MA676: Bayesian Methods – Exercise Sheet 1

1. Suppose that $X_1, \ldots, X_n$ is a random sample from the distribution with pdf

$$f(x|\theta) = \begin{cases} \theta x^{\theta-1} & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Suppose also that the value of the parameter θ is unknown ($\theta > 0$) and that the prior distribution of θ is a gamma distribution with parameters α and β ($\alpha > 0$ and $\beta > 0$). Determine the posterior distribution of θ and hence obtain the Bayes estimator of θ under a squared error loss function.

Solution:

$$\text{Likelihood: } f(\mathbf{x}|\theta) = \theta^n (x_1 \cdots x_n)^{\theta-1}$$

$$\text{Prior: } \pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \theta^{n+\alpha-1} e^{-\beta\theta + \theta \sum \log x_i} \\ &= \theta^{n+\alpha-1} e^{-\theta(\beta - \sum \log x_i)} \end{aligned}$$

Hence $\theta|\mathbf{x} \sim \text{Gamma}(n+\alpha, \beta - \sum \log x_i)$.

Therefore the Bayes estimator under squared error loss is

$$E(\theta|\mathbf{x}) = \frac{n+\alpha}{\beta - \sum \log x_i}.$$

□

2. A Bayes estimator is required for θ under the loss function

$$L(a, \theta) = e^{c(a-\theta)} - c(a-\theta) - 1,$$

where c is a positive constant. As the constant c varies, the loss function varies from very asymmetric to almost symmetric. This is called the LINEX (LINear-EXponential) loss.

By minimising the expected loss, show that the Bayes estimator is

$$\hat{\theta} = \frac{-1}{c} \log E\left(e^{-c\theta}|\mathbf{x}\right),$$

where the expectation is under the posterior distribution $\pi(\theta|\mathbf{x})$.

Suppose that $X_1, \ldots, X_n$ is a random sample from the normal distribution $N(\theta,\ \sigma^2)$ where σ^2 is known. A priori θ follows the normal distribution with mean μ and variance τ^2 where both μ and τ^2 are known. Find the Bayes estimators of θ under the above loss function and under the squared error loss. Compare the two estimators. **Hint:** Recall the definition and the expression for the moment generating function of a normal distribution.

Solution:

(a)

We first find the expected loss.

$$\begin{aligned}
E\left[L(a,\theta)\right] &= \int_{-\infty}^{\infty} L(a,\theta)\pi(\theta|\mathbf{x})d\theta \\
&= \int_{-\infty}^{\infty} \left\{e^{c(a-\theta)} - c(a-\theta) - 1\right\}\pi(\theta|\mathbf{x})d\theta \\
&= \int_{-\infty}^{\infty} e^{c(a-\theta)}\pi(\theta|\mathbf{x})d\theta - c(a-\theta)\pi(\theta|\mathbf{x})d\theta - 1 \\
&= e^{ca}E\left[e^{-c\theta}|\mathbf{x}\right] - c\left\{a - E(\theta|\mathbf{x})\right\} - 1 \\
&= g(a), \quad \text{say}
\end{aligned}$$

It is easy to see that

$$\begin{aligned}
g'(a) &= ce^{ca}E\left[e^{-c\theta}|\mathbf{x}\right] - c \quad \text{and} \\
g''(a) &= c^2e^{ca}E\left[e^{-c\theta}|\mathbf{x}\right].
\end{aligned}$$

Hence

$$\begin{aligned}
& g'(a) &&= 0 \\
\Longrightarrow\ & ce^{ca}E\left[e^{-c\theta}|\mathbf{x}\right] - c &&= 0 \\
\Longrightarrow\ & e^{ca}E\left[e^{-c\theta}|\mathbf{x}\right] &&= 1 \\
\Longrightarrow\ & ca + \log E\left[e^{-c\theta}|\mathbf{x}\right] &&= 0 \\
\Longrightarrow\ & a &&= -\tfrac{1}{c}\log E\left[e^{-c\theta}|\mathbf{x}\right].
\end{aligned}$$

Therefore the Bayes estimator is

$$\hat{\theta} = -\frac{1}{c}\log E\left[e^{-c\theta}|\mathbf{x}\right].$$

(b)

From notes we have,

$$\theta|\mathbf{x} \sim N\left(\lambda = \frac{n\bar{x}/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2}, \rho^2 = \frac{1}{n/\sigma^2 + 1/\tau^2}\right).$$

Note that

$$E\left[e^{-c\theta}|\mathbf{x}\right] = \text{ the moment generating function function of } N(\lambda,\rho^2) \text{ evaluated at } t = -c.$$

However, we know that

$$M_X(t) = E(e^{tX}) = e^{\mu t + \sigma^2 t^2/2} \quad \text{if } X \sim N(\mu,\sigma^2).$$

Hence

$$\begin{aligned}
E\left[e^{-c\theta}|\mathbf{x}\right] &= e^{\lambda(-c) + \rho^2 c^2/2} \\
&= e^{-\lambda c + \rho^2 c^2/2}
\end{aligned}$$

Therefore,

$$\log E\left[e^{-c\theta}|\mathbf{x}\right] = -\lambda c + \rho^2 c^2/2,$$

and the Bayes estimator is

$$\hat{\theta} = \lambda - \rho^2 c/2 = \frac{n\bar{x}/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2} - \frac{c}{2(n/\sigma^2 + 1/\tau^2)}.$$

(c)

The Bayes estimator under squared error loss is

$$\tilde{\theta} = \frac{n\bar{x}/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2}.$$

Here we see that $\hat{\theta} < \tilde{\theta}$ since $c > 0$.

$\square$

3. A random variable X has a gamma distribution gamma(m, β) with pdf

$$f(x) \quad = \quad \frac{\beta^m}{\Gamma(m)} x^{m-1} \exp(-\beta x) \qquad x > 0.$$

Show that if $Y = 1/X$ then Y has p.d.f.

$$f(y) \quad = \quad \frac{\beta^m}{\Gamma(m)} \frac{1}{y^{m+1}} \exp(-\beta/y) \qquad y > 0.$$

This is the inverse gamma distribution.

A particular measuring device has normally distributed error with mean zero and unknown variance σ^2. In an experiment to estimate σ^2, n independent evaluations of this error are obtained.

If the prior distribution for σ^2 is inverse gamma with parameters β and m, show that the posterior distribution is also inverse gamma, with parameters β^* and m^*, and derive expressions for β^* and m^*. Show that the predictive distribution for the error, X_{n+1}, of a further observation made by this device has p.d.f.

$$f(x_{n+1}) \quad \propto \quad \left(1 + \frac{x_{n+1}^2}{2\beta^*}\right)^{-m^*-\frac{1}{2}}, \qquad x_{n+1} \in \mathbb{R}.$$

Solution:

(a)

Here $y = 1/x \Longrightarrow x = 1/y = r(y)$, say.

