# Problem 1

## Part (a)

We want to evalutate $\int_0^1 x^2 dx$ through Monte Carlo technique. Here we sample $n$ independent uniform(0,1) random variables $X_1, \ldots, X_n$. Then the integral is estimated by

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

And the estimated variance of $\hat{\mu}$ is estimated by

$$\hat{Var}(\hat{\mu}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (X_i^2 - \hat{\mu})^2.$$

In my implementation, $n = 100000$, $\hat{\mu} = 0.3339044$ and standard error $\sqrt{\hat{Var}(\hat{\mu})} = 0.000943176$. Note that the exact value of the integral is 1/3 and we can see that the exact value lies within two standard error of the estimated integral.

## Part (b)

First of all,

$$\int_0^1 \int_{-2}^2 x^2 \cos(xy) dx dy = \int_0^1 \int_{-2}^2 4x^2 \cos(xy) f(x,y) dx dy$$

where $f(x,y) = 1/4$ for $-2 < x < 2$ and $0 < y < 1$. Note that $f$ is the joint probability density function of independent Uniform(-2,2) and Uniform(0,1). To estimate the integral, we we sample $n$ independent random vectors $(X_1, Y_1), \ldots, (X_n, Y_n)$ from $f$. Then the integral is estimated by

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \left\{ 4X_i^2 \cos(X_i Y_i) \right\}.$$

And the estimated variance of $\hat{\mu}$ is estimated by

$$\hat{Var}(\hat{\mu}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left\{ 4X_i^2 \cos(X_i Y_i) - \hat{\mu} \right\}^2.$$

In my implementation, $n = 100000$, $\hat{\mu} = 3.480555$ and standard error $\sqrt{\hat{Var}(\hat{\mu})} = 0.01128749$. Note that the exact value of the integral is $2\sin(2) - 4\cos(2) = 3.483182$ and we can see that the exact value lies within two standard error of the estimated integral.

## Part (c)

Note that the integral is from zero to infinity. Thus we have to sample from distribution with domain including at least the whole positive real line. One natural choice is exponential distribution.

$$\int_0^\infty \frac{3}{4} x^4 e^{-x^3/4} dx = \int_0^\infty \frac{3}{4} x^4 e^{-x^3/4 + x} e^{-x} dx$$

Here we sample $n$ independent Exp(1) random variables $X_1, \ldots, X_n$. Then the integral is estimated by

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \left( X_i^4 e^{-X_i^3/4 + X_i} \right).$$

And the estimated variance of $\hat{\mu}$ is estimated by

$$\hat{Var}(\hat{\mu}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (X_i^4 e^{-X_i^3/4 + X_i} - \hat{\mu})^2.$$

In my implementation, $n = 100000$, $\hat{\mu} = 2.293257$ and standard error $\sqrt{\hat{Var}(\hat{\mu})} = 0.01152466$. Note that the exact value of the integral is $(4 \times 2^{1/3}\Gamma(2/3))/3 = 2.274776$ and we can see that the exact value lies within two standard error of the estimated integral.

Here, instead of exponential random variable, we can also use Gamma random variable with scale 1 and shape 5. And now

$$\hat{\mu} = \frac{3\Gamma(5)}{2^7} \exp(-x^3/4 + 2x)$$

If $n = 100000$, we have $\hat{\mu} = 2.273836$ and standard error $\sqrt{\hat{Var}(\hat{\mu})} = 0.006031039$ in my implementation.

## Problem 2

Note that

$$\begin{aligned} I &= \frac{1}{\sqrt{2\pi}} \int_1^2 e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} f(x) dx \end{aligned}$$

where $f$ is the pdf of Uniform(1,2). As a result,

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \frac{f(x)}{g(x)} g(x) dx$$

where $g$ is the pdf of $N(1.5, \nu^2)$. Here,

$$w(x) = \frac{f(x)}{g(x)} = \begin{cases} 1/g(x) & \text{if } 1 < x < 2 \\ 0 & \text{otherwise.} \end{cases}$$

Thus denote

$$m(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{f(x)}{g(x)} = \begin{cases} \nu \exp(-\frac{x^2}{2} + \frac{(x-1.5)^2}{2\nu^2}) & \text{if } 1 < x < 2 \\ 0 & \text{otherwise.} \end{cases}$$

To estimate $I$ through importance sampling, we first sample $n$ independent random variables $X_1, \ldots, X_n$ from $g$. Then the integral is estimated by

