Lect. 17 → Renamed Lect 7 (2023dec03) | 2021nov22 | (updated 2023dec21) | 1

Bayesian Analysis (BA)

1) Preamble: BA is path to truth quantified
a scientific method quantified. (p. 2

2) Unquantified or Qualitative Bayesian Analysis
Priors & Posteriors (p. 3)

3) Bayes Theorem & a bit of History (p. 6

4) Proof of Bayesian Analysis in the ideal limit — which can be approach arbitrarily closely (p. 11
~~5) Truth~~ 5) Bayes Odds Ratio & Bayes Factor (p. 21

7) Multinomial Theorem & Multinomial Probability (p. 23)

8) An Example of Bayesian Analysis
A toy example: The Dice Problem

9) Marginalization & Occam's Razor

shortcuts to Bayesian analysis

10) BIC = Bayesian Information Criteria
AIC = Akaike Information Criterion

2)

1) Preamble

(Note in this lecture I give a proof of Bayesian analysis. Not an explication of how to do it in practice — immense amount of tricks, procedures, formulas ... But there are computer packages)

I argue ~~that~~ Bayesian Analysis (BA = AKA Bayesian inference or Bayesian statistics) is a true theory of finding truth or i.e., a true theory in an ideal limit ~~that~~ can be approached arbitrarily closely ideally (at least improvement, The Scientific Method Quantified)

~~Many~~ Many important theories are like this.

E.g., Newtonian Physics,
↳ It is believed to be ~~the~~ macroscopic, limit of QM
low velocity limit of Special relativity
weak gravity limit of General relativity

Another perspective is it an approximate theory

But yours truly (following a herd) would call it a true emergent theory → exactly true in the classical limit which can be approached arbitrarily closely and is approached very closely in everyday life.

Other examples 2nd law of Thermodynamics

Natural selection evolution

unlike Newtonian Physics ~~these~~ can be proven by mathematical logic

— and so can Bayesian Analysis

aside from pure rigorist or philosophical quibbling.

2) Qualitative Bayesian Analysis

⇒ We do it all the time and so all other conscious beings to one degree or another.

All of life experience gives ~~vague~~ probabilities about ~~how~~ what thing will happen in ~~many~~ in a given situation

e.g., impossible, virtually impossible, highly unlikely, unlikely, possibly likely, very likely, virtually certain, certain

+) These are your prior probabilities
or priors
Then New experience happens
in ~~that~~ situation
and you vaguely update
your priors
to your posteriors {i.e. posterior probabilities
(which sounds
awful and often is)
Your updating is also
qualitative and of variable
accuracy,
but it works well enough
mostly in everyday life
~~to imp~~ to your advantage

[Nature via Natural selection
evolution has its own
way improving success rate
relative to local conditions.
Of course, things like major impactors
can change the rules of success
suddenly → like what happened to
non-avian dinosaurs]

Ex. Job interviews
→ especially if you are new at the game — or the 'rules' have changed → each interview causes you to update your priors to posteriors often in an intuitive (but still useful) way.

To some degree 'qualitative Bayesian" analysis
trial & error
improvement if just restricted to each new experience ~~of a unique situation~~ ~~from~~ has very restricted generality.

~~But #~~

But as more general conclusions about updates occur "qualitative Bayesian analysis approaches the scientific method.

6) Bayesian analysis itself is the sci. Method Quantified (I argue)

3) Bayes Theorem

Root of Bayesian analysis is Bayes theorem — which is really simple and simple to prove. Easier to prove than to remember.

Discovered by Thomas Bayes (c.1701–1761) and published posthumously in 1763. (Wik)

Pierre-Simone Laplace (1749–1827) independently discovered it (and published?) it in 1774 (Wik)

Consider 3 events A, B, K.

K is background knowledge that I introduce as extra item because I need it later. It isn't needed for the proof Bayes Theorem itself.

I use ABK as joint event as a shorthand for union = $\cup$ symbol in math which is too klutzy for me in this context.

order has no meaning

Conditional probability

$$P(A|K) = \frac{N_{AK}}{N_K}$$

Probability of A given K

A frequentist definition of probability (Trotta p. 5)

Trotta p. 5 argues you don't need this for probability formalism. But I argue that you need it for probability formalism to have any meaning outside itself.

True N_{AK} and N_K may be vague often but they must exist in some ideal sense. Idealization is the rule as Galileo argued for understanding.

