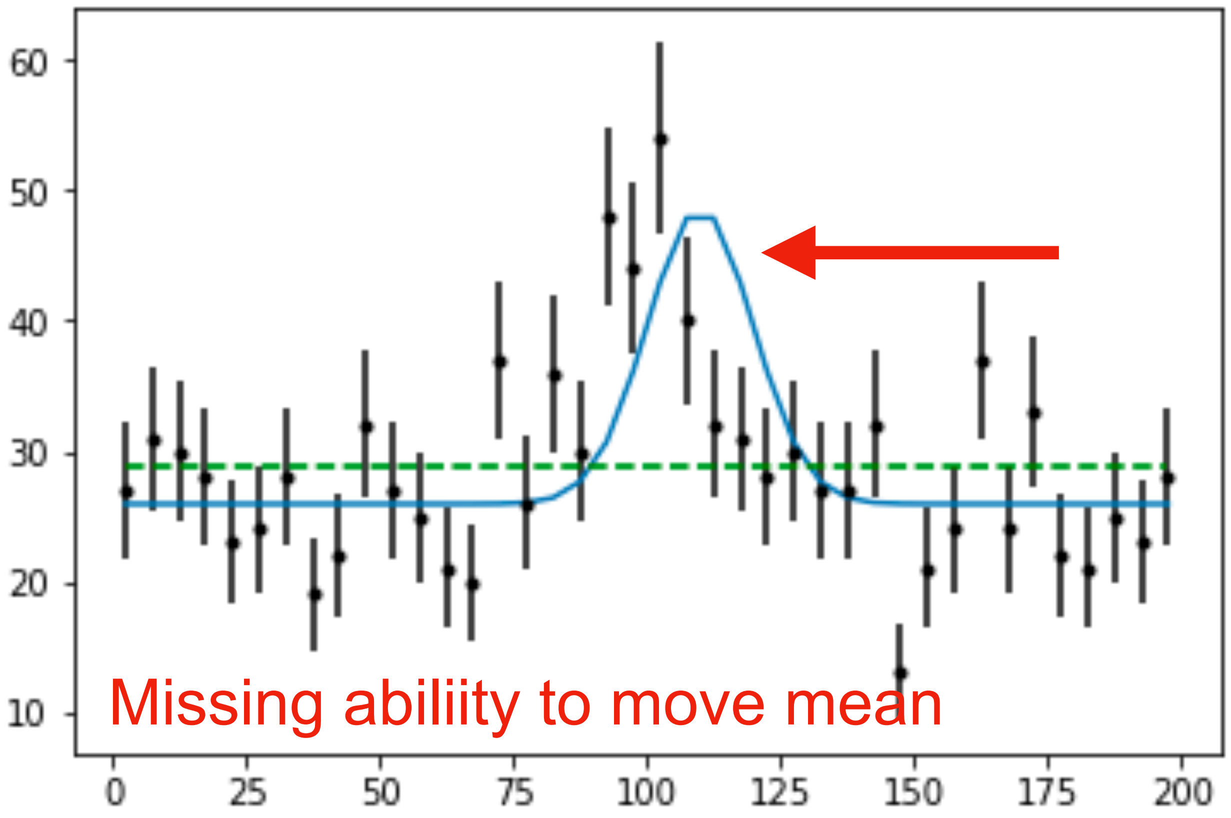
Bayesian Mean=0 and Std Dev=1, Frequentist Mean=2.002+/-0.06,StdDev=1.02+/-0.03

