STA623 - Bayesian Data Analysis - Practical 1 & 2 (Solutions)

Marc Henrion

2023-08-28

Practicals 1 & 2

Notation

- X, Y, Z - random variables

- x, y, z - measured / observed values

- $\bar{X}, \bar{Y}, \bar{Z}$ - sample mean estimators for X, Y, Z

- $\bar{x}, \bar{y}, \bar{z}$ - sample mean estimates of X, Y, Z

- $\hat{T}, \hat{t}$ - given a statistic T, estimator and estimate of T

- $P(A)$ - probability of an event A occuring

- $f_X(.), f_Y(.), f_Z(.)$ - probability mass / density functions of X, Y, Z; sometimes $p_X(.)$ etc. rather than $f_X(.)$

- p(.) - used as a shorthand notation for pmfs / pdfs if the use of this is unambiguous (i.e. it is clear which is the random variable)

- $X \sim F$ - X distributed according to distribution function F

- $E[X], E[Y], E[Z], E[T]$ - the expectation of X, Y, Z, T respectively

Exercise 1

A company has developed a new diagnostic test, for a disease D with prevalence 10%, which can give a positive (T) or negative ($\bar{T}$) result. There is also an already existing test, which can also give a positive (T_{old}) or negative ($\bar{T}_{old}$) result.

The company is very proud that the new test has much better sensitivity ($P(T|D) = 0.99$) than the old one ($P(T_{old}|D) = 0.8$). However this increase in sensitivity comes at the cost of specificity: $P(\bar{T}|\bar{D}) = 0.9$ compared to the existing test $P(\bar{T}_{old}|\bar{D}) = 0.95$.

What is the probability of a patient having the disease if the test result for the new test is positive?

For the existing test?

We can assume that, given the disease state (D or $\bar{D}$), the two tests are independent as they use very different technologies and are not run on the same biological sample (each test takes a new sample from the patient).

What is the probability for a patient having the disease if both the new and the existing test give a positive result?

Hint: define a virtual test which is positive if both the new and existing tests return a positive results and gives a negative result otherwise. Compute the sensitivity and specificity for this compound test, then proceed as before.

Exercise 1 (Solution)

For the new test:

$P(D) = 0.1$, $P(T|D) = 0.99$, $P(\bar{T}|\bar{D}) = 0.9$ and so, by Bayes' rule:

$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})P(\bar{D})} \tag{1}$$

$$= \frac{0.99 \cdot 0.1}{0.99 \cdot 0.1 + (1 - 0.9) \cdot (1 - 0.1)} \tag{2}$$

$$= 0.5238 \tag{3}$$

And for the existing test, $P(D) = 0.1$, $P(T_{old}|D) = 0.8$, $P(\bar{T}_{old}|\bar{D}) = 0.95$:

$$P(D|T_{old}) = \frac{P(T_{old}|D)P(D)}{P(T_{old}|D)P(D) + P(T_{old}|\bar{D})P(\bar{D})} \tag{4}$$

$$= \frac{0.8 \cdot 0.1}{0.8 \cdot 0.1 + (1 - 0.95) \cdot (1 - 0.1)} \tag{5}$$

$$= 0.64 \tag{6}$$

Let's define a new test which can be positive (event U) or negative ($\bar{U}$) where $U = T \cap T_{old}$. Then:

$$P(U|D) = P(T \cap T_{old}|D) \tag{7}$$

$$= P(T|D) \cdot P(T_{old}|D) \tag{8}$$

$$= 0.99 \cdot 0.8 \tag{9}$$

$$= 0.792 \tag{10}$$

Likewise:

$$P(U|\bar{D}) = P(T \cap T_{old}|\bar{D}) \tag{11}$$

$$= P(T|\bar{D}) \cdot P(T_{old}|\bar{D}) \tag{12}$$

$$= (1 - P(\bar{T}|\bar{D})) \cdot (1 - P(\bar{T}_old|\bar{D})) \tag{13}$$

$$= (1 - 0.9) \cdot (1 - 0.95) \tag{14}$$

$$= 0.005 \tag{15}$$

Equipped with this:

$$P(D|T \cap T_{old}) = P(D|U) = \frac{P(U|D)P(D)}{P(U|D)P(D) + P(U|\bar{D})P(\bar{D})} \tag{16}$$

$$= \frac{0.792 \cdot 0.1}{0.792 \cdot 0.1 + 0.005 \cdot (1 - 0.1)} \tag{17}$$

$$= 0.9462 \tag{18}$$

Exercise 2

Suppose $\Pi \sim \text{Beta}(a, b)$. Compute $E[\Pi]$.

