QUEEN MARY, UNIVERSITY OF LONDON

MTH6102: Bayesian Statistical Methods

Solutions of exercise sheet 10

2023-2024

This exercise sheet 10 is assessed and counts for 4% of the module total. The deadline for submission is **Monday the 11th December at 11am**.

Submit the R code used as an R script file (with extension .R). But you need to write the answers in a separate file. This can be a Word document, pdf or a clearly legible image of hand-written work. So you need to submit two files.

1. **50 marks.** Let the observed data be $y = (6, 4, 9, 2, 0, 3)$, a random sample from the Poisson distribution with mean λ, where $\lambda > 0$ is unknown. Suppose that we assume a Gamma$(1, 1)$ prior distribution for λ. The posterior density, $p(\lambda \mid y)$, for λ is Gamma$(1+S, 1+n)$, where $S = \sum_{i=1}^{6} y_i$ and $n = 6$. Suppose that you want to construct a symmetric Metropolis-Hastings on the log-scale to generate a sample from this posterior distribution by using a normal proposal distribution with standard deviation $b = 0.2$.

   (a) Write down the steps in this symmetric Metropolis-Hastings (on the log-scale) to simulate realisations from the posterior density $p(\lambda \mid y)$.

   (b) Implement the algorithm in R and plot the observations as a function of the iterations. Use $M = 5000$ for the number of iterations.

   (c) To assess the accuracy compare the empirical distribution of the sample with the exact posterior density, Gamma$(1 + S, 1 + n)$.

   (d) Rerun the algorithm in R using a smaller $b = 0.01$ and a larger $b = 20$. What are the effects on the behaviour of the algorithm of making b smaller? What are the effects of making it larger?

   (e) Add code to count how many times the proposed value for λ was accepted. Rerun the algorithm using values of $b = 0.01$, $b = 0.2$ and $b = 20$, and each time calculate the proportion of steps that were accepted. Then plot this acceptance probability against b. Examine how the acceptance probability for this algorithm depends b.

   Solution:

   (a) The observed data is $y = (y_1, \ldots, y_n) = (6, 4, 9, 2, 0, 3)$ where $n = 6$, and each $y_i \sim$ Poisson(λ) with pmf

   $$p(y_i \mid \lambda) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}.$$

   The log likelihood is

   $$\log p(y \mid \lambda) = \log \prod_{i=1}^{6} p(y_i \mid \lambda) = \sum_{i=1}^{6} \log p(y_i \mid \lambda).$$

The prior for λ, $p(\lambda)$, is Gamma(1, 1) with pdf $p(\lambda) = e^{-\lambda}$, $\lambda > 0$.
Define

$$\mathcal{L}(\lambda) = \log\left(p(\lambda)\, p(y \mid \lambda)\right) = \log\left(p(\lambda)\right) + \log\left(p(y \mid \lambda)\right),$$

the log of the posterior density (up to a constant).

Start with λ_1 randomly. For each $i > 1$:

i. Generate $\psi \sim N(\lambda_{i-1}, b^2)$, for $b = 0.2$.

ii. Compute the probability of acceptance

$$\delta = \min\left(0, \mathcal{L}(\psi) - \mathcal{L}(\lambda_{i-1})\right).$$

iii. Generate $U \sim U[0, 1]$. Set

$$\lambda_i = \begin{cases} \psi, & \text{if } \log(U) < \delta \\ \lambda_{i-1}, & \text{otherwise} \end{cases}$$

(b) The chain moves up and down quickly through the parameter space.

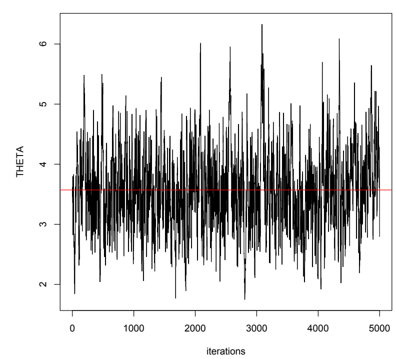

Figure 1: Sample paths when $b = 0.2$

(c) From Figure 2 the empirical distribution of the simulated values is very close to the true posterior distribution for λ

(d) In this question we examine the choices $b = 0.01$ and $b = 20$. Figure 3 shows the sample paths of a single run of the corresponding symmetric Metropolis-Hastings under two different proposal standard deviations b. Table 1 shows the probability of acceptance. Choosing b too small yields, a very high probability of acceptance, however at the price of a chain that is hardly moving. Choosing b too large allows the chain to move fast and make large jumps, however, most of the proposed values are rejected, so the chain remains for a long time at each accepted values. We also see that $b = 0.2$ is not the optimal choice as the probability of acceptance is very high (0.80). Theoretically, it has been shown that the optimal acceptance rate is around 0.234-(an asymptotic result). But experience suggests that an acceptance rate of around 20%-30%. Thus, the standard deviation b should be tuned to get an acceptance rate of around this level.