Now

$$P(AB|K) = \frac{N_{ABK}}{N_K} = \frac{N_{ABK}}{N_{BK}}\frac{N_{BK}}{N_K}$$

$$= P(A|BK)P(B|K)$$

a factorization rule which is proven by frequentist approach. But you could assume it as just an axiom — but I think that is pointless.

Clearly

$$P(AB|K) = P(B|AK)P(A|K)$$

∴ Bayes Theorem in symmetric form (easy to remember)

$$P(A|BK)P(B|K) = P(B|AK)P(A|K)$$

Asymmetric forms $P(A|B)P(B) = P(B|A)P(A)$ dropping K — or rather just leaving it implicit.

$$P(A|BK) = \frac{P(B|AK)P(A|K)}{P(B|K)} \qquad P(B|AK) = \frac{P(A|BK)P(B|K)}{P(A|K)}$$

or suppressing K

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

QED.

8] That's all: a profound and profoundly simple logical rule of reality

Root of Bayesian analysis

→ Bayesian analysis was greatly elaboraten in 1939 ~~on~~.
(Scientific Inference 3rd ed 1973
by Harold Jeffreys (1891 – 1989)
(no relation – but he manages to look like my father anyways)

What is Bayesian analysis?
It's a path to true theories
by using Bayes theorem to update prior probabilities for posterior probabilities for theories until one true theory is left

In toy cases this is easy.
In real cases, hard, because there is a lot of slop in assigning priors.
No one much used Bayesian analysis before 1990s. Since then it ~~has~~ had growing vogue.

Why was it unused before?

Well, its real power is in dealing with statistical theories (i.e., theories that only give probabilities) and interesting cases in cosmology, epidemiology, social sciences only came with vast data sets and vast computing power to manipulate them. But since 1990s ~~or~~ so on, we have enough of both.

So nowadays, Bayesian analysis is an ubiquitous tool (though less wonderful than one can hope)

You may ask how can a theory have a probability of being true, Isn't it just true or false

10

speaking in an absolute sense
or having some truth and some falseness definitely in less absolute sense.

But the probability of truth is NOT in the absolute sense.

It's probability to our knowledge

Example

I've a coin in my hand.
Is it heads or tails?

It's one or the other in absolute sense.

But to your knowledge it's 50% - 50%

See, Nothing in my hand
→ So you should include the theory of deception.

9) Proof of Bayesian Analysis in Ideal limit that is path to truth

Say we have background or initial knowledge K_0 about some aspect of reality and a set theories $\{T_i\}$ about that aspect.

{I of the[...]

We assume theories discrete for simplicity. Often they form continuum & they are disting[uished] by free parameter values. ~~separated~~ In fact such continua of theories are usually cons[idered] ou[r] thee[ories] us[...]

[Of course, the set is another source part of K_0]

The $\{T_i\}$ is exhaustive and one is true, others false. (can ideal case)

And we have probabilities of being true (to our knowledge)

$$P(T_i \mid K_0)$$ our initial priors

free para[meters]. But we won't consider that complica[tion] now. Later.

How do you assign $P(T_i|K_0)$

— maybe K_0 gives you an idea

— the barest idea is the Principle of indifference (Kass 1995, p. 776, Not explicit with)

all $P(T_i|K_0)$ are equal.

12

Then you acquire data in a sequence $D_1, D_2, D_3 \ldots D_\ell$

And your knowledge increase

$$K_1 = K_0 D_1$$

I mean union by product $K_0 \cup D$ for simplicity (Wiki union set

$$K_2 = K_1 D_2 = K_0 D_1 D_2$$

$$\downarrow$$

$$K_\ell = K_{\ell-1} D_\ell$$

$$\vdots$$

K_∞ = You know everything but don't be disappointed if you don't get there.

Each data acquisition allows you to update priors to posteriors which then become priors for the next data acquisition

Since we are thinking ideally we assume $\{T_i\}$ is complete.

i.e, $P = \sum_i P(T_i \mid K_0) = 1$

↳ if we are wrong we ~~may~~ ~~ok~~ may not find out.