$$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} m(X_i).$$

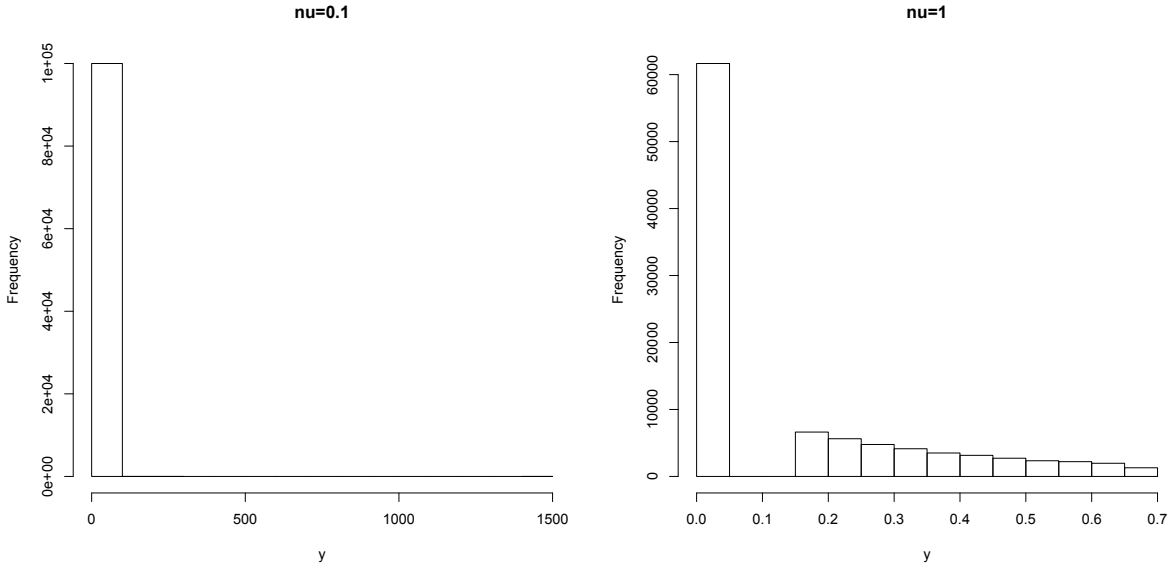
And the estimated variance of $\hat{I}$ is estimated by

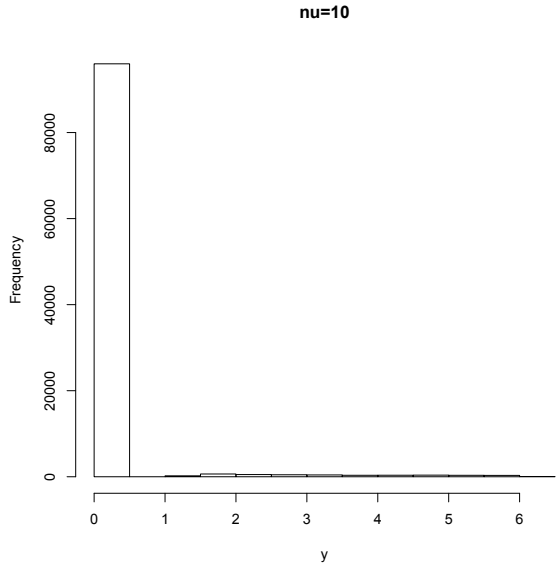
$$\hat{Var}(\hat{I}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (m(X_i) - \hat{\mu})^2.$$

In all of my implementations, $n = 100000$.

- $\nu = 0.1$: $\hat{I} = 0.1239123$ and standard error $\sqrt{\hat{Var}(\hat{I})} = 0.01566348$.
- $\nu = 1$: $\hat{I} = 0.1360533$ and standard error $\sqrt{\hat{Var}(\hat{I})} = 0.0006187184$.
- $\nu = 10$: $\hat{I} = 0.1367735$ and standard error $\sqrt{\hat{Var}(\hat{I})} = 0.002289549$.

And the histogram of $m(X_i)$'s are shown as follows.






Here, you can see that there are outliers for the case $\nu = 0.1$ and $\nu = 10$. Also, the outliers seem to be more extreme in the case $\nu = 0.1$. You can see that the frequency of the first bar is around $1 \times 10^5$. That means most of the simulated values lie there and it can have a few very large value, with the largest around 1500. Actually, this phenomenon is reflected in their corresponding standard errors. And thus $g$ with $\nu = 1$ gives the best estimate among the three since its has the smallest standard error.

# Problem 3

## Part (a)

The estimated value is 0.6951.

## Part (b)

$E(1 + U) = 1 + 0.5 = 1.5$. Note that $Var[c(U)] = Var(1 + U) = 1/12$. and $Cov[h(U), c(U)] = 1 - (3/2)\ln 2$. Thus $b^* = -0.4767$.

## Part (d)

Here, the absolute value of the correlation between $h(U)$ and $c(U)$ is the key. We would want it to be large. $\sqrt{U}$ is one of the choice that has larger correlation than the $1 + U$.

# Problem 4

## Part (a)

Suppose $e_{ij}$ have independent and identical double exponential distributions with pdf

$$f(x) = \frac{1}{2\theta} \exp\left(-\frac{|x|}{\theta}\right), \quad -\infty < x < \infty,$$

where $\theta > 0$.

Here:

- The null hypothesis is $H_0 : \alpha_i = 0$, $\forall i$ and $\theta > 0$.
- The alternative hypothesis is $H_1 : \alpha_i \neq 0$ for at least one $i$ and $\theta > 0$.

