BAYESIAN APPROACH

Recall the Bayes Theorem for events A and B:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

Where: $P(A)$ = Prior degree of belief on A before knowing that B has happened

$P(A|B)$ = Posterior degree of belief on A after knowing that B has happened

- Bayesian probability and likelihood = parameter estimation

Suppose we have n measurements of a variable x with a given PDF which depends on some parameters $\vec{\vartheta} \in \Omega$.

The probability of obtaining the ~~data~~ (exactly) the data $\vec{x}$ given all possible choices of the parameters $\vec{\vartheta}$ is:

$$P(\vec{x}|\vec{\vartheta}) = \mathcal{L}(\vec{x}|\vec{\vartheta})$$

Prior to the measurement, our degree of belief on the parameters is $\pi(\vec{\vartheta})$

Therefore we can use the Bayes Theorem to find the posterior probability of the parameters $\vec{\vartheta}$ given the measurement $\vec{x}$:

$$P(\vec{\vartheta}|\vec{x}) = \frac{\mathcal{L}(\vec{x}|\vec{\vartheta})\,\pi(\vec{\vartheta})}{\underbrace{\int_{\Omega} \mathcal{L}(\vec{x}|\vec{\vartheta})\,\pi(\vec{\vartheta})\,d\vec{\vartheta}}}$$

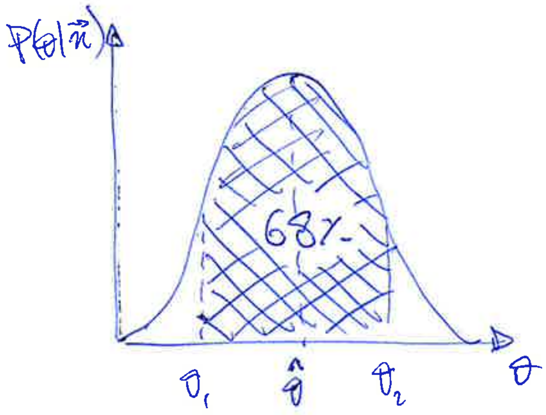
→ depends only on $\vec{\vartheta}$ ($\vec{x}$ is measured)

→ same dimension of $\vec{\vartheta}$

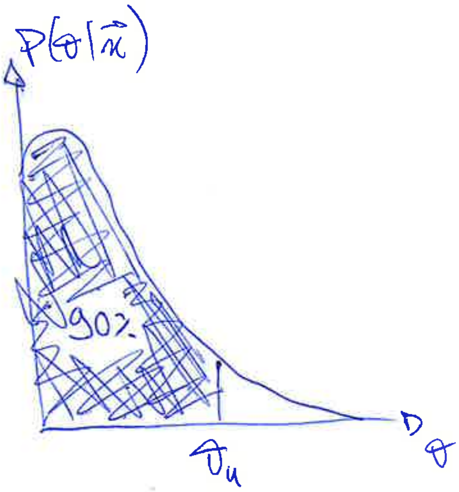
↳ This is a constant, so it's just a normalization term.

- Credible intervals

The posterior PDF of θ will look like:



or



From $P(\theta|\vec{x})$ we can quote:

a) The mode $\hat{\theta}$ (which could also be at the boundary of the physical region)

b) The central / shortest interval $[\theta_1, \theta_2]$ which contains the ~~free~~ most probable value of θ with 68% probability

c) If θ is near one of the borders, we can quote an upper or lower limit (typically at 90% coverage).

These intervals are called "credible intervals".
They tell us that, based on our current knowledge, the true value of θ will be contained in that range with the specified probability.

↳ The concept of coverage here applies to the true parameter and is a property of the interval, not of the method

↳ A Bayesian 68% credible interval might have 0% frequentist coverage, and this is not a problem.

We will quote as a result: $\theta = \hat{\theta}^{+(\theta_2-\hat{\theta})}_{-(\hat{\theta}-\theta_1)}$

<u>Example</u>: Poisson posterior for rate

In such case, we should repeat the fit twice:

1) With the background only model H_0
2) With the signal + background model H_1, which includes the parameter ϑ.

At this point, we need a criterium to ~~understand wether~~ ~~the~~ ~~we~~ compare H_0 and H_1, and see if we actually need the new signal term in H_1.