We have seen that if

$$\begin{cases} \Pi \sim \text{Beta}(a, b) \\ Y|\pi \sim \text{Bin}(n, \pi) \end{cases}$$

then $\Pi|Y = k \sim \text{Beta}(a + k, b + n - k)$.

Show that

$$E[\Pi|Y = k] = \frac{a + b}{a + b + n} E[\Pi] + \frac{n}{a + b + n} \bar{y}$$

Where $\bar{y} = k/n = \sum_i y_i/n$, the average of n Bernoulli(π) trials.

What does this tell you if $n \to \infty$?

Exercise 2 (solution)

We have seen that if $X \sim \text{Beta}(\alpha, \beta)$, then $E[X] = \frac{\alpha}{\alpha + \beta}$.

Here we have:

- prior: $\Pi \sim \text{Beta}(a, b) \Rightarrow E[\Pi] = \frac{a}{a+b}$

- posterior: $\Pi|k \sim \text{Beta}(a + k, b + n - k) \Rightarrow E[\Pi|k] = \frac{a+k}{a+b+n}$

We see that

$$E[\Pi|k] = \frac{a + k}{a + b + n} \tag{19}$$

$$= \frac{a}{a + b + n} + \frac{k}{a + b + n} \tag{20}$$

$$= \frac{a}{a + b + n} \cdot \frac{a + b}{a + b} + \frac{k}{a + b + n} \cdot \frac{n}{n} \tag{21}$$

$$= \frac{a + b}{a + b + n} \cdot \frac{a}{a + b} + \frac{n}{a + b + n} \cdot \frac{k}{n} \tag{22}$$

$$= \frac{a}{a + b + n} E[\Pi] + \frac{n}{a + b + n} \cdot \bar{y} \tag{23}$$

$$\tag{24}$$

4

As $n \to \infty$, $k \to \infty$ (since $E[Y|\pi] = \pi n$).

At the same time:

- $\frac{a+b}{a+b+n} \to 0$

- $\frac{n}{a+b+n} \to 1$

We see that

$$\lim_{n \to \infty} E[\Pi|k] = 0 \cdot E[\Pi] + 1 \cdot \bar{y} = \bar{y}$$

Asympotically, $E[\Pi|k]$ converges to the data average; i.e. the likelihood dominates the prior as $n \to \infty$.

Exercise 3

Suppose $Y_1, \dots, Y_n \sim_{\text{iid}} \text{Pois}(\lambda)$ and that you observe data $\{y_i\}_{i=1}^n$.

Derive the Jeffreys prior for this sampling model.

Is this a proper prior?

What posterior distribution do you get? Is this a valid distribution?

Exercise 3 (Solution)

We have seen that $\sum_i Y_i$ is a sufficient statistic for λ in a Poisson sampling model.

Let's write $k = \sum_i y_i$, and recall:

$$p(y_1, \dots, y_n | \lambda) = \prod_i \frac{e^{-\lambda} \lambda^{y_i}}{y_i!} = \frac{1}{\prod_i y_i!} e^{-n\lambda} \lambda^k$$

The Jeffreys prior is given by

$$p_J(\lambda) \propto \sqrt{-E\left[\frac{\delta^2}{\delta \lambda^2} \log p(y_1, \dots, y_n | \lambda) \right]}$$

$$\frac{\delta}{\delta \lambda} \log p(y_1, \dots, y_n | \lambda) = \frac{\delta}{\delta \lambda} \log \frac{e^{-n\lambda} \lambda^k}{\prod_i y_i!} \tag{25}$$

$$= \frac{\delta}{\delta \lambda} \left(-n\lambda + k \log \lambda - \sum_i \log y_i! \right) \tag{26}$$

$$= -n + \frac{k}{\lambda} \tag{27}$$

$$= \frac{k - n\lambda}{\lambda} \tag{28}$$

And so

$$\frac{\delta^2}{\delta \lambda^2} \log p(y_1, \dots, y_n | \lambda) = \frac{-k}{\lambda^2}$$

From this we get that

$$p_J(\lambda) \propto \sqrt{E\left[\frac{k}{\lambda^2}\Big|\lambda\right]} = \sqrt{\frac{E[k|\lambda]}{\lambda^2}} \tag{29}$$

$$\propto \sqrt{\frac{n\lambda}{\lambda^2}} = \sqrt{\frac{n}{\lambda}} \tag{30}$$

$$\propto \lambda^{-1/2} \tag{31}$$

Note that $\int_0^\infty \lambda^{-1/2} d\lambda = \infty$.

