Think Bayes

Chapter 1 - Bayes Theorem

→ probability - chance that something will happen
↳ 0 - impossible
1 - certain.
0.5 - outcome is as likely as not.

→ conditional probability
↳ background information is given

$$p(\text{first heart attack}) = \frac{7.85 \times 10^5}{3.11 \times 10^8} = 0.3\%$$

↳ each person has attributes that make event more or less likely

- high cholesterol
- age

vs.

- low blood press.
- ? smoker

$\boxed{P(A|B)}$ = probability of A given B

→ conjoint probability

↳ How likely is it that two events happen together?

$$p(A \cup B) = \underbrace{p(A)\,p(B)}$$

↳ given A & B are independent.

or $P(B|A) = P(B)$

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

∴ it is really

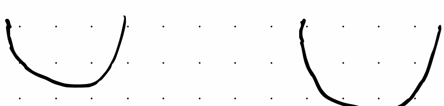
these are equal

$$p(A \cup B) = P(A)\,P(B|A)$$

$$P(B \cup A) = P(B)\,P(A|B)$$

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

→ eg



cookie flav	A	B
vanilla	30	20
chocolate	10	20

if you draw vanilla which bowl is most likely?

$$P(v|A) = \frac{30}{30+10} \qquad P(v|B) = \frac{20}{20+20}$$

$$\therefore P(v|A) > P(v|B)$$

now think of the problem in terms of bayes theorem.

$$P(A \mid V) = \frac{P(V \mid A)\, P(A)}{P(V)} \quad \begin{cases} P(V \mid A) = \frac{3}{4} \\ P(V) = \frac{30+20}{2(40)} = \frac{5}{8} \\ P(A) = \frac{1}{2} \end{cases}$$

$$= \frac{(1/2)(3/4)}{(5/8)}$$

$$= 3/5$$

Possible to use B.T. to get $P(A \mid B)$ from $P(B \mid A)$

→ <u>Diachronic Interpretation</u>

update hypothesis given data.

$$\underbrace{P(H \mid X)}_{\text{posterior}} = \frac{\overbrace{P(X \mid H)}^{\text{likelihood}}\, \overbrace{P(H)}^{\text{prior}}}{\underbrace{P(X)}_{\text{normalizing}}}$$

- likelihood can be difficult to calculate re prob. of data given any hypothesis

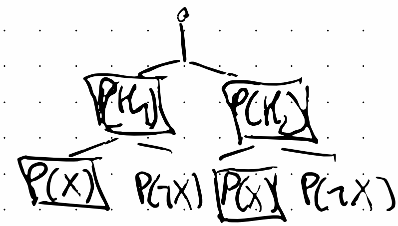
solution:

→ mutual exclusive hypothesis
→ collective exhaustion

} suite

If these cases are true then

$$P(X) = P(X \mid H_1)P(H_1) + P(X \mid H_2)P(H_2)$$



→ **The M&M problem**

M&Ms change over time

	Brown	yellow	red	green	tan	blue	orange
< 1998	0.3	0.2	0.2	0.1	0.1	0	0.1
> 1995	0.13	0.14	0.13	0.2	0	0.24	0.16

↳ friend gives two bags, one of each kind

↳ take 1 m&m from each bag

we get: yellow, green. what is $P(y \mid <1995)$?

$$P(y \mid <1995) = \frac{P(<1995 \mid y)\, P(y)}{P(<1995)}$$

$$= \frac{\frac{0.2}{0.34} \quad \frac{0.34}{2}}{\frac{1}{2}} = 0.2?$$

you can also make tables showing the outcomes

hypothesis	prior $p(H)$	likelihood	$p(H)p(P\mid H)$	Posterior
A	1/2	(20)(20)	200	20/27
B	1/2	(14)(10)	70	7/27

Σ of these is norm. = 270

→ The monty hall problem

↳ 3 doors → 1 car, 2 goats

you pick door A, monty shows you a goat in B

Should you change to C.

hypothesis	prior	likelihood	$P(H)P(X\mid H)$	Posterior
A	$\frac{1}{3}$	$\frac{1}{2}$	1/6	1/3
B	$\frac{1}{3}$	0	0	0
C	$\frac{1}{3}$	1	$\frac{1}{3}$	$\boxed{2/3}$

car is in A, B, or C!

