

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 occurred.

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

example : We have a standard 6 sided die

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

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

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

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

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.

example : 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

$$\begin{aligned} \text{marginal} : P(B) &= P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c) \\ &= 0.99 \cdot 0.01 + 0.05 \cdot 0.99 \\ &\approx 0.0594 \end{aligned}$$

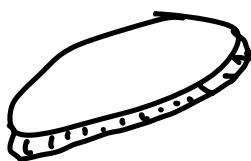
$$\text{posterior } P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{0.99 \cdot 0.01}{0.0594} \approx \underline{\underline{0.16}}_2$$

a bit surprising.

Bayesian inference

1. Start with a prior distribution
2. Collect data
3. Compute the likelihood
4. Compute the marginal probability
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

$$P(H_0 | TTT) = \frac{P(TTT | H_0) P(H_0)}{P(TTT)}$$

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

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

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

Fundamentals of Bayesian Inference

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

Quizz exercise:

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

Likelihood :

$$P(\text{5H} | H_0) = \left(\frac{1}{2}\right)^5 = 1/32 = 0.03125$$

$$P(\text{5H} | H_1) = \int_0^1 p^5 dp = \left[\frac{p^6}{6}\right]_0^1 = 1/6$$

marginal : $P(\text{5H}) = P(\text{5H} | H_0) \cdot P(H_0) + P(\text{5H} | H_1) \cdot P(H_1)$

$$= 0.03125 \cdot 0.66 + \left(\frac{1}{6}\right) \cdot 0.34$$

$$\approx \underline{0.078}$$

Posterior:

$$P(H_0 | \text{5H}) = \frac{P(\text{5H} | H_0) \cdot P(H_0)}{P(\text{5H})}$$

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

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

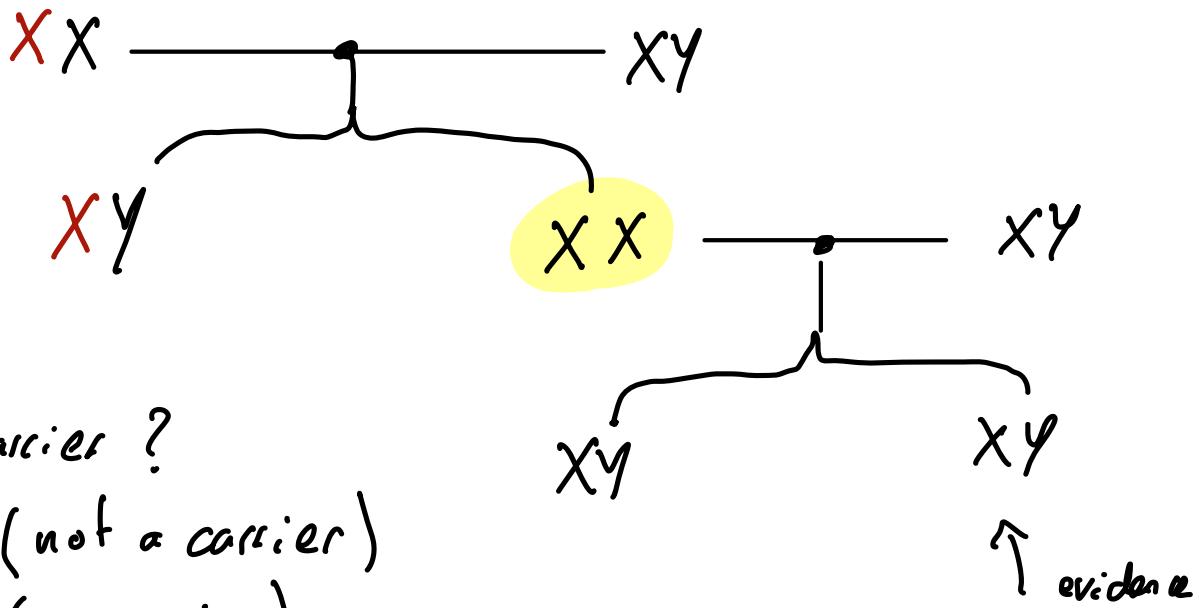
$p(x)$: pdf of x

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

Hemophilia example:

from the book

X-linked trait



Yellow is a carrier?

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

$H_1 \theta = 1$ (a carrier)

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

$$\text{prior : } (50/50) = 0.5$$

$$\text{likelihood : } p(\bar{y} | \theta = 1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

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

$$\begin{aligned} \text{marginal (for scaling)} : p(\bar{y}) &= p(\bar{y} | \theta = 1) \cdot p(\theta = 1) + \\ &\quad p(\bar{y} | \theta = 0) \cdot p(\theta = 0) \\ &= \frac{1}{4} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \underline{\underline{\frac{3}{8}}} \end{aligned}$$

posterior :

$$p(\theta = 1 | \bar{y}) = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{5} = \underline{\underline{0.2}}$$