(e)

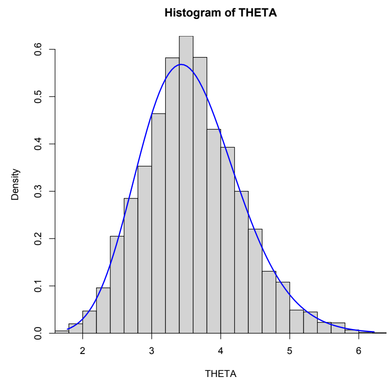

Figure 2: Empirical distribution of the chain vs true Gamma posterior distribution when $b = 0.2$

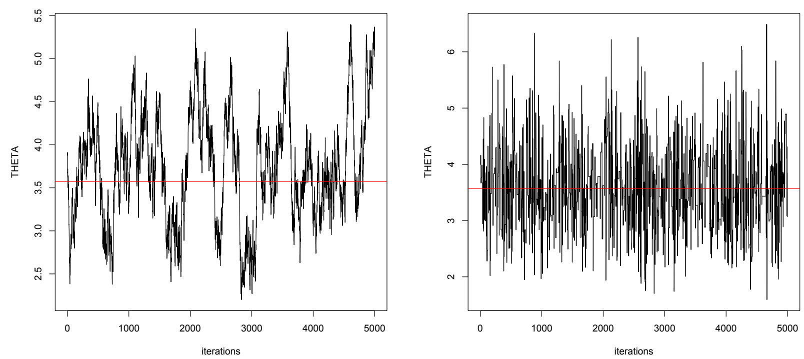

Figure 3: Markov chains under two different proposal standard deviations b. Going from left to right, the values of b are 0.01 and 20, respectively.

	Probability of acceptance
$b = 0.01$	0.9448
$b = 0.2$	0.7972
$b = 20$	0.1856

2. **50 marks.** Let $y_1, \ldots, y_n$ be a sample from a Poisson distribution with mean λ, where λ is given a Gamma(α, β) prior distribution.

   (a) It is observed that $y_1 = y_2 = \cdots = y_n = 0$, and we take $\alpha = 1, \beta = 1$.

      i. What is the posterior distribution for λ?

      ii. What is the posterior mean?

      iii. What is the posterior median and an equal tail 95% credible interval for λ (without using R)?

   (b) Show that if a new data-point x is generated from the same Poisson distribution,

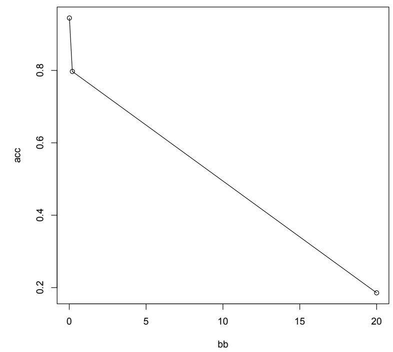

Figure 4: Plot of acceptance probability as a function of the proposal standard deviation b.

the posterior predictive probability that $x = 0$ is

$$p(x = 0 \mid y) = \frac{n+1}{n+2}.$$

(c) Now suppose that we have general $y_1, \ldots, y_n$, α and β; and that again x is a new data-point from the same Poisson distribution.

i. Find the mean and variance of x.

ii. Derive the full posterior predictive distribution for x.

Solution:

(a) For general $y_1, \ldots, y_n$, α and β, the posterior distribution is

$$\lambda \sim \text{Gamma}(S + \alpha, n + \beta), \; S = \sum_{i=1}^{n} y_i.$$

i. If $S = 0$ and $\alpha = 1, \beta = 1$, the posterior distribution is $\text{Gamma}(1, n+1)$ with pdf

$$p(\lambda \mid y) = (n+1)e^{-(n+1)\lambda}, \; \lambda \geq 0.$$

ii. The posterior mean is

$$E(\lambda \mid y) = \frac{1}{n+1}.$$

iii. The posterior pdf is

$$p(\lambda \mid y) = (n+1)e^{-(n+1)\lambda}, \; \lambda \geq 0.$$

Integrating this gives cdf

$$F(\lambda) = \int_0^{\lambda} (n+1)e^{-(n+1)\lambda'} \, d\lambda' = 1 - e^{-(n+1)\lambda}.$$

