Bayesian Confidence Limits and Intervals

Roger Barlow

SLUO Lectures on Statistics

August 2006

Probability Revisited

I say:

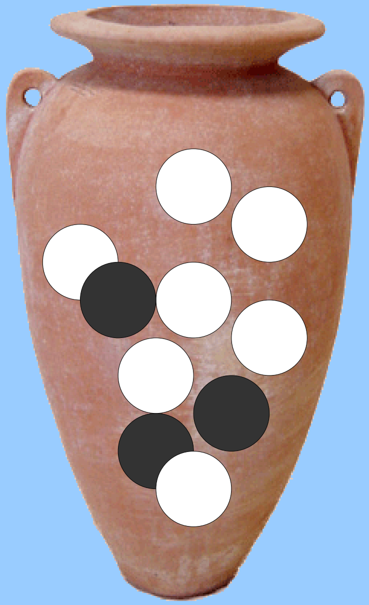
"The probability of rain tomorrow is 70%"

I mean:

I regard 'rain tomorrow' and 'drawing a white ball from an urn containing 7 white balls and 3 black balls' as equally likely.

By which I mean:

If I were offered a choice of betting on one or the other, I would be indifferent.



This is Subjective (Bayesian) probability

GOOD

I can talk about – and do calculations on – the probability of anything:

- Tomorrow's weather
- The Big Bang
- The Higgs mass
- The existence of God
- The age of the current king of France

BAD

There is no reason for my probability values to be the same as your probability values

I may use these probabilities for my own decisions, but there is no call for me to impose them on others

Bayes' Theorem

$$P(B|A)P(A) = P(A\&B) = P(A|B)P(B)$$

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

Conventional application

$$P(disease|symptom) = \frac{P(symptom|disease)P(disease)}{P(symptom)}$$

"Bayesian" application

$$P(Theory|Data) = \frac{P(Data|Theory)P(Theory)}{P(Data)}$$

Bayes at work

Dr. A Sceptic thinks that Global Warming is probably a myth. P=10%

Data arrives showing loss of Antarctic ice coverage. Global warming said this would definitely happen (P=1). But it could happen as part of natural cyclical fluctuations (P=20%)

All numbers totally fictitious

Use Bayes Theorem

$$P_G' = \frac{P(melt|G)\,P_G}{P(melt|G)\,P_G + P(melt|\bar{G})\,\bar{P}_G} = \frac{0.1}{0.1 + 0.2\text{x}0.9} = 0.36$$

Misinformation abounds...

Fun Fact!	**Q. What is the Bayesian Conspiracy?** A. The Bayesian Conspiracy is a multinational, interdisciplinary, and shadowy group of scientists that controls publication, grants, tenure, and the illicit traffic in grad students. The best way to be accepted into the Bayesian Conspiracy is to join the Campus Crusade for Bayes in high school or college, and gradually work your way up to the inner circles. It is rumored that at the upper levels of the Bayesian Conspiracy exist nine silent figures known only as the Bayes Council.

http://yudkowsky.net/bayes/bayes.html

Priors and Posteriors

Can regard the function P(M) as a set of different probabilities for theories about M (now a parameter of the model)

$$P(M)' = \frac{P(R|M)\,P(M)}{P(R)}$$

Posterior distribution for M

Probability distribution for R given M

distribution for R anyway

Prior distribution for M

General notation:
Model parameter(s) M
Experiment result(s) R

Probability and The Likelihood Function

P(R|M) is the probability of what can be a a whole set of results R, as a function of the model parameter(s) M

Also known as the likelihood L(R,M)

It is always tempting to think of it as L(M,R): Probability for model parameter(s) M given result(s) R

For frequentists this is rubbish. For Bayesians it follows if the prior is uniform.

'Inverse Probability'

Bayes theorem says:

$$P(M|R) \propto P(R|M)$$

Call this 'inverse probability'. Probability distribution for a model parameter M given a result R. Just normalise

Seems easy.

But:

P(M) is meaningless/nonexistent in frequentist probability

Working with P(M) and $P(M^2)$ and P(ln M) will give different and incompatible answers

Integrating the Likelihood function L(R,M)

For Bayesians:

Go ahead.