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

$$\bar{y} = (0)$$
$$p(\theta=1 | y) = \frac{p(y|\theta=1) \cdot p(\theta=1)}{p(y)}$$
$$= \frac{\left(\frac{1}{2}\right) \cdot \left(\frac{1}{5}\right)}{\frac{1}{2} \cdot \frac{1}{5} + 1 \cdot \frac{4}{5}} = \frac{1}{9} \approx 0,11$$

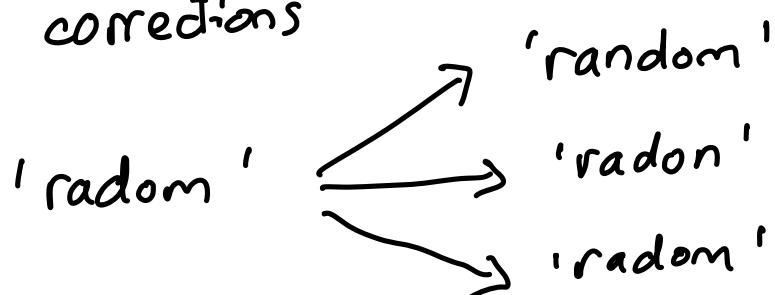
Exchangeability

order of observations doesn't matter

Subjectivity \circlearrowleft objectivity
=> prior

Example from the book

Spelling corrections



data : $y = \text{radom}$

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

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

posterior :

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

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

Bayesian Computation

length in millimeters

off by -1mm or 1mm

$$\theta = 1 : y \sim N(1, 1)$$

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

prior

0.5

0.5

Likelihood:

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

marginal:

$$p(y=0.5) = p(y=0.5 | \theta=1) p(\theta=1) + p(y=0.5 | \theta=-1) p(\theta=-1) \\ \approx 0.09605$$

posterior:

$$p(\theta=1 | y=0.5) = \frac{0.07022}{0.09605} \approx 0.73$$

Stan package in R

PyStan in python

example in R : estimation of a distribution

Faulty caliper problem
from the book

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
 \overline{TTT} ?

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

$$p(y=1, n=3 | \theta = 0.75) \approx 0.14$$

{TTH or THT or HTT}

example 2

θ : proportion of female birth

y : number of female birth in n recorded births

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

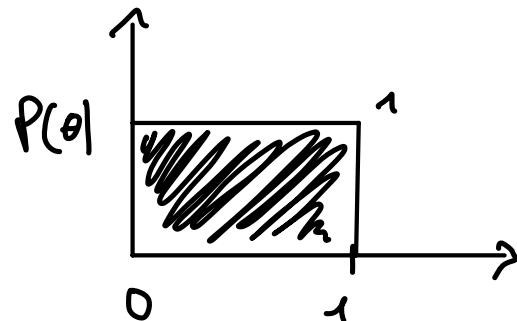
What is the posterior distribution?

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:

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

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

once normalized

$$p(\theta|y) = \frac{\binom{n}{y}}{\alpha + \beta} \quad \alpha = y, \beta = n-y$$

Beta distribution

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

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

so $\theta \sim \text{Beta}$ and no longer Uniform

$$P(\tilde{y} | y)$$

$$\text{So } P(\tilde{y} | y) = \int P(\tilde{y}, \theta | y) d\theta$$

$$= \int P(\tilde{y} | \theta y) P(\theta | y) d\theta$$

$$= \left\{ \underbrace{P(\tilde{y} | \theta)}_{\text{prediction}} \underbrace{P(\theta | y)}_{\text{posterior}} \right\} d\theta$$

chain
rule of
probability

posterior
predictive
distribution

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

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

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

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

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

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

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

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

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

$$= \binom{m}{\tilde{y}} \binom{n}{y} \left[\frac{m+n}{y+\tilde{y}} \right]^{-1} \frac{n+1}{m+n+1}$$

n = # of data points

y = # of successes

m = # of predicted points

\tilde{y} = # of predicted successes

prediction : next birth is female ?

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

$$\begin{aligned} p(\tilde{y}=1, m=1 | y) &= 1 \cdot \binom{n}{y} \left[\frac{n+1}{y+1} \right]^{-1} \frac{n+1}{n+2} \\ &= \frac{n!}{y!(n-y)!} \cdot \frac{(y+1)!(n-y)!}{(n+1)!} \frac{n+1}{n+2} \\ &\approx \frac{y+1}{n+2} \end{aligned}$$

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

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

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

\Rightarrow What is the mean of the posterior distribution?

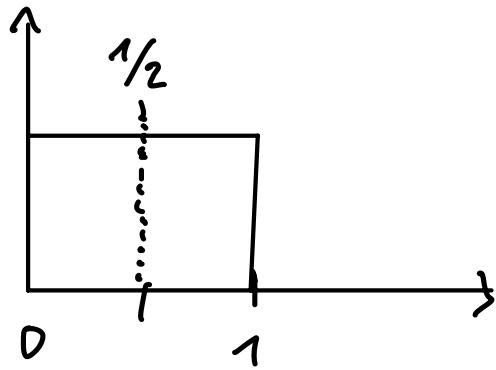
$$p(\theta | y) = \text{Beta}(y+1, n-y+1)$$

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

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

As n gets large,

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



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

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

If data is large

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

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

we observe that

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

Priors

Reflect initial information

Informative \textcircled{vs} Noninformative priors

The prior should encompass all possible values

Example:

defect rate : assumed 5% with variance 0.25%.