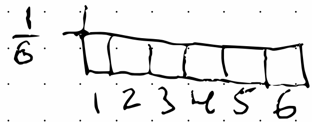
$P(\text{goat} \mid B)$

$$\Sigma = \frac{1}{6} + \frac{1}{3} = \frac{3}{6} = \frac{1}{2}$$

Better off switching

Chapter 2 - Computational Statistics

- **Distribution**: values & their probabilities
  - ↳ 6-sided dice



uniform distribution

- Otherwise this chapter is mostly implementations of problems in the last chapter.
- check github repo for notebooks.

Chapter 3 - Estimation

example: box w/ d4, d6, d8, d12 & d20 dice

select dice at random & roll

↳ what is prob for each die?

→ Approach:
- develop hypothesis
- choose representation for data
- write likelihood function

Try encoding this problem on a notebook

→ The Locomotive Problem

→ locomotives named in order 1 ... N

↳ see locomotive w/ number 60

How many locomotives are there

1. What do we know about N before? (Prior)

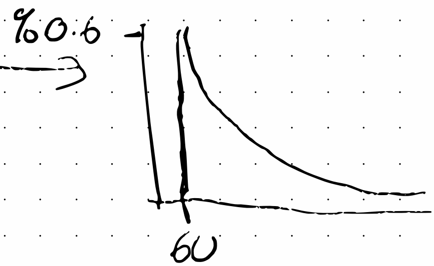
$n_{loc} \geq 1$

2. What is the likelihood of seeing the data? (Likelihood)

→ Assume 1000 hypothesis, all equally likely

→ likelihood is $\frac{1}{hypo}$ for hypo ≥ 60 else 0

posterior →



Still, seems unlikely

Avg(hypo * prob.) = 333

↳ minimises sq. error over time

→ What about the prior

In last problem we chose 1000 trains

↳ Changing changes the mean of our posterior !!

What to do? 1. more data 2. more info

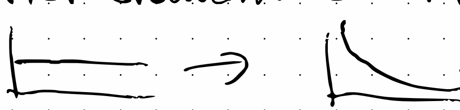
→ Alternative prior

↳ gather more background info

Find out distribution of trains using the internet!

observation: n trains follows power law.

prior shouldn't be uniform!



$\leftarrow PMF \propto \left(\frac{1}{x}\right)^{\alpha}$

$$\frac{p(600 \mid H_2)}{p(600 \mid H_1)} = 3$$ ← 600 became much more likely

power law is less sensitive to upper bound to boot!

→ Credible intervals

single point estimates: mean
median
mode
max likelihood

"credible"
~~confidence~~ intervals?

↳ $P(X \leq 95\% \mid \text{posterior}) - P(X \leq 5\% \mid \text{posterior})$

re w/ locomotive problem 90% C.I. is (91, 243)

Shows we are still quite uncertain.

→ German Tank Problem

↳ American embassy estimates german tanks.

info → locations, logs, inventories, repairs, etc.

Stats incorporating this data was much more accurate.

Chapter 4: More Estimation

→ The euro problem

When spun on edge, coin tosses 140 heads & 110 tails. p val is 0.07. Is coin biased?

$p(x) = 0.5$, x = coin lands heads.

→ Step 1: define priors.

↳ range (0, 100) values for x

→ Step 2: likelihood.

↳ p(x | biased coin), ie p(outcome)

posterior is a bell curve slightly above 50%

↳ max likelihood of 56

↳ mean 55.95, mode 56

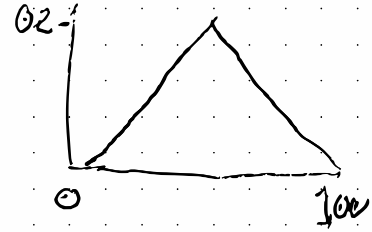
↳ CI of (51, 61)

↖ 50% just outside of CI but 101 hypothesis was arbitrary

Note: uniform prior might not make sense

↳ coin isn't as biased as 0.1 or 0.9

What about a triangular prior?



↳ we get a similar dist. maybe the coin is a little biased.

→ Beta distribution

↳ continuous dist. from 0 to 1, good for representing distributions.

- if likelihood is binomial like w/ coin toss then it is a conjugate prior.
- depends on parameters α & β.
  ↳ update w/ $\alpha + H$ & $\beta + t$ ($\alpha = 1, \beta = 1$ uninformed prior)

we can update priors in one step now
↳ converges on solution above.
(we have a lot of data so this is expected.)

