Fitting and statistical inference in science and modelling

- Models in physics have 'free' or unknown parameters

⇒ • Q: Please give examples

- How do we compare models with unknown parameters to data?
- We need techniques (and tools) to 'fit' models to data, and to make statistical inferences

Interlude – Bernoulli (Lindley + Phillips, 1976)

- I will be the scientist; you are statisticians and data nalysts
- I took this pin spoon! . I practised tossing it so it spun many times in flight. I let it land
- I tossed it 12 times. 9 times it landed up! 3 times it pointed downwards

⇒ • Q: what's the chance of up on my 13th toss?

⇒ • Before you answer, you are free to ask any questions you like

...

• 3U, D, U, D, 5U, D [9U, 3D]

YOU DON'T HAVE ENOUGH INFO FOR ANY CLASSICAL STATISTICAL INFERENCE!

<u>was pin unbiased?</u>

P-value : Probability of obtaining a result as or more extreme than observed result under null hypothesis

$$p = \Pr(\lambda \geq \lambda_{obs} \mid H_0)$$

• Let's take H_0: pin unbiased, $p(up) = p(down) = 0.5$, and investigate whether $p(up) > p(down)$

• we observed 3 down, 9 up: as or more extreme: $d = 0, 1, 2, 3,$

⇒ Q : Compute p-value, analytically and/or by MC

• Hint : PTO × 2

Stop - why 12?

- Why was 10 U and 2 D possible?

MAYBE I TOSSED UNTIL I OBSERVED 3 DOWN!

- Maybe I stopped when my son started crying! We need a clearly stated plan!

$\Rightarrow$ • Q: Compute p-value if I stopped after 3 down results.

- We needed to know the stopping rule

Likelihood principle

- Another way of saying it is that we needed to know the sample space - all the outcomes we could have observed
- Why are things that might have happened, but did not, relevant?

| Likelihood Principle: inferences depend on observed data, not sample space of unobserved data

Estimating p

- Now suppose I want to estimate $p(up)$.

> Unbiased estimate: An estimator is unbiased, if, on average it equals the true value in a long-run of repeated experiments
>
> $$E[\hat{p}] = p$$

⇒ Q: compute unbiased estimate of p when $n = 12$ fixed and when $n(down) = 3$ fixed?

⇒ Q: show that they are unbiased through simulation

> - Hints: we can use scipy.stats,
>
>   pip install scipy
>
> - The relevant distributions are binomial and <u>negative</u> binomial - binom and nbinom

- The unbiased estimators are

→ $\hat{p} = \frac{u}{n}$ (n fixed)

→ $\hat{p} = 1 - \frac{d-1}{u+d-1}$ (d fixed)

A Bayesian analysis - modelling

- Let's model the pin tosses using Bayesian statistics
- Use probability to describe reasonable degree of belief
- Probabilities will be updated using Bayes' theorem:

$$p(A|B) = p(B|A) \cdot \frac{p(A)}{p(B)}$$

Exchangeability and de Finetti

- Suppose we consider the tosses to be exchangable: position and order and irrelevant.

UUDUDU $\Leftrightarrow$ any 4 U's and 2 D's

- de Finetti's theorem: the probability must have the form

$$p(\overbrace{u, n}^{\text{any order}}) = \int \overbrace{\theta^{u} (1-\theta)^{n-u}}^{\text{Bernoulli}} p(\theta)\, d\theta$$

- θ is the unknown $\lim_{n \to \infty} \frac{u}{n} = \theta$, but has the interpretation of frequencies of u in long run

- Now, return to question, what is probability of up on 13^{th} toss?

$$P \equiv p(\text{up on } 13 \mid 9U, 3D) = p(10U, 3D) / p(9U, 3D)$$

by Bayes' theorem

- Using de Finetti,

$$P = \int \theta \{ \theta^9 (1-\theta)^3 \, p(\theta) / p(9U, 3D) \} \, d\theta$$

noting that

$$p(\theta \mid 9U, 3D) = \frac{p(9U, 3D \mid \theta) \quad p(\theta)}{p(9U, 3D)}$$

$$P = \int \theta \, p(\theta \mid 9U, 3D) \, d\theta$$

- We have our answer! Alas it depends on a prior $p(\theta)$

Bernoulli on Stan / PyMC etc

Prior – our belief about the parameter before seeing the data.

=) Q : Before seeing the data, what did you think about the pin toss? What would you pick for $p(\theta)$?

MCMC – Markov Chain Monte Carlo. Algorithms for Bayesian computation. Sample-based representation of posterior using chain of correlated steps through possible parameter values.

