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: $P(B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)$

$$= 0.99 \cdot 0.01 + 0.05 \cdot 0.99$$

$$\approx 0.0594$$

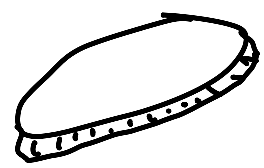
posterior $P(A|B) = \dfrac{P(B|A) \cdot P(A)}{P(B)} = \dfrac{0.99 \cdot 0.01}{0.0594} \approx \dfrac{0.1\bar{6}}{2}$

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

Notation:

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

y : observed data

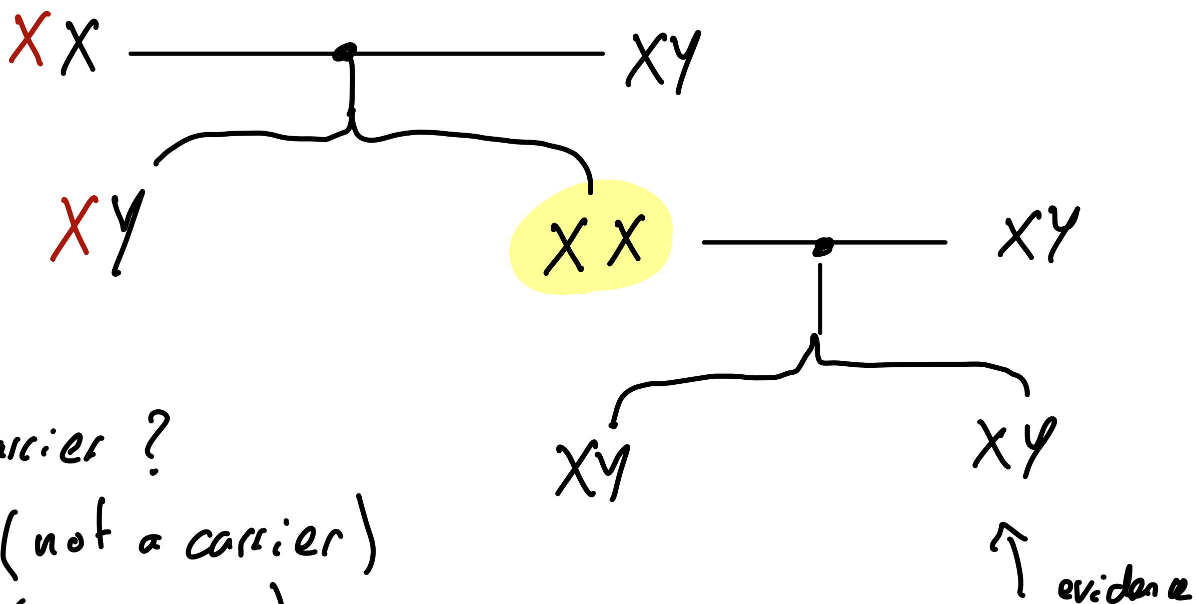
$\tilde{y}$: unknown but potentially observable data

$p(x)$: pdf of x

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

Hemophilia example:
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 : $p(\bar{y} \mid \theta = 1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

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

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

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

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' → 'random'

'radom' → 'radon'

'radom' → '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	$p(\text{'radom'} \mid \theta)$
random	0.00193
radon	0.000143
radom	0.975

posterior:

θ	$p(\theta)\, p(\text{'radom'} \mid \theta)$	$p(\text{'radom'} \mid \theta)$
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 ?

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

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

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

example 2

θ : proportion of female birth

y : number of female birth in n recorded births

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

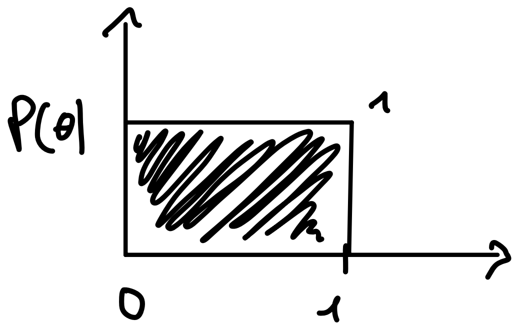
What is the posterior distribution?

Binomial likelihood.

