Bayesian Computational Statistics

reference book: Bayesian Data Analysis; Gelman et al.

Bayes' rule and its consequences

conditional probabilities

$P(A|B)$ probability of event A given that event B has occured.

$$P(A|B) = P(A \cap B) / P(B)$$

example: We have a standard 6 sided die

$$A = \{5\}, \quad B = \{1, 3, 5\}$$

$$P(A|B) = P(A \cap B) / P(B) = 1/3$$

$$P(A) = 1/6$$

$C = \{2, 4, 6\}$ so $P(A \cap C)/P(C) = 0/3 = 0$

Note : $P(A|B) \neq P(B|A)$

Bayes' rule relates $P(A|B)$ to $P(B|A)$:

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

It gives us a framework for making and updating estimates of $P(A|B)$ based on evidence.
Updating our beliefs in the face of new information.

<u>example</u>: medical testing

event A: having the disease
event B: testing positive

prior $P(A)$: 0.01 (1%)

likelihood $P(B|A)$: 0.99

false positive rate
to compute the marginal 0.05

marginal: $$P(B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)$$
$$= 0.99 \cdot 0.01 + 0.05 \cdot 0.99$$
$$\approx 0.0594$$

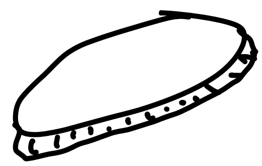
posterior $$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{0.99 \cdot 0.01}{0.0594} \approx \frac{0.1\overline{6}}{2}$$

a bit surprising.

Bayesian inference

1. Start with a prior distribution
2. Collect data
3. Compute the likelihood
4. Compute the marginal probabilit
5. Compute the posterior via Bayes' rule (updated beliefs)

example: Is it a fair coin?



H_0: It is fair

H_1: It is biased

prior: 0.5

Data: T, T, T

$$P(H_0 \mid TTT) = \frac{P(TTT \mid H_0)\, P(H_0)}{P(TTT)}$$

$$P(TTT) = P(TTT \mid H_0) \cdot P(H_0) + P(TTT \mid H_1) \cdot P(H_1)$$

$$P(TTT \mid H_1) = \int_0^1 p^3 \, dp = \frac{p^4}{4}\Big|_0^1 = 1/4, \quad \text{so}$$

$$P(H_0 \mid TTT) = \underline{1/3}$$

Fundamentals of Bayesian Inference

Bayes' Rule
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Quizz exercise:

		prior	data
H_0 :	coin is fair	0.66	HHHHH
H_1 :	coin is biased	0.34	(5H)

likelihood:

$$P(5H \mid H_0) = \left(\tfrac{1}{2}\right)^5 = 1/32 = 0.03125$$

$$P(5H \mid H_1) = \int_0^1 p^5 \, dp = \left[p^6/6 \right]_0^1 = 1/6$$

marginal:
$$P(5H) = p(5H \mid H_0) \cdot p(H_0) + p(5H \mid H_1) \cdot p(H_1)$$
$$= 0.03125 \cdot 0.66 + (1/6) \cdot 0.34$$
$$\approx \underline{0.078}$$

posterior:
$$p(H_0 \mid 5H) = \frac{p(5H \mid H_0) \cdot p(H_0)}{p(5H)}$$

$$= \frac{0.03125 \cdot 0.66}{0.078} \approx \underline{0.268}$$

Bayesian Inference:

process of fitting a probability model to a set of data using Bayes' rule

<u>Notation</u>:

θ : parameter, scalars or vectors
e.g $\theta = (\beta_0, \beta_1)$

y : observed data

$\tilde{y}$: unknown but potentially observable data

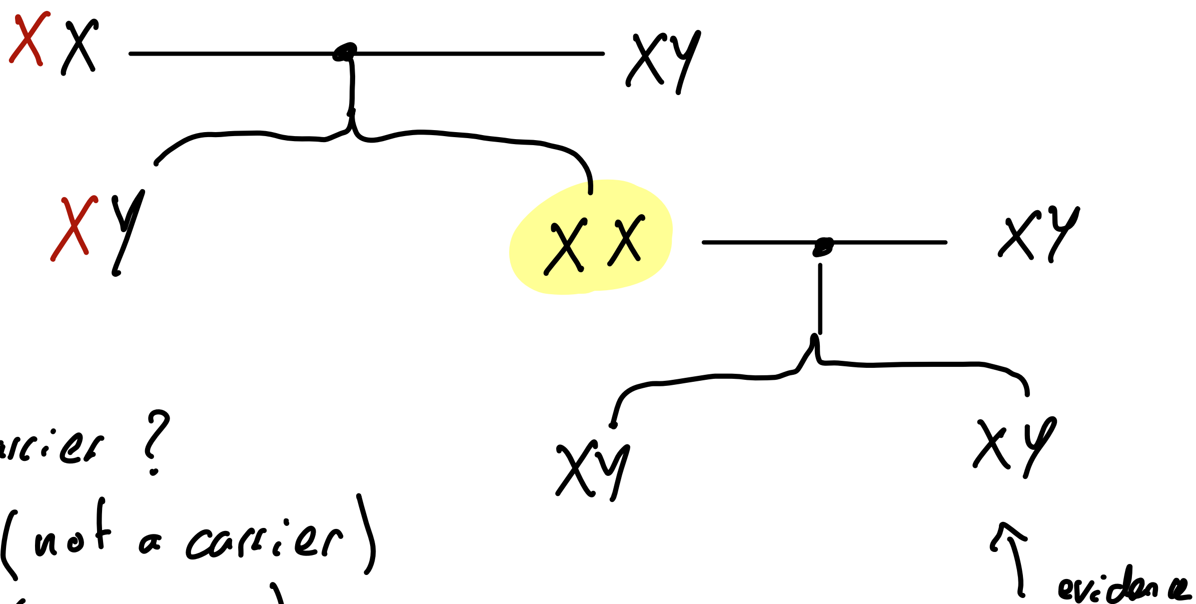
$p(x)$: pdf of x

$p(x, y)$: joint distribution of x and y

<u>Hemophilia example</u>:

from the book

X-linked trait



Yellow is a carrier ?

$H_0 \quad \theta = 0$ (not a carrier)

$H_1 \quad \theta = 1$ (a carrier)

data : $\bar{y} = (0, 0)$

prior : $(50/50) = 0.5$

likelihood : $p(\bar{y} \mid \theta = 1) = 1/2 \cdot 1/2 = 1/4$

$p(\bar{y} \mid \theta = 0) = 1 \cdot 1 = 1$

marginal (for scaling) : $p(\bar{y}) = p(\bar{y} \mid \theta = 1) \cdot p(\theta = 1) + p(\bar{y} \mid \theta = 0) \cdot p(\theta = 0)$

$= 1/4 \cdot 1/2 + 1 \cdot 1/2 = \underline{5/8}$

posterior :

$p(\theta = 1 \mid \bar{y}) = \frac{1/8}{5/8} = 1/5 = \underline{0.2}$

What if there is a third child who is also XY and not afflicted. (update)

$$\tilde{y} = (0)$$

$$p(\theta = 1 \mid y) = \frac{p(y \mid \theta = 1) \cdot p(\theta = 1)}{p(y)}$$

$$= \frac{(1/2) \cdot (1/5)}{\frac{1}{2} \cdot \frac{1}{5} + 1 \cdot 4/5} = 1/9 \approx \underline{0{,}11}$$

Exchangeability

order of observations doesn't matter

subjectivity vs objectivity

$\Rightarrow$ prior

Example from the book

Spelling corrections

'radom' $\rightarrow$ 'random'
'radom' $\rightarrow$ 'radon'
'radom' $\rightarrow$ 'radom'

data : y = radom

$$p(\theta \mid y) \propto p(\theta)\, p(y \mid \theta)$$

scaling can be done easily at the end

prior:

θ	rel. freq	prob
random	$7.6 \cdot 10^{-5}$	
radon	$6.05 \cdot 10^{-6}$	
radom	$3.12 \cdot 10^{-7}$	