- Problem defined by

$$\text{target} \propto \text{likelihood} \times \text{prior}$$

- In our case, likelihood is the Bernoulli

$$\theta^{u} \cdot (1-\theta)^{n-u}$$

by de Finetti

MCMC analysis

- Markov Chain $(i+1)$-th state depends only on i-th state

- states are accepted in Metropolis - Hastings with a probability

$$\alpha = \min\left(1, \frac{\pi(\theta_{i+1})}{\pi(\theta_i)} \frac{Q(\theta_{i+1}, \theta_i)}{Q(\theta_i, \theta_{i+1})}\right)$$

pdf in 2nd var

- "Downhill" moves are possible

- Build chain of correlated samples that are asymptotically marginal draws from target
  - stationary distn = target
- How to pick θ_{i+1}? many of these tools use Hamiltonian Monte Carlo — exploiting Hamilton's eq'ns and Liouville th'm

- We wanted

$$p = \int \theta \, p(\theta \mid 9V, 3D) \, d\theta$$

⇒ Q: Find posterior mean of θ.

<u>R-hat and ESS, trace plots</u>

⇒ Q: How many steps was enough?

- ESS – effective sample size. Remember, samples are not independent.

$$\text{errors} \propto \frac{1}{\sqrt{ESS}}$$

- sample until have enough ESS ($\gtrsim 400$ or so usually ok) for p.o.i
- $\hat{R}$ – Gelman-Rubin diagnostic. Compares inter- and intra-chain variance. $\hat{R} < 1.001$ for p.o.i.
- Hint: we will use arviz library

pip install arviz

Summary + refs.

- Bayesian and frequentist methods differ as latter depends on sample space and stopping rule (Lindley + Phillips, 1976)

- we computed p-value, probability data as or more extreme than observed, for 2 stopping rules

- Bayesian approach needed a prior.
- Update prior with data using MCMC (Hogg + Foreman-Mackey, 2017, 1710.06068)

- Explored emcee, pymc, numpyro

  emcee → ensemble

  pymc, numpyro: HMC, derivatives, PPL

- Analysed results using arviz — including diagnostics (Martin et al, 2024, 2110.13017)

Bonus tracks

CL_s

- If you look at LHC results, you will often see "CL_s".

- Step back. What is a confidence interval?

Informally, the range of parameter values "allowed" by an experiment.

- We often want upper limits - how big is allowed?

> An $X\%$ upper limit for a parameter θ is a value V generated by an experiment that in repeated sampling has an $X\%$ probability of being greater than the true value of θ

- Suppose true value of new effect $\mu = 0$ (no effect). We will obtain $V = 0$ at a rate $(1-x)\%$

- If that bothers you, use CL_s!
- It might bother you if you consider.

$$b = 100 \, , \quad s = 10$$
$$\lambda = b + \mu s$$

→ background fluctuates downwards, observe o

70 events. Exclude all values of μ!

- Including $\mu = 10^{-100}$, which surely your experiment was never sensitive to!

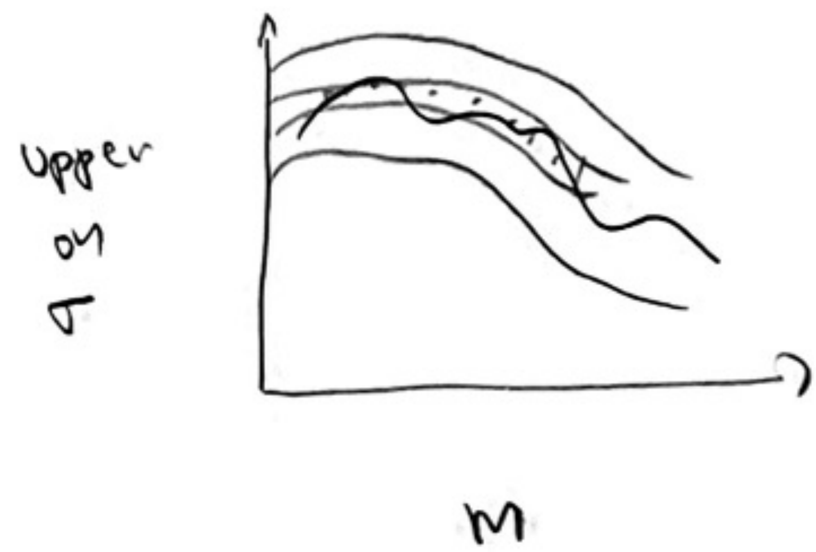
- Enter CL_s! Modify upper limit s.t.

→ "frequentist" definition violated

→ never exclude $\mu = 0$, even if big downward fluctuation in background

- Curiously, CL_s equivalent to a Bayesian procedure in some cases

Brazil Band Plot



- observed upper limit as function of mass
- Brazil band - what do we expect if $\sigma = 0$? simulate it!
- we would get a different result in each simulation
- plot quantiles (~2.5%, ~16%, 50%, ~84%, ~97.5%)
  THIS IS BRAZIL BAND
- Does <u>not</u> depend on observed data; shows our expectations if $\sigma = 0$