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

uniform prior:

$$p(\theta) = 1 \quad \text{for } \theta \in [0,1]$$



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

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

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

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

So $$p(\tilde{y} \mid y) = \int p(\tilde{y}, \theta \mid y)\, d\theta$$

$$= \int p(\tilde{y} \mid \theta y)\, p(\theta \mid y)\, d\theta$$ chain rule of probability

$$= \int \underbrace{p(\tilde{y} \mid \theta)}_{\text{prediction}}\, \underbrace{p(\theta \mid y)}_{\text{posterior}}\, d\theta$$ posterior predictive distribution

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

$$p(\theta \mid y) = \text{Beta}(y+1, n-y+1)$$

$$= \frac{\Gamma(n+2)}{\Gamma(y+1)\,\Gamma(n-y+1)} \theta^{y} (1-\theta)^{n-y}$$

$$p(\tilde{y} \mid y) = \int \frac{\Gamma(n+2)}{\Gamma(y+1)\,\Gamma(n-y+1)} \binom{m}{\tilde{y}} \theta^{y+\tilde{y}} (1-\theta)^{n+m-(\tilde{y}+y)}$$

$$= \binom{m}{\tilde{y}} \frac{\Gamma(n+2)}{\Gamma(y+1)\,\Gamma(n-y+1)} \frac{\Gamma(y+\tilde{y}+1)\,\Gamma(m+n-\tilde{y}-y+1)}{\Gamma(m+n+2)}$$

$$\cdot \underbrace{\cancel{\int \frac{\Gamma(m+n+2)}{\Gamma(y+\tilde{y}+1)\,\Gamma(m+n-\tilde{y}+y+1)}\, \theta^{y+\tilde{y}} (1-\theta)^{m+n-(\tilde{y}+y)}\, d\theta}}_{\text{Beta distribution}} = 1$$

$$= \binom{m}{\tilde{y}} \frac{\Gamma(n+2)}{\Gamma(y+1)\,\Gamma(n-y+1)} \frac{\Gamma(y+\tilde{y}+1)\,\Gamma(m+n-\tilde{y}-y+1)}{\Gamma(m+n+2)}$$

$$= \binom{m}{\tilde{y}} \frac{(n+1)!\,(y+\tilde{y})!\,(m+n-\tilde{y}+y)!}{y!\,(n-y)!\,\underbrace{(m+n+1)!}_{(m+n+1)(m+n)!}}$$

$$= \binom{m}{\tilde{y}} \frac{n+1}{m+n+1} \frac{n!}{y!\,(n-y)!} \frac{(y+\tilde{y})!\,(m+n-(y+\tilde{y}))!}{(m+n)!}$$

$$= \binom{m}{\tilde{y}} \binom{n}{y} \begin{bmatrix} m+n \\ y+\tilde{y} \end{bmatrix}^{-1} \frac{n+1}{m+n+1}$$

n = # of data points

y = # of successes

m = # of predicted points

$\tilde{y}$ = # of predicted successes

prediction : next birth is female ?

$$m = \tilde{y} = 1$$

$$p(\tilde{y}=1, m=1 \mid y) = 1 \cdot \binom{n}{y} \begin{bmatrix} n+1 \\ y+1 \end{bmatrix}^{-1} \frac{n+1}{n+2}$$

$$= \frac{n!}{y!(n-y)!} \cdot \frac{(y+1)!(n-y)!}{(n+1)!} \frac{n+1}{n+2}$$

$$= \frac{y+1}{n+2}$$

$\frac{y}{n}$: data

$\frac{1}{2}$: mean of our posterior

$\text{Beta}(\alpha, \beta)$ have a mean of $\frac{\alpha}{\alpha+\beta}$

$\Rightarrow$ What is the mean of the posterior distribution ?

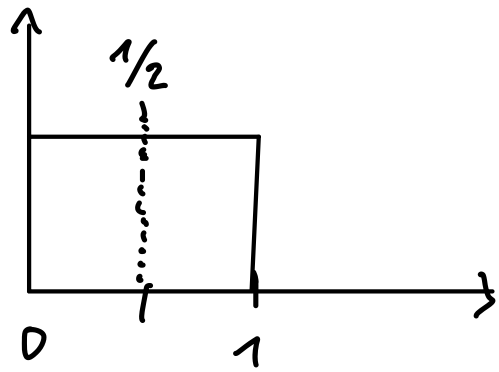
$$p(\theta \mid y) = \text{Beta}(y+1, n-y+1)$$

$$E[\theta \mid y] = \frac{y+1}{n-y+1+y+1} = \frac{y+1}{n+2}$$

If n is close to 0, then the average is more influenced by the average of the posterior distribution.