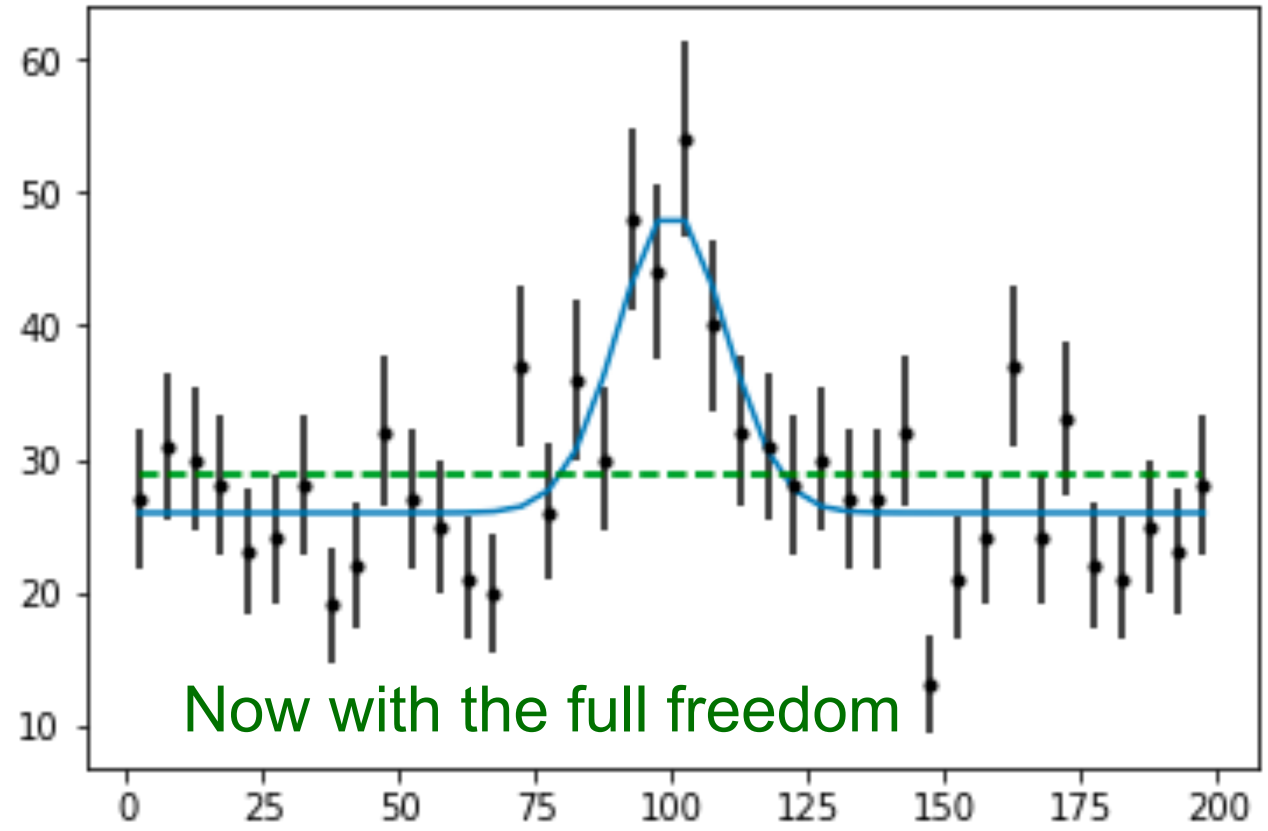
How to View it



What if models is incomplete?

- https://www.youtube.com/watch?v=Ku42Iszh9KM
- Sometimes there are other parameters not included in our model



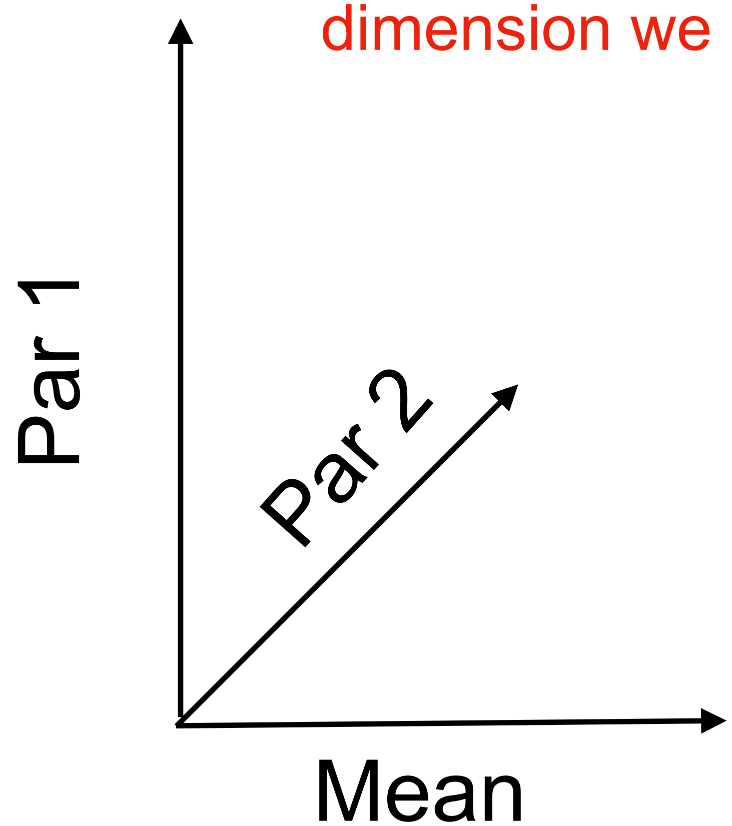


What if models is incomplete?