Be aware that if you reparametrise M then your results will change

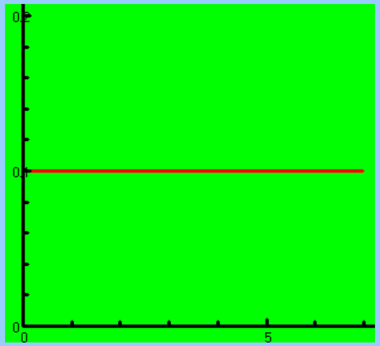
Unless you specify a prior and reparametrise that

For Frequentists, integrating L wrt M is unthinkable.

Integrating/summing over R is fine. Use it to get expectation values

If you integrate a likelihood then you're doing something Bayesian.
If you are a Bayesian, this is not a problem.
If you claim to be a frequentist, you have crossed a borderline somewhere

Uniform prior



- Often take P(M) as constant ("flat prior")
- Strictly speaking P(M) should be normalised: $\int P(M)\, dM = 1$
- Over an infinite range this makes the constant zero...
- Never mind! Call it an "Improper Prior" and normalise the posterior
- A prior flat in M is not flat in M' (ln M, M^2, ..)

Bayesian Confidence Intervals

Trivial!

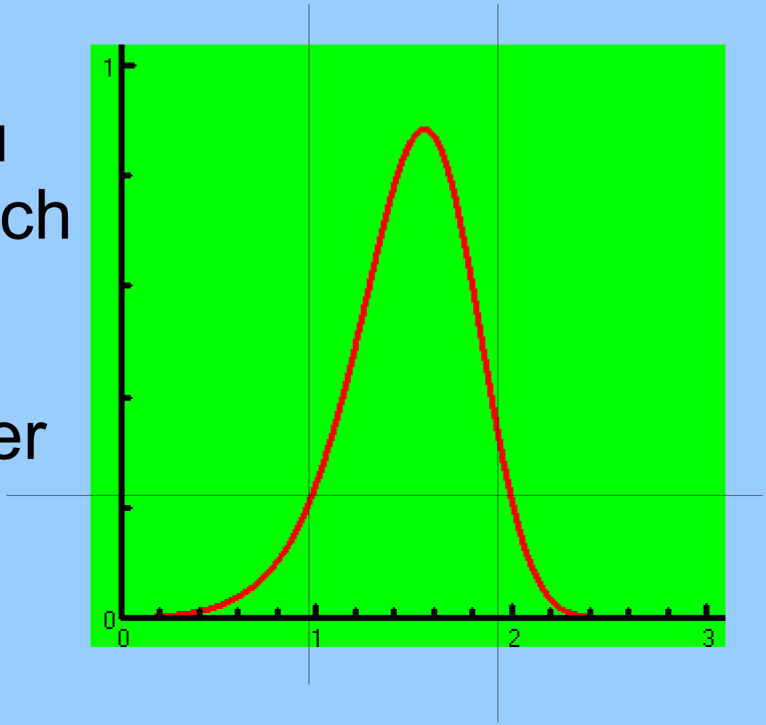
- Given the posterior P'(M|R) you choose a range [M_{lo},M_{hi}] for which

$$\int_{M_{lo}}^{M_{hi}} P'(M|R)\,dM = CL$$

Choice of strategies: central, upper limit lower limit, etc.

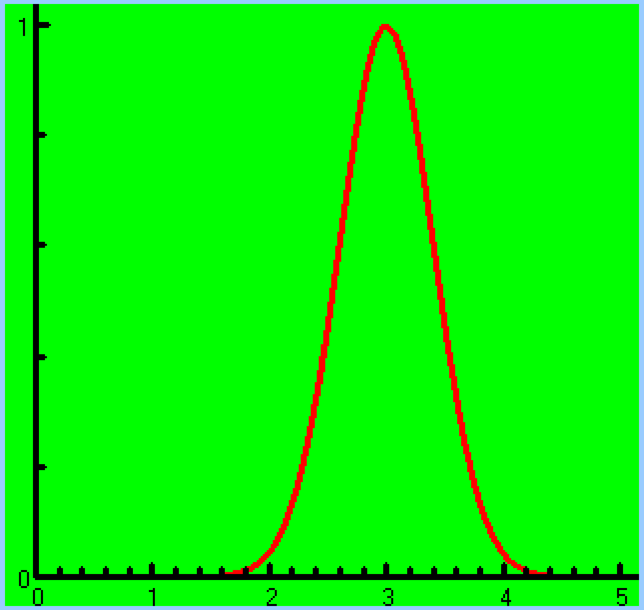
Includes HPD (Highest Posterior Density) – gives shortest range