Therefore, $\frac{dx}{dy} = -\frac{1}{y^2}$ and the pdf of Y is

$$\begin{aligned} f(y) &= f(r(y)) \left|\frac{dx}{dy}\right| \\ &= \frac{\beta^m}{\Gamma(m)} \frac{1}{y^{m-1}} \exp(-\beta/y) \frac{1}{y^2}, \\ &= \frac{\beta^m}{\Gamma(m)} \frac{1}{y^{m+1}} \exp(-\beta/y), \quad y > 0. \end{aligned}$$

(b)

We have $X_1, \ldots X_n \overset{iid}{\sim} N(0, \sigma^2)$. Therefore,

$$\begin{aligned} f(\mathbf{x}|\sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} x_i^2} \\ &= \frac{1}{(2\pi)^{n/2} (\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2} \end{aligned}$$

Prior for σ^2 is

$$\pi(\sigma^2) = \frac{\beta^m}{\Gamma(m)} \frac{1}{(\sigma^2)^{m+1}} e^{-\beta/\sigma^2}.$$

Therefore, the posterior density is:

$$\begin{aligned} \pi(\sigma^2|\mathbf{x}) &\propto \frac{1}{(\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2} \frac{1}{(\sigma^2)^{m+1}} e^{-\beta/\sigma^2}, \\ &= \frac{1}{(\sigma^2)^{n/2+m+1}} e^{-\frac{1}{\sigma^2}(\beta + \frac{1}{2}\sum_{i=1}^n x_i^2)}, \quad \sigma^2 > 0. \end{aligned}$$

Clearly this is the density of the inverse gamma distribution with parameters $m^* = n/2 + m$ and $\beta^* = \beta + \frac{1}{2}\sum_{i=1}^n x_i^2$.

(c) We proceed as follows for the posterior predictive distribution.

$$\begin{aligned} f(x_{n+1}|\mathbf{x}) &= \int_{-\infty}^{\infty} f(x_{n+1}|\theta)\pi(\theta|\mathbf{x}) d\theta \\ &= \int_0^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} x_{n+1}^2} \frac{(\beta^*)^{m^*}}{\Gamma(m^*)} \frac{1}{(\sigma^2)^{m^*+1}} e^{-\beta^*/\sigma^2} d\sigma^2, \\ &\propto \int_0^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{(\sigma^2)^{m^*+\frac{1}{2}+1}} e^{-\frac{1}{\sigma^2}(\beta^* + \frac{1}{2}x_{n+1}^2)} d\sigma^2. \end{aligned}$$

Now the integrand looks like the inverse gamma density with $\tilde{m} = m^* + \frac{1}{2}$ and $\tilde{\beta} = \beta^* + \frac{1}{2}x_{n+1}^2$

$$\begin{aligned} f(x_{n+1}|\mathbf{x}) &\propto \frac{\Gamma(\tilde{m})}{(\tilde{\beta})^{\tilde{m}}} \\ &= \frac{\Gamma(m^* + \frac{1}{2})}{(\beta^* + \frac{1}{2}x_{n+1}^2)^{m^* + \frac{1}{2}}} \\ &\propto (\beta^* + \tfrac{1}{2}x_{n+1}^2)^{-m^* - \frac{1}{2}} \\ &\propto \left(1 + \frac{x_{n+1}^2}{2\beta^*}\right)^{-m^* - \frac{1}{2}}. \end{aligned}$$

□

4. Assume $Y_1, Y_2, \ldots, Y_n$ are independent observations which have the normal distribution with mean βx_i and variance σ^2, where the x_is and σ^2 are known constants, and β is an unknown parameter, which has a normal prior distribution with mean β_0 and variance τ^2, where β_0 and τ^2 are known constants.

(a) Derive the posterior distribution of β.

(b) Show that the mean of the posterior distribution is a weighted average of the prior mean β_0, and the maximum likelihood estimator of β.

(c) Find the limit of the posterior distribution as $\tau^2 \to \infty$, and discuss the result.

(d) How would you predict a future observation from the population $N(\beta x_{n+1}, \sigma^2)$, where x_{n+1} is known?

Solution:

We have $Y_1, \ldots, Y_n \overset{iid}{\sim} N(\beta x_i, \sigma^2)$.

Therefore,

$$\begin{aligned} f(y_1, \ldots, y_n|\beta) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2} \\ &\propto e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta x_i)^2}. \end{aligned}$$

Prior is

$$\pi(\beta) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}(\beta - \beta_0)^2}$$

Therefore, posterior is:

$$\begin{aligned} \pi(\beta|\mathbf{y}) &\propto e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta x_i)^2 - \frac{1}{2\tau^2}(\beta - \beta_0)^2} \\ &= e^{-\frac{1}{2}\left\{\frac{1}{\sigma^2}\sum_{i=1}^n (y_i - \beta x_i)^2 + \frac{1}{\tau^2}(\beta - \beta_0)^2\right\}} \\ &= e^{-\frac{1}{2}M}, \text{ say.} \end{aligned}$$

Now

$$\begin{aligned} M &= \frac{\sum y_i^2}{\sigma^2} - 2\beta\frac{\sum y_i x_i}{\sigma^2} + \beta^2\frac{\sum x_i^2}{\sigma^2} + \beta^2\frac{1}{\tau^2} - 2\beta\frac{\beta_0}{\sigma^2} + \frac{\beta_0^2}{\tau^2} \\ &= \beta^2\left(\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2}\right) - 2\beta\left(\frac{\sum y_i x_i}{\sigma^2} + \frac{\beta_0}{\tau^2}\right) + \frac{\sum y_i^2}{\sigma^2} + \frac{\beta_0^2}{\tau^2} \\ &= \beta^2\left(\frac{1}{\sigma_1^2}\right) - 2\beta\frac{\beta_1}{\sigma_1^2} + \frac{\sum y_i^2}{\sigma^2} + \frac{\beta_0^2}{\tau^2} \\ &= \frac{(\beta - \beta_1)^2}{\sigma_1^2} - \frac{1}{\sigma_1^2}\left(\frac{\sum y_i x_i}{\sigma^2} + \frac{\beta_0}{\tau^2}\right)^2 + \frac{\sum y_i^2}{\sigma^2} + \frac{\beta_0^2}{\tau^2} \end{aligned}$$

where

$$\sigma_1^2 = \frac{1}{\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2}} \quad \text{and} \quad \beta_1 = \sigma_1^2\left(\frac{\sum y_i x_i}{\sigma^2} + \frac{\beta_0}{\tau^2}\right).$$

Clearly,

$$\beta|\mathbf{y} \sim N(\beta_1, \sigma_1^2).$$

(b)