- https://www.youtube.com/watch?v=Ku42Iszh9KM

- Sometimes there are other parameters not included in our model

Bayesian view:
Each parameter adds a dimension we probe

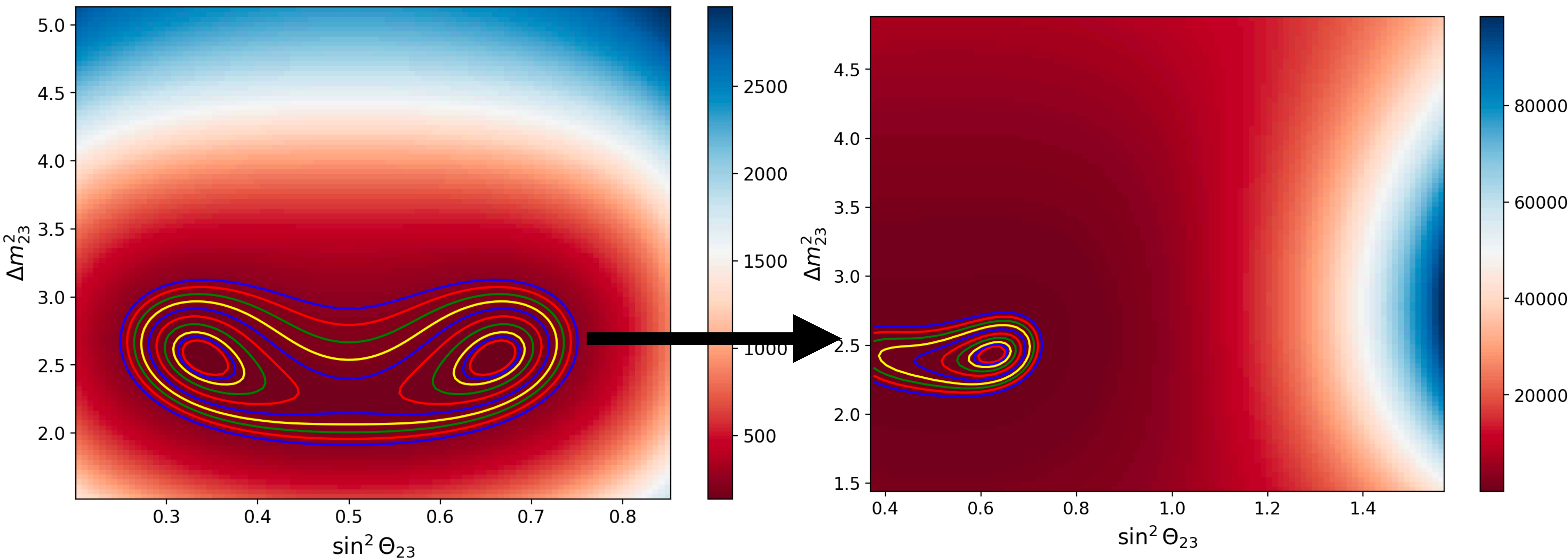


Frequestist View:
Make sure your model is complete

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right).$$

Float this parameter

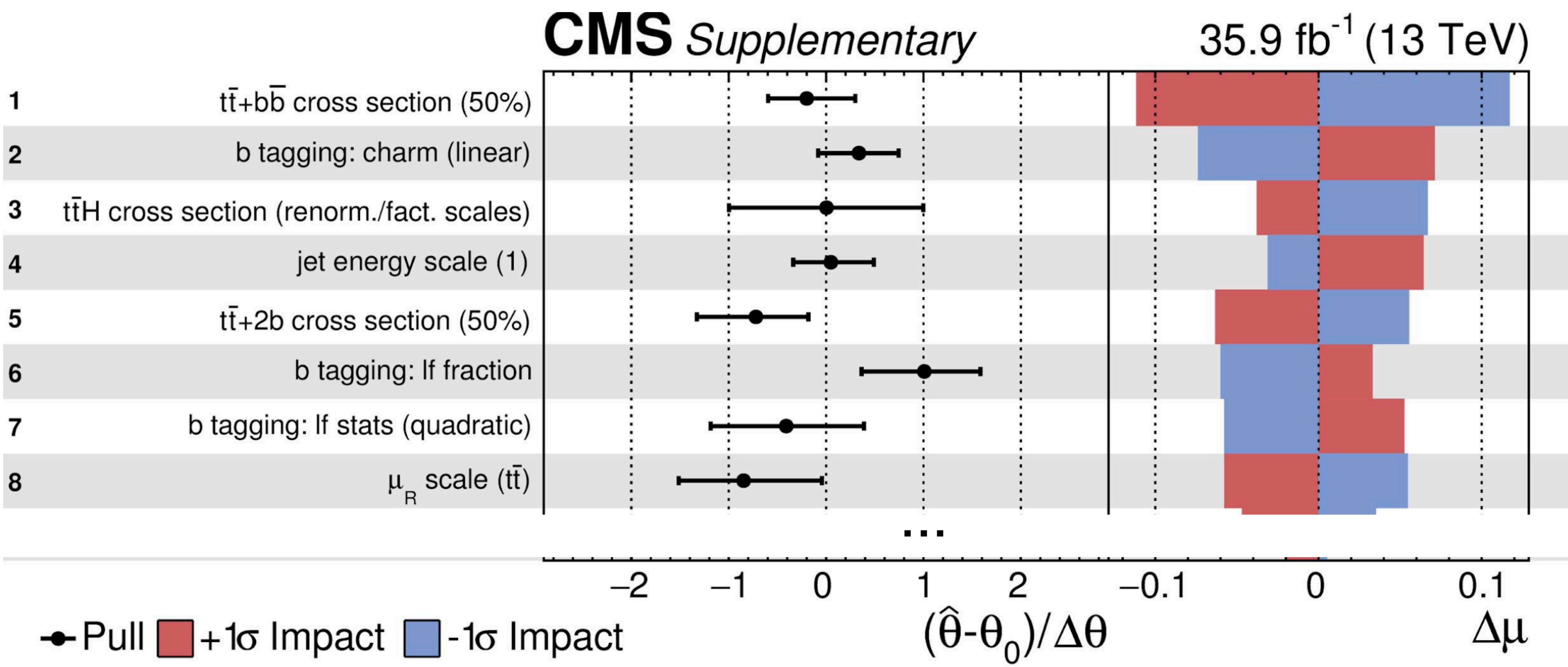
Nuisance



External Nuisances

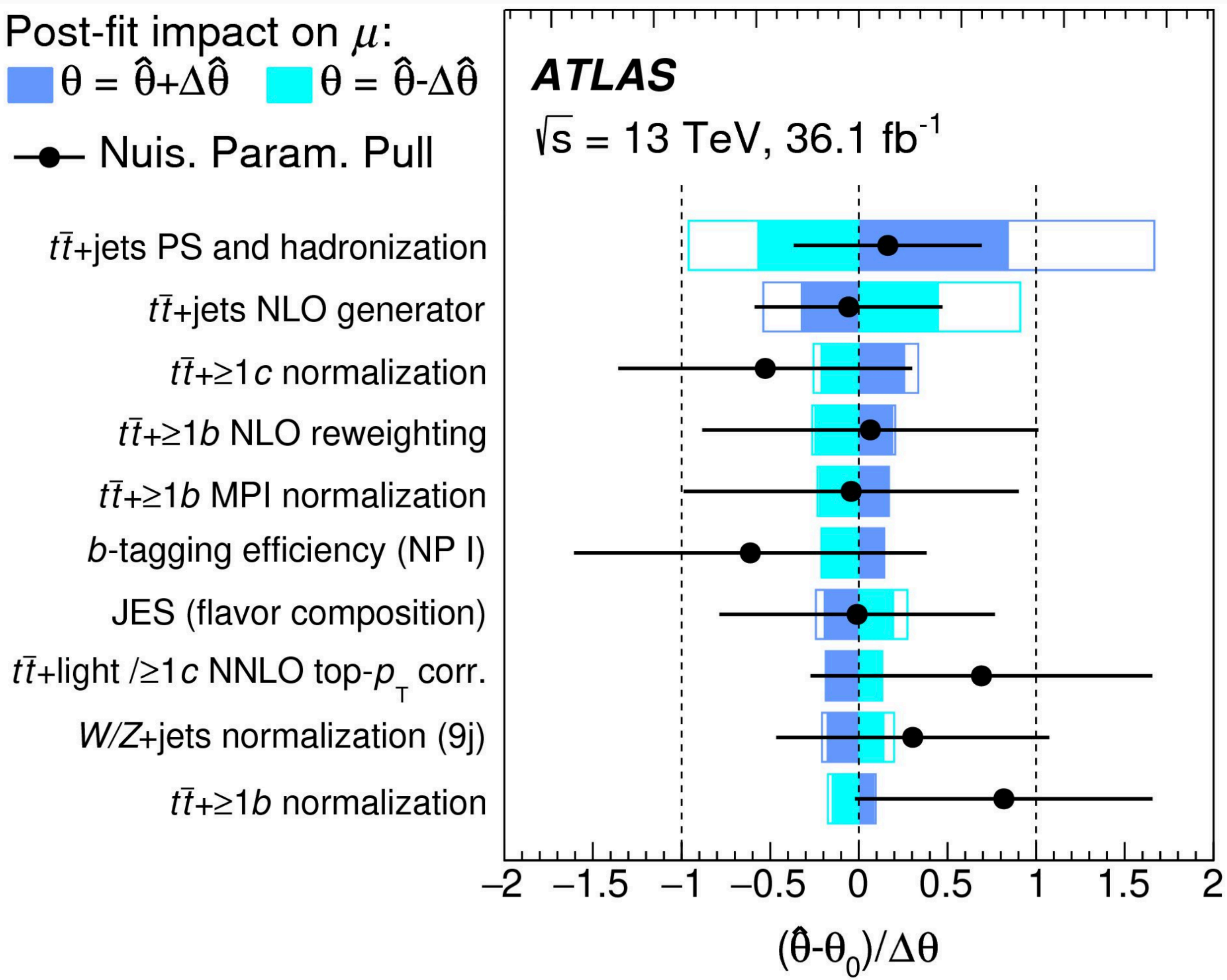
$$-2\log\left(\mathcal{L}(x|\vec{\theta})\right) = -2\log\left(\mathcal{L}(x|\vec{\theta})\right)_{\text{original}} + \frac{\left(\sin\theta_{23} - \sin\theta_{23}^{\text{best}}\right)^2}{\sigma^2_{\sin\theta_{23}}} + \frac{\left(\Delta m^2_{23} - \Delta m^{2\,\text{best}}_{23}\right)^2}{\sigma^{2\,\text{best}}_{\Delta m^2_{12}}}$$

Nuisance



CMS Supplementary
35.9 fb-1 (13 TeV)
1 tt̄+bb̄ cross section (50%)