Consider data acquisition ℓ completed and recall Bayes Thm
(on p. 7)

likelihood but not maximum likelihood

Prior

$$P(T_i \mid K_\ell) = P(T_i \mid K_{\ell-1} D_\ell)$$

Posterior

$$= \frac{P(D_\ell \mid T_i K_{\ell-1})\, P(T_i \mid K_{\ell-1})}{P(D_\ell \mid K_{\ell-1})}$$

We assume you can calculate the $P(D_\ell \mid T_i K_{\ell-1})$ — the probability of D_ℓ given theory T_i and $K_{\ell-1}$

What is this?

$$P(D_\ell \mid K_{\ell-1})$$

$$= \sum_j P(D_\ell \mid T_j K_{\ell-1}) P(T_j \mid K_{\ell-1})$$

$$= \sum_j \frac{N_{D_\ell T_j K_{\ell-1}}}{N_{T_j K_{\ell-1}}} \frac{N_{T_j K_{\ell-1}}}{N_{K_{\ell-1}}}$$

Using ~~the~~ Frequentist definition of probability. But what does ~~that~~ mean here,

You ~~went~~ ~~did~~ a trial (a complex trial of reality) $N_{K_{\ell-1}} \to \infty$ times and got ~~...~~ the distribution of theories

~~N_{K_0}~~ ~~...~~ you assumed by principle of indifference $\frac{1}{\Sigma}$

Your theories were exhaustive

14 2021 nov 22

Lim a vague implicit and hard to calculate sense you get $\frac{N_{T_i K_{e-1}}}{N_{K_{e-1}}}$

and ~~the~~ ratio is easier by far ~~than~~ trying to analyze ~~what~~ the individual N's are ~~probably~~ ~~problem~~ ~~too~~. But you could

N, T, ~~Predicted~~ likelihood

think in a sort of multiverse sense

$N_{K_{e-1}}$ universes and ~~$N_{T_i K_{e-1}}$~~ ~~theory~~ T_j is true in $N_{T_i K_{e-1}}$ of them ~~times~~

Do I believe this. Sort of, but it really takes a more careful analysis

$$P(T_i \mid K_e) = \frac{P(D_e \mid T_i K_{e-1})\, P(T_i \mid K_{e-1})}{\sum_j P(D_e \mid T_j K_{e-1})\, P(T_j \mid K_{e-1})}$$

$$= \frac{P(D_e \mid T_i K_{e-1})\, P(T_i \mid K_{e-1})}{\langle P(D_e \mid T_j K_{e-1}) \rangle}$$

This we identify as the weighted average of getting data D_e: The average likelihood not the maximum likelihood

Recall we assume the original

step 1 or base case of a proof by induction

$$\sum_i P(T_i | K_0) = 1$$

$P(T_j|K_0)$ were normalized.

And our theories were exhaustive

step 2 We assume $P(T_i | K_0)$ are normalized

step 3 $$\sum_i P(T_i|K_\ell) = \frac{\langle \;.\;.\;.\; \rangle}{\langle \;.\;.\;.\; \rangle} = 1$$

QED.

Normalization is conserved

Now $$\underbrace{P(T_i|K_\ell)}_{\text{Posterior}} = \underbrace{\frac{P(D_\ell|T_i K_{\ell-1})}{\langle P(D_\ell|T_j K_{\ell-1}) \rangle}}_{\text{update factor}} \underbrace{P(T_i|K_{\ell-1})}_{\text{Prior}}$$

~~if a~~ [illegible]

$P(D_\ell | T_i K_{\ell-1})$

greater/lesser than ~~[illegible]~~ implies means increase/decrease from prior to posterior.

Now you continue getting new sets of data D_ℓ until

~~and~~ $$P(T_j|K_\ell) = \begin{cases} 1 & \text{for theory } i \\ 0 & \text{for theory } j \neq \end{cases}$$

You might be satisfied with $P(T_j|K_\ell) = 1 - \varepsilon$ with $\varepsilon << 1$.

16

You are guaranteed to get to this point:

a) Your theory set was exhaustive

b) Only 1 was true

c) You can gather new sets until you've exhausted the universe for the aspect of reality that the theories address.

Of course, $l \to \infty$ for exhaustion in some cases but since the universe is finite probably not if you choose data cleverly ~~all wisely~~.

If you choose badly, it might take a long time. {In fact, the probabilities of false theories could increase ~~and~~ true one decrease

Ernest Rutherford aphorism!

= "If you need statistics, you did the wrong ~~ex~~ experiment."

(Actually all aphorisms are both true and false including this one)

The point of Rutherford's aphorism is choose decisive ~~to~~ experiments, but in many area (e.g., cosmology) we don't have many of those — so we are ~~down~~ down ~~to~~ statistics

Also ~~what~~ note what you mean by data can be a broad category.

Data could be results from some completely different aspect of reality from the one you are considering.

It could be a new theory about something else.

It could be a new result in math or logic — absolutely true if axioms are true.

8

[2019 nov 26]

19

The iteration stalls till you complete the cycle

Example of a case where you may never get to the "truth" in practice.