Define

$$\hat{\beta} = \frac{\sum y_i x_i}{\sum x_i^2}$$

which is the maximum likelihood estimate of β. We have

$$\begin{aligned}
\beta_1 &= \sigma_1^2\left(\frac{\sum y_i x_i}{\sigma^2}+\frac{\beta_0}{\tau^2}\right)\\
&= \frac{\frac{\sum y_i x_i}{\sigma^2}+\frac{\beta_0}{\tau^2}}{\frac{\sum x_i^2}{\sigma^2}+\frac{1}{\tau^2}}\\
&= \frac{\tau^2\sum y_i x_i+\sigma^2\beta_0}{\tau^2\sum x_i^2+\sigma^2}\\
&= \frac{\tau^2\frac{\sum y_i x_i}{\sum x_i^2}+\frac{\sigma^2}{\sum x_i^2}\beta_0}{\tau^2+\frac{\sigma^2}{\sum x_i^2}}\\
&= \frac{w_1\hat{\beta}+w_2\beta_0}{w_1+w_2}
\end{aligned}$$

where

$$w_1=\tau^2 \text{ and } w_2=\frac{\sigma^2}{\sum x_i^2}.$$

(c)

As $\tau^2\to\infty$, $\sigma_1^2\to\frac{\sigma^2}{\sum x_i^2}$. That is

$$\begin{aligned}
\beta|\mathbf{y} &\sim N\left(\hat{\beta},\ \frac{\sigma^2}{\sum x_i^2}\right)\\
i.e.\quad \beta|\mathbf{y} &\sim N\left(\hat{\beta},\ \mathrm{var}(\hat{\beta})\right).
\end{aligned}$$

Hence inference for β using the posterior will be same as that based on the maximum likelihood estimate.

(c) We want

$$f(y_{n+1}|\mathbf{y})=\int_{-\infty}^{\infty}f(y_{n+1}|\beta)\pi(\beta|\mathbf{y})d\beta.$$

Although this can be derived from the first principles, we take a different approach to solve this.

We use two results on conditional expectation:

$$E(X)=EE(X|Y),\quad \mathrm{var}(X)=E\mathrm{var}(X|Y)+\mathrm{var}E(X|Y).$$

We take $X=Y_{n+1}$ and $Y=\beta$. We also have $Y_{n+1}|\beta\sim N(\beta x_{n+1},\sigma^2)$ and $\beta|\mathbf{y}\sim N(\beta_1,\sigma_1^2)$. Now

$$E(Y_{n+1}|\mathbf{y})=E(\beta x_{n+1})=\beta_1 x_{n+1}.$$

$$\begin{aligned}
\mathrm{var}(Y_{n+1}|\mathbf{y}) &= E[\mathrm{var}(Y_{n+1}|\beta)]+\mathrm{var}[E(Y_{n+1}|\beta)]\\
&= E[\sigma^2]+\mathrm{var}[\beta x_{n+1}]\\
&= \sigma^2+x_{n+1}^2\sigma_1^2
\end{aligned}$$

Also we can write

$$Y_{n+1}|\mathbf{y}=x_{n+1}\beta|\mathbf{y}+\epsilon$$

where $\beta|\mathbf{y}$ follows $N(\beta_1, \sigma_1^2)$ and ϵ follows $N(0, \sigma^2)$ independently. Hence $Y_{n+1}|\mathbf{y}$ follows a normal distribution. Therefore,

$$Y_{n+1}|\mathbf{y} \sim N(\beta_1 x_{n+1}, \sigma^2 + x_{n+1}^2 \sigma_1^2).$$

$\square$

5. Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of independent, identically distributed random variables with probability density function

$$f(y|\lambda) \;=\; \begin{cases} \lambda e^{-\lambda y}, & y > 0 \\ 0, & \text{otherwise} \end{cases}$$

where λ is an unknown, positive parameter with a gamma(m, β) prior distribution (see above).

(a) Show that the posterior distribution of λ given $Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n$ is gamma$(n+m, \beta+t)$ where $t = \displaystyle\sum_{i=1}^{n} y_i$.

(b) Show that the (predictive) density of Y_{n+1} given the n observations $Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n$ is

$$\pi(y_{n+1}|y_1, \ldots, y_n) \;=\; \frac{(n+m)(\beta+t)^{n+m}}{(y_{n+1}+\beta+t)^{n+m+1}}.$$

(c) Find the joint (predictive) density of Y_{n+1} and Y_{n+2} given $Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n$.

Solution:

(a) Here

$$f(\mathbf{y}|\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n y_i}$$

and

$$\pi(\lambda) = \frac{\beta^m}{\Gamma(m)} \lambda^{m-1} e^{-\beta\lambda}$$

The posterior is

$$\begin{aligned} \pi(\lambda|\mathbf{y}) &\propto f(\mathbf{y}|\lambda) \times \pi(\lambda) \\ &\propto \lambda^{m+n-1} e^{-\lambda(\beta + \sum_{i=1}^n y_i)}. \end{aligned}$$

Clearly $\lambda|\mathbf{y} \sim Gamma(m+n, \beta + \sum_{i=1}^n y_i)$.

(b)

We have $f(y_{n+1}|\lambda) = \lambda e^{-\lambda y_{n+1}}$. Now

$$\begin{aligned} f(y_{n+1}|\mathbf{y}) &= \int_0^\infty f(y_{n+1}|\lambda)\, \pi(\lambda|\mathbf{y}) d\lambda \\ &= \int_0^\infty \lambda e^{-\lambda y_{n+1}} \tfrac{(\beta+t)^{m+n}}{\Gamma(m+n)} \lambda^{m+n-1} e^{-(\beta+t)\lambda} \\ &= \tfrac{(\beta+t)^{m+n}}{\Gamma(m+n)} \int_0^\infty \lambda^{m+n+1-1} e^{-(\beta+t+y_{n+1})\lambda} \\ &= \tfrac{(\beta+t)^{m+n}}{\Gamma(m+n)} \tfrac{\Gamma(m+n+1)}{(\beta+t+y_{n+1})^{m+n}} \\ &= \tfrac{(n+m)(\beta+t)^{n+m}}{(y_{n+1}+\beta+t)^{n+m+1}}, \end{aligned}$$

where $y_{n+1} > 0$.

We have $f(y_{n+2}, y_{n+1}|\lambda) = \lambda^2 e^{-\lambda(y_{n+2}+y_{n+1})}$, since Y_{n+2} and Y_{n+1} are conditionally independent given λ. Now

$$\begin{aligned} f(y_{n+2}, y_{n+1}|\mathbf{y}) &= \int_0^\infty f(y_{n+2}, y_{n+1}|\lambda)\, \pi(\lambda|\mathbf{y}) d\lambda \\ &= \int_0^\infty \lambda^2 e^{-\lambda(y_{n+2}+y_{n+1})} \tfrac{(\beta+t)^{m+n}}{\Gamma(m+n)} \lambda^{m+n-1} e^{-(\beta+t)\lambda} \\ &= \tfrac{(n+m+1)(n+m)(\beta+t)^{n+m}}{(y_{n+2}+y_{n+1}+\beta+t)^{n+m+2}}, \end{aligned}$$

where $y_{n+2} > 0$ and $y_{n+1} > 0$.