(but not invariant under changes of variable)



Examples: Gaussian

Gaussian Measurement



$$P(R,M) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(R-M)^2/2\sigma^2}$$

Taking P(M) and P(R) as flat, this is also the probability distribution for M

This can act as the prior for a second measurement. Gaussians multiply and give post-posterior Gaussian with expected mean and width

Confidence intervals for M are just like confidence intervals for R (1σ,2 σ, etc)

Examples: Poisson

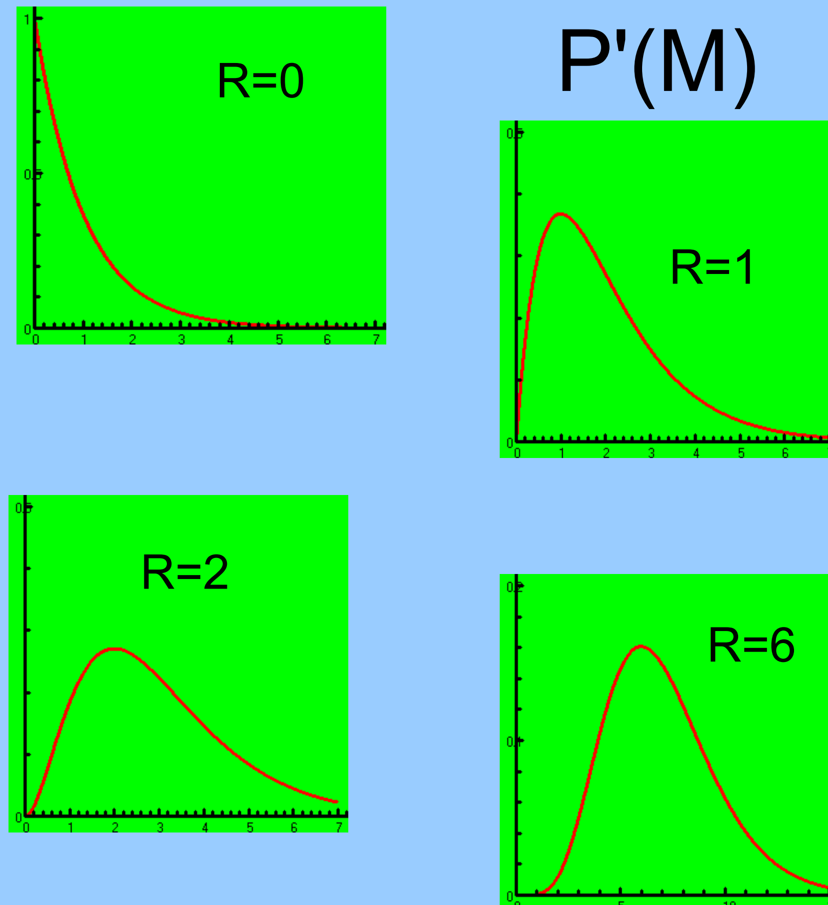
$$P(R,M)=e^{-M}\frac{M^R}{R!}$$

We can regard this as a posterior for M

(assuming uniform prior)

Shown for various small N results

Read off intervals...



Upper and lower limits

Upper limit

$$\int_0^{x_{hi}} e^{-x} \frac{x^N}{N!} dx = 1-\alpha$$

Repeated integration by parts:

$$\sum_0^N e^{-x_{hi}} \frac{x_{hi}^r}{r!} = \alpha$$

Same as frequentist limit (coincidence!)