Good one

Complete determin... ac to source ↓ but complete random to recei...

Random number generator

Random or completely deterministic

But in some respects, it does NOT matter

It may not matter. But a profound question. Yes profound

In other respects, it does because we want to know truth

(maybe → are we ready for the truth)

Bill Press, Numerical Recipes (1992?)

p. 191 — 199 but seems to have omitted some of the great discussion I recall

for cycle repeat

$Y \cdot 4.3 \times 10^{6001}$ flops

X 10^{19} flops $\times$ $\frac{3 \times 10^{16} s}{GYr}$

Good ones

$\sim \frac{Y}{X} \times 10^{5970}$ Gyr

Random Number Generators

maybe a quantum computer

peta flops

N flops by rule of Random

4.3×10^{6001}, N

X 10^{15} flops

$\cong 10^{6000}$ Y_a / X_p

Y 1 Gyr 3×10^{16}

→ completely deterministic relative to source & will recycle

but random relative to receiver

Maybe a quantum computer can solve the cycle?

Mersenne Twister (MK) pseudo Random

→ $2^{19937} - 1 \triangleq 4.3 \times 10^{6001}$

20

Say you just ~~were~~ were given like string of numbers

After many statistical tests → maybe a brilliant specially designed test could pull out the determinism, but for most purposes it wouldn't matter

You conclude they are random

in truth you're wrong → absolutely deterministic

but for virtually any purpose → Monte Carlo etc.

they work just as if they were.

Absolutely wrong

effectively right

Something can be completely deterministic as to source but random as to all reasonable effects as to receiver

Philosophically

→ Reality completely deterministic

or with intrinsic random element as in QM

Random in time or random all of time in "initial" conditions

How can you tell?

Could make any difference to universe evolution or human history?

Maybe Not

5) Bayes Odds Ratio + Bayes Factor

In fact getting the sum of probabilities of theories

Bayesian evidence Trotta-56

$$P(D_e|K_{e-1}) = \sum_j P(D_e|T_j K_{e-1})\, P(T_j|K_{e-1})$$

(see p. 4)

is hard,

~~But~~ ~~I think~~ average likelihood NOT Max.

Since really one seldom exhaustively knows all theories (~~or all relevant the~~ Well doesn't want to test all anyway)

One often just wants to compare two or more competing theories ← Relative Probabilities (Whatever covered)

→ So Odds ratio

Practically Often only 1 formal iteration ~~No~~ $\ell=1$

likelihoods → often maximized to fit parameters (But in Bayesian inference marginalize over param)

$$\frac{P(T_i|K_e)}{P(T_j|K_e)} = \frac{P(D_e|T_i K_{e-1})}{P(D_e|T_j K_{e-1})} \frac{P(T_i|K_{e-1})}{P(T_j|K_{e-1})}$$

Posterior Odds ratio (Web) (Wik: Odds Ratio but not explicit)

Kass p.776

Trotta-57

Bayes Factor or Bayes Factor (Wik) or likelihood ratio

LukeBarnes p.6

Prior odds ratio (Web)

often 1 by principle of indifference

22

Jeffrey's ~~log~~ scale ~~scale~~ for k factor theory i to theory j — Kass p.776

based on some empirical analysis.

< 0	negative for theory i
$0 - \frac{1}{2}$	barely worth mention
$\frac{1}{2} - 1$	significant
$1 - \frac{3}{2}$	strong
$\frac{3}{2} - 2$	very strong
$2 \rightarrow$ up	decisive

Why? Subject to later revision you do not expect and admit perhaps seeking in a direct sense.

4) Multinomial Probability Distribution & The Multinonomial Theorem

Say you had I bins from which to draw events or in statistical mechanics terms in which to put ~~parti~~ classical partic... There probabilities are $P_1, P_2, \ldots$

(classical but identical)

Probability of B_i 1	P_1	P_2	P_3	P_4 ...	P_I
selection 2	P_1	P_2	P_3	P_4	P_I
selection 3	P_1	P_2	P_3	P_4	P_I
⋮					
N					

Do N selections with replacement in stats jargon With ... samp...

(Pointon-II calls these arrangements but these are equal probability)

You get sequences: e.g.

sequence 1: $P_1 P_1, P_1 \ldots P_1 = P_1^N$ probability of that one sequence

seq 2: $P_1 P_2 P_1 \ldots\ldots P_1 = P_1^N P_2$

seq: $P_2 P_1 P_1 \ldots\ldots P_1 = P_1^N P_2$

There are 2 sequences of equal probability

~~interchanging~~ permuting the identical subscripts does not give a new sequence

~~interchanging~~ permuting the 1 and 2 indices of sequence 2 gives sequence 3 which is different sequence.