$\square$

MA676: Bayesian Methods – Exercise Sheet 2

1. Let $X_1, X_2, \ldots, X_6$ be a sequence of independent, identically distributed Bernoulli random variables with parameter θ, and suppose that $x_1 = x_2 = x_3 = x_4 = x_5 = 1$ and $x_6 = 0$.

   Derive the posterior model probabilities for Model $0 : \theta = \frac{1}{2}$ and Model $1 : \theta > \frac{1}{2}$, assuming the following prior distributions:

   (a) $P(M_0) = 0.5$, $P(M_1) = 0.5$, $\pi_1(\theta) = 2$; $\quad \theta \in \left(\frac{1}{2}, 1\right)$.

   (b) $P(M_0) = 0.8$, $P(M_1) = 0.2$, $\pi_1(\theta) = 8(1-\theta)$; $\quad \theta \in \left(\frac{1}{2}, 1\right)$.

   (c) $P(M_0) = 0.2$, $P(M_1) = 0.8$, $\pi_1(\theta) = 48\left(\theta - \frac{1}{2}\right)(1-\theta)$; $\quad \theta \in \left(\frac{1}{2}, 1\right)$.

   Solution:

   Recall that

   $$P(M_i|\mathbf{x}) = \frac{P(M_i) f(\mathbf{x}|M_i)}{P(M_0) f(\mathbf{x}|M_0) + P(M_1) f(\mathbf{x}|M_0)}$$

   where

   $$f(\mathbf{x}|M_i) = \int f(\mathbf{x}|\theta, M_i) \times \pi_i(\theta) d\theta$$

   and $P(M_i)$ is the prior probability of model i. We have

Model 0	Model 1
$\theta = \frac{1}{2}$	$\frac{1}{2} < \theta < 1$
$f(\mathbf{x}\|\theta = \frac{1}{2}) = (\frac{1}{2})^6$	$f(\mathbf{x}\|\theta) = \theta^5(1-\theta)$
$f(\mathbf{x}\|M_0) = (\frac{1}{2})^6$	$f(\mathbf{x}\|M_1) = \int_{\frac{1}{2}}^{1} \theta^5(1-\theta)\, \pi_1(\theta)\, d\theta$

   Now we calculate the model probabilities.

Model 0	Model 1	$P(M_0\|\mathbf{x})$
$P(M_0) = \frac{1}{2}$ $f(\mathbf{x}\|M_0) = \frac{1}{64}$	$P(M_1) = \frac{1}{2}$ $f(\mathbf{x}\|M_1) = \int_{\frac{1}{2}}^{1} \theta^5(1-\theta)\, 2\, d\theta = \frac{5}{112}$	$P(M_0\|\mathbf{x}) = 0.26$
$P(M_0) = 0.8$ $f(\mathbf{x}\|M_0) = \frac{1}{64}$	$P(M_1) = 0.2$ $f(\mathbf{x}\|M_1) = \int_{\frac{1}{2}}^{1} \theta^5(1-\theta)\, 8(1-\theta)\, d\theta = \frac{73}{1792}$	$P(M_0\|\mathbf{x}) = 0.60$
$P(M_0) = 0.2$ $f(\mathbf{x}\|M_0) = \frac{1}{64}$	$P(M_1) = 0.8$ $f(\mathbf{x}\|M_1) = \int_{\frac{1}{2}}^{1} \theta^5(1-\theta)\, 48(\theta - \frac{1}{2})(1-\theta)\, d\theta = 0.051$	$P(M_0\|\mathbf{x}) = 0.07$

   □

2. Suppose that:

   $$M_0 : X_1, X_2, \ldots, X_n|\theta_0 \sim f_0(x|\theta_1) = \theta_0(1-\theta_0)^x, \quad x = 0, 1, \ldots.$$

   $$M_1 : X_1, X_2, \ldots, X_n|\theta_1 \sim f_1(x|\theta_1) = e^{-\theta_1}\theta_1^x/x!, \quad x = 0, 1, \ldots.$$

Suppose that θ_0 and θ_1 are both unknown. Assume that $\pi_0(\theta_0)$ is the beta distribution with parameters α_0 and β_0 and $\pi_1(\theta_1)$ is the Gamma distribution with parameters α_1 and β_1. Compute the (prior) predictive means under the two models. Obtain the Bayes factor. Hence study the dependence of the Bayes factor on prior data combinations. Calculate numerical values for $n = 2$ and for two data sets $x_1 = x_2 = 0$ and $x_1 = x_2 = 2$ and two sets of prior parameters $\alpha_0 = 1, \beta_0 = 2$, $\alpha_1 = 2, \beta_1 = 1$ and $\alpha_0 = 30, \beta_0 = 60$, $\alpha_1 = 60, \beta_1 = 30$.

Solution:

Here

$$E(X_i|\theta_0) = \frac{1-\theta_0}{\theta_0}, \; E(X_i|\theta_1) = \theta_1.$$

We have

$$\theta_0 \sim Beta(\alpha_0, \; \beta_0), \quad \theta_1 \sim Gamma(\alpha_1, \beta_1)$$

Therefore,

$$\begin{aligned} E(X_i|M_0) &= \int_0^1 \tfrac{1-\theta_0}{\theta_0} \tfrac{1}{B(\alpha_0,\beta_0)} \theta_0^{\alpha_0-1}(1-\theta_0)^{\beta_0-1} d\theta_0 \\ &= \tfrac{1}{B(\alpha_0,\beta_0)} \int_0^1 \theta_0^{\alpha_0-1-1}(1-\theta_0)^{\beta_0+1-1} d\theta_0 \\ &= \tfrac{B(\alpha_0-1,\beta_0+1)}{B(\alpha_0,\beta_0)} \\ &= \tfrac{\beta_0}{\alpha_0-1}. \end{aligned}$$

Now $E(X_i|M_1) = E(\theta_1)$ where $\theta_1 \sim Gamma(\alpha_1, \beta_1)$. Therefore, $E(X_i|M_1) = \frac{\alpha_1}{\beta_1}$.

Two predictive means are equal if

$$\frac{\beta_0}{\alpha_0 - 1} = \frac{\alpha_1}{\beta_1}.$$

The Bayes factor for Model 0 is

$$B_{01}(\mathbf{x}) = \frac{f(\mathbf{x}|M_0)}{f(\mathbf{x}|M_1)}$$

where $f(\mathbf{x}|M_i)$ is the marginal likelihood under Model i.