So we will not be able to find a normalising constant so that $\int_0^\infty p_J(\lambda) d\lambda = 1$.

The Jeffreys prior for the Poisson sampling model is *improper*.

With the Jeffreys prior, the posterior distribution is given by:

$$p(\lambda|y_1,\dots,y_n) \propto e^{-n\lambda}\lambda^k \cdot \lambda^{-1/2} \tag{32}$$

$$\propto e^{-n\lambda}\lambda^{k+1/2-1} \tag{33}$$

So this yields a $\Gamma\left(\sum_i y_i + \frac{1}{2}, n\right)$ distribution, which is a valid distribution.

Exercise 4

Derive the posterior predictive distribution $p(\tilde{y}|y_1, \dots, y_n)$ for

$$
\begin{cases}
\Lambda \sim \Gamma(a, b) & \text{(prior)} \\
Y_i|\lambda \sim_{\text{iid}} \text{Pois}(\lambda) & i = 1, \dots, n \quad \text{(sampling model)}
\end{cases}
$$

Exercise 4 (Solution)

From lectures, we have seen

$$
\Lambda|y_1, \dots, y_n \sim \Gamma\left(a + \sum_i y_i, b + n\right)
$$

Therefore, the posterior predictive distribution is given by

$$
p(\tilde{y}|y_1, \dots, y_n) = \int_0^\infty p(\tilde{y}, \lambda|y_1, \dots, y_n)d\lambda \tag{34}
$$

$$
= \int_o^\infty p(\tilde{y}|\lambda, y_1, \dots, y_n)\, p(\lambda|y_1, \dots, y_n)d\lambda \tag{35}
$$

$$
= \int_0^\infty \frac{e^{-\lambda}\lambda^{\tilde{y}}}{\tilde{y}!} \frac{(b+n)^{a+\sum_i y_i}}{\Gamma(a + \sum_i y_i)}\lambda^{a + \sum_i y_i - 1}e^{-(b+n)\lambda}d\lambda \tag{36}
$$

$$
= \frac{1}{\tilde{y}!}\frac{(b+n)^{a+\sum_i y_i}}{\Gamma(a + \sum_i y_i)}\int_0^\infty \lambda^{a + \sum_i y_i + \tilde{y} - 1}e^{-(b+n+1)\lambda}d\lambda \tag{37}
$$

We can recognise that the term inside the integration looks like a $\Gamma(a + \sum_i y_i + \tilde{y}, b + n + 1)$ density. Knowing that this density needs to integrate to 1, we know that the integral needs to integrate to the inverse of the constant term for that pdf.

We therefore get:

$$
p(\tilde{y}|y_1, \dots, y_n) = \frac{1}{\tilde{y}!}\frac{(b+n)^{a+\sum_i y_i}}{\Gamma(a + \sum_i y_i)}\frac{\Gamma(a + \sum_i y_i + \tilde{y})}{(b+n+1)^{a+\sum_i y_i + \tilde{y}}} \tag{38}
$$

Rewriting $\tilde{y}! = \Gamma(\tilde{y} + 1)$ and simplifying the above, we get:

$$p(\tilde{y}|y_1,\ldots,y_n) = \tag{39}$$

$$\frac{\Gamma(a+\sum_i y_i+\tilde{y})}{\Gamma(\tilde{y}+1)\Gamma(a+\sum_i y_i)}\left(\frac{b+n}{b+n+1}\right)^{a+\sum_i y_i}\left(\frac{1}{b+n+1}\right)^{\tilde{y}} \tag{40}$$

for $\tilde{y} \in \{0, 1, 2, \ldots\}$.

This is a $\text{NegBin}(a+\sum_i y_i, (b+n)/(b+n+1))$ distribution.