$\Rightarrow$ rewriting

θ	rel. freq.	prob
random	$760 \cdot 10^{-7}$	0.923
radon	$60.5 \cdot 10^{-7}$	0.073
radom	$3.12 \cdot 10^{-7}$	0.004

likelihood

θ	$p('radom' \mid \theta)$
random	0.00193
radon	0.000143
radom	0.975

posterior:

θ	$p(\theta)\, p('radom' \mid \theta)$	$p('radom' \mid \theta)$
random	$1.47 \cdot 10^{-7}$ $(1470 \cdot 10^{-10})$	~ 0.325
radon	$8.65 \cdot 10^{-10}$	~ 0.002
radom	$3.04 \cdot 10^{-7}$ $(3040 \cdot 10^{-10})$	~ 0.673

example where we don't need marginals because we can scale the results at the very end.

Bayesian Computation

Faulty caliper problem from the book

length in millimeters
off by -1mm or 1mm

		prior
$\theta = 1$	$y \sim N(1,1)$	0.5
$\theta = -1$	$y \sim N(-1,1)$	0.5

likelihood:

$$p(Y=0.5 \mid \theta=1) = \frac{1}{2\pi} e^{-\left(\frac{(y-1)^2}{2}\right)} \approx \underline{0.1405}$$

marginal:

$$p(y=0.5) = p(y=0.5 \mid \theta=1)\, p(\theta=1) + p(y=0.5 \mid \theta=-1)\, p(\theta=-1)$$

$$\approx \underline{0.09605}$$

posterior:

$$p(\theta=1 \mid Y=0.5) = \frac{0.07022}{0.09605} \approx \underline{0.73}$$

Stan package in R

PyStan in python

example in R : estimation of a distribution

General approach to Bayesian Computation

Binomial and Posterior Distributions

Binary data 0, 1 Bernoulli outcomes

$$p(y|\theta) = \binom{n}{y} \theta^{y} (1-\theta)^{n-y}$$

y: number of success
θ: proportion of success

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}$$

Biased $\theta = 0.75$

$H = 1$, $T = 0$

What is the probability of TTT ?

$$p(y=0, n=3 \mid \theta = 0.75) = \binom{3}{0} 0.75^{0} \, 0.25^{3} \approx \underline{0.016}$$

$$p(y=1, n=3 \mid \theta = 0.75) \approx \underline{0.14}$$

$\{$TTH or THT or HTT$\}$

<u>example 2</u>

θ : proportion of female birth

y : number of female birth in n recorded births

$\theta \sim U_{[0,1]}$

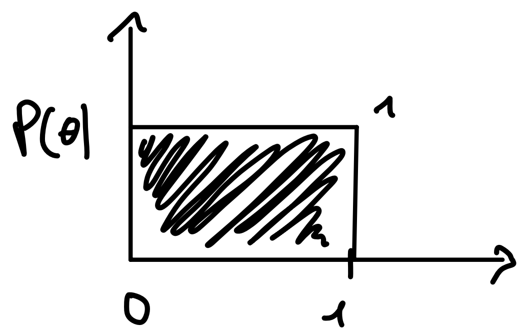
What is the posterior distribution?

Binomial likelihood.

$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

uniform prior:

$p(\theta) = 1$ for $\theta \in [0,1]$



posterior:

$$p(\theta|y) \propto p(\theta)\, p(y|\theta)$$
$$= \cancel{\binom{n}{y}} \theta^y (1-\theta)^{n-y} \qquad \theta \in [0,1]$$

$$p(\theta|y) \propto \theta^y (1-\theta)^{n-y}$$

once normalized

$$p(\theta|y) = B(\underbrace{y+1}_{\alpha}, \underbrace{n-y+1}_{\beta})$$

Beta distribution

$\tilde{y}$: predictions with the m next births

$$\tilde{y} \sim Bin(m, \theta)$$

so $\theta \sim$ Beta and no longer Uniform

$$p(\tilde{y} \mid y)$$