Let $t = \sum x_i$. Here

$$\begin{aligned} f(\mathbf{x}|M_0) &= \int_0^1 \theta_0^n (1-\theta_0)^t \tfrac{1}{B(\alpha_0, \; \beta_0)} \theta_0^{\alpha_0-1}(1-\theta_0)^{\beta_0-1} d\theta_0 \\ &= \tfrac{1}{B(\alpha_0, \; \beta_0)} \int_0^1 \theta_0^{n+\alpha_0-1}(1-\theta_0)^{t+\beta_0-1} d\theta_0 \\ &= \tfrac{B(n+\alpha_0, \; t+\beta_0)}{B(\alpha_0, \; \beta_0)} \end{aligned}$$

For the Poisson model

$$\begin{aligned} f(\mathbf{x}|M_1) &= \int_0^\infty \tfrac{e^{-n\theta_1}\theta_1^t}{\prod_{i=1}^n x_i!} \; \tfrac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \theta_1^{\alpha_1-1} e^{-\beta_1\theta_1} d\theta_1 \\ &= \tfrac{1}{\prod_{i=1}^n x_i!} \tfrac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \int_0^\infty \theta_1^{t+\alpha_1-1} e^{-\theta(n+\beta_1)} d\theta_1 \\ &= \tfrac{1}{\prod_{i=1}^n x_i!} \tfrac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \tfrac{\Gamma(t+\alpha_1)}{(n+\beta_1)^{t+\alpha_1}} \end{aligned}$$

We now calculate numerical values of the Bayes factor.

	$\alpha_0 = 1, \beta_0 = 2$ $\alpha_1 = 2, \beta_1 = 1$	$\alpha_0 = 30, \beta_0 = 60$ $\alpha_0 = 60, \beta_0 = 30$
$x_1 = x_2 = 0$	1.5	2.7
$x_1 = x_2 = 2$	0.29	0.38

□

3. Suppose that $X_1, \ldots, X_n$ is a sample from the negative binomial distribution which has the probability mass function,

$$f(x|r, \theta) = \binom{r + x - 1}{x} \theta^r (1 - \theta)^x, \quad x = 0, 1, \ldots; \; 0 < \theta < 1,$$

where $r > 0$ is a known integer. Suppose also that a-priori θ has a beta(α, β) distribution with the pdf,

$$\pi(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}, \; 0 < \theta < 1.$$

(a) Find the posterior distribution of θ.

(b) Suppose further that $r = 2, n = 1$, and we observe that $x_1 = 1$. Of the two hypotheses $H_1 : \theta \leq 0.5$ and $H_2 : \theta > 0.5$, which has greater posterior probability under the uniform prior.

(c) What is the Bayes factor in favor of H_2? Does it suggest strong evidence in favor of this hypothesis?

Solution:

Here

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^n \tbinom{r+x_i-1}{x_i} \theta^r (1-\theta)^{x_i} \\ &= \theta^{nr} (1-\theta)^{\sum x_i} \prod_{i=1}^n \tbinom{r+x_i-1}{x_i} \end{aligned}$$

Let $t = \sum x_i$. The posterior distribution is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &\propto \theta^{nr} (1-\theta)^t \theta^{\alpha-1} (1-\theta)^{\beta-1} \\ &= \theta^{nr+\alpha-1} (1-\theta)^{t+\beta-1} \end{aligned}$$

That is $\theta|\mathbf{x} \sim Beta(nr + \alpha, t + \beta)$.

We have $n = 1, r = 2, \alpha = \beta = 1$ and $x_1 = 1$.

Therefore, $\theta|\mathbf{x} \sim Beta(3, 2)$.

Now

$$\begin{aligned} Pr\{H_1 \text{ is true}|\mathbf{x}\} &= \int_0^{\frac{1}{2}} \tfrac{1}{B(3,2)} \theta^{3-1} (1-\theta)^{2-1} d\theta \\ &= = 0.3125. \end{aligned}$$

Since, H_1 and H_2 are complementary

$$Pr\{H_2 \text{is true}|\mathbf{x}\} = 1 - Pr\{H_1 \text{ is true}|\mathbf{x}\} = 0.6875.$$

Therefore, H_2 has greater posterior probability.

We first find the marginal likelihoods. Here

$$\begin{aligned} f(\mathbf{x}|H_1) &= \int_0^{\frac{1}{2}} f(x_1|\theta)\ \pi_1(\theta)\ d\theta \\ &= \int_0^{\frac{1}{2}} \theta^2(1-\theta)2\ 2\ d\theta \\ &= 4\int_0^{\frac{1}{2}} \theta^2(1-\theta)\ d\theta \end{aligned}$$

$$\begin{aligned} f(\mathbf{x}|H_2) &= \int_{\frac{1}{2}}^1 f(x_1|\theta)\ \pi_2(\theta)\ d\theta \\ &= \int_{\frac{1}{2}}^1 \theta^2(1-\theta)2\ 2\ d\theta \\ &= 4\int_{\frac{1}{2}}^1 \theta^2(1-\theta)\ d\theta \end{aligned}$$

The Bayes factor in favor of H_2 is

$$B_{21}(x_1) = \frac{f(\mathbf{x}|H_2)}{f(\mathbf{x}|H_1)} = \frac{0.6875}{0.3125} = 2.2.$$

Therefore, there is some evidence in favor of H_2.

□

4. In an experiment to compare two measuring devices n_1 objects are measured with the first device, the mearsurements errors being recorded as $x_1, \ldots, x_{n_1}$, and n_2 objects are measured with the second device, the mearsurements errors being recorded as $x_{n_1+1}, \ldots, x_{n_1+n_2}$. It is assumed that measurement errors are normally distributed with zero mean. Two models are proposed.

   The first model assumes that the measuring devices are identical and variance of both devices is ϕ (unknown). That is $x_1, \ldots, x_{n_1+n_2}$ is a sample of i.i.d. observations from $N(0, \phi)$.

   The second model allows for a difference between the variances, with $x_1, \ldots, x_{n_1}$ being a sample of i.i.d. observations from $N(0, \phi_1)$ and $x_{n_1+1}, \ldots, x_{n_1+n_2}$ from $N(0, \phi_2)$.

   Assume that the prior distributions for ϕ, ϕ_1 and ϕ_2 are all inverse gamma distributions

$$\pi(\phi) = \frac{\beta^m}{\Gamma(m)} \frac{1}{\phi^{m+1}} \exp(-\beta/\phi), \qquad \phi > 0$$

   with the same m and β for each case.

   Obtain the Bayes factor for comparing the models.