As n gets large,

$$E[\theta | y] \sim \frac{y}{n}$$



Given y, what is the expected value of θ, $E[\theta|y]$ in this problem?

$$E[\theta | y] = \frac{y+1}{n+2}$$

If data is large

$$E[\theta | y] = \frac{y}{n}$$

$$\mathrm{var}(\theta) = E[\mathrm{var}(\theta | y)] + \mathrm{var}(E[\theta | y])$$

we observe that

$$E[\mathrm{var}(\theta | y)] < \mathrm{var}(\theta)$$

Priors

Reflect initial information

Informative vs Noninformative priors

The prior should encompass all possible values

Example:

defact rate : assumed 5% with variance 0.25%

What could be a good informative prior?

defective

non defective

$\Rightarrow$ likelihood is Binomial

conjugate prior :

$$\theta \sim \text{Beta}(a, b) \qquad \mu = \frac{a}{a+b}$$

$$\sigma^2 = \frac{ab}{(a+b)^2(a+b+1)}$$

solution :

$$\frac{a}{a+b} = 0.05 \quad \Rightarrow \quad 19a = b$$

$$0.0025 = \frac{ab}{(a+b)^2(a+b+1)}$$

$$= \frac{19}{400(20a+1)}$$

$$\Rightarrow \quad a = \frac{18}{20} = 0.9$$

$$b = 17.1$$

So the prior is $\theta \sim B(0.9, 17.1)$

Binomial distribution

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

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

informative prior $\quad$ Beta(a,b)

$$p(\theta|y) \propto \underbrace{\theta^y (1-\theta)^{n-y}}_{\text{likelihood}} \; \underbrace{\theta^{a-1} (1-\theta)^{b-1}}_{\text{prior}}$$

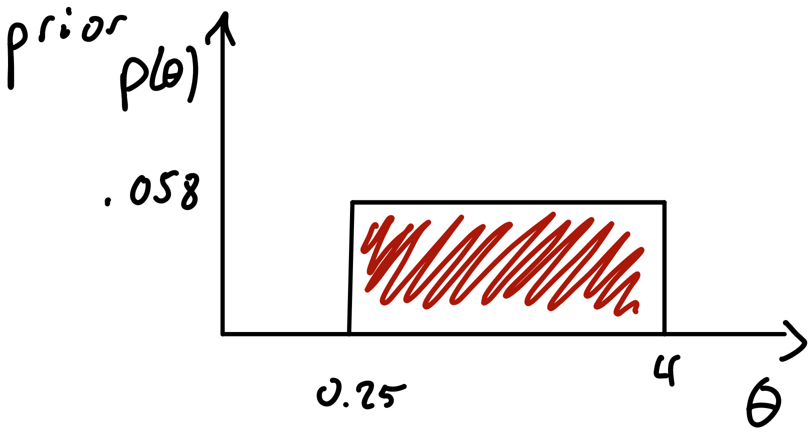
$$= \theta^{y+a-1} (1-\theta)^{n-y+b-1}$$

$$\theta | y \sim \text{Beta}(y+a, n-y+b)$$

$\Rightarrow$ conjugacy

Nonconjugate prior distributions

Normal likelihood and Uniform prior



Uniform

$$p(\theta) = \begin{cases} 0.308 & \text{if } \theta \in [0.25, 4] \\ 0 \end{cases}$$

posterior

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

$$\propto \begin{cases} \exp\left(-\frac{(y-\mu)^2}{2}\right) & \theta \in [0.25, 4] \\ 0 & \text{otherwise} \end{cases}$$

truncated Normal distribution (non conjugate)