2 b tagging: charm (linear)
3 tt̄H cross section (renorm./fact. scales)
4 jet energy scale (1)
5 tt̄+2b cross section (50%)
6 b tagging: lf fraction
7 b tagging: lf stats (quadratic)
8 μR scale (tt̄)
...
−2 −1 0 1 2
−0.1 0 0.1
(θ̂-θ0)/Δθ
Δμ
Pull
+1σ Impact
-1σ Impact

Nuisance



Post-fit impact on μ:
$\theta = \hat{\theta}+\Delta\hat{\theta}$
$\theta = \hat{\theta}-\Delta\hat{\theta}$
Nuis. Param. Pull
$t\bar{t}$+jets PS and hadronization
$t\bar{t}$+jets NLO generator
$t\bar{t}+\geq 1c$ normalization
$t\bar{t}+\geq 1b$ NLO reweighting
$t\bar{t}+\geq 1b$ MPI normalization
b-tagging efficiency (NP I)
JES (flavor composition)
$t\bar{t}$+light /$\geq 1c$ NNLO top-p_T corr.
W/Z+jets normalization (9j)
$t\bar{t}+\geq 1b$ normalization
ATLAS
$\sqrt{s}$ = 13 TeV, 36.1 fb^{-1}
–2 –1.5 –1 –0.5 0 0.5 1 1.5 2
$(\hat{\theta}-\theta_0)/\Delta\theta$