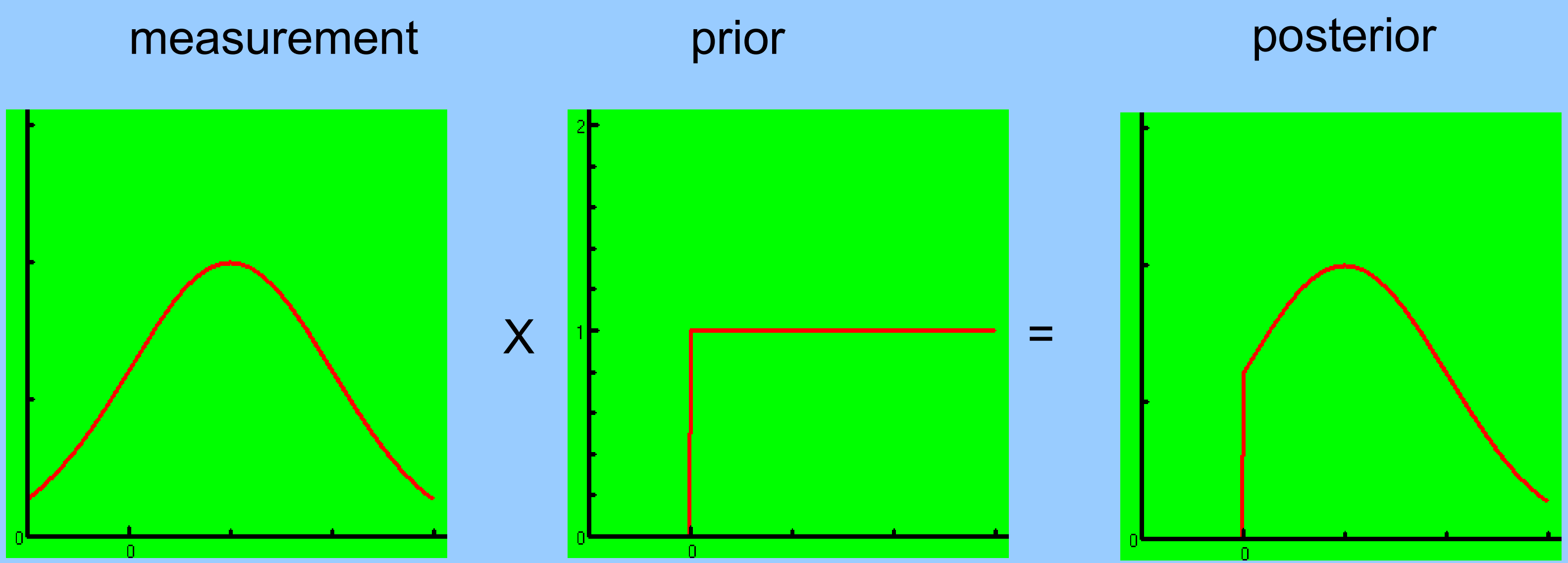
Lower Limit

$$\sum_0^N e^{-x_{lo}} \frac{x_{lo}^r}{r!} = 1-\alpha$$

Not (quite) the same: includes r=N term

Incorporating Constraints: Gaussian

- Say we know M>0
- Prior is now a step function.



Normalise and read off properties

Incorporating Constraints: Poisson

Work with total source strength (s+b) you know is greater than the background b

Need to solve

$$\alpha = \frac{\sum_0^n e^{-(s+b)}(s+b)^r / r!}{\sum_0^n e^{-b} b^r / r!}$$

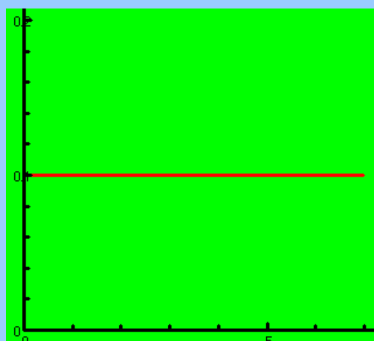
Formula not as obvious as it looks.

Robustness

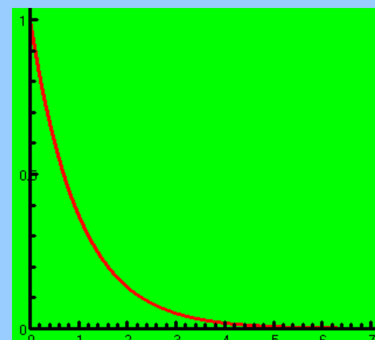
- Result depends on chosen prior
- More data reduces this dependence
- Statistical good practice: try several priors and look at the variation in the result
- If this variation is small, result is robust under changes of prior and is believable
- If this variation is large, it's telling you the result is meaningless

Example: Poisson with no events 90% Upper Limit?