   Solution:

   Consider Model 1 first. We have $X_1, \ldots, X_{n_1+n_2} \overset{iid}{\sim} N(0, \phi)$.

$$\begin{aligned} f(\mathbf{x}|\phi) &= \frac{1}{(2\pi)^{(n_1+n_2)/2}} \frac{1}{\phi^{(n_1+n_2)/2}} e^{-\frac{1}{2\phi}\sum_{i=1}^{n_1+n_2} x_i^2} \\ &= \frac{1}{(2\pi)^{(n_1+n_2)/2}} \frac{1}{\phi^{(n_1+n_2)/2}} e^{-\frac{1}{\phi}\frac{S^2}{2}} \end{aligned}$$

where $S^2 = \sum_{i=1}^{n_1+n_2} x_i^2$. The prior is:

$$\pi(\phi) = \frac{\beta^m}{\Gamma(m)} \frac{1}{\phi^{m+1}} e^{-\beta/\phi}.$$

Therefore

$$\begin{aligned}
f(\mathbf{x}|M_1) &= \int_0^\infty f(\mathbf{x}|\phi)\pi(\phi)\ d\phi \\
&= \int_0^\infty \frac{1}{(2\pi)^{(n_1+n_2)/2}} \frac{1}{\phi^{(n_1+n_2)/2}} e^{-\frac{1}{\phi}\frac{S^2}{2}} \frac{\beta^m}{\Gamma(m)} \frac{1}{\phi^{m+1}} e^{-\beta/\phi}\ d\phi \\
&= \frac{1}{(2\pi)^{(n_1+n_2)/2}} \frac{\beta^m}{\Gamma(m)} \int_0^\infty \frac{1}{\phi^{m+(n_1+n_2)/2+1}} e^{-\frac{1}{\phi}(\beta+\frac{S^2}{2})}\ d\phi \\
&= \frac{1}{(2\pi)^{(n_1+n_2)/2}} \frac{\beta^m}{\Gamma(m)} \frac{\Gamma(m+(n_1+n_2)/2)}{\left(\beta+\frac{S^2}{2}\right)^{m+(n_1+n_2)/2}}.
\end{aligned}$$

Now consider Model 2. Here $X_1, \ldots, X_{n_1} \stackrel{iid}{\sim} N(0, \phi_1)$, and $X_{n_1+1}, \ldots, X_{n_1+n_2} \stackrel{iid}{\sim} N(0, \phi_2)$ and the two sets are independent. Now

$$\begin{aligned}
f(x_1, \ldots, x_{n_1+n_2}|\phi_1, \phi_2) &= \frac{1}{(2\pi)^{(n_1+n_2)/2}} \frac{1}{\phi_1^{n_1/2}} e^{-\frac{1}{2\phi_1}\sum_{i=1}^{n_1} x_i^2} \frac{1}{\phi_2^{n_2/2}} e^{-\frac{1}{2\phi_2}\sum_{i=1}^{n_1} x_{n_1+i}^2} \\
&= \frac{1}{(2\pi)^{(n_1+n_2)/2}} \frac{1}{\phi_1^{n_1/2}} e^{-\frac{1}{\phi_1}\frac{S_1^2}{2}} \frac{1}{\phi_2^{n_2/2}} e^{-\frac{1}{\phi_2}\frac{S_2^2}{2}}
\end{aligned}$$

where $S_1^2 = \sum_{i=1}^{n_1} x_i^2$ and $S_2^2 = S^2 - S_1^2$. The prior is:

$$\pi(\phi_1, \phi_2) = \frac{\beta^{2m}}{\Gamma^2(m)} \frac{1}{\phi_1^{m+1}\phi_2^{m+1}} e^{-\beta/\phi_1 - \beta/\phi_2}.$$

Therefore

$$\begin{aligned}
f(\mathbf{x}|M_2) &= \int_0^\infty \int_0^\infty f(\mathbf{x}|\phi_1, \phi_2)\pi(\phi_1)\pi(\phi_2)\ d\phi_1 d\phi_2 \\
&= \frac{1}{(2\pi)^{(n_1+n_2)/2}} \frac{\beta^{2m}}{\Gamma^2(m)} \frac{\Gamma(m+n_1/2)}{\left(\beta+\frac{S_1^2}{2}\right)^{m+n_1/2}} \frac{\Gamma(m+n_2/2)}{\left(\beta+\frac{S_2^2}{2}\right)^{m+n_2/2}}.
\end{aligned}$$

Hence the Bayes factor for Model 1 is:

$$B_{12}(\mathbf{x}) = \frac{\Gamma(m)}{\beta^m} \frac{\Gamma(m+(n_1+n_2)/2)}{\Gamma(m+n_1/2)\Gamma(m+n_2/2)} \frac{\left(\beta + \frac{S_1^2}{2}\right)^{m+n_1/2} \left(\beta + \frac{S_2^2}{2}\right)^{m+n_2/2}}{\left(\beta + \frac{S^2}{2}\right)^{m+(n_1+n_2)/2}}.$$

□

5. Suppose that $Y_1, \ldots, Y_n$ are independently distributed as $N(\beta x_i, \sigma^2)$ where σ^2 and x_i's are known constants. Assume that β follows $N(0, \tau^2)$ a-priori. Find the Bayes factor where one model corresponds to $\beta = 0$ and the other model does not specify any particular value of β. Hence, show that the Bayes factor is a function of the classical test statistic for testing $H_0 : \beta = 0$.

   Solution:

We have

$$f(y_1,\ldots,y_n|\beta) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum(y_i-\beta x_i)^2}$$

and

$$\pi(\beta) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}\beta^2}.$$

Model 0 assumes that $\beta = 0$, hence

$$f(y_1,\ldots,y_n|M_0) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum y_i^2}.$$

Model 1 leaves β unspecified. Let

$$\sigma_1^2 = \frac{1}{\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2}} \quad \text{and} \quad \beta_1 = \frac{\sigma_1^2}{\sigma^2}\sum y_i x_i.$$

Therefore,

$$\begin{aligned}
f(\mathbf{y}|M_1) &= \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum(y_i-\beta x_i)^2} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}\beta^2} d\beta \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \frac{1}{\sqrt{2\pi\tau^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\sum(y_i-\beta x_i)^2 - \frac{1}{2\tau^2}\beta^2} d\beta \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \frac{1}{\sqrt{2\pi\tau^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left\{\frac{(\beta-\beta_1)^2}{\sigma_1^2} + \frac{\sum y_i^2}{\sigma^2} - \frac{\beta_1^2}{\sigma_1^2}\right\}} d\beta \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\sigma^2}\sum y_i^2 + \frac{1}{2\sigma_1^2}\beta_1^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma_1^2}(\beta-\beta_1)^2} d\beta \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\sigma^2}\sum y_i^2 + \frac{1}{2\sigma_1^2}\beta_1^2} \sqrt{2\pi\sigma_1^2} \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \sqrt{\frac{\sigma_1^2}{\tau^2}} e^{-\frac{1}{2\sigma^2}\sum y_i^2 + \frac{1}{2\sigma_1^2}\beta_1^2}.
\end{aligned}$$