Weakly informative priors

Quizz

(2) $p(\theta) \sim B(2,2)$

$p(y|\theta) \sim Bin(10, 1/2)$

$y = 3H$

$y = 3 \quad n = 10 \quad a = 2, \; b = 2$

$p(\theta|y) \sim Beta(a+y, n-y+b)$

$\sim Beta(5, 9)$

(1) $p(\theta) \sim B(2,2)$

$y < 3, \; n = 10$

$p(y|\theta) \sim Bin(n, K)$

$p(\theta|y) \sim \theta(1-\theta)^9(1+8\theta+36\theta^2)$

Other single-parameter models

example:

likelihood $p(y|\theta) \propto e^{-(y-\theta)^2/2\sigma^2}$

prior: $p(\theta) \propto e^{A\theta^2 + B\theta + C}$

aim: show that this prior is a normal distribution

$$A\theta^2 + B\theta + C = A\left(\theta + B/2A\right)^2 - B^2/4A^2 + C$$

$$p(\theta) \propto e^{A\theta^2 + B\theta + C} \propto e^{A(\theta + B/2A)^2}$$

let $A = -1/2\tau_0^2$ $\quad -B/2A = \mu_0$

$$\boxed{\begin{array}{c} p(\theta) \propto e^{-1/2\tau_0^2 (\theta - \mu_0)^2} \\ \theta \sim N(\mu_0, \tau_0^2) \end{array}}$$

prior

aim: show that the posterior is also normal

$$p(\theta \mid y) \propto p(\theta)\, p(y \mid \theta)$$
$$= e^{-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2} e^{-(y-\theta)^2/2\sigma^2}$$
$$= \exp\left(-\frac{1}{2}\left(\frac{(y-\theta)^2}{\sigma^2} + \frac{(\theta - \mu_0)^2}{\tau_0^2}\right)\right)$$
$$= \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma^2}(y^2 - 2\theta y + \theta^2) + \frac{1}{\tau_0^2}(\theta^2 - 2\theta\mu_0 + \mu_0^2)\right)\right)$$

$$\propto \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\tau_0^2}\right)\theta^2 - 2\left(\frac{y}{\sigma^2}+\frac{\mu_0}{\tau_0^2}\right)\theta\right)$$

$$= \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\tau_0^2}\right)\left(\theta^2 - 2\left(\frac{y/\sigma^2 + \mu_0/\tau_0^2}{1/\sigma^2 + 1/\tau_0^2}\right)\theta\right)\right)$$

$$\frac{1}{\tau_1^2} = \underbrace{\frac{1}{\sigma^2}}_{\text{likelihood}} + \underbrace{\frac{1}{\tau_0^2}}_{\text{prior}} \qquad \mu_1 = \frac{y/\sigma^2 + \mu_0/\tau_0^2}{1/\sigma^2 + 1/\tau_0^2}$$

$$\propto \exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right)$$

$$\theta \mid y \sim N(\mu_1, \tau_1^2) \qquad \text{posterior}$$

The inverse variance is called precision.

Now, for new data points $\tilde{y}$

$$p(\tilde{y} \mid y) = \int p(\tilde{y} \mid \theta)\, p(\theta \mid y)\, d\theta$$

$$\propto \int \exp\left(-\frac{(\tilde{y}-\theta)^2}{2\sigma^2}\right)\exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_1^2}\right) d\theta$$

$$\propto \exp\left(A\tilde{y} + B\tilde{y} + C\right)$$

$$\Rightarrow \tilde{y}|y \sim N(\cdot\,,\cdot)$$

$$E[\tilde{y}|y] = E[E[\tilde{y}|\theta, y]|y]$$
$$= E[E[\tilde{y}|\theta]|y]$$
$$= E[\theta|y]$$
$$= \mu_1$$

$$var(\tilde{y}|y) = E[var(\tilde{y}|\theta)|y] + var(E[\tilde{y}|\theta]|y)$$
$$= E[\sigma^2|y] + var(\theta|y)$$
$$= \sigma^2 + \tau_1^2$$

$$\tilde{y}|y \sim N(\mu_1, \sigma^2 + \tau_1^2)$$

How do we handle multiple data points?

Prior: $p(\theta) \propto \exp\left(-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2\right)$

likelihood:

$$p(y|\theta) \propto \prod_{i=1}^{n} \exp\left(-\frac{(y_i-\theta)^2}{2\sigma^2}\right)$$

$$= \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right)$$

posterior:

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

$$\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2 - \frac{1}{\tau_0^2}(\theta-\mu_0)^2\right)$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} y_i^2 - 2\theta\sum_{i=1}^{n} y_i + n\theta^2\right) - \frac{1}{2\tau_0^2}\left(\theta^2 - 2\theta\mu_0 + \mu_0^2\right)\right)$$