24

I have never found a good way of explaining this, but permuting identical indices gives no new sequences but permuting non-identical ones does.

All I can do is show the example.

But also permuting non-identical indices gives sequences of the same probability and ~~those sequences~~ all put particles ~~in~~ the same ~~state~~,

∴ we'd like to ~~collect~~ sum probabilities of get ~~the same~~ ~~distribution~~ ~~arrangement~~ of particles in the states. → $\{n_i\}$ (set of particles n_1 in state 1, n_2 in state 2 etc.)

a ~~arrangement~~ (combinations w.r.t.)

In statistics jargon, Pointon-Ira calls it a configuration

The trick is $N!$ ways of permuting a sequence

$$\underbrace{N!}_{\text{all permutations in sequence}} = \underbrace{C(\{n_i\})}_{\text{All permutations that give a new sequence}} \underbrace{\prod_i n_i!}_{\text{all permutations of those just in state → these give no new sequence}}$$

n_i

Then $C(\{n_i\}) = \dfrac{N!}{\prod_i n_i!}$ which must be an integer

The probability of any one sequence giving $\{n_i\}$ is $\prod_i p_i^{n_i}$

$\therefore$ The probability of getting the combination ~~arrangement~~, ~~distribution~~ is

$$P(\{n_i\}) = \underbrace{C(\{n_i\})}_{\substack{\text{number of sequence}\\ \text{giving the distribution}}} \underbrace{\prod_i p_i^{n_i}}_{\substack{\text{Probability of getting}\\ \text{any one sequence}}}$$

Probability distribution of ~~the distribution~~ combination

$$= \frac{N!}{\prod_i n_i!} \prod_i p_i^{n_i}$$

$$P(\{n_i\}) = N! \prod_i \frac{p_i^{n_i}}{n_i!}$$

As we know from statistical mechanics this is the distribution for classical identical particles.

The probability distribution for $\underbrace{\text{bosons}}_{\text{Bose-Einstein statistics}}$ & $\underbrace{\text{fermions}}_{\text{Fermi-Dirac statistics}}$ are different

we often say this are because the particles are identical but I think we just mean quantum identical not classical identical

26 — and the particle statistics arise from the symmetrization principle in QM → which nature demands to prevent infinite degeneracy or so it seems.

Normalization?

Proof

$$1 = \sum_i P_i$$ ← sequences of 1 which we demand for putting particles into bin

$$1 = \left(\sum_i P_i\right)^2$$ — sequences of 2 = Binomial theorem

$$1 = \left(\sum_i P_i\right)^N = \text{multinomial theorem}$$ ← sequences of N

Wiki polynomial one coefficient and variables ax^2+bx+c etc. (Wiki)

if $i = 2$ you have the binomial theorem

$$(P_1 + P_2)^N = \underbrace{\sum_{\ell=0}^{N} \binom{N}{\ell}}_{\text{Binomial coefficient}} P_1^{\ell} P_2^{N-\ell}$$

~~and~~ $P(\{n_i\}) = \binom{N}{\ell} P_1^{\ell} P_2^{N-\ell}$

for higher multinomials ~~the formula~~ are harder to get ~~formula~~ easy formula for. Are no easy ones I think.

The key point is that in ~~multiply out the~~ expanding ~~into and collect~~ you get just the sequences ~~and~~ and collecting the ~~coefficient the~~ like terms, the multinomial probability distribution

e.g., $(P_1 + P_2)^2 = P_1 + P_1P_2 + P_2P_1 + P_2^2 = P_1^2 + 2P_1P_2 + P_2^2$

Of course, if you actually started collecting sequences of putting N particles into I ~~states~~, it is only in the limit of $l \to \infty$ sequences ~~that~~ you would recover the multinomial probability distribution

Omit: Digression

Binomial Thm
Special case: Binomial theorem & Binomial probability distribution,

Prove by induction if afflicted by paranoia.

$$1 = (p+q)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

for $q = 1-p$

$$= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

the n choose k formula

Just called binomial coefficient (MiK)

Moments of the binomial distribution.