Therefore,

$$B_{01}(\mathbf{x}) = \sqrt{\frac{\tau^2}{\sigma_1^2}} \exp\left\{-\frac{1}{2}\frac{\beta_1^2}{\sigma_1^2}\right\}.$$

The first term in the above expression does not involve the observations $y_1,\ldots,y_n$. Recall that

$$\hat{\beta} = \frac{\sum y_i x_i}{\sum x_i^2}, \quad \text{and} \quad \text{var}(\hat{\beta}) = \frac{\sigma^2}{\sum x_i^2}.$$

Now consider the exponent.

$$\begin{aligned}
\frac{\beta_1^2}{\sigma_1^2} &= \sigma_1^2 \left(\frac{\sum y_i x_i}{\sigma^2}\right)^2 \\
&= \frac{\sigma_1^2}{\sigma^4} \left(\sum y_i x_i\right)^2 \\
&= \frac{\sigma_1^2}{\sigma^4} \left(\sum x_i^2\right)^2 \left(\frac{\sum y_i x_i}{\sum x_i^2}\right)^2 \\
&= \sigma_1^2 \left(\frac{\sum x_i^2}{\sigma^2}\right)\left(\frac{\sum x_i^2}{\sigma^2}\right)\hat{\beta}^2 \\
&= \sigma_1^2 \frac{\sum x_i^2}{\sigma^2} \frac{(\hat{\beta})^2}{\text{var}(\hat{\beta})} \\
&= \frac{\frac{\sum x_i^2}{\sigma^2}}{\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2}} \frac{(\hat{\beta})^2}{\text{var}(\hat{\beta})}.
\end{aligned}$$

The second term varies with the observations, $y_1, \ldots, y_n$ and $B_{01}(\mathbf{x})$ will be small if $\frac{(\hat{\beta})^2}{\mathrm{var}(\hat{\beta})}$ is large.

This is the connection with the classical test of hypothesis. In their setup one rejects H_0 if $\frac{(\hat{\beta})^2}{\mathrm{var}(\hat{\beta})}$ is large.

MA676: Bayesian Methods – Exercise Sheet 3

Suppose the problem is to simulate from $\pi(y)$. This $\pi(y)$ is to be seen as the likelihood times the prior in the Bayesian setup. The Metropolis-Hastings algorithm makes a transition as follows.

Suppose $X^{(t)} = x$ is the current value of the chain.

1. Generate y from $q(y|x)$.
2. Calculate $\alpha(x,y) = \min\left\{1, \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}\right\}$.
3. Generate u from the uniform distribution in (0, 1).
4. If $u < \alpha(x,y)$ then set $x^{(t+1)} = y$ otherwise set $x^{(t+1)} = x$.

We need to note two special cases. The first is called the Metropolis algorithm. This corresponds to the case $q(y|x) = q(x|y)$. The acceptance probability of Metropolis algorithm is given by:

$$\alpha(x,y) = \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\}.$$

The second case is a special case of the independence sampler. Suppose that the target posterior distribution is

$$\pi(\theta) \propto L(\theta) \times \pi_0(\theta),$$

where $L(\theta)$ is the likelihood function and $\pi_0(\theta)$ is the prior distribution. Suppose further that we take the proposal distribution to be the prior distribution, i.e. $q(\theta|\phi) = \pi_0(\theta)$. Then

$$\begin{aligned}
\alpha(\phi,\theta) &= \min\left\{1, \frac{\pi(\theta)q(\phi|\theta)}{\pi(\phi)q(\theta|\phi)}\right\} \\
&= \min\left\{1, \frac{\pi(\theta)\pi_0(\phi)}{\pi(\phi)\pi_0(\theta)}\right\} \\
&= \min\left\{1, \frac{L(\theta)\times\pi_0(\theta)\pi_0(\phi)}{L(\phi)\times\pi_0(\phi)\pi_0(\theta)}\right\} \\
&= \min\left\{1, \frac{L(\theta)}{L(\phi)}\right\}.
\end{aligned}$$

That is, **if the proposal distribution is taken as the prior distribution then the Metropolis-Hastings acceptance ratio is the ratio of the likelihood function.**

1. Code (write computer programme) the Metropolis algorithm for obtaining samples from (i) $N(5, 1.5^2)$, (ii) Gamma$(\alpha = 0.5, \beta = 1)$. Study the sensitivity of the algorithm with respect to the chosen proposal scaling.

   Solution:

   We only give the details for part (i). Here $\pi(x) \propto \exp\left\{-\frac{1}{2(1.5)^2}(x-5)^2\right\}$. We take $q(y|x) = N(x, \sigma^2)$, where σ^2 is called the proposal scaling. You can try different choices, e.g. 1, 2 etc. Now

$$\begin{aligned}
\alpha(x,y) &= \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} \\
&= \min\left\{1, \frac{\exp\left\{-\frac{1}{2(1.5)^2}(y-5)^2\right\}.}{\exp\left\{-\frac{1}{2(1.5)^2}(x-5)^2\right\}}\right\}
\end{aligned}$$

which can be further simplified. Work out the gamma example yourself. See Splus program on the web.

□

2. Suppose that $x_1, \ldots, x_n$ are i.i.d. observations from a Bernoulli distribution with mean θ. A logistic normal prior distribution is proposed for θ (a normal distribution for $\log \frac{\theta}{1-\theta}$). Show that if the prior mean and variance for $\log \frac{\theta}{1-\theta}$ are 0 and 1 respectively then the prior density function for θ is

$$\pi(\theta) \quad = \quad \frac{1}{\sqrt{2\pi}\theta(1-\theta)} \exp\left(-\tfrac{1}{2}\left(\log \tfrac{\theta}{1-\theta}\right)^2\right)$$

As this prior distribution is not conjugate, the Bayes estimator $E(\theta|x_1, \ldots, x_n)$ is not directly available. It is proposed to estimate it using a Monte Carlo sample generated by the Metropolis-Hastings method. One possible algorithm involves generating proposals from the prior distribution, independently of the current observation. Suppose $n = 10$, $\sum_{i=1}^{10} x_i = 8$. Write a Splus programme and run it to obtain the posterior mean.

Solution:

Let $\phi = \log \frac{\theta}{1-\theta}$. It is given that $\phi \sim N(0, 1)$. Therefore

$$\pi(\phi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\phi^2\right)$$

The question asks us to find the pdf of θ. We calculate the Jacobian

$$\frac{d\,\text{old}}{d\,\text{new}} = \frac{d\phi}{d\theta} = \frac{1}{\theta(1-\theta)}.$$

Therefore

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}\theta(1-\theta)} \exp\left(-\frac{1}{2}\left(\log\frac{\theta}{1-\theta}\right)^2\right),$$

if $0 < \theta < 1$.