If yes → quote the mode and 68% c.i. shortest interval

If no → still use the fit with H_1 model, but quote 90% c.i. limit on ϑ.

The criterion on how to compare the models H_0 and H_1 will be discussed next week in the session on hypothesis testing.

- Bayesian upper limits for Poisson counts

Suppose we measure a ~~process where w~~ something where we have a signal contribution with s expected signal events on b expected bkg events.

Assume a flat prior on s and b,

→ Zero background case

If $b=0$, so if we are sure that we have no background, and if we measure $n=0$ events, the posterior for s (assuming a flat prior on s and b) is:

$$P(s|n) = \frac{s^n e^{-s}}{n!}$$

$$P(s|n=0) = e^{-s}$$

The 90% c.i. limit is: $s < 2.3$ counts

• Parameter ranges

How should I choose my parameter range?

Well, it depends on the case...

a) If there is a physical constrain in the parameter θ, use it!

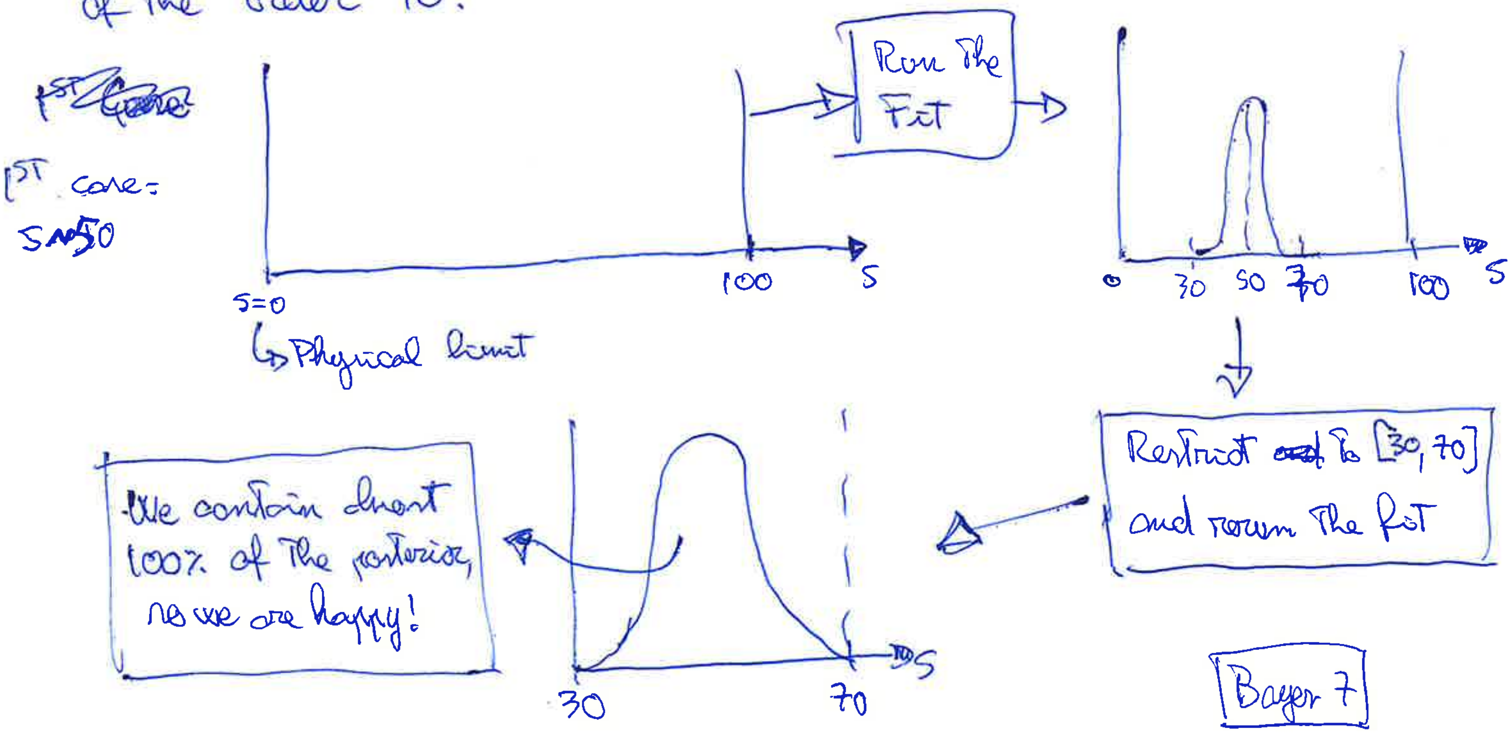
For example, if $\theta = m = \text{mass}$, use $m \geq 0$.