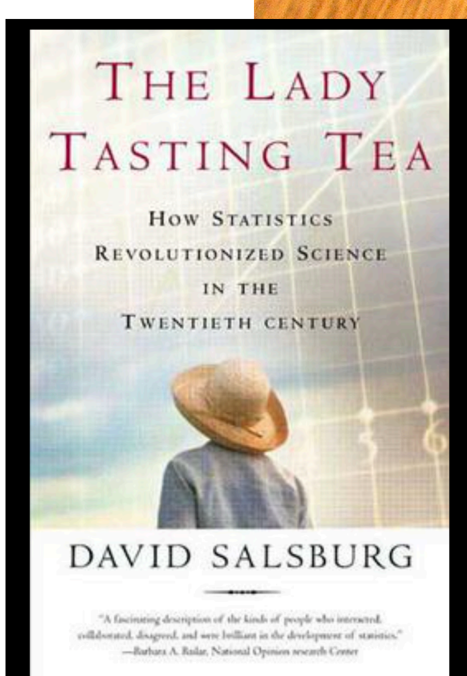
Hypothesis Test

- Null Hypothesis :
  - **$p_a(x) = p_b(x)$**
- Alternative Hypothesis
  - **$p_a(x) \neq p_b(x)$**
- Answers these questions:
  - **Is my data consistent?**
  - **Can I detect an effect?**
  - **Did I just make a discovery?**