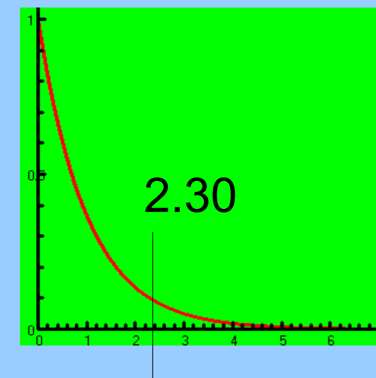
Prior flat in λ



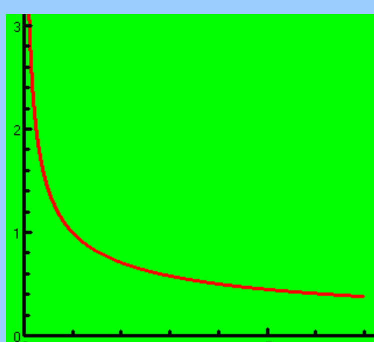
x



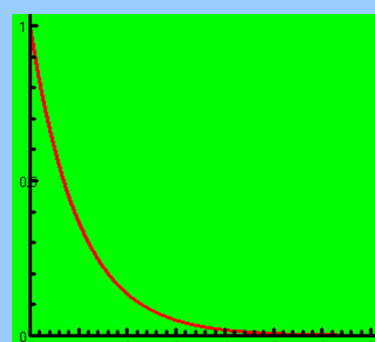
=



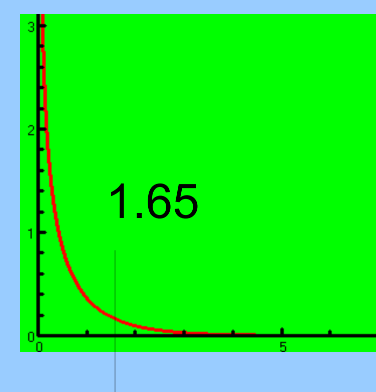
Prior flat in √λ



x



=



Objective Priors

Introduced by Jeffreys

Transform to a variable M' for which the Fisher Information $I = -\langle \frac{d^2 \ln L}{d M'^2} \rangle$ is constant

Equivalently: take prior proportional to √I

I is the curvature of the likelihood function at the peak. Describes how much a measurement will tell you. If it's not constant then some regions are 'better' than others.

Common Cases

- For a location parameter   L(R,M)=f(R-M)
  a uniform prior is objective

- For a scale parameter L(R,M)=f(r x M)
  the objective prior is 1/M, or equivalently work with M'= ln M and a uniform prior

- For a Poisson mean, the objective prior is $1/\sqrt{M}$

When a physicist says 'a flat prior' they mean a uniform prior. When a statistician says 'a flat prior' they mean a Jeffreys' prior

Why didn't it catch on?

It is 'objective' in the sense that everyone can agree on it. But they don't.

- It's more work than a uniform prior
- There are cases where it diverges and gives posterior functions that can't be normalised
- It does not work in more than one dimension (valiant attempts are being made to do this generalisation, under the name of Reference Priors)
- It depends on the form of L(R,M) which depends on the experiment. If you have an initial degree-of-belief prior function for (say) the Higgs mass, that should not depend on the measurement technique

Are Systematics Bayesian?

- A systematic error is an uncertainty in an effect
- Sometimes this is well understood and determined experimentally – e.g. Energy calibration
- Often (more often?) they are estimates - “Theory Errors”
- These are intrinsically Bayesian. Can/must be treated as such

Systematic Errors =Nuisance Parameters

Suppose the result of an experiment depends on a parameter of interest M and a 'nuisance parameter' N

$$P'(M,N|R) \propto L(R|M,N)\ P_M(M)\ P_N(N)$$

We are interested in

$$P'(M|R)=\int P'(M,N|R)\ dN \propto P_M(M) \int L(R|M,N)\ \ P_N(N)\ dN$$

This is called Marginalisation. Frequentists cannot do it as it involves integrating the Likelihood. For Bayesians it's obvious. (Depends on prior $P_N(N)$)

Application to Poisson

Cousins and Highland: Signal strength λ=As+b

A is 'sensitivity', b is background. Uncertainties on both values give systematic errors

Fully Bayesian treatment requires prior for source strength s. Tricky

Partial Bayesian treatment uses Gaussians for b and A and marginalises (integrates)

Prior dependence for A – 10% uncertainty in A gives limits ~1% different for different priors