Note that

- $E[\tilde{Y}|y_1,\ldots,y_n] = \frac{a+\sum_i y_i}{b+n} = E[\Lambda|y_1,\ldots,y_n]$

- $Var(\tilde{Y}|y_1,\ldots,y_n) = \frac{a+\sum_i y_i}{b+n} \cdot \frac{b+n+1}{b+n} = E[\Lambda|y_1,\ldots,y_n] \cdot \frac{b+n+1}{b+n}$

where $E[\Lambda|y_1,\ldots,y_n]$ is both the sampling mean and the sampling variability (since this is a Poisson sampling model).

Note that $(b+n+1)/(b+n) > 1$.

Exercise 5

In the 1990s, the General Social Survey gathered data on the eductional attainment and number of cildren of 155 women who were 40 years old at the time of their participation in the survey.

We will compare women with college degrees to those without.

Let $Y_{1,1}, \ldots, Y_{n_1,1}$ be the number of children for the n_1 women without a college degree and $Y_{1,2}, \ldots, Y_{n_2,2}$ for those with degrees.

Assume the following sampling models and priors:

- $Y_{1,1}, \ldots, Y_{n_1,1} \sim_{\text{iid}} \text{Poisson}(\theta_1)$
- $Y_{1,2}, \ldots, Y_{n_2,2} \sim_{\text{iid}} \text{Poisson}(\theta_2)$
- $\theta_1, \theta_2 \sim_{\text{iid}} \Gamma(2, 1)$

The data from the survey can be summarised by:

- $n_1 = 111, \sum_{i=1}^{n_1} y_{i,1} = 217, \bar{y}_1 = 1.95$
- $n_2 = 44, \sum_{i=1}^{n_2} y_{i,2} = 66, \bar{y}_1 = 1.50$

Derive the posterior distributions for $\theta_1 | \sum_i y_{i,1}, \theta_2 | \sum_i y_{i,2}$.

Compute $P(\theta_1 > \theta_2 | \sum y_{i,1} = 217, \sum y_{i,2} = 66)$.

Derive and compare the posterior predictive distributions for each group.

Exercise 5 (Solution)

See Hoff P. D. (2009), pp.48-50, for a fuller discussion of this example.

From lectures, we know that the posterior distributions will be

$$
\begin{cases}
\theta_1 | n_1 = 111, \sum_i y_{i,1} = 217 & \sim \Gamma(2 + 217, 1 + 111) = \Gamma(219, 112) \\
\theta_2 | n_2 = 44, \sum_i y_{i,1} = 66 & \sim \Gamma(2 + 66, 1 + 44) = \Gamma(68, 45)
\end{cases}
$$

$$P(\theta_1 > \theta_2 | \ldots) = \int_0^\infty \int_{\theta_2}^\infty p(\theta_1, \theta_2 | \ldots) d\theta_1 d\theta_2 \tag{41}$$

$$= \int_0^\infty \gamma(\theta_2; 68, 45) \int_{\theta_2}^\infty \gamma(\theta_1; 219, 112) d\theta_1 d\theta_2 \tag{42}$$

$$= 0.9726 \tag{43}$$

where

- $\gamma(.; a, b)$ is shorthand for the pdf of a gamma distribution with parameters a, b.
- ... is shorthand for $n_1 = 111, \sum_i y_{i,1} = 217, n_2 = 44, \sum_i y_{i,2} = 66$

There is substantial evidence that $\theta_1 > \theta_2$.

You can compute the last integral in `R`:

```
fun1<-function(x){dgamma(x,219,112)}
fun2<-function(x){dgamma(x,68,45)}
middleInt<-function(x){integrate(fun1,x,Inf)$value}
outerInt<-function(x){fun2(x)*sapply(x,middleInt)}
integrate(outerInt,0,Inf)
```

```
0.9725601 with absolute error < 2.4e-05
```

The posterior predictive distributions are both negative binomial distributions (see Exercise 4):

$$\tilde{Y}_{n_1+1,1} | n_1 = 111, \sum_i y_{i,1} = 217 \sim \text{NegBin}(2 + 217, \frac{1 + 111}{1 + 111 + 1})$$

$$\tilde{Y}_{n_2+1,2} | n_2 = 44, \sum_i y_{i,2} = 66 \sim \text{NegBin}(2 + 66, \frac{1 + 44}{1 + 44 + 1})$$

