~ lecture 1 ~

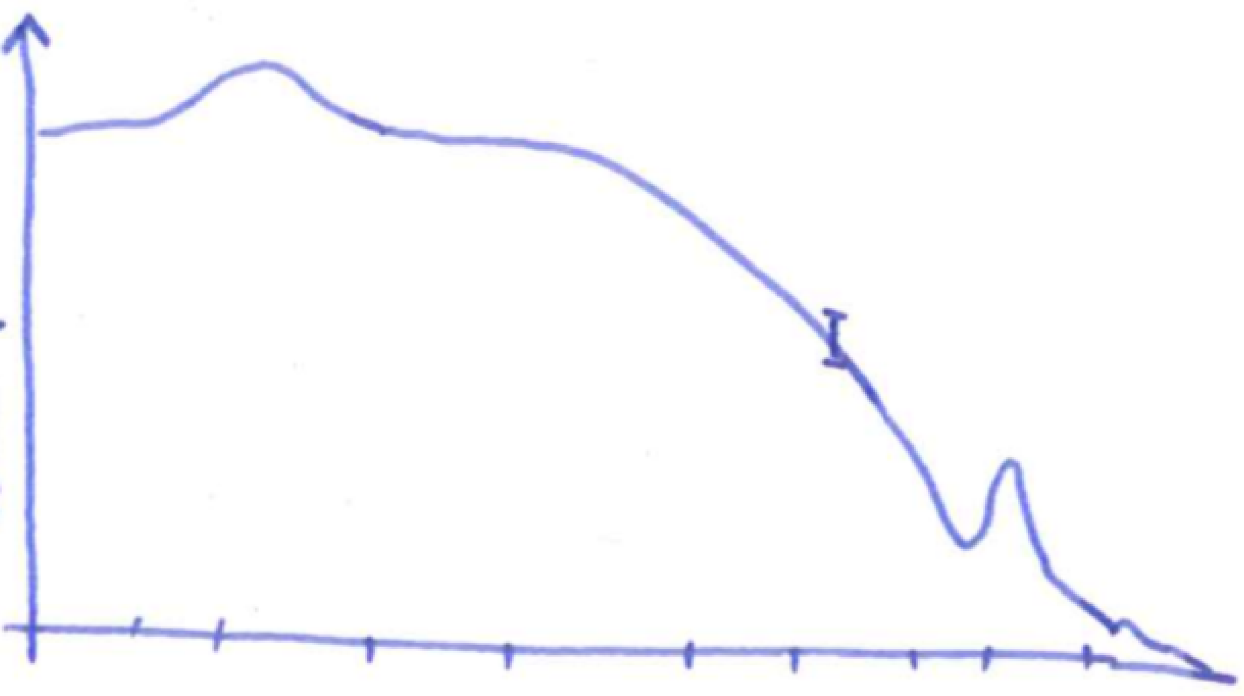
Why Baysian statistics?

one way of seeing it...

- After staring at the same spot for 4 weeks, your telescope comes up with the following Spektrum:

some flux unit

doesn't tell you how bright your object is without knowing the redshift

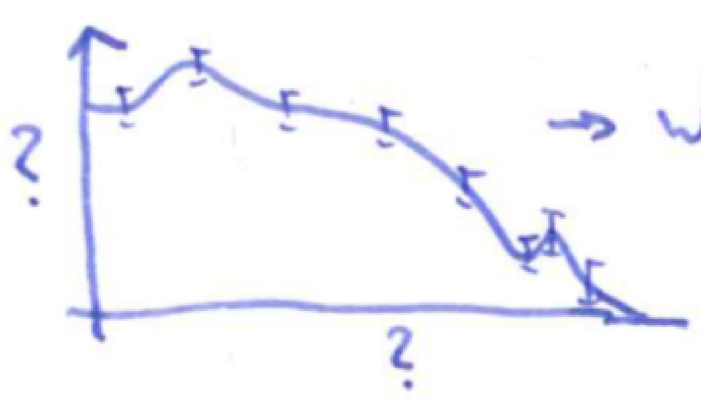


What is it?

chandra energy channels ≠ „true energy" ⇝ redshift

since we have no idea what it is, let's ask a statistician

⇝ ask a Frequentist:

- 

→ what is this?

F: which physical model produced that data?