Let $\bar{y}$ be the sample mean, so $n\bar{y} = \sum_{i=1}^{n} y_i$

$$= \exp\left(\theta^2\left(-\frac{n}{2\sigma^2} - \frac{1}{2\tau_0^2}\right) - 2\theta\left(\frac{n\bar{y}}{\sigma^2} + \frac{1}{\tau_0^2}\right)\right)$$

$$\tau_n^2 = \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$

$$= \exp\left(-\frac{1}{2\tau_n^2}\left(\theta^2 - 2\tau_n^2\left(\frac{n\bar{y}}{\sigma^2} + \frac{1}{\tau_0^2}\right)\theta\right)\right)$$

$$\mu_n = \tau_n^2\left(\frac{n\bar{y}}{\sigma^2} + \frac{1}{\tau_0^2}\right)$$

$$\propto \exp\left(-\frac{1}{2\tau_n^2}(\theta - \mu_n)^2\right)$$

$$\Rightarrow \quad \theta \mid y \sim N(\mu_n, \tau_n^2)$$

<u>example</u>: screws of 5 mm with variations

mean ? variance: $\sigma^2 = 0.5$

Data: $y = \{5.1, 4.9, 5.2, 5.3, 5\}$

$n = 5$ $\quad \tau_0^2 = 1$

$\bar{y} = 5.1$ $\quad \mu_0 = 5$

$\sigma^2 = 0.5$

$$\tau_n^2 = \left(\frac{1}{1} + \frac{5}{0.5}\right)^{-1} = \frac{1}{11} \approx 0.11$$

$$\mu_n = \frac{1}{11}\left(\frac{25.5}{0.5} + \frac{5}{1}\right) = \frac{56}{11} \approx 5.1$$

$$\theta | y \sim N(5.1, 0.1)$$

easy to compute with conjugacy

Normal distribution with unknown variance

But mean is known

Review: Scaled inverse χ^2 distribution

$$\theta \sim \mathrm{Inv}\chi^2(\underbrace{\nu}_{\text{deg. of freedom}}, \underbrace{s^2}_{\text{scale}})$$

$$p(\theta) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} s^\nu \theta^{-(\nu/2+1)} e^{\frac{-\nu s^2}{2\theta}}$$

$$\propto \theta^{-(\nu/2+1)} e^{-\frac{\nu s^2}{2\theta}}$$

likelihood:

$$p(y|\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_i-\theta)^2/2\sigma^2}$$

$$\propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \theta)^2\right.$$

Define : $v := \frac{1}{n} \sum_{i=1}^{n} (y_i - \theta)^2$ sample variance

$$p(y \mid \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left(\frac{nv}{2\sigma^2}\right)$$

prior : $$p(\sigma^2) \propto (\sigma^2)^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left(\frac{-\nu_0 \sigma_0^2}{2} \cdot \frac{1}{\sigma^2}\right)$$

posterior :

$$\begin{aligned} p(\sigma^2 \mid y) &\propto p(\sigma^2)\, p(y \mid \sigma^2) \\ &\propto (\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{\nu_0 \sigma_0^2}{2} \cdot \frac{1}{\sigma^2}\right) \\ &\quad \cdot (\sigma^2)^{-n/2} \exp\left(-\frac{nv}{2\sigma^2}\right) \\ &= (\sigma^2)^{-\frac{1}{2}(\nu_0+n+2)} \exp\left(-\frac{1}{2\sigma^2}(\nu_0 \sigma_0^2 + nv)\right) \end{aligned}$$

$$\sigma^2 \sim \text{Inv}\chi^2\left(\nu_0+n, \frac{\nu_0 \sigma_0^2 + nv}{\nu_0}\right)$$

Poisson distributions

θ : average number of events

y : actual observations

review: Gamma(a, b) a: shape
b: inverse scale

mean: a/b

variance: a/b^2

$$p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}$$
$$\propto \theta^{a-1} e^{-b\theta}$$

poisson likelihood
$$p(y|\theta) = \prod_{i=1}^{n} \frac{1}{y_i!} \theta^{y_i} e^{-\theta} \propto \theta^{n\bar{y}} e^{-n\theta}$$

prior:
$$p(\theta) \propto e^{-b_0\theta} \theta^{a_0 - 1}$$

posterior: $p(\theta|y) \propto p(\theta)\, p(y|\theta)$
$$\propto \theta^{n\bar{y} + a_0 - 1} e^{-n\theta - b_0\theta}$$