We can also use `R` to plot the posterior predictive distributions. These distributions overlap quite a bit: note the difference between comparing parameters (θ_1, θ_2) and future data $(\tilde{Y}_1, \tilde{Y}_2)$.

```
y<-0:10
p1<-dnbinom(y,size=2+217,p=(1+111)/(1+111+1))
p2<-dnbinom(y,size=2+66,p=(1+44)/(1+44+1))
dat<-data.frame(y=c(y,y),
```

```
                p=c(p1,p2),
                group=c(rep("without degree",11),rep("with degree",11))
                )

ggplot(data=dat,mapping=aes(x=y,y=p,fill=group)) +
  geom_bar(position="dodge",stat="identity",width=0.5) +
  xlab(expression(y[n+1])) +
  ylab(expression(paste(sep="","p(",y[n+1],"|",y[1],",....,",y[n],")"))) +
  ggtitle("Posterior predictive distributions for both groups of women")
```


Posterior predictive distributions for both groups of women
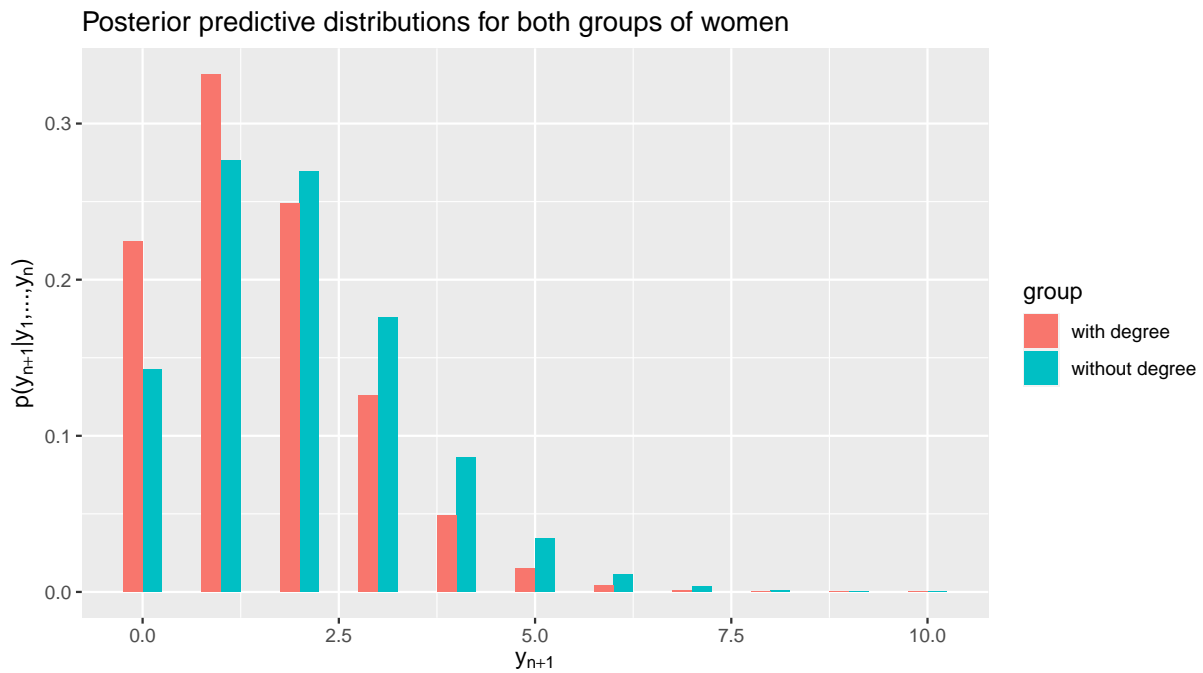
Exercise 6

Let $\phi = g(\theta)$, where $g(.)$ is a monotone function, and let $g^{-1}(.)$ be the inverse of $g(.)$ so that $\theta = g^{-1}(\phi)$. If $p_\theta(\theta)$ is the probability density of θ, then the probability density of ϕ induced by p_θ is given by $p_\phi(\phi) = p_\theta(g^{-1}(\phi)) \cdot \left| \frac{dg^{-1}}{d\phi} \right|$.

Using this result, show that $\text{Beta}(a,b)$ is conjugate for $\text{Bin}(n,\pi)$ using exponential family notation.

Make this prior as weakly informative as possible for π. Do you recognise this prior?

