Part I: Bayes vs chi-squared

Andrew Fowlie

Recap chi-squared method

- Model with parameters $\vec{p} = (p_1, p_2, \ldots)$
- Experimental data, $\vec{d} = (d_1, d_2, \ldots)$
- Build a chi-squared function:

$$\chi^2 = \sum_i (O_i(\vec{p}) - d_i)^2 / \sigma_i^2$$

- Minimise that function with respect to the model parameters.

Recap chi-squared method

- Find, amongst other things:
- A best-fit point, confidence intervals for the parameters.
- Confidence intervals, e.g.

$$m_h = 125.8 \pm 0.6 \, \mathrm{GeV}$$

Confidence interval

- What does it mean? If experiment and analysis repeated many times, 68% of the time, that interval contains the “true” parameter.
- This is “frequentist” statistics.
- Probability related to frequency with which data is obtained.

Frequentist

- Chi-squared is related to a probability - the "likelihood":

$$p(\vec{d}|\vec{p}, m) = N e^{-\chi^2/2}$$

- Probability of data, given model point.
- Minimizing chi-squared is equivalent to maximising likelihood, because the exponential is monotonic.

Bayesian!

- Is that what we really want? Don't we want probability of model point given data, $p(\vec{p}|\vec{d}, m)$
- They are ***not*** the same! E.g.

$$p(\text{pregant}|\text{woman}) \neq p(\text{woman}|\text{pregant})$$

- We see that frequentist stats is constructed "in the data".
- Let's try to construct "in the model".

Bayesian statistics

- We need ***Bayesian statistics***.
- ***Probability no longer related to frequencies***.
- ***Probability is a numerical measure of our belief in a proposition***.
- Bayesian statistics is a "***calculus of beliefs***".
- Won't tell you what they ought to be.
- Will tell you how to update your beliefs in light of data in a logical, consistent way.

Bayes Theorem

- Find what we really want with Bayes theorem:

$$p(\vec{p}|\vec{d}, m) = \frac{p(\vec{d}|\vec{p}, m) \times p(\vec{p}|m)}{p(\vec{d}|m)}$$

- Now we have probability of model point, given data.
- But we’ve had to introduce two new quantities. What are they?

Evidence $p(\vec{d}|m)$

- The evidence quantifies “naturalness”.
- I will return to it.
- At the moment, see it as a normalisation constant.

Prior $p(\vec{p}|m)$

- Prior is source of much controversy & misunderstanding.
- Prior belief, before seeing the data!
- If the data is "strong" enough, different (but fair, honest) investigators should make identical conclusions, regardless of their choices of priors.

What should we choose?

- If you don’t know anything, pick a uniform prior?
- But a uniform prior weights successive decades 10 times more.
- Probability “piles” up at large values.
- Pick a scale invariant prior! Log prior.
- If you do already believe you know the scale, uniform prior.

There *ARE* wrong choices of prior

- There is no “right” choice of prior.
- But there are wrong/dishonest choices. E.g. a delta-function for a parameter that you know nothing about.
- *Repeat*: Bayesian statistics is a calculus of beliefs. It cannot tell you what your prior beliefs should be.

What is the posterior $p(\vec{p}|\vec{d}, m)$

- Main object of interest.
- Probability density of model's parameter space, given the data and the model.
- Probability density function. If you change variables, you need a Jacobian.
- Problem: what if we want to know about one or two parameter? not the whole parameter space?

Marginalisation

- Because it is a pdf, we integrate ("marginalise") parameters we are not interested in, e.g.,

$$p(p_1|\vec{d}, m) = \int p(\vec{p}|\vec{d}, m)\, \mathrm{d}p_2 \mathrm{d}p_3 \ldots$$

- "Volume effect" : parameters that satisfy data without fine-tuning are favoured.
- NB volume effect is occasionally (*wrongly*) considered a fault in the HEP literature.

Credible regions

- Want to present “best” regions of parameter space.
- Regions that contain given fractions, e.g. 68% or 95% of posterior probability.
- There are infinitely many ways of defining such regions!
- Must pick an “ordering rule”

1D credible region

- Tempting to pick the smallest/most dense region that contains the fraction.
- But “smallest” is not parameterisation invariant.
- Pick a symmetric ordering rule:

$$\int_{-\infty}^{L} p(p_1|\vec{d},m)\,\mathrm{d}p_1 = \int_{U}^{\infty} p(p_1|\vec{d},m)\,\mathrm{d}p_1 = (1-0.68)/2$$

2D credible region

- No way to extend "symmetric" rule to 2D.
- This time do pick the smallest/most dense region possible.

$$\int_A p(p_1, p_2|\vec{d}, m)\, \mathrm{d}p_1 \mathrm{d}p_2 = 0.68$$

$$A \text{ is such that } \int_A \mathrm{d}p_1 \mathrm{d}p_2 \text{ is minimised.}$$

2D credible region

- Found numerically, pseudo-code e.g.:
- Sum regions of low density until 1-0.68 of pdf is found

```
crit_density = 0
while area < 1-0.68:
      crit_density += 0.001
      area = sum(pdf, where density < crit_density)
```

2D credible region

- Disadvantage rarely (*never!*) mentioned in HEP literature.
- ***The 2D credible regions are not invariant under parameter transformations***.
- E.g. make credible region on (x,y) plane.
- There will ***not*** be a nice (many-to-one) correspondence with credible region on (x^2, y^2) plane!

