Applied Bayesian Methods

Solutions

1 (a) Let $C = 0$ be the event that a baby is born in a developing country, $C = 1$ be that a baby is born in a developed country.
Let $D = 0$ be the event that a baby survives the first four weeks, $D = 1$ be that a baby dies in the first four weeks.
Given that $P(C = 0 \mid D = 1) = 0.99$ and $P(C = 0) = 0.9$, it follows that

$$\begin{aligned}\frac{P(D = 1 \mid C = 0)}{P(D = 1 \mid C = 1)} &= \frac{P(C = 0 \mid D = 1)P(D = 1)}{P(C = 0)} \frac{P(C = 1)}{P(C = 1 \mid D = 1)P(D = 1)} \\ &= \frac{0.99}{0.9}\frac{0.1}{0.01} = 11 \ .\end{aligned}$$

(b) Let $T_1 = 1$ be the event that a woman tests positive in the first test, $T_1 = 0$ be test-negative; similarly T_2 is used for the second test.
Let $W = 1$ be the event that a woman is pregnant, $W = 0$ be non-pregnant.
So we have the following conditions:
$P(T_1 = 1 \mid W = 1) = \theta$, $P(T_1 = 1 \mid W = 0) = \eta$,
$P(T_2 = 1 \mid W = 1) = \theta$, $P(T_2 = 1 \mid W = 0) = \eta$, and
$P(W = 1) = \lambda$.
It follows that the predictive probability $P(T_2 = 0 \mid T_1 = 1)$ is

$$\begin{aligned} & P(T_2 = 0 \mid T_1 = 1) \\ = \ & P(T_2 = 0, W = 1 \mid T_1 = 1) + P(T_2 = 0, W = 0 \mid T_1 = 1) \\ = \ & P(T_2 = 0 \mid W = 1)P(W = 1 \mid T_1 = 1) + P(T_2 = 0 \mid W = 0)P(W = 0 \mid T_1 = 1) \ .\end{aligned}$$

Because

$$\begin{aligned} & P(W = 1 \mid T_1 = 1) \\ = \ & \frac{P(T_1 = 1 \mid W = 1)P(W = 1)}{P(T_1 = 1)} \\ = \ & \frac{P(T_1 = 1 \mid W = 1)P(W = 1)}{P(T_1 = 1 \mid W = 1)P(W = 1) + P(T_1 = 1 \mid W = 0)P(W = 0)} \\ = \ & \frac{\theta\lambda}{\theta\lambda + \eta(1 - \lambda)} \ ,\end{aligned}$$

we have

$$P(T_2 = 0 \mid T_1 = 1)$$

$$
\begin{aligned}
&= (1-\theta)\frac{\theta\lambda}{\theta\lambda+\eta(1-\lambda)} + (1-\eta)\left\{1-\frac{\theta\lambda}{\theta\lambda+\eta(1-\lambda)}\right\} \\
&= \frac{(1-\theta)\theta\lambda+(1-\eta)\eta(1-\lambda)}{\theta\lambda+\eta(1-\lambda)} \ .
\end{aligned}
$$

(c) Under zero-one loss, an optimal estimator $T(y)$ of θ is the posterior mode of θ for which $p(\theta|y)$ equals its maximum.

Let us assume a loss function $L(\theta, T(y))$ as:

$$
L(\theta, T(y)) = \begin{cases} 0 & \text{if } |\theta - T(y)| \le \epsilon \ , \\ 1 & \text{if } |\theta - T(y)| > \epsilon \ . \end{cases}
$$

with ϵ very small. Hence the posterior expected loss $E_{\theta|y}[L(\theta, T(y))]$ is

$$
\begin{aligned}
& \int_{-\infty}^{\infty} L(\theta, T(y))\, p(\theta \mid y) d\theta \\
=\ & \int_{-\infty}^{T(y)-\epsilon} 1 \cdot p(\theta \mid y) d\theta + \int_{T(y)+\epsilon}^{\infty} 1 \cdot p(\theta \mid y) d\theta \\
=\ & P(\theta < T(y) - \epsilon \mid y) + P(\theta > T(y) + \epsilon \mid y) \\
=\ & 1 - P(T(y) - \epsilon \le \theta \le T(y) + \epsilon \mid y) \ ,
\end{aligned}
$$

which is minimised when $P(T(y) - \epsilon \le \theta \le T(y) + \epsilon \mid y)$ is maximised, i.e. when $T(y)$ is the posterior mode of θ.