Exercise 6 (Solution)

Let $Y \sim \text{Bin}(n, \pi)$. Then we can write:

$$p(Y = k|\pi) = \binom{n}{k} \pi^k (1 - \pi)^{n-k} \tag{44}$$

$$= \binom{n}{k} \left(\frac{\pi}{1 - \pi} \right)^k (1 - \pi)^n \tag{45}$$

$$= \binom{n}{k} (1 - \pi)^n \exp \left(k \log \frac{\pi}{1 - \pi} \right) \tag{46}$$

Let's write $\phi = \log \frac{\pi}{1-\pi} \Rightarrow \pi = \frac{e^\phi}{1+e^\phi}$. Then:

$$p(Y = k|\phi) = \binom{n}{k} (1 + e^\phi)^{-n} e^{k\phi} \tag{47}$$

So we see that this is in exponential family form with $h(k) = \binom{n}{k}$, $c(\phi) = (1 + e^\phi)^{-1}$ (note that the exponent $-n$ in the likelihood is just due to the fact that we have n observations) and $\phi = \log \frac{\pi}{1-\pi}$, the log odds.

From this we then get the conjugate prior for ϕ:

$$p(\phi|n_0, t_0) \propto (1 + e^\phi)^{-n_0} e^{n_0 t_0 \phi}$$

where t_0 represents the prior expectation of the binomial probability parameter π and n_0 is the equivalent prior sample size.

To get the conjugate prior for π we need to use the change of variable formula:

$$\frac{d}{d\pi} \log \frac{\pi}{1-\pi} = \frac{1}{\pi(1-\pi)}$$

We then get:

$$p(\pi|n_0, t_0) \propto \left(1 + \frac{\pi}{1-\pi}\right)^{-n_0} \left(\frac{\pi}{1-\pi}\right)^{n_0 t_0} \frac{1}{\pi(1-\pi)} \tag{48}$$

$$\propto \pi^{n_0 t_0 - 1}(1-\pi)^{n_0(1-t_0)-1} \tag{49}$$

This is a $\text{Beta}(n_0 t_0, n_0(1-t_0))$ distribution.

This can be made weakly informative by setting $t_0 = \frac{1}{2}$ (our prior guess of both binary outcomes being equally likely) and $n_0 = 1$, the smallest possible equivalent sample size for the prior.

But this is then a $\text{Beta}(1/2, 1/2)$ distribution - the Jeffrey's prior for the binomial sampling model we saw in lectures. So we see that this prior is weakly informative.

Exercise 7

Find the conjugate prior for

    i. a Geometric(π) sampling model

    ii. an Exponential(λ) sampling model

Exercise 7 (Solution)

Geometric(π) sampling model

Sampling model:

$$Y_1, \dots, Y_n | \pi \sim \text{Geo}(\pi) \Rightarrow p(y_1, \dots, y_n | \pi) = \prod_{i=1}^{n} (1-\pi)^{y_i} \pi = (1-\pi)^{\sum_i y_i - n} \pi^n$$

The posterior is found by

$$p(\pi | y_1, \dots, y_n) \propto \qquad\qquad p(\pi)\, p(y_1, \dots, y_n | \pi) \qquad\qquad (50)$$

$$\propto \qquad\qquad p(\pi)(1-\pi)^{\sum_i y_i - n} \pi^n \qquad\qquad (51)$$

For $p(\pi)$ to be conjugate, we see that it needs to be of the form

$$(1-\pi)^{c_1} \pi^{c_2}$$

We also need $\pi \in [0, 1]$ since it needs to be a probability parameter.

The Beta($a, b,$) distribution satisfies this, so let's try:

$$\Pi \sim \text{Beta}(a, b)$$

Indeed, if $\Pi \sim \text{Beta}(a, b)$, then $p(\pi) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1}(1-\pi)^{b-1}$.

$$\Rightarrow p(\pi | y_1, \dots, y_n) \propto \qquad\qquad \pi^{a-1}(1-\pi)^{b-1}(1-\pi)^{\sum_i y_i - n} \pi^n \qquad\qquad (52)$$

$$\propto \qquad\qquad \pi^{a+n-1}(1-\pi)^{b+\sum_i y_i - n - 1} \qquad\qquad (53)$$

$$\Rightarrow \Pi | y_1, \dots, y_n \sim \qquad\qquad \text{Beta}(a+n, b + \sum_y y_i - n) \qquad\qquad (54)$$

We see that $\text{Beta}(a, b)$ is conjugate for a $\text{Geometric}(\pi)$ sampling model.

Exponential(λ) sampling model

Let $Y_1, \dots, Y_n \sim_{\text{iid}} \text{Exp}(\lambda)$ and suppose we observe data $y_1, \dots, y_n$.