$L(D|M(\vec{\theta}))$: propability of getting the data for a model M

$\vec{\theta}$: some parameters

⇒ if I gave the frequentist the model, he would tell me the constraints on its parameters

⇝ but I dont have the model

- Just to illustrate... let's agree to invent some completely unmotivated model

width + position of bump (1,2)

sharpness of edge (6)

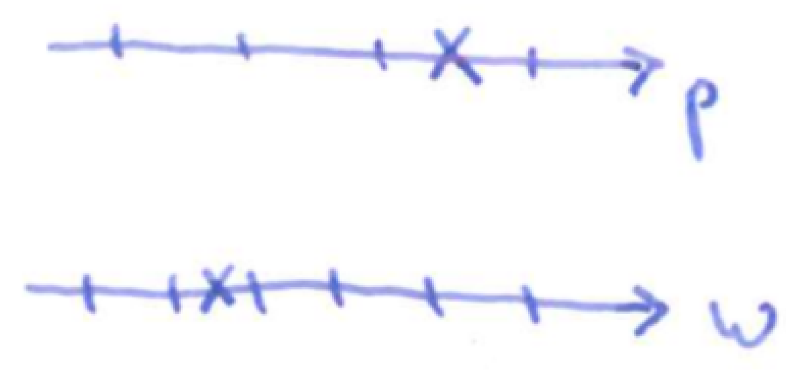
width + position of bump (4,5)

length of plateau (3)

⇒ 6 parameters → it will fit somehow

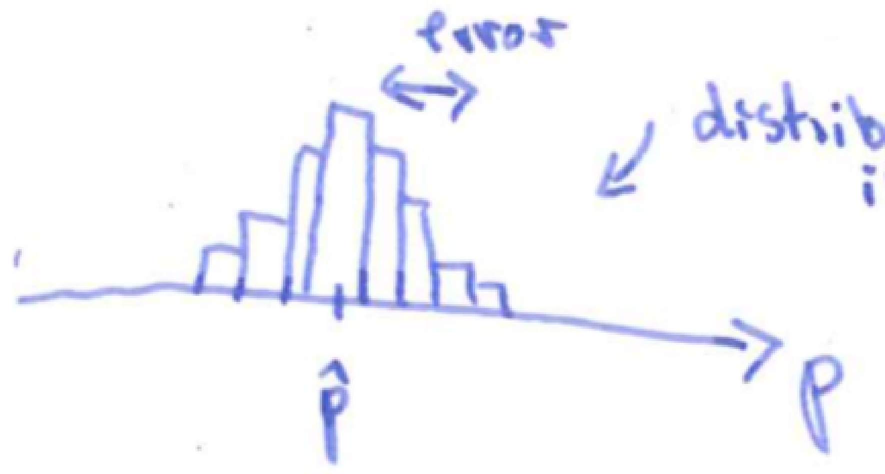
~> Frequentist: $L(D \mid M(\vec{\theta}))$ maximieren $\Rightarrow$ fitting, best fit parameters

· histogram:



• at this point the Frequentist will ask for new data
... which you can't give him (your colleagues have the telescope for 4 weeks...)

$\Rightarrow$ Frequentist will complain: look Frequentist statistics works like this:



estimators: $\mu(p) = \frac{1}{N} \sum_i p_i$

$\mathrm{Var}(p) = \frac{1}{N-1} \sum_i (p_i - \mu)^2$

good and unbiased estimators for $N \to \infty$

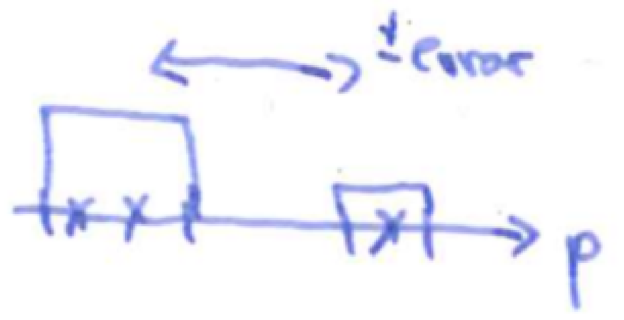
but for $N = 1$ .... ?

So he wants to have the sampling step  in order to estimate the error on the parameters

~> partition available data set into subdata sets, repeat analysis ... (is this bootstrap or jackknife?)