$$\theta|y \sim \text{Gamma}(a_0 + n\bar{y}, b_0 + n)$$

example: $3/200{,}000$ died of asthma (1 year)

world: $0.6/100{,}000$ prior

data

prior: mean of 0.6 $\Rightarrow \frac{a}{b} = 0.6$

suppose: $a = 3$, $b = 5$

posterior: $\text{Gamma}(a_0 + y, b_0 + x) = \text{Gamma}(6, 7)$

mean: $6/7 \approx 0.86$ variance: $6/7^2$

Exponential distributions

$$p(y \mid \theta) = \theta^n e^{-n\bar{y}\theta}$$

waiting time or time between events θ.

Conjugate prior: Gamma.

multiparameter models

Nuisance parameters

example

$$y \sim N(\mu, \sigma^2)$$

Jeffrey's prior (noninformative) for μ and σ^2

Jeffrey's invariance principle

$$p(\theta) \propto [J(\theta)]^{1/2} \quad \text{where}$$

$$J(\theta) = -E\left[\frac{\partial^2 \log p(y|\theta)}{\partial\theta^2} \,\Big|\, \theta\right]$$

$$\theta = (\mu, \sigma^2)$$

$$L := p(y|\theta) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)$$

$$\log(L) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y-\mu)^2$$

$$= -\frac{1}{2}\log(2\pi) - \log(\sigma) - \frac{1}{2\sigma^2}(y-\mu)^2$$

$$\frac{\partial \log(L)}{\partial\mu} = \frac{1}{\sigma^2}(y-\mu)$$

$$\frac{\partial \log(L)}{\partial\sigma} = -\frac{1}{\sigma} + \frac{1}{\sigma^3}(y-\mu)^2$$

$$\frac{\partial^2 \log(L)}{\partial\mu^2} = -\frac{1}{\sigma^2}$$

$$\frac{\partial \log(L)}{\partial\mu\, d\sigma} = -\frac{2(y-\mu)}{\sigma^3}$$

$$\frac{\partial \log(L)}{(\partial\sigma)^2} = \frac{1}{\sigma^2} - \frac{3}{\sigma^4}(y-\mu)^2$$

$$J(\theta) = -\in\left[\det\begin{vmatrix} -1/\sigma^2 & -\frac{2(y-\mu)}{\sigma^3} \\ \frac{-2(y-\mu)}{\sigma^3} & \frac{1}{\sigma^2}-\frac{3}{\sigma^4}(y-\mu)^2 \end{vmatrix}\right]^{1/2}$$

$$= \left(-\in\left[-1/\sigma^4 + \frac{3}{\sigma^6}(y-\mu)^2 - \frac{4(y-\mu)^2}{\sigma^6}\right]\right)^{1/2}$$

remember $\in[(y-\mu)^2] = \sigma^2$ so,

$$= \left(\frac{1}{\sigma^4} + \frac{1}{\sigma^4}\right)^{1/2}$$

$$= \left(\frac{2}{\sigma^4}\right)$$

$$= \frac{\sqrt{2}}{\sigma^2} \propto \frac{1}{\sigma^2}$$

$p(\theta) \propto \frac{1}{\sigma^2}$ Noninformative prior

posterior:

$$p(\theta \mid y) \propto p(y \mid \theta)\, p(\theta) \qquad \theta = (\mu, \sigma^2)$$

$$\propto (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right) \cdot \frac{1}{\sigma^2}$$

$$\propto (\sigma^2)^{-n/2-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right)$$

$$\sum_{i=1}^{n} (y_i - \mu)^2 = \sum_{i=1}^{n} \left((y_i - \bar{y}) + (\bar{y} - \mu)\right)^2$$

$$= \sum_{i=1}^{n} \left[(y_i - \bar{y})^2 + 2(y_i - \bar{y})(\bar{y} - \mu) + (\bar{y} - \mu)^2\right]$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 + 2(\bar{y} - \mu) \underbrace{\cancel{\sum_{i=1}^{n} (y_i - \bar{y})}}_{=0}$$

$$+ n(\bar{y} - \mu)^2$$

$$\Rightarrow p(\mu, \sigma^2 \mid y) \propto (\sigma^2)^{-n/2-1} \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right)\right)$$