Since $X_1, \ldots, X_n \sim \mathit{Bernoulli}(\theta)$ we have the likelihood

$$L(\theta) = \theta^t(1-\theta)^{n-t}$$

where $t = \sum x_i$.

Since the proposal distribution is the prior distribution the Metropolis-Hastings acceptance ratio is the ratio of the likelihood function, i.e.

$$\alpha(x, y) = \min\left\{1, \frac{y^t(1-y)^{n-t}}{x^t(1-x)^{n-t}}\right\}.$$

□

3. Assume that $X_1, X_2, \ldots, X_n$ are independent identically distributed $N(\theta, 1)$ observations. Suppose that the prior distribution for θ is Cauchy with density

$$\pi(\theta) \;\; = \frac{1}{\pi}\frac{1}{1+\theta^2} \quad -\infty < \theta < \infty.$$

Derive, upto a constant of proportionality, the posterior density of θ. Suppose that the importance sampling distribution is the prior distributions given above. Obtain the acceptance probability for the rejection method and the Metropolis-Hastings independence sampler. Suppose that $n = 10$ and $\bar{x} = 1.5$. Code the two methods and find the Bayes estimate for θ under squared error loss.

Solution:

Here

$$\begin{aligned} L(\theta) &= \tfrac{1}{(2\pi)^{n/2}} \exp\left(-\tfrac{1}{2}\sum(x_i-\theta)^2\right) \\ &= \exp\left(-\tfrac{n}{2}(\theta-\bar{x})^2\right), \end{aligned}$$

and the prior is

$$\pi(\theta) = \frac{1}{\pi}\frac{1}{1+\theta^2}.$$

Let us write $a = \bar{x}$ and $x = \theta$ then we have the posterior

$$\pi(x) \propto \exp\left(-\frac{n}{2}(x-a)^2\right) \times \frac{1}{\pi}\frac{1}{1+x^2}.$$

In the rejection method $g(x) = \frac{1}{\pi}\frac{1}{1+x^2}$. Therefore

$$\begin{aligned} M &= \sup_{-\infty<x<\infty} \tfrac{\pi(x)}{g(x)} \\ &= \sup_{-\infty<x<\infty} \exp\left(-\tfrac{n}{2}(x-a)^2\right) \\ &= 1, \end{aligned}$$

since the supremum is achieved at $x = a$. The acceptance probability of the rejection method is

$$\frac{1}{M}\frac{\pi(x)}{g(x)} = \exp\left[-\frac{n}{2}(x-a)^2\right].$$

Now we consider the Metropolis-Hastings algorithm. Since the proposal distribution is the prior distribution, the Metropolis-Hastings acceptance ratio is the ratio of the likelihood function, i.e.

$$\alpha(x, y) = \min\left\{1, \frac{\exp\left[-\frac{n}{2}(y-a)^2\right]}{\exp\left[-\frac{n}{2}(x-a)^2\right]}\right\}.$$

□

4. Assume that $X_1, X_2, \ldots, X_n$ are independent identically distributed $N(\theta, \sigma^2)$ observations. Suppose that the joint prior distribution for θ and σ^2 is

$$\pi(\theta, \sigma^2) \;\; = \frac{1}{\sigma^2}.$$

(a) Derive, upto a constant of proportionality, the joint posterior density of θ and σ^2.

(b) Derive the conditional posterior distributions of θ given σ^2 and σ^2 given θ.

(c) Derive the marginal posterior density of θ.

(d) Write a Splus programme for Gibbs sampling from the joint posterior distribution of θ and σ^2. Hence obtain the estimates of $E(\theta|x_1, \ldots, x_n)$ and $\text{Var}(\theta|x_1, \ldots, x_n)$. For your own data set verify that the estimates are close to the true values.

Solution:

Here

$$L(\theta, \sigma^2) \propto \frac{1}{(\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum (x_i - \theta)^2\right]$$

Therefore,

$$\begin{aligned}
\pi(\theta, \sigma^2|\mathbf{x}) &\propto \tfrac{1}{(\sigma^2)^{n/2}} \exp\left[-\tfrac{1}{2\sigma^2} \sum (x_i - \theta)^2\right] \tfrac{1}{\sigma^2} \\
&= \tfrac{1}{(\sigma^2)^{n/2+1}} \exp\left[-\tfrac{1}{2\sigma^2} \sum (x_i - \theta)^2\right]
\end{aligned}$$

Since

$$\sum (x_i - \theta)^2 = \sum (x_i - \bar{x})^2 + n(\theta - \bar{x})^2$$

we have

$$\theta|\sigma^2, \mathbf{x} \sim N(\bar{x}, \sigma^2/n).$$

Also

$$\sigma^2|\theta, \mathbf{x} \sim IG\left(m = n/2, \beta = \sum (x_i - \theta)^2/2\right),$$

IG denote the inverse gamma distribution.

$$\begin{aligned}
\pi(\theta|\mathbf{x}) &= \int_0^\infty \pi(\theta, \sigma^2|\mathbf{x}) d\sigma^2 \\
&\propto \int_0^\infty \tfrac{1}{(\sigma^2)^{n/2+1}} \exp\left[-\tfrac{1}{2\sigma^2} \sum (x_i - \theta)^2\right] d\sigma^2 \\
&= \tfrac{\Gamma(n/2)}{\left[\sum (x_i - \theta)^2/2\right]^{n/2}} \\
&\propto \left[\sum (x_i - \theta)^2\right]^{-n/2} \\
&= \left[\sum (x_i - \bar{x})^2 + n(\theta - \bar{x})^2\right]^{-n/2} \\
&\propto \left[1 + \tfrac{n(\theta - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right]^{-n/2} = \left[1 + \tfrac{t^2}{\alpha}\right]^{-\frac{\alpha+1}{2}},
\end{aligned}$$

where $\alpha = n - 1$ and

$$t^2 = \frac{n(n-1)(\theta - \bar{x})^2}{\sum (x_i - \bar{x})^2} = \frac{n(\theta - \bar{x})^2}{s^2},$$

and now

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2.$$

Clearly we see that

$$t = \frac{\theta - \bar{x}}{s/\sqrt{n}}$$

follows the Student t-distribution with $n - 1$ df.

Therefore $\theta|\mathbf{x} \sim t$-distribution with $n-1$ df and

$$E(\theta|\mathbf{x}) = \bar{x}$$

and

$$\text{var}(\theta|\mathbf{x}) = \frac{s^2}{n}\text{var}(t_{n-1}) = \frac{s^2}{n}\frac{n-1}{n-3}, \quad \text{if } n > 3.$$

$\square$