(d) Jeffreys' prior for success probability θ of $\text{NegBin}(r, \theta)$ can be derived as:

$$
\begin{aligned}
p(y|\theta) &= \binom{y+r-1}{y} \theta^r (1-\theta)^y \\
\log p(y|\theta) &= r\log\theta + y\log(1-\theta) + \text{const} \\
\frac{d}{d\theta}\log p(y|\theta) &= \frac{r}{\theta} - \frac{y}{1-\theta} \\
\frac{d^2}{d\theta^2}\log p(y|\theta) &= -\frac{r}{\theta^2} - \frac{y}{(1-\theta)^2} \\
I(\theta) &= -E_{Y|\theta}\left[-\frac{r}{\theta^2} - \frac{Y}{(1-\theta)^2}\right] \\
&= \frac{r}{\theta^2} + \frac{E(Y)}{(1-\theta)^2} \\
&= \frac{r}{\theta^2} + \frac{r(1-\theta)/\theta}{(1-\theta)^2} = \frac{r}{\theta^2} + \frac{r}{(1-\theta)\theta} \\
&= \frac{r(1-\theta)+r\theta}{\theta^2(1-\theta)} = \frac{r}{\theta^2(1-\theta)} \\
I(\theta)^{\frac{1}{2}} &= \left(\frac{r}{\theta^2(1-\theta)}\right)^{\frac{1}{2}} \propto \theta^{-1}(1-\theta)^{-\frac{1}{2}} \ .
\end{aligned}
$$

So, Jeffreys' prior for θ is $\text{Beta}\left(0, \frac{1}{2}\right)$.

(e) There are at least two ways for the students to solve this problem: one is via factorising the full-conditional distribution, and the other is via simple Bayesian inference for Normal distributions with only a single parameter unknown. The answers are:

$$\begin{aligned}
\mu|\tau,\mathbf{y} &\sim \text{Normal}\left(\frac{n\tau}{n\tau+10^{-6}}\bar{y},\ (n\tau+10^{-6})^{-1}\right) ,\\
\tau|\mu,\mathbf{y} &\sim \text{Gamma}\left(0.001+\frac{n}{2},\ 0.001+\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2\right) .
\end{aligned}$$

2 (a) Here, an acceptable hierarchical Bayesian model contains

(i) an appropriate distribution for the data Y_i with mean $t_i\theta_i$, where Y_i is a discrete random variable with values of $0, 1, 2, \ldots$; e.g. $Y_i|\theta_i \sim \text{Poisson}(t_i\theta_i)$ (credits to any reasonable alternatives);

(ii) an appropriate prior distribution for the parameter θ_i, where θ_i is positive and can be larger than 1; e.g. $\theta_i \sim \text{Gamma}(\alpha, \beta)$ (credits to any reasonable alternatives); and

(iii) appropriate distributions for hyper-parameters; for example, hyper-priors for the positive hyper-parameters α and β in $\text{Gamma}(\alpha, \beta)$ (credits to any reasonable alternatives).

(b) Choose M such that the chain has reached the highest posterior region of the stationary distribution. Choose N such that the sample is adequate for required estimates of posterior summaries of interest (mention thinning and possible storage problems). Sort the data in the sample in ascending order and find the 5%, 50% and 95% quantiles (denoted by $\theta_{0.05}$, $\theta_{0.5}$ and $\theta_{0.95}$, respectively). The posterior median is $\theta_{0.5}$; the 90% posterior credible interval is $[\theta_{0.05}, \theta_{0.95}]$.

The highest posterior density (HPD) interval of θ_6 is an interval such that the posterior density $p(\theta_6|\mathbf{y})$ at any point inside the interval is greater than that at any point outside the interval. One way to estimate the 50% HPD interval of θ_6 is to first build a binned histogram of $\theta_6|\mathbf{y}$ from the sample (or equivalently count the frequencies of binned values of θ_6 from the sample), then add the bins of highest frequencies into the interval until half of the sample have been included. There can be other ways; credits to any reasonable suggestions.

(c) Taking into account the autocorrelation in the sample of θ_6, a 'batching' method for the MCSE of the posterior mean of θ_6 can be described by the following steps:

(i) Divide the sequence $\theta_6^{(M+1)}, \ldots, \theta_6^{(N)}$ into Q batches, $b_1, \ldots, b_Q$; each batch is of a sufficiently large length L.

(ii) Calculate $\mu_q = \frac{1}{L}\sum_{i \in b_q} \theta_6^{(i)}$ for $q = 1, \ldots, Q$, and $\bar{\mu} = \frac{1}{Q}\sum_{q=1}^{Q} \mu_q$.

(iii) Estimate the required MCSE by

$$\widehat{\text{SE}}(\bar{\theta}_6) = \sqrt{\frac{1}{Q(Q-1)}\sum_{q=1}^{Q}(\mu_q - \bar{\mu})^2}\ .$$

(d) In order to use a Gibbs sampler to obtain samples from the predictive distribution $p(\tilde{y}|\mathbf{y})$, we recall that $p(\tilde{y}|\mathbf{y}) = \int p(\tilde{y}, \eta|\mathbf{y})d\eta$, where η includes all the unknown parameters (θ_i, $\tilde{\theta}$ and all hyper-parameters) and $\mathbf{y}$ includes all the observed values (y_i, t_i and $\tilde{t}$), in which $\tilde{\theta}$ is the unknown parameter corresponding to $\tilde{y}$, and $\tilde{t} = 50$.

Essentially, we can add $\tilde{y}$ and $\tilde{\theta}$ into the original hierarchical Bayesian model, treat $\tilde{y}$ and $\tilde{\theta}$ as two new unknown parameters, derive the full conditional distributions for them, and update the Gibbs sampler. Then we can use the updated Gibbs sampler to obtain samples for each unknown 'parameter' from its posterior distribution, including those for $\tilde{y}$ from $p(\tilde{y}|\mathbf{y})$.

End of Paper