A old trick

Define function

$$f(x) = (x+q)^n = \sum_{k=0}^{n} \binom{n}{k} x^k q^{n-k}$$

$$M_l = \left(x\frac{d}{dx}\right)^l f \Big|_{x=p} = \sum_{k=0}^{n} k^l \binom{n}{k} x^k q^{n-k} \Big|_{x=p} = \left(x\frac{d}{dx}\right)^l (x+q)^n \Big|_{x=p}$$

28)

$\ell=0 \qquad M_0 = 1$

$\ell=1 \qquad M_1 = x\,n(x+q)^{n-1}\Big|_{x=p} = np$

$\ell=2 \qquad M_2 = n x \frac{d}{dx}\left[\boxed{x}(x+q)^{n-1}\right]\Big|_{x=p}$

$= np\left[(\cancel{x}\,x+q)^{n-1} + x(n-1)(x+q)^{n-2}\right]\Big|_{x=p}$

$= np\left[1 + \cancel{x}\,p(n-1)\right]$

$= pn + p^2 n(n-1)$

$\sigma^2 = \langle (x-\bar{x})^2 \rangle = \langle x^2 - 2x\bar{x} + \bar{x}^2 \rangle$

$= \langle x^2 \rangle - \cancel{\bar{x}}^2 = pn + p^2n(n-1) - p^2n^2$

$= \cancel{p^3 n(n-1)} = pn - p^2 n$

$= pn(1-p)$ BeV-53

correct

208

$p_i = \frac{1}{(\ldots)}$ $\sum p_i = 1$ as it should

Then sampled N events and got $\{n_i\}$

29

Theory 1

$$P(\{n_i\}) = \frac{N!}{n_1! \ldots n_I!} \prod_{i=1}^{I} \left(\frac{1}{I}\right)^{n_i} \qquad \sum_i n_i = N$$

$$= \frac{N!}{n_1! \ldots n_I!} \left(\frac{1}{I}\right)^N$$

This is one I ever [illegible] $P_i \propto 1$ i odd $P_i \propto 2$ Total $= \frac{I}{2} \cdot 1 + \frac{I}{2} \cdot 2 = \frac{3}{2} I$

Theory 3

$$P(\{n_i\}) = \frac{N!}{n_1! \ldots n_I!} \prod_{i=1}^{I} \frac{i^{n_i}}{\left(\frac{I(I+1)}{2}\right)^{n_i}}$$

$$= \frac{N!}{n_1! \ldots n_I!} \frac{1}{\left(\frac{I(I+1)}{2}\right)^N} \prod_{i=1}^{I} i^{n_i}$$

As so often in Astronomy, God has told us there are only 2 possible true theories

Theory 2

$$P(\{n_i\}) = \frac{N!}{n_1! \ldots n_I!} \frac{1}{\left(\frac{3}{2}I\right)^N} \prod_{i=1}^{I} [2 - \mathrm{mod}(i,2)]^{n}$$

Very easy to "postdict" $P(D|T)$ not like complex system with huge data

5) Die Problem for Zeus II

Normal problem

$I = 6$ Theory 1 $P_i = \frac{1}{6}$

six items

Theory 2 $P_i = \frac{2 - \mathrm{mod}(i,2)}{9}$ $\begin{cases} \frac{1}{9} & i \text{ odd} \\ \frac{2}{9} & i \text{ even} \end{cases}$

$P_1 + P_2 + \ldots P_6$

$$\left.\begin{aligned} P(T_1) &= \tfrac{1}{2} \\ P(T_2) &= \tfrac{1}{2} \end{aligned}\right\}$$

Prior odds $P(T_1)/P(T_2) = 1$

Principle of indifference (Barnes p. 6)

So 10 Throws complex enough, but decisive?

(N = 10)

I Really did this in 2019

1	2	3	4	5	6	7	8	9	10
1	5	1	1	2	5	2	6	3	4

$$P(D|T_1) = \frac{10!}{3!\,2!\,2!\,1!\,1!\,1!} \cdot \left(\tfrac{1}{6}\right)^{10} = 2.50057\ldots \times 10^{-3}$$

$$P(D|T_2) = \ldots \left(\tfrac{1}{9}\right)^{10} 1^6 \cdot 2^4 = 6.938\ldots \times 10^{-4}$$

math

3Q)

$K = 3.604$

Probability given your knowledge — Before even, No assumption

So Prior odds = 1

Bayes K factor { $k = 3.604$
Posterior odds = 3.604

If ~~they~~ two theories really are exhaustive, one can calculate their probabilities

$$P(1) = \frac{3.604}{4.604} = 0.7828\ldots, \quad P(2) = \frac{1}{4.604} = 0.21799\ldots$$

If one carried on the experiment to $N \Rightarrow \infty$ for a die,

$P(1) \Rightarrow 1$

$P(2) \Rightarrow 0$

But my die is not perfect and it can be loaded Face 4 opens!!!