The likelihood for this sampling model is given by:

$$p(y_1, \dots, y_n | \lambda) = \prod_i p(y_i | \lambda) \tag{55}$$

$$= \lambda^n e^{-\lambda \sum_i y_i} \tag{56}$$

The posterior is then obtained from:

$$p(\lambda | y_1, \dots, y_n) \propto p(\lambda)\, p(y_2, \dots, y_n | \lambda) \propto p(\lambda) \lambda^n e^{-\lambda \sum_i y_i}$$

For $p(\lambda)$ to be conjugate it needs to be of the form $\lambda^{c_1} e^{-c_2 \lambda}$ and we also require $\lambda > 0$ as it is the parameter of an exponential distribution.

The $\Gamma(a, b)$ distribution satisfies this requirement, so let's try:

$$\Lambda \sim \Gamma(a, b)$$

Indeed, if this is the case:

$$p(\lambda | y_1, \dots, y_n) \propto p(\lambda)\, p(y_1, \dots, y_n | \lambda) \tag{57}$$

$$\propto \lambda^{a-1} e^{-b\lambda} \lambda^n e^{-\lambda \sum_i y_i} \tag{58}$$

$$\propto \lambda^{a+n-1} e^{-(b + \sum_i y_i)\lambda} \tag{59}$$

$$\Rightarrow \Lambda | y_1, \dots, y_n \sim \Gamma(a + n, b + \sum_i y_i) \tag{60}$$

So we see that $\Gamma(a, b)$ is conjugate for an Exponential(λ) sampling model.

Exercise 8

Show that the conjugate prior for the mean parameter μ in a Gaussian distribution is itself a Gaussian distribution (assume that the variance σ^2 of the sampling model is known and fixed).

Exercise 8 (Solution)

Let's suppose $Y_1, \ldots, Y_n \sim_{\text{iid}} \mathcal{N}(\mu, \sigma^2)$ and we observe data $y_1, \ldots, y_n$. Consider σ^2 to be known and fixed (all derivations below will be conditional on σ^2).

Suppose we have a normal distribution prior for $\mu | \sigma^2$: $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$.

Then the posterior $p(\mu | \sigma^2, y_1, \ldots, y_n)$ is given by

$$p(\mu | \sigma^2, y_1, \ldots, y_n) \propto p(\mu | \sigma^2)\, p(y_1, \ldots, y_n | \mu, \sigma^2) \tag{61}$$

$$\propto \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right) \exp\left(-\frac{1}{2\sigma^2}\sum_i (y_i - \mu)^2\right) \tag{62}$$

Let's focus only on the terms inside the exponentials and let's ignore the factor $-\frac{1}{2}$ for now.

$$\frac{1}{\sigma_0^2}\left(\mu^2 - 2\mu\mu_0 + \mu_o^2\right) + \frac{1}{\sigma^2}\left(\sum_i y_i^2 - 2\mu \sum_i y_i + n\mu^2\right) = a\mu^2 - 2b\mu + c$$

where $a = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}$, $b = \frac{\mu_0}{\sigma_0^2} + \frac{\sum_i y_i}{\sigma^2}$ and $c = c(\mu_0, \sigma_0^2, \sum_i y_i, \sigma^2)$.

With this notation, we can write:

$$p(\mu | \sigma^2, y_1, \ldots, y_n) \propto \exp\left(-\frac{1}{2}(a\mu^2 - 2b\mu)\right) \tag{63}$$

$$\propto \exp\left(-\frac{1}{2}a\left(\mu^2 - 2\frac{b}{a}\mu + \frac{b^2}{a^2}\right) + \frac{b^2}{2a^2}\right) \tag{64}$$

$$\propto \exp\left(-\frac{1}{2}a\left(\mu - \frac{b}{a}\right)^2\right) \tag{65}$$

$$\propto \exp\left(-\frac{1}{2}\left(\frac{\mu - b/a}{1/\sqrt{a}}\right)^2\right) \tag{66}$$

This is a normal distribution with posterior mean

$$\mu_n = \frac{b}{a} = \frac{\frac{1}{\sigma_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$$

and posterior variance

$$\sigma_n^2 = \frac{1}{a} = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$$

We conclude that the normal distribution is the conjugate prior for μ in a normal sampling model, assuming σ^2 known and fixed.