$\Rightarrow$ say you had 30 energy channels

$\Rightarrow$ partition into 3×10 subdata sets



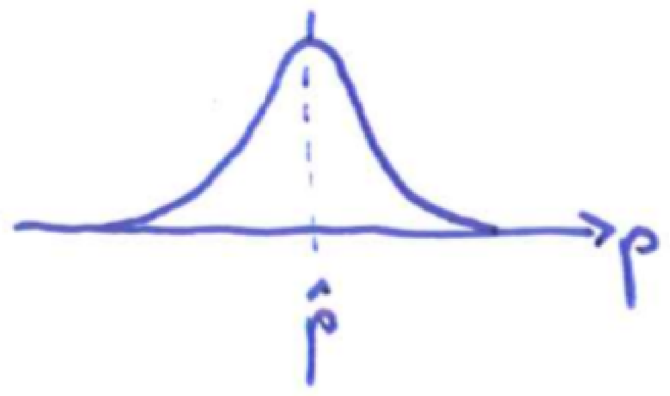
$\Rightarrow$ note: error: not σ
$\Rightarrow$ with this histogramm, we don't know whether it is Gaussian

~> this allows to assess by how much your estimates jump around.
but this procedure has to stop when the number of data points in subset = number of parameters

$\Rightarrow$ You quickly get stuck with frequentist statistics, because:

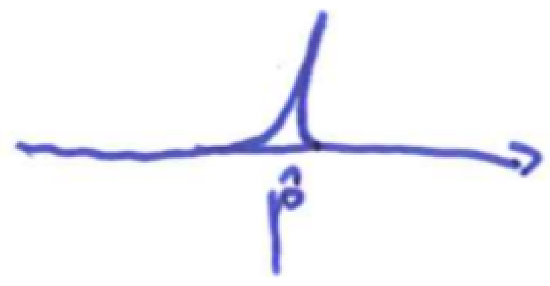
- you don't know the model
- speaking about „frequencies" is a strong, strong euphemism: $3 \ll \infty$
  $\Rightarrow$ lousy averaging ~> strong biases

Frequentist inference:

at some point, for large N, the distribution  stabilizes

⇝ the maximum doesn't move around
⇝ the shape doesn't change
} „significantly" when more measurements are added

⇒ it has converged to the „true" distribution

⇝ could resample it to get 

⇒ $\hat{p}$ is the „true" value in the sense „I have measured $\hat{p}$ 10^{many} times, I always find $\hat{p} \pm$ error, if I measure once more, I think I will again measure $\hat{p} \pm$ an error which I understand"

Bayesian inference

- I don't have enough data / I don't want to have enough data to talk about „frequencies"
  - ⇝ nuclear reactor accidents + epidemology: you really don't want to remeasure... ⇒ yet it is clear there is a causal connection between things
    ⇒ rely on your general understanding and use physics to come up with a credible argumentation line
  - ⇝ take your data as fixed and assess ~~not the~~ „probability" not as „frequency" but as „credibility" instead

- if you don't know the model ⇒ come up with multiple physically motivated models
  - ⇝ admit that credibility does not only come from good data, but can also stem from an „extremely powerfull theory"
    = „a theory that explains a lot"
    {⇝ Priors}
  - ⇝ which one is the best model? {⇝ Evidence}
    best ↓ „most natural explanation"

- mathematically:

Frequentist: $L(\vec{x} \mid \vec{\Theta}(\mathcal{M}))$ ⇒ likelihood of getting the data x, for some value of the parameters $\vec{\Theta}$, which belong to the Model $\mathcal{M}$

Bayesian: Data $\vec{x}$ are given ⇒ $P(\vec{\Theta}(\mathcal{M}) \mid \vec{x})$ · Bayes theorem here

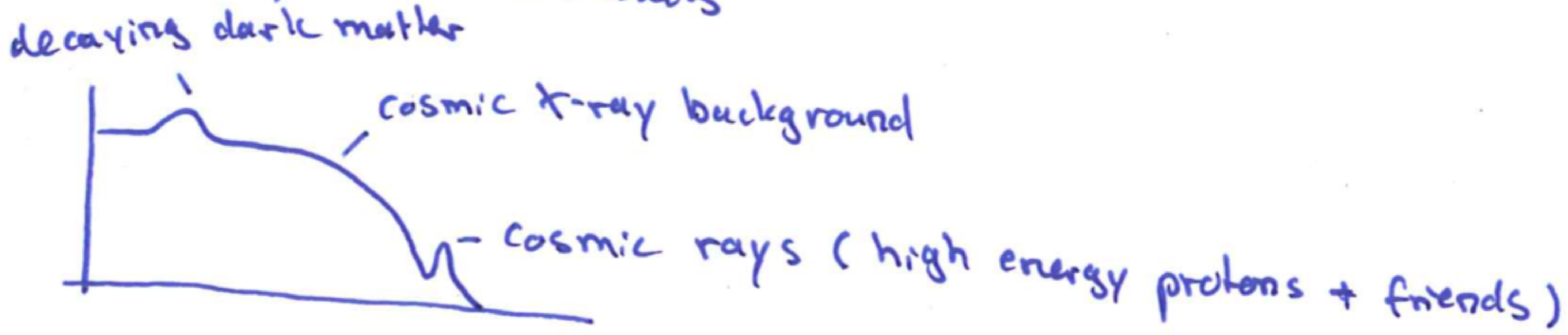
$L(\vec{x} \mid \vec{\Theta}(\mathcal{M}))$ and $P(\vec{\Theta}(\mathcal{M}) \mid \vec{x})$ are related by Bayes theorem:

$$\underset{\downarrow \atop \text{„Posterior"}}{P(\vec{\Theta}(\mathcal{M}) \mid \vec{x})} = \frac{L(\vec{x} \mid \vec{\Theta}(\mathcal{M}))\ \overbrace{\rho(\vec{\Theta}(\mathcal{M}))}^{\text{Prior}}}{\underbrace{\mathcal{E}(\vec{x}(\mathcal{M}))}_{\text{Evidence = marginal likelihood}}}$$

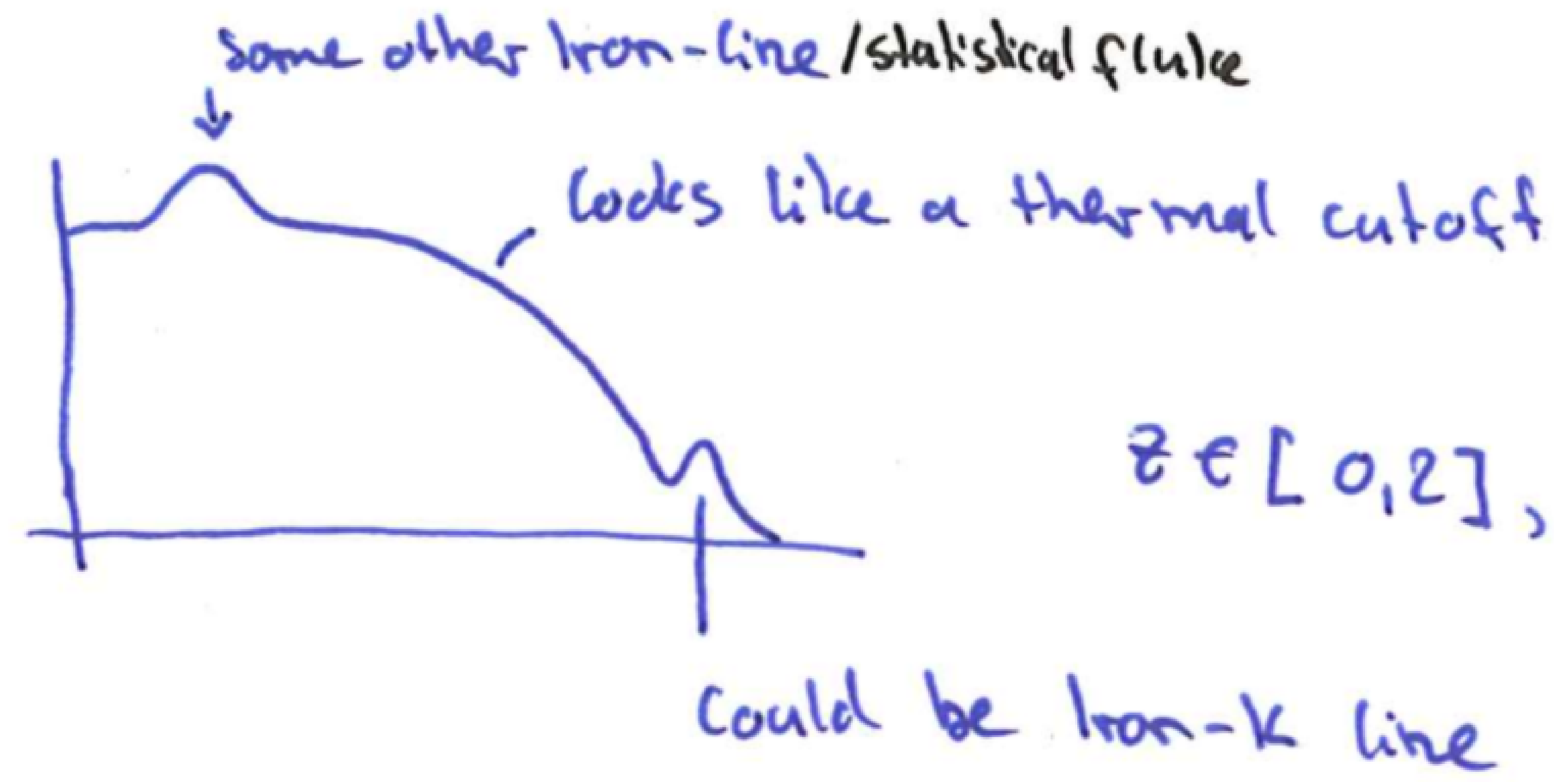
Here, with the notation $Q(\mathcal{M})$, we mean that the quantity Q is evaluated in the framework of the chosen Model $\mathcal{M}$