I can load it. It's a real vegas die

$P_1(1) = 0.7828\ldots$ $P_2(2) = .217799\ldots$ the posteriors → priors

Do another set of 10

10·9·1·7·6·5·4

1	2	3	4	5	6	7	8	9	10
6	1	3	6	6	2	1	5	3	5

The House never loses. And we are the House

$$P(D|T_1) = \frac{10!}{3!2!2!1!2!} \left(\frac{1}{6}\right)^{10}$$

$$P(D|T_2) = \ldots\ldots \left(\frac{1}{9}\right)^{10} (1)^6 (2)^4$$

But posterior odds
supp.?
$= k \times \left(\frac{P(T_1|K_1)}{P(T_2|K_1)}\right) = k^2$
≈ 13

So in math 2
we get $k = 3.604$ again
because same number of even & odd
But favours theory 1 twice in a row

$P_2(1k_2) = .9285$ ... $P_2(2k_2) = .0748$ ...

So do we keep advancing to truth by more & more trials?

But no! Instead of rolling die,
we investigate its symmetry

The ideal die Not my imperfect one

$D_3 \equiv$ ~~6~~ fold symmetry

Rutherford Rule

$P(D_3 | 1k_2) = 1, \quad P(D_3 | 2k_2) = 0$

$$P(1|K_3) = \frac{P(D_3|1k_2)\, P(1|k_2)}{P(1|k_2)\cdot P(D_3|1k_2) + P(2|k_2)\, P(D_3|2k_2)}$$

$P(1,k_2)$ & $P(2,k_2)$ are now irrelevant

$= 1$

leap & but factor in other orders!

What if you can't or don't know how or it's just harder to do the basic single experiment.

$P(2|k_3) = P(D_3|2k_2) * P(2|K_2)$

$= 0$

So Rutherford was right!! Tvo Ha v.1
"If you need statistics, you did the wrong experiment"

32) But if the system (2019 nov 26)
is intrinsically
probablistic,
(Die is)
and you can not
find ~~its~~ distribution
directly (as you can for Die)
~~then~~ Bayesian analysis
makes sense.

~~Perhaps~~ - Cosmology
- social science
- epidemiology

maybe be common choice?
Trotta's choice for set of free param

avg individual error

5) Marginalization

Say your ~~theory~~ theory
~~has~~ a free parameter θ
it's a continuum infinity
of theories.
But best to regard as ~~one~~ one theory
with free parameters

If you think your theory

density
$P_x = \frac{dx}{g}$

$D_e | T, k_{e-} \rangle \langle T, k_{e-} | dx$

$T(x)$
integrate

Set of ...

Marginalization

2021nov22

What if your theories are NOT discrete: i.e. theories $\{T_i\}$ have free parameters collectively symbolized Θ?

Bayes Odds ratio

$$\frac{P(T_i(\Theta)|K_\ell)}{P(T_j(\Theta)|K_\ell)} = \frac{P(D_\ell|T_i(\Theta)K_{\ell-1})}{P(D_\ell|T_j(\Theta)K_{\ell-1})} \frac{P(T_i(\Theta)|K_{\ell-1})}{P(T_j(\Theta)|K_{\ell-1})}$$

Bayes K factor

Prior ratio

You don't believe You have all theories, even all interesting theories, and so the denominator of Bayesian Analysis is ~~probably~~ usually ~~meaningless and~~ impossible to know — but that exists in principle is vital to Bayesian analysis being ~~a~~ true theory.

So you just do Bayes Odds Ratio.

Almost always $\ell = 0$ and these are the initial priors since one only does one explicit Bayesian Analysis iteration. All past knowledge K_ℓ is implicit earlier iterations.

But what do you do about free parameters and the priors for a theory with them?

Do you choose the parameters ~~by~~ by maximum likelihood

$\stackrel{?}{=} P(D_e | T_i(\theta_{max}) k_{e-1})$; $\mathcal{L}_{max}$ ~~$\mathcal{L}(\theta_{max}, D_e)$~~ ?

~~for data in hand~~

~~If a theory is true, maximum~~

Maximum likelihood gives best choice assuming a theory is true,

$$\lim_{D_e \to \infty} \{\theta_{max}\} = \theta_{true}$$

but not if the theory is not true.