Chapter 5: Odds & Addends

"Odds": ratio expression of probability.

e.g. $P(x) = \frac{1}{2}$, odds$(x) = 1:1$

↳ $1/(1+1) = 0.5$

→ Odds in bayes theorem

Normal Bayes theorem: $P(H|D) = \frac{P(H)P(D|H)}{P(D)}$

↳ 2 hypothesis $= \frac{P(A|D)}{P(B|D)} = \frac{P(A)P(D|A)}{P(B)P(D|B)}$

Note: No more need for $P(D)$

If A & B are Mutually Excl. & Collectively Exhaustive

↳ $P(A) = 1 - P(B)$

then $o(A|D) = o(A)\frac{P(D|A)}{P(D|B)}$, $o(x)$ = odds of x.

Applied to cookie problem:

$\Rightarrow o(\text{bowl A} \mid \text{vanilla}) = o(1:1)\,\frac{3/4}{1/2} = 3/5$

	V	C
Bowl A	3	1
Bowl B	2	2

→ Oliver's Blood

Blood at scene of a crime by 2 people
↳ Oliver has 'O' type (suspect)

traces were 'O' & 'AB', population rate of 60% & 1% respectively.

↳ Is 'O' proof that oliver was at the scene

↳ equation: $\frac{O(A|D)}{O(A)} = \boxed{\frac{P(D|A)}{P(D|B)}}$ ← Bayes factor

Oliver at CS

↳ responsible for O

P 'AB' $= 0.01$

Oliver ¬at CS

↳ ('O' & 'AB' or 'AB' 'O')

$2(0.6)(0.01) = \boxed{0.012}$

Likelihood is higher if oliver is not there ∴ 'O' blood is evidence against Oliver's guilt!

↳ More ways for 'O' to arrive at CS.

→ Addends

in DnD dice are used to generate player stats.

↳ {str. int. wis. dex. const. char.}
1 2 3 4 5 6

↳ each set by $\boxed{\text{sum of } 3\times d6}$

↳ What is this dist?

Compute by
1) Simulation
2) Enumeration

→ Simulation: roll a bunch of electronic dice & check experimental result

→ Enumeration: Exhaustively find all combinations

→ Maxima

↳ What is a character's best attributes?

Found by sim. & enum. as before...

→ Exponentiation

↳ convert Pmf → Cdf

Why do we do this conversion?

↳ Simple algorithm for finding the max of a cdf.

reason:

$$CDF(x) = P(X \leq x)$$

↑ number chosen at random from dist

if I draw from random from 2 dist. (Z)

$CDF_1 \rightarrow X$

$Z = \max(X, Y)$

$CDF_2 \rightarrow Y$

↑ this value will have its own distribution.

if the distribution is the same

$$CDF_k(Z) = CDF_1(Z)^k$$

↳ this means we can enumerate the probabilities & raise them to the kth power

→ Mixtures

↳ Now, if we have a mixture of dices, how can we estimate the posterior distribution

↳ each dice has a weight associated with it. (the likelihood it is drawn)

↳ we can see that all dice can roll a 4 & so it has the largest probability of being drawn.

Chapter 6 - Decision Analysis

↳ Skipping chapter because of difficulty reading data & implementing code.

Chapter 7 - prediction

↳ The Boston Bruins Problem

VC	vs	BB
1	-	0
3	-	2
1	-	8
4	-	0

4 games

↳ best of 7 wins, who wins turnament?

Assumptions

↳ Scoring a goal is a poisson process

↳ score from 1st 4 games determine λ

↳ use posterrer λ for each team to estimate victory distributions

→ n goals per game is normally dist. w/ mean 2.8 w/ std 0.3

→ Poisson processes

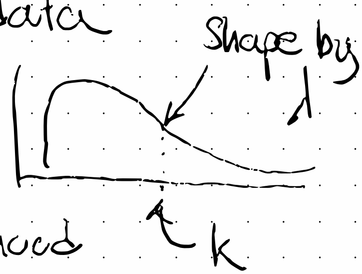
"process" → stochastic model of a physical sys
↳ "has randomness"

Poisson ← events happen w/ equal likelihood in "time" w/ rate prop to λ

→ Posteriors

↳ compute likelihood for each data

λ ← hypothesis k = data



↳ priors were gaussian & when multiplied by likelihood then gaussian curve shifts.

→ The distribution of goals

↳ we can do this by evaluating a pmf

Not finishing this book due to bad.