Example – CKM fitter
Le Diberder, T'Jampens and others

Sad Story

Fitting CKM angle α from $B \to \rho\rho$

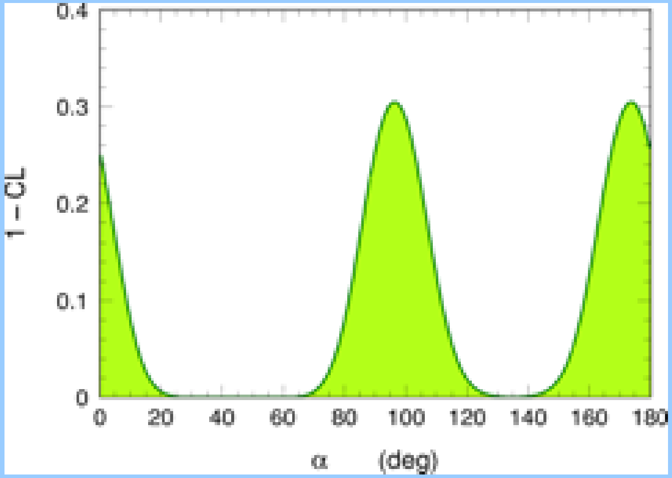
6 observables

3 amplitudes: 6 unknown parameters (magnitudes, phases)

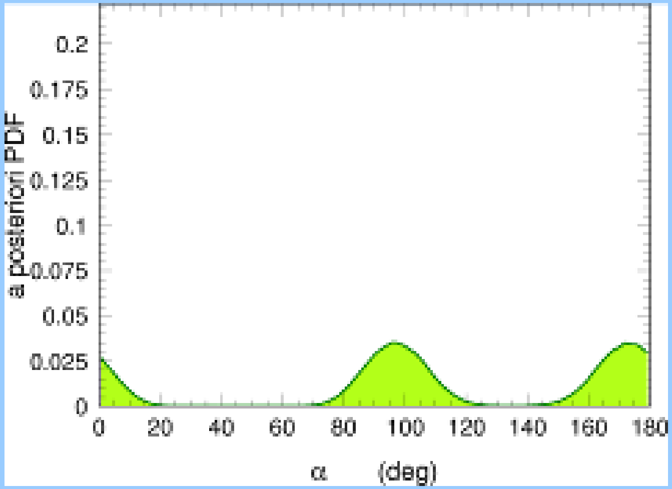
α is the fundamentally interesting one

Results

Frequentist



Bayesian
Set one phase to zero
Uniform priors in other two phases and 3 magnitudes