If you assume equal priors for two theories, i and j, one with far more free parameters (theory i) then

$$k_{factor} = \frac{P(D_e | T_i(\theta_{max}^{\text{large number of parameters}}) k_{e-1})}{P(D_e | T_j(\theta_{max}^{\text{small number of parameters}}) k_{e-1})}$$

might be spuriously large since the θ_{max}'s are fitted to the D_e.

If you could correct by choosing unequal priors to put theories on an equal footing but what is your guidance.

Usually none. In any case, θ_{max} may be useless choice of θ if T_i is not true

The path is marginalization. (Implements Occam's razor by effectively eliminating plausibility of parameters of uncertain ~~value~~ use and uncertain value.)

Nuisance param... don't ca... abo...

Expand

$$P(D_e \mid T_i\, K_{e-1})$$

$$= \int P(D_e \mid T_i(\theta)\, K_{e-1})\, P(\theta)\, d\theta$$

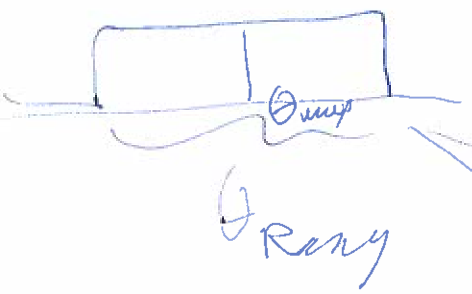
where $P(\theta)$ is a probability density for the priors

Usually, one just guesses

Can do more complicated things but need guidance

$$P(\theta) = \begin{cases} \dfrac{1}{\theta_{range}} & \text{for } \theta \text{ in } \theta_{range} \\ 0 & \text{for } \theta \text{ out of } \theta_{range} \end{cases}$$

The range that you think at all possible



Maximum likelihood parameters are Not favored, since you don't know T_i is True

36

— as long as you ~~set~~ your range fairly theories with ~~unequal~~ equal sets of parameters are on an equal footing and you can use equal priors again.

So marginalization implements Occam's razor effectively. It puts 'interesting' theories on an equal footing.

↳ eliminating pluralities

Of course for complex theories with many free parameters θ and petabytes or exabytes of data to fit/reproduce ~~Make an unf~~

$$P(D_e | T_i k_o)$$

$$= \int P(D_e | T_i(\theta) k_e)\, \rho(\theta)\, d\theta$$

can be an immense supercomputer problem.

So it's not exactly choosing formal prior $P(T_i | k_o)$ that is the uncertain part — Just set equal for interesting theories. The uncertain part is choosing $\rho(\theta)$ (a sort of prior)

Make an unfair choice
and your Bayes K factor
could be off by orders of magnitude

→ This why the Jeffreys's scale
sets $\log K = 2$, ($K=100$)
as decisive

and in cosmology $K \approx 10$
is considered completely undecisive

I'd guess.

2019 nov 26

39

If you have vast sets of data

Cosmological peta bytes

1 pb = 10^{19} bytes

LHC ~~in~~ 15 pb per year

couple process 24 pb per day in 2009

— Integration can be huge even for our modern standard just doing ~~[illegible]~~ K factors

— Are shortcuts of varying utility of which I no for know little

a) Paxysgical Information crit (Wilk (Wilk = DIC eg Schwarz crit

Finding likelihood $P(D|T_i(\theta_i))$

for the problem simple but petabytes produce & [illegible] a complex theory

I intuit that here you need Markov chain Monte Carlo MCMC (Votta-35 and a lost skill.

dense focus Bayes an approx

40

derive from information theory

1) Akaike Information Crit. (Wik

AIC

relative

quality of model

Not absolute

(from information theory

smaller better

$$AIC = 2k - 2\ln L$$

for number of parameters the same

$[\ln(\frac{L_{max}}{L_{max}, AIC\ in\ min})$

$\cong k$ or ~~other~~

author order

$k\uparrow$ AIC$\uparrow$

number of estimated parameter

max likelihood of model

$L\uparrow, AIC\downarrow$

by someone proof

Best of set ~~of All~~

$[AIC_{~~min~~} - AIC]/2$

~~$= [\ldots 2\ln L]/2$~~

~~$= 2\ln(\ldots)/2$~~

like k or posterior order

Probably better than comparing

here 1, other lower

BIC NOT

similarly to ~~Bayesian~~

some different weighting

somehow related

Chi-Square

Bev-314 which seems choice $p < 0.05$ or more

~~Theory~~ BIC & AIC must work well in some ~~ideal~~ limits & so useful

but can't be universally used or people