Bayesian posterior mean

- In chi-squared methods, identify “best-fit” point, that minimizes chi-squared.
- With posterior pdf, we could find the mode, the point that maximises pdf.
- But that is ***not*** parameterisation invariant,

$$\mathrm{mode}[f(x)] \neq \mathrm{mode}[f(x^2)]$$

$$\bar{x} = \int x \cdot p(x|\vec{d})\,\mathrm{d}x$$

- Instead use posterior mean:

Posterior mean

- It is the expectation for parameter.
- Disadvantage:
- Suppose distribution has many modes, posterior mean might lie between modes!
- Might be a very bad point, e.g. unphysical point with tachyons or incorrect EWSB etc.

Recap, pros & cons vs chi-squared

- Calculate proposition of interest.
- Include prior beliefs, in a formal way.
- Penalise fine-tuning, in a formal way.
- Constructed “in the model” i.e. we think about only the data we have, not pseudo-data from imaginary experiments!

Cons

- Less understood in HEP community.
- Suspicion about “subjective” nature of priors.
- Problems with parameterisation independence.

Part II: Naturalness

EW fine-tuning etc

Naturalness in SUSY

- For ~30 years, theorists have worried about naturalness in SUSY.
- Especially after LEP-II.
- And even more so after LHC 7 & 8 TeV.
- What are they worried about?

EWFT

- Z-boson mass from EWSB is a function of SUSY breaking and preserving, superpotential parameters.
- Unless they are of ~MZ, we need cancellations between large numbers.
- I.e. we need fine-tuning.

EWFT

- EWFT quantified by Barbieri & Giudice:

$$\Delta_{\rm BG} \propto \left| \frac{\partial \ln M_Z}{\partial \ln p_i} \right|$$

- Sensible measure for tuning.
- But it makes no connection with probabilities.
- Since BG, others have made their own measures.

Bayes and fine-tuning

- The fine-tuning measure in Bayesian statistics is the evidence!
- Clear in Trotta, Ruiz et al Balazs et al, and others.

Evidence

- Evidence is probability of data given model.
- Similar to likelihood. It updates prior beliefs about model: $p(m|d) \propto p(\vec{d}|m) \times p(m)$

$$\mathcal{Z} = p(\vec{d}|m) = \int p(\vec{d}|\vec{p}, m)p(\vec{p}|m) \prod \mathrm{d}p$$

- If small, model is fine-tuned - agrees with data only in small part of its parameter space.

Evidence

- Best to consider a ratio of evidences ("Bayes factor")
- A ratio of evidences tells you how you how to update your prior beliefs about 2 models.

$$p(m_A|\vec{d})/p(m_B|\vec{d}) = \mathcal{B} \times p(m_A)/p(m_B)$$

- If Bayes >>1, and you have P(A)/P(B)~1, in light of data, you should now strongly prefer model A.

Evidence applied to EWFT

- In EWFT, the data is the measurement of M_Z. The precision is such that the likelihood is ~a Dirac function.
- Our SUSY model has parameter μ, in the superpotential, and b, soft-breaking bilinear.
- We ought to formulate our priors in μ and b.

Evidence applied to EWFT

- But numerically, tricky to work with μ and b. Switch to M_Z and tan β via EWSB conditions.
- There is an associated Jacobian:

$$J_{ij} = \left| \frac{\partial(m_Z, \tan\beta)}{\partial(\mu, b)} \right|$$

- Proportional to BG measure.

EWFT

- We see that the BG fine-tuning measure is similar to the penalty from Bayesian statistics.
- The penalty drops-out from principles of Bayesian statistics; it is not arbitrary, unlike BG measure.

However

- The BayesFIT analyses in the past parameterised in M_Z *(not μ)* but included no Jacobian.
- Equivalent to having a prior for μ that is always just the μ that gives the right M_Z!
- i.e. not a “fair” prior.

Related note: the μ-problem

- Look again at our prior for M_Z, with a change of
$$p(M_Z) = p(\mu) \times \left[\frac{\partial M_Z}{\partial \mu}\right]^{-1}$$
- EWFT problem is (amongst other things) that that the factor in brackets [] is very big for MZ equal to its measured value.
- i.e. $\mu >> MZ$,
$$\frac{\partial M_Z}{\partial \mu} = -2\frac{\mu}{M_Z}$$

μ-problem

- The μ-problem is that *a priori* μ is unrelated to the EW and SUSY breaking scales.
- What should our prior be? Perhaps logarithmic?
- The evidence for such a model will be diluted! Only very particular values of μ acceptable.
- Solution NMSSM? μ generated dynamically, "naturally" of order SUSY breaking scale.

μ-problem → $p(\mu)$; EWFT → $\left[\frac{\partial M_Z}{\partial \mu}\right]^{-1}$

$$p(M_Z) = p(\mu) \times \left[\frac{\partial M_Z}{\partial \mu}\right]^{-1}$$

Final note: don’t have full picture...

- Can we judge naturalness if we don’t understand the whole model? Motivates e.g. split-SUSY,
- *“...in the landscape picture, the measure is dominated by the requirement of getting a small enough CC … can dwarf the tuning required to keep the Higgs light.”* Arkani-Hamed & Dimopoulos
- Maybe a model with EWFT has less fine-tuning elsewhere