Prior **$p_a(x)$**

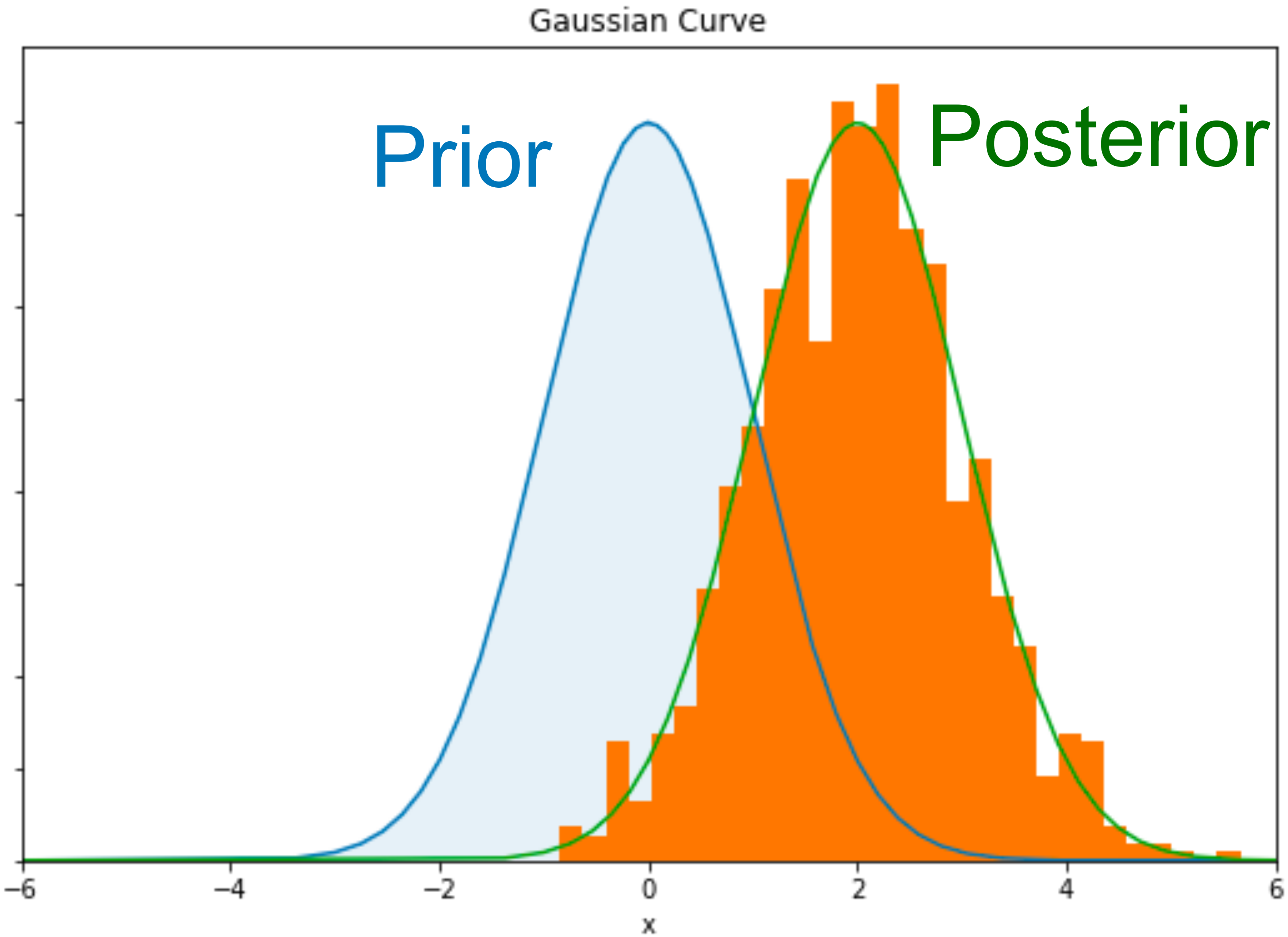
Posterior **$p_b(x)$**

Hypothesis Test



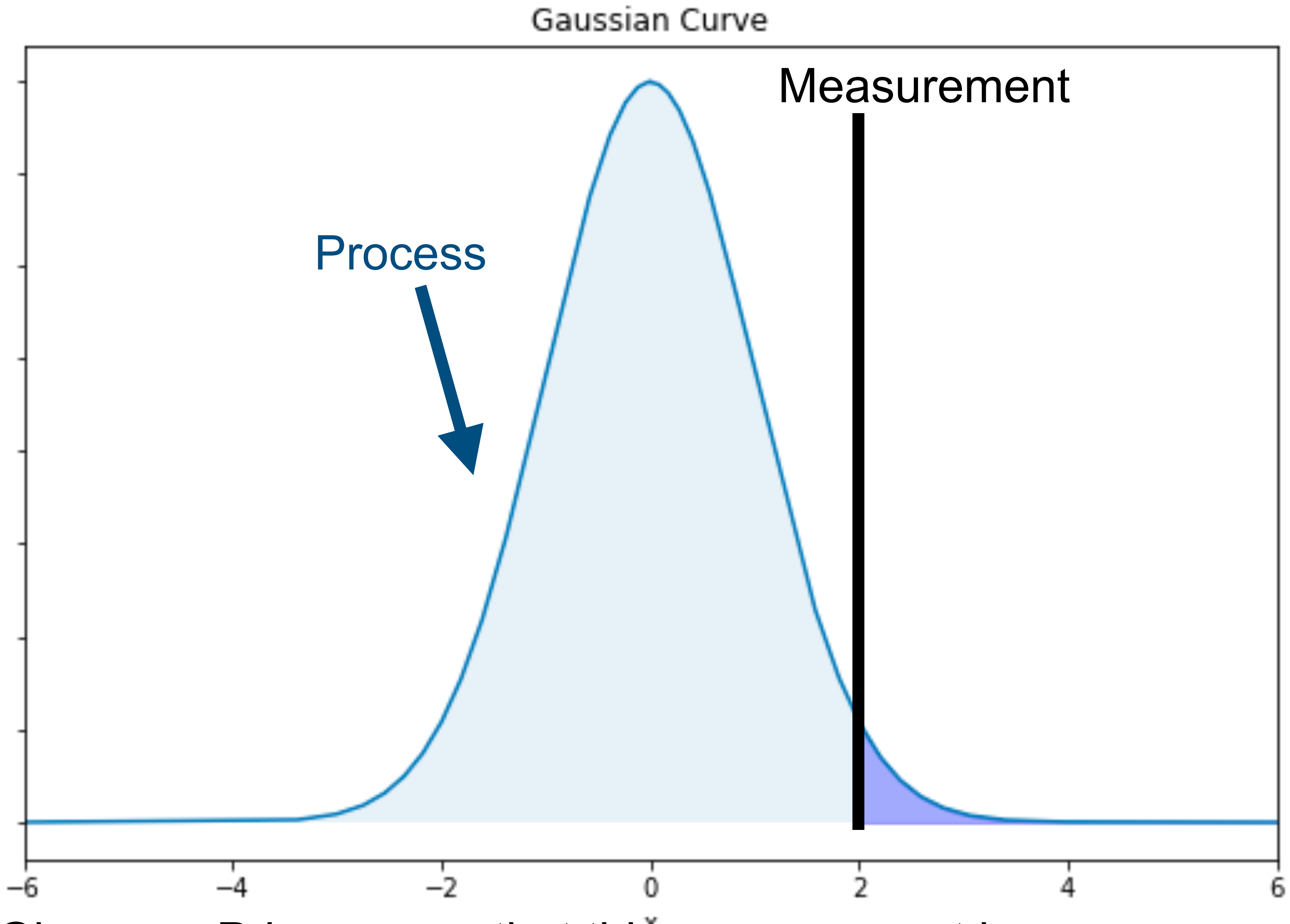
Lets Get Back to the original

- How do I tell if my measurements are consistent?



What is a measurement?