⇝ $\mathcal{M}$ could be: · the standard model of cosmology
· the Torkado-theory or any other crazy stuff about which people always write me Emails at my outreach adress

- In our case: Come up with 2 Models

1) 

↑ whole spectrum is integrated over z, the redshift

2) 

$z \in [0,2]$, some concretely located thing

⇝ ad hoc ① and ② are $\underbrace{\text{equally crazy}}_{= \text{„equally likely"}}$ (Iron-K, excuse me...)

$\boxed{\Rightarrow \text{Let's Bayes it}}$

Parameters:

1): modelling decaying dark matter : $p_1 = E_d$ Energy of decay radiation

$p_2 = \sigma_d$ width of the decay spectrum

modelling cosmic X-ray background: $p_3 = S$, some spectral tilt

$p_4 = A$, some Amplitude

modelling cosmic rays:

$p_5 = E_r$, mean energy of rays

$p_6 = \sigma_r$, width of CR-peak

$\Rightarrow \vec{\theta} = (E_d, \sigma_d, S, A, E_r, \sigma_r)$

2) Energy of iron lines are known from earth laboratories $\Rightarrow$ no parameters

parameters:
- cosmic redshift z
- amount of iron / metallicity Z
- temperature T

$\vec{\theta} = (z, Z, T)$

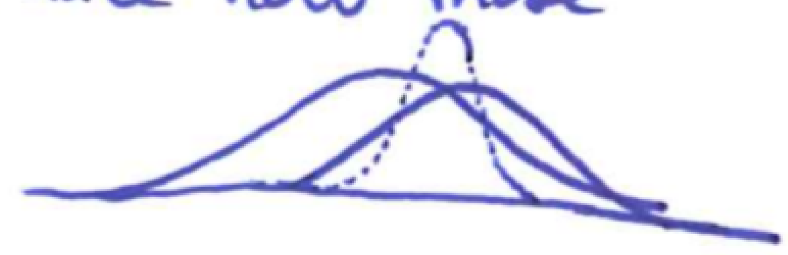
Likelihoods: clear ✓

$\Rightarrow$ what is this prior and Evidence stuff?

• Priors

• yes, one can fool oneself with priors, if one isn't carefull.
• yes, there are well motivated priors, that don't fool you.

⇝ where there is nothing to discuss about priors: Data driven priors
- prior = likelihood from previous measurements
  likelihood = new statistical power from new data set
  ⇒ posterior = more sharply peaked because in total you have now more data



⇝ where there is something to discuss about priors: Theory-driven priors
(my use of the word „prior")

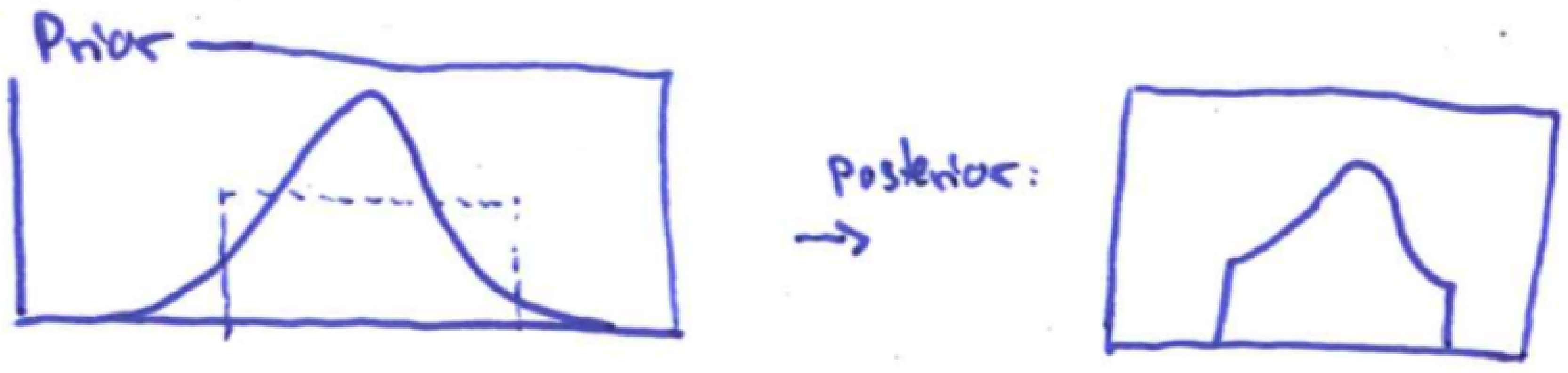
all data — additional theoretical knowledge