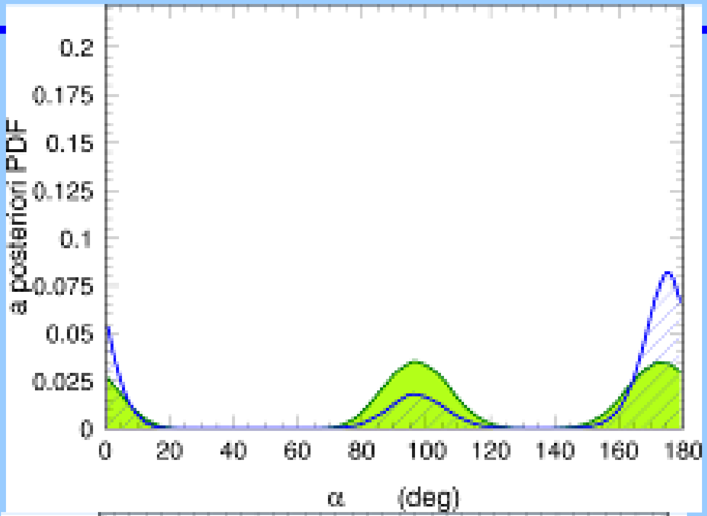
More Results

Bayesian
Parametrise Tree and Penguin amplitudes

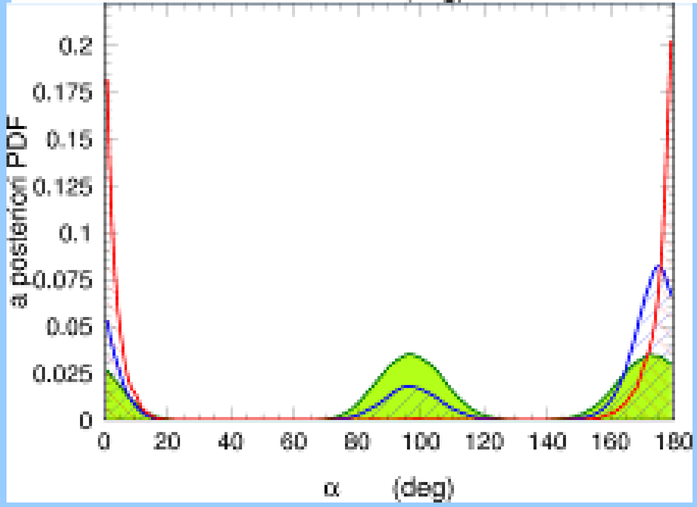
$$A^{+-} = -Te^{-i\alpha} + Pe^{i\delta_P}$$

$$A^{+0} = -\frac{1}{\sqrt{2}} e^{-i\alpha}\left(T + T_C e^{i\delta_{T_C}}\right)$$

$$A^{00} = -\frac{1}{\sqrt{2}}\left(e^{-i\alpha} T_C e^{i\delta_{T_C}} + Pe^{i\delta_P}\right)$$



Bayesian
3 Amplitudes:
3 real parts, 3 Imaginary parts

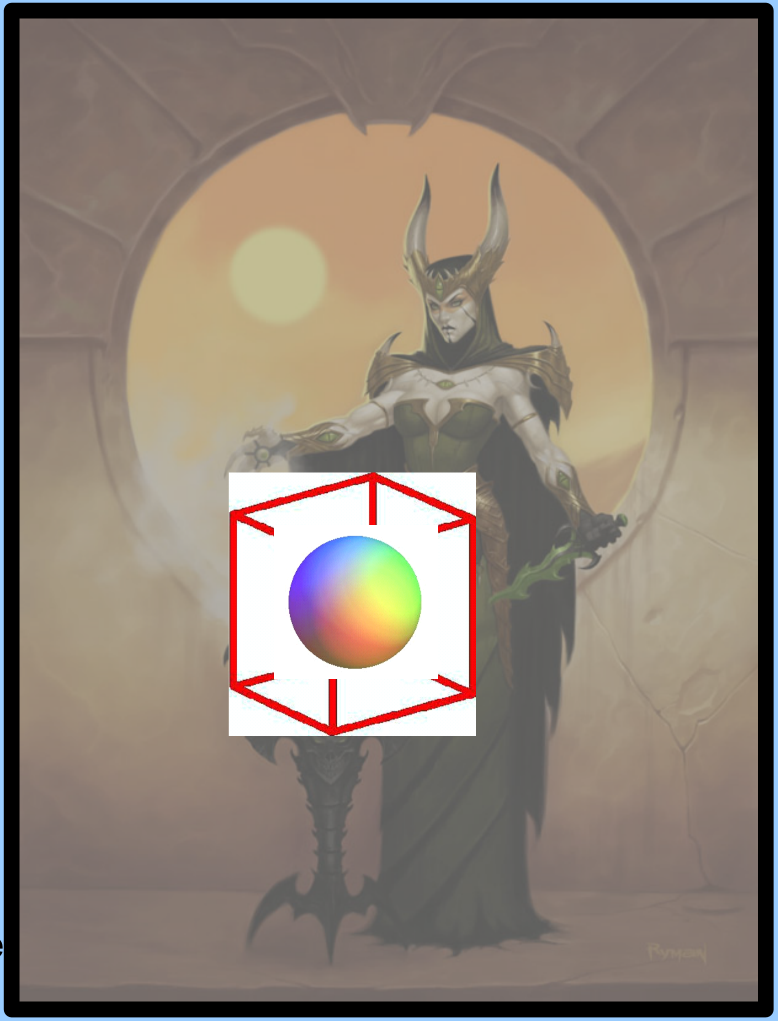


Interpretation

- B→ππ shows same (mis)behaviour
- Removing all experimental info gives similar *P(α)*
- ***The curse of high dimensions is at work***

Uniformity in *x,y,z* makes *P(r)* peak at large *r*

This result is not robust under changes of prior



Conclusions

Bayesian Statistics are

- Illuminating
- Occasionally the only tool to use
- Not the answer to everything
- To be used with care
- Based on shaky foundations ('house built on sand')
- Results depend on choice of prior/choice of variable

Always check for robustness by trying a few different priors. Serious statisticians do.