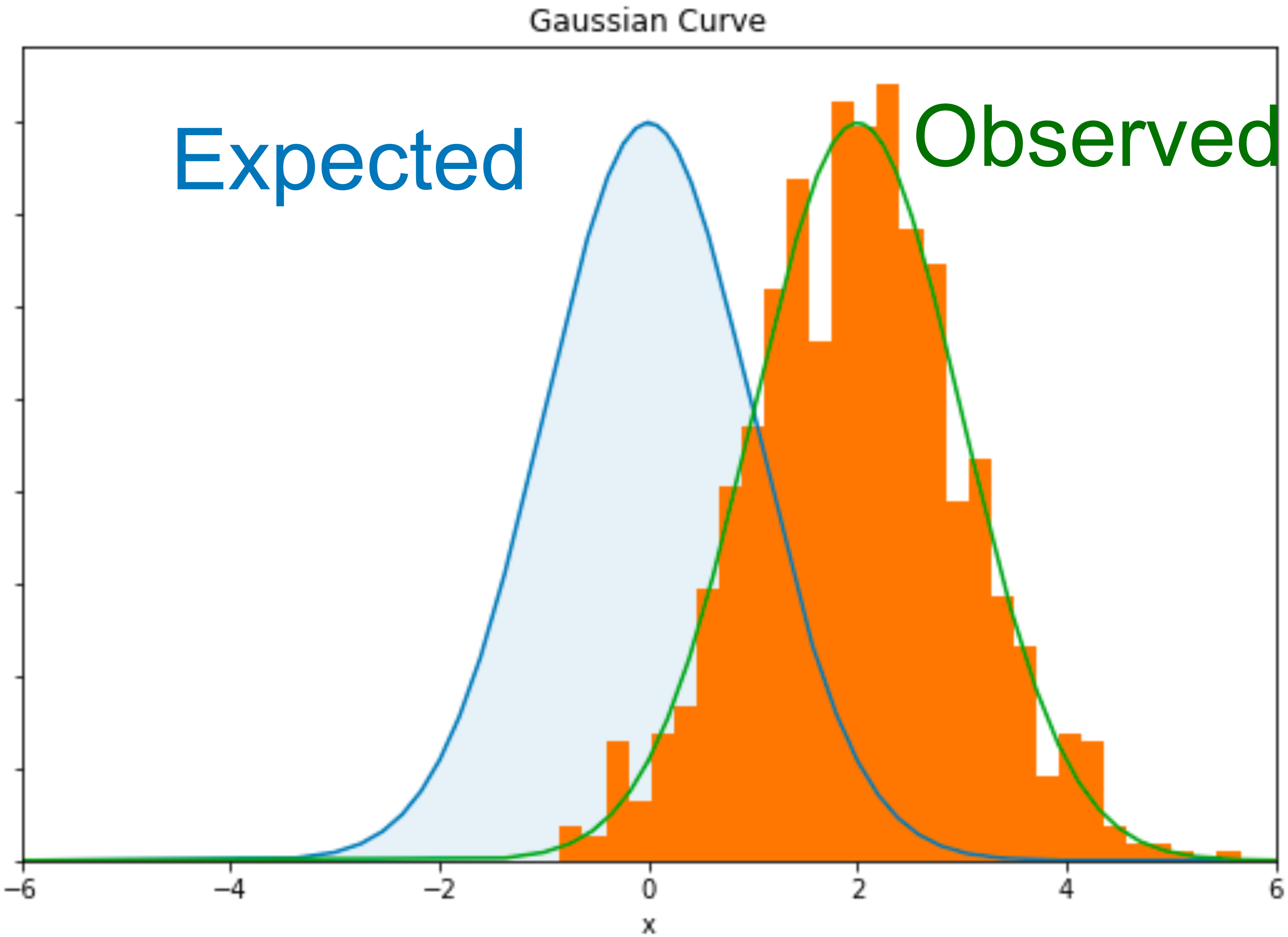
- We have assumed a model for this process
  - With one measurement we can test it against a prior



Given our Prior we say that this measurement has a z-score of 2

Think Like a Frequentist

- How do I tell if my measurements are consistent?



Frequentist: chi2-test

- To test if two histograms are consistent
- $$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = N \sum_{i=1}^{n} \frac{(O_i/N - p_i)^2}{p_i}$$

where

χ^2 = Pearson's cumulative test statistic, which asymptotically approaches a χ^2 distribution.

O_i = the number of observations of type *i*.