$$P(\vec{\theta}(M) \mid \vec{x}) = \frac{L(\vec{x} \mid \vec{\theta}(M))\; p(\vec{\theta}(M))}{\sum p(\vec{x}(M))}$$

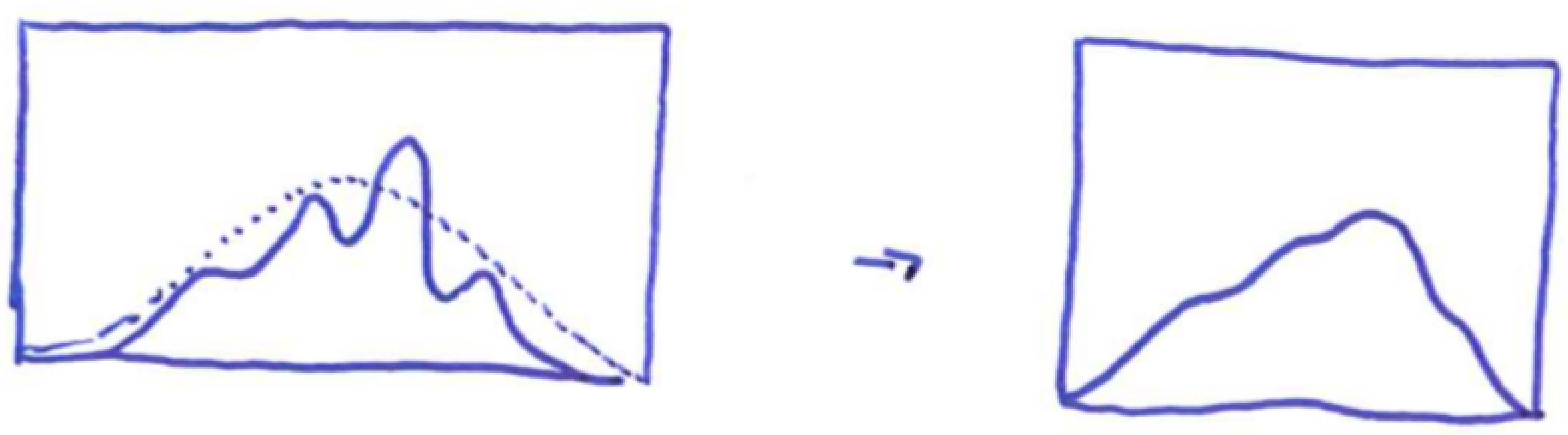
Examples:
- upper/lower bounds of a parameter
- distributions of nuisance parameters