b) If there is a prior measurement $\theta = \hat{\theta} \pm \sigma_\theta$, a good choice is to use $[\hat{\theta} - k\sigma, \hat{\theta} + k\sigma]$ with $k \geq 5$,

so that we are sure that:
1) The new measurement is likely to be well contained in the parameter range.
2) The posterior distribution is likely to be fully contained in the range, with the tails not hitting the borders.

c) Otherwise, ~~use a range~~ start with a large range, run the fit, check the posterior, ~~and~~ restrict it to $\pm(5\text{-}10)\sigma$, then repeat the fit. If you hit a border that is not justified by physics, enlarge it.

→ Suppose we have a parameter s describing some number of counts, and we have no prior knowledge on it, but we expect s to be of the order 10.



Bayes 7

- Bayes fits: practical implementation

Let's ~~go~~ back to the Bayes Theorem for parameter estimation:

$$P(\vec{\vartheta}|\vec{x}) = \frac{\mathcal{L}(\vec{x}|\vec{\vartheta})\,\pi(\vec{\vartheta})}{\int \mathcal{L}(\vec{x}|\vec{\vartheta})\,\pi(\vec{\vartheta})\,d\vec{\vartheta}}$$

In principle, we need to:

Map the posterior $P(\vec{\vartheta}|\vec{x})$ and integrate it over the nuisance parameters.

However: a) $P(\vec{\vartheta}|\vec{x})$ might be very complicated to integrate over $d\vec{\nu}$

b) $\mathcal{L}\cdot\pi$ might be too complicated to integrate over $d\vec{\vartheta}$

~~Solutions~~ ~~Observations:~~

~~b) $\int \mathcal{L}\cdot\pi\, d\vec{\vartheta}$ is a constant that does not depend on $\vec{\vartheta}$~~

~~a) We don't ...~~

Solution:

1) Map $\mathcal{L}\cdot\pi$ with a Markov Chain (e.g. Metropolis-Hastings) with n tested samples

2) Forget about the denominator $\int \mathcal{L}\,\pi\, d\vec{\vartheta}$, because it's just a constant that does not depend on $\vec{\vartheta}$.

3) Renormalize the posterior $P(\vec{\vartheta}|\vec{x})$ by $\frac{1}{n}$

4) To build the marginalized of some parameter ϑ_i, just put the ϑ_i values of all accepted samples of the Markov-Chain into a histogram, and normalize it by $1/n$!

5) Extract $\hat{\vartheta}_i$ and σ_i from the marginalized histogram.

c) Flat on $\log_{10}$ of some parameter or observable

↳ Sometimes called "scale-invariant" prior

↳ Corresponds to saying that we give the same prior probability to the parameter to be e.g. of order 10 or of order 10^3

Notice however that a) all these priors ~~have~~ are not invariant under reparametrization

b) none of them is a ~~good~~ truly "uninformative" prior.

In fact, there is an intrinsic arbitrariness in the method that we cannot avoid.

→ Example: religious belief

- Jeffrey's Priors

Jeffrey's priors are a ~~set~~ class of "uninformative" priors that can be used in case we have no knowledge about the parameters, and that are invariant under reparametrization.

Jeffrey's prior are of the type: $\pi(\vec{\vartheta}) = \sqrt{I_F(\vec{\vartheta})}$ → Fisher information matrix

where: $I_F(\hat{\vartheta}) = \det\left[E\left[\dfrac{\partial \ln \mathcal{L}}{\partial \vartheta_i} \dfrac{\partial \ln \mathcal{L}}{\partial \vartheta_j}\right]\right]$

For example, we ~~can~~ have the following Jeffrey's priors for these PDFs:

PDF	Jeffrey's prior
Poisson mean s	$1/\sqrt{s}$
Poisson signal s + background b	$1/\sqrt{s+b}$
Gaussian mean μ	uniform
Gaussian std σ	$1/\sigma$
Binomial efficiency ε	$1/\sqrt{\varepsilon(1-\varepsilon)}$
Exponential decay constant λ	$1/\lambda$