The inverse cdf function (quantile function) is $Q(u)$, found by setting $F(\lambda) = u$.

$$1 - e^{-(n+1)\lambda} = u$$

$$\begin{aligned}
e^{-(n+1)\lambda} &= 1-u \\
-(n+1)\lambda &= \log(1-u) \\
Q(u) = \lambda &= -\frac{\log(1-u)}{n+1}
\end{aligned}$$

The posterior median is $Q(0.5) = -\dfrac{\log(0.5)}{n+1}$.
The limits for a 95% credible interval are given by

$$(Q(0.025), Q(0.975)) = \left(-\frac{\log(0.975)}{n+1}, -\frac{\log(0.025)}{n+1}\right).$$

(b) For a given value of λ, the probability that a new data-point x is zero is the Poisson probability mass function

$$P(x=0 \mid \lambda) = e^{-\lambda}.$$

The posterior predictive probability that x is zero is

$$\begin{aligned}
P(x=0 \mid y) &= \int_0^\infty P(x=0 \mid \lambda)\, p(\lambda \mid y)\, d\lambda \\
&= \int_0^\infty e^{-\lambda}\, (n+1)\, e^{-(n+1)\lambda}\, d\lambda \\
&= \int_0^\infty (n+1)\, e^{-(n+2)\lambda}\, d\lambda \\
&= \frac{n+1}{n+2}.
\end{aligned}$$

(c) The general posterior distribution for λ is Gamma(a, b), where $a = S + \alpha$ and $b = n + \beta$.

i. For a given value of λ, if a new data-point $x \sim$ Poisson(λ), independently of y, then

$$E(x \mid \lambda, y) = E(x \mid \lambda) = \lambda,\ \mathrm{Var}(x \mid \lambda, y) = \mathrm{Var}(x \mid \lambda) = \lambda.$$

Putting these together and by the law of iterated expectation, the predictive mean for x is

$$E(E(x \mid \lambda, y)) = E(\lambda) = \frac{\alpha}{\beta},$$

So $E(x) = \frac{\alpha}{\beta}$. By the law of total variance,

$$E(\mathrm{Var}(x \mid \lambda, y)) + \mathrm{Var}(E(x \mid \lambda, y)) = E(\lambda) + \mathrm{Var}(\lambda) = \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2}.$$

so $\mathrm{var}(x) = \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2}$

ii. The posterior predictive distribution is

$$\begin{aligned}
p(x \mid y) &= \int_0^\infty p(x \mid \lambda)\, p(\lambda \mid y)\, d\lambda \\
&= \int_0^\infty \frac{\lambda^x e^{-\lambda}}{x!}\, \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)}\, d\lambda \\
&= \frac{b^a}{x!\Gamma(a)} \int_0^\infty \lambda^{a+x-1} e^{-(b+1)\lambda}\, d\lambda
\end{aligned}$$

For a Gamma(a, b) density, we know

$$1 = \int_0^\infty \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} d\theta, \quad a > 0, b > 0.$$

This means

$$\int_0^\infty \theta^{a-1} e^{-b\theta} d\theta = \frac{\Gamma(a)}{b^a}, \quad a > 0, b > 0.$$

Now substitute $a + x = \alpha + S + x$ instead of a and $1 + b = \beta + n + 1$ instead of b to get

$$\int_0^\infty \lambda^{a+x-1} e^{-(b+1)\lambda} \, d\lambda = \frac{\Gamma(\alpha + S + x)}{(\beta + 1 + n)^{\alpha+S+x}}.$$

Then,

$$p(x \mid y) = \frac{(\beta + n)^{S+\alpha}}{x!\Gamma(S + \alpha)} \frac{\Gamma(\alpha + S + x)}{(\beta + 1 + n)^{\alpha+S+x}}.$$

iii. R code to check the part (a) is on QMPlus.

```
> alpha = 1
> beta = 1
> S = 0
> n = 11
> a = S + alpha
> b = n + beta
>
> qgamma(0.5, shape=a, rate=b)
[1] 0.05776227
> -log(0.5)/(n+1)
[1] 0.05776227
>
> qgamma(0.025, shape=a, rate=b)
[1] 0.002109817
> qgamma(0.975, shape=a, rate=b)
[1] 0.3074066
> -log(0.975)/(n+1)
[1] 0.002109817
> -log(0.025)/(n+1)
[1] 0.3074066
```