Fun with priors + how to fool yourself

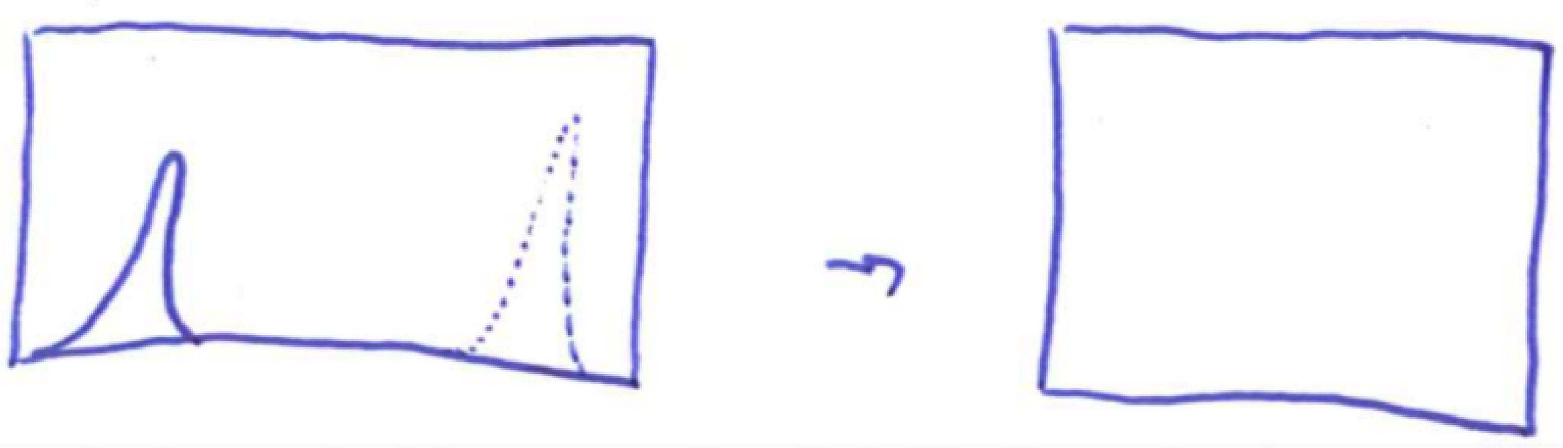
— Likelihood
····· Prior



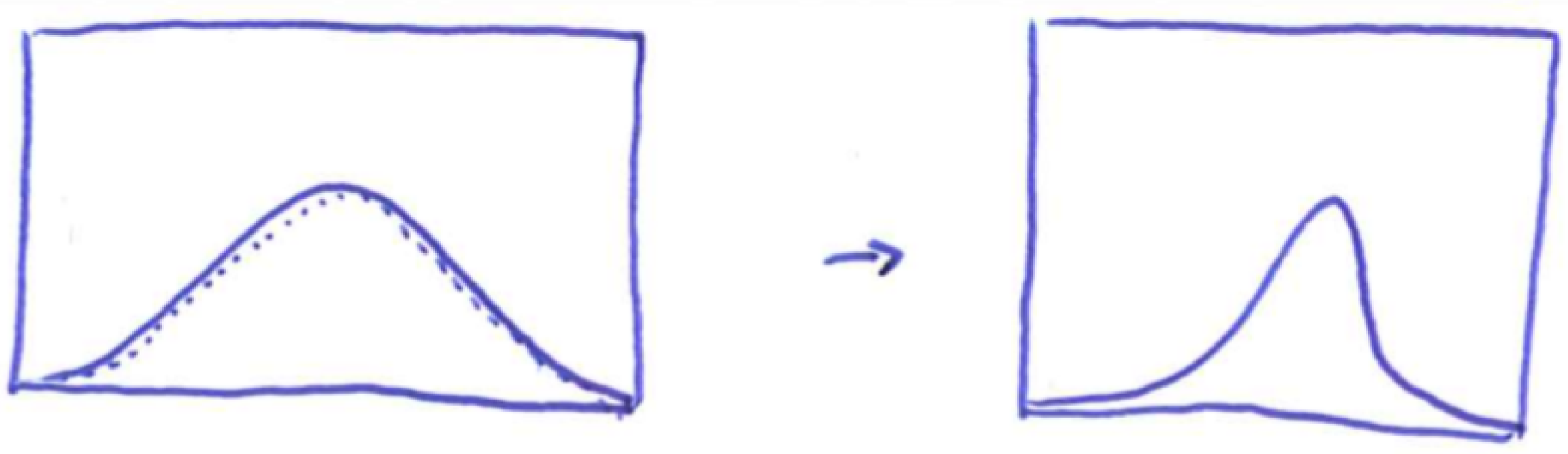
cutting away unphysical values



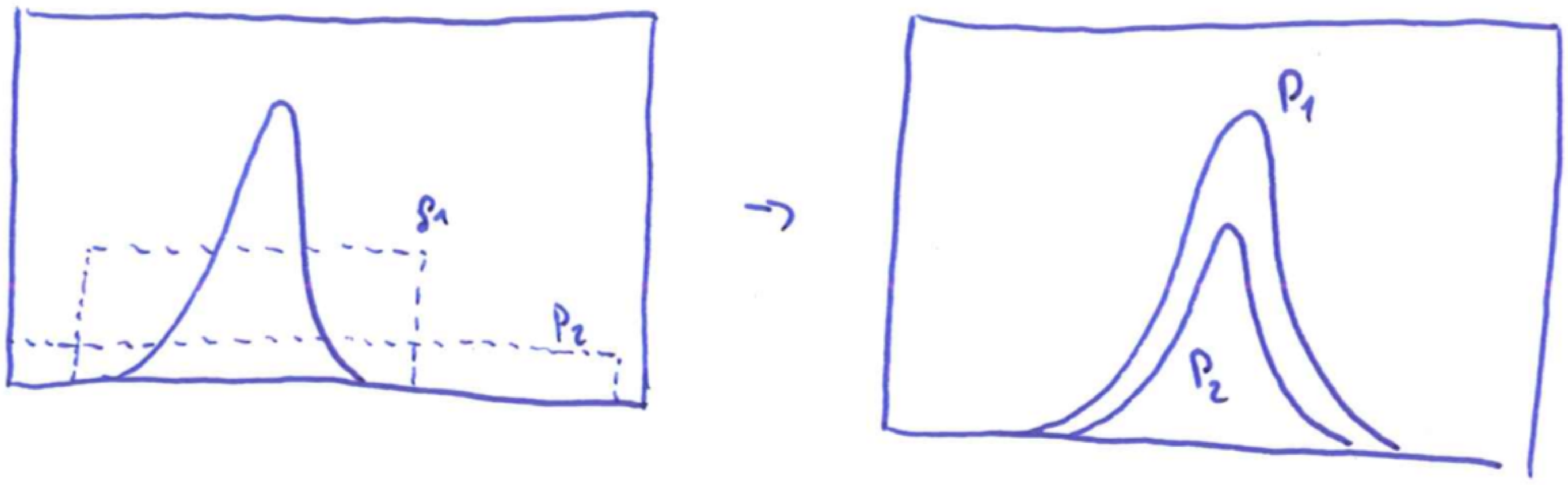
„Gaussianizing" a posterior



theory and observation at mismatch



using data twice

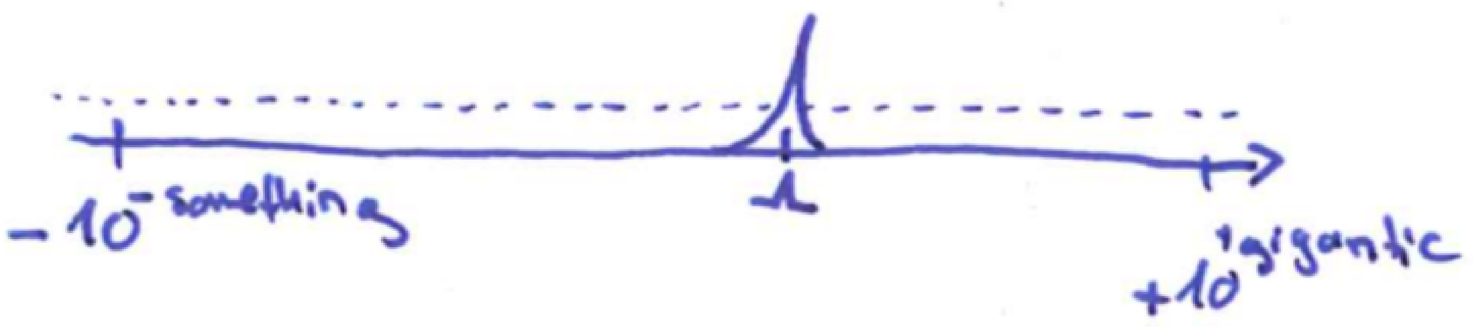


$\rightsquigarrow$ whatever you do: state your priors explicitely!

Don't decide on theoretical priors **after** the arrival of data!

Typical examples of priors in Bayesian cosmology:

•



$\rightsquigarrow$ why does Λ have the value it has?

• assume your data set depends on some parameters $\vec{\Theta}$, which you want to know, and a nuisance parameter p_n

$\rightsquigarrow$ maybe you have some understanding of the nuisance?

E.g.: • p_n comes from a fluktuating quantity(*) $\Rightarrow$ p_n has mean zero

• maybe it is additionally known to be Gauss distributed, with unknown σ:

$$p_n \sim \mathcal{N} \exp\left(-\frac{1}{2}\frac{p^2}{\sigma^2}\right)$$

$\rightsquigarrow$ prior $\rightsquigarrow$ marginalize out σ

(*) e.g. per mass conservation

• the width of the prior Δp, in comparison to your measurement errors of a parameter ($p_1 \pm \Delta p_1$) will decide whether you consider your parameter measured.

$\frac{\Delta p_1}{\Delta p} \gg 1$ oder $\gtrsim 1$: parameter not measured

$\frac{\Delta p_1}{\Delta p} \ll 1$: parameter measured