What could be a good informative prior?

→ defective \Rightarrow likelihood is Binomial
→ non defective

conjugate prior :

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

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

solution :

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

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

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

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

$$b = 17.1$$

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 $Beta(a, b)$

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

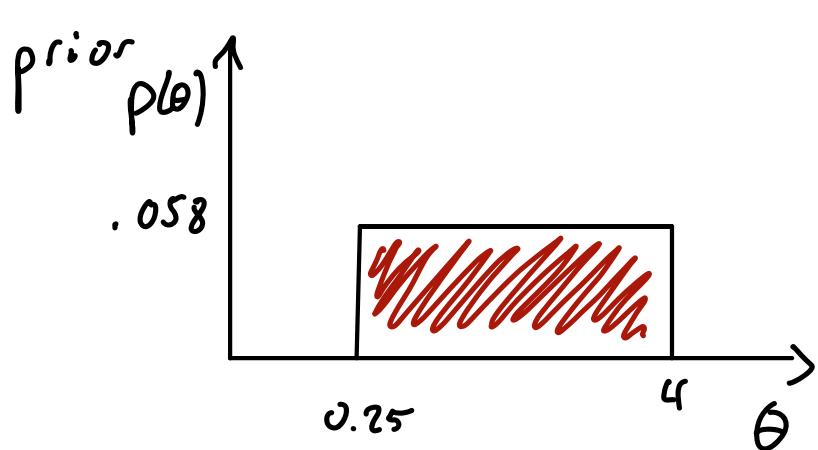
$$= \theta^{y+a-1} (1-\theta)^{n-y+b-1}$$

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

\Rightarrow conjugacy

Nonconjugate prior distributions

Normal likelihood and Uniform prior



Uniform

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

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

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

$$p(y|\theta) \sim \text{Bin}(10, \gamma_2)$$

$$y = 3H$$

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

$$\begin{aligned} p(\theta|y) &\sim \text{Beta}(a+y, n-y+b) \\ &\sim \text{Beta}(5, 9) \end{aligned}$$

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

$$y < 3, n = 10$$

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

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

Other single-parameter models

example :

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

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

aim: show that this prior is a normal distribution

$$A\theta^2 + B\theta + C = A(\theta + \frac{B}{2A})^2 - \frac{B^2}{4A^2} + C$$

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

$$\text{let } A = -1/2\tau_0^2 \quad -\frac{B}{2A} = \mu_0$$

$$p(\theta) \propto e^{-\frac{1}{2}\tau_0^2(\theta - \mu_0)^2}$$

$$\theta \sim N(\mu_0, \tau_0^2)$$

prior

aim: show that the posterior is also normal

$$\begin{aligned}
 p(\theta | y) &\propto p(\theta) p(y | \theta) \\
 &= e^{-\frac{1}{2}\tau_0^2(\theta - \mu_0)^2} e^{-(y - \theta)^2 / 2\sigma^2} \\
 &= \exp \left(-\frac{1}{2} \left(\frac{(y - \theta)^2}{\sigma^2} + \frac{(\theta - \mu_0)^2}{\tau_0^2} \right) \right) \\
 &= \exp \left(-\frac{1}{2} \left(\frac{1}{\sigma^2} (y^2 - 2y\theta + \theta^2) + \right. \right. \\
 &\quad \left. \left. \frac{1}{\tau_0^2} (\theta^2 - 2\theta\mu_0 + \mu_0^2) \right) \right)
 \end{aligned}$$

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

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

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

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

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

The inverse variance is called precision.

Now, for new data points \tilde{y}

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

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

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

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

$$\begin{aligned} E[\tilde{y}|y] &= E[E[\tilde{y}|\theta, y]|y] \\ &= E[E[\tilde{y}|\theta]|y] \\ &= E[\theta|y] \\ &= \mu_1 \end{aligned}$$

$$\begin{aligned} \text{var}(\tilde{y}|y) &= E[\text{var}(\tilde{y}|\theta)|y] + \text{var}(E[\tilde{y}|\theta]|y) \\ &= E[\sigma^2|y] + \text{var}(\theta|y) \\ &= \bar{v}^2 + \bar{\tau}_\theta^2 \end{aligned}$$

$$\tilde{y}|y \sim N(\mu_1, \bar{v}^2 + \bar{\tau}_\theta^2)$$

How do we handle multiple data points?

Prior: $p(\theta) \propto \exp\left(-\frac{1}{2\bar{\tau}_0^2} (\theta - \mu_0)^2\right)$

1: 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 :

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

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

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

$$-\frac{1}{2\tau_0^2} (\theta^2 - 2\theta\mu_0 + \mu_0^2)$$

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

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

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

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

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

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

$$\Rightarrow \theta | y \sim N(\mu_n, \gamma_n^2)$$

example : 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

$$\theta \sim \text{Inv} \chi^2(v, s^2)$$

↑ ↑
deg. of freedom scale

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

Likelihood :