N = total number of observations

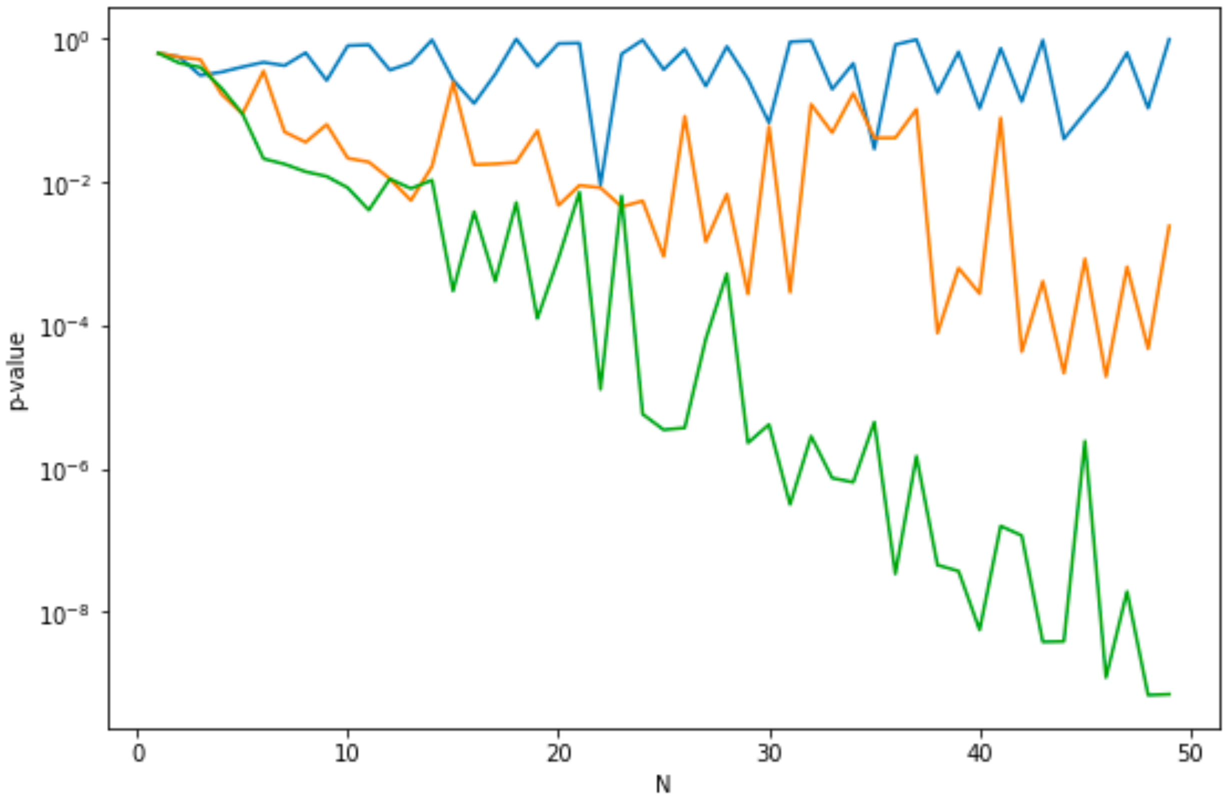
$E_i = Np_i$ = the expected (theoretical) count of type *i*, asserted by the null hypothesis that the fraction of type *i* in the population is p_i

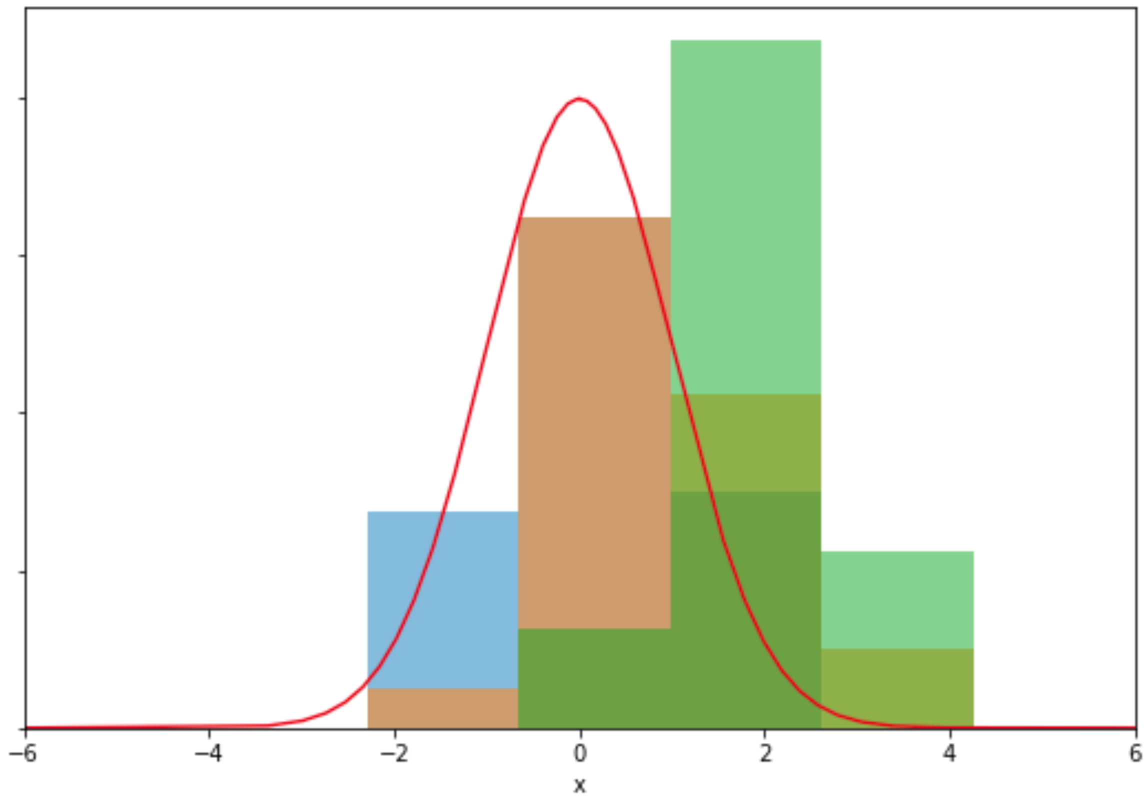
n = the number of cells in the table.

If we have a continuous distribution we can turn it into a histogram with zero uncertainty, there are a lot of ways to do this

chi2 consistency

- Variation as a function of sample size
  - Compute chi2 consistency between histogram and observed
  - Larger p-value with larger n-samples





KS-test

- Komolgorov-Smirnov Test
  - Estimate the maximum difference between observed and predicted cumulative distribution functions and compare with expectations.
  - An unbinned way to check histograms are equal

$$\widehat{F}_n(t) = \frac{\text{number of elements in the sample} \leq t}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \leq t},$$

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|,$$

Supremum Function (ie max difference)