Suppose there are $k$ groups and we have $n_i$ observations from the $i$th group. Also, let $\mu_i = \mu + \alpha_i$. In order to derive a test statistic, we use the idea of likelihood ratio test and form the likelihood ratio as

$$\begin{aligned}\lambda &= \frac{\sup_{\mu_i, i=1,\ldots,k \text{ and } \theta} (2\theta)^{-\sum_{i=1}^k n_i} \exp\{-\sum_{i=1}^k \sum_{j=1}^{n_i} |y_{ij} - \mu_i|/\theta\}}{\sup_{\mu,\theta} (2\theta)^{-\sum_{i=1}^k n_i} \exp\{-\sum_{i=1}^k \sum_{j=1}^{n_i} |y_{ij} - \mu|/\theta\}} \\ &= \left(\frac{\hat{\theta}_0}{\hat{\theta}}\right)^{\sum_{i=1}^k n_i}\end{aligned}$$

where

$$\hat{\theta} = \frac{1}{\sum_{i=1}^k n_i} \sum_{i=1}^k \sum_{j=1}^{n_i} |y_{ij} - m_i| \qquad \text{and} \qquad \hat{\theta}_0 = \frac{1}{\sum_{i=1}^k n_i} \sum_{i=1}^k \sum_{j=1}^{n_i} |y_{ij} - m|.$$

Here, $m_i$ is the median of $\{y_{i1}, \ldots, y_{in_i}\}$ and $m$ is the median of $\{y_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, k\}$. And thus

$$\lambda = \left(\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} |y_{ij} - m|}{\sum_{i=1}^k \sum_{j=1}^{n_i} |y_{ij} - m_i|}\right)^{\sum_{i=1}^k n_i} \tag{1}$$

Note that although we know the asymptotic null distribution of $\log \lambda$, (Note that the way I define $\lambda$ may be different from the usual way) we may still want to perform an exact test for the case of finite sample size. Here, if $\lambda$ is large, we reject $H_0$. Note that under $H_0$,

$$\lambda = \left(\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} |e_{ij} - m'|}{\sum_{i=1}^k \sum_{j=1}^{n_i} |e_{ij} - m'_i|}\right)^{\sum_{i=1}^k n_i} \tag{2}$$

where $m_i$ is the median of $\{e_{i1}, \ldots, e_{in_i}\}$ and $m$ is the median of $\{e_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, k\}$. To proceed, we may want to perform the Monte Carlo test and the idea bases on that we can simulate $\lambda$ under the $H_0$. However, difficulties arise since the null hypothesis does not specify only one

distribution, but a set of distribution characterized by $\theta$. Now, we make use of the scale invariance property of $\lambda$ under $H_0$. Under $H_0$, one may see that the value of $\lambda$ is

$$\lambda = \left( \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} |e_{ij}/\theta - m''|}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} |e_{ij}/\theta - m_i''|} \right)^{\sum_{i=1}^{k} n_i}$$

where $m_i$ is the median of $\{e_{i1}/\theta, \ldots, e_{in_i}/\theta\}$ and $m$ is the median of $\{e_{ij}/\theta, j = 1, \ldots, n_i, i = 1, \ldots, k\}$. And $e_{ij}/\theta$ has pdf (does not depend on $\theta$):

$$f(x) = \frac{1}{2} \exp\left(-|x|\right), \quad -\infty < x < \infty. \tag{3}$$

Thus $\lambda$ is independent of $\theta$ under $H_0$.

Thus, we can perform a Monte Carlo test as follows.

1. Sample $e_{11}, \ldots, e_{1n_1}, \ldots, e_{k1}, \ldots, e_{kn_k}$ independently from double exponential distribution (3).
2. Compute the $\lambda$ by (2).
3. Repeat Step 1 and 2, say, 999 times.
4. If the $\lambda$ computed from sample through (1) is amongst the largest $\alpha\%$ of these $\lambda$'s (from Step 1 to 3), we reject $H_0$. ($\alpha\%$ is the significance of the test.)

### Part (b)

One of the tests that we could use is the permutation test (or randomized test), which is introduced by Fisher. The tests can be performed in the following way:

1. First, merge $y_{11}, \ldots, y_{1n_1}, \ldots, y_{k1}, \ldots, y_{kn_k}$ to form a sample of $n_1 + \cdots + n_k$ data points.
2. Draw without replacement $n_i$ points to group $i$, for $i = 1, \ldots, k-1$ (such that no duplicated points within and between groups). And the remaining $n_k$ points to form group $k$.
3. Compute the test statistic $T = \sum_{i>j} |\hat{\mu}_i - \hat{\mu}_j|$, where $\hat{\mu}_i$ is the group mean of group $i$.
4. Repeat Step 1 to Step 3 for, say, 999 times.
5. If $T$ from the sample is amongst the largest $\alpha\%$ of these $T$'s (from Step 1 to 4), we reject $H_0$. ($\alpha\%$ is the significance of the test.)

## Problem 5

### Part (a)

The generated sample is as follows.

```
[1] 0 0 0 0 0 0 1 0 0 0 0 0 5 0 1 0 4 1 0 3 0 0 6 4 2 0 0 0 0 1 0 2 0 0 4 0 4 2 0 0 0
[42] 0 0 0 0 0 0 0 5 0 2 0 0 0 1 0 0 0 1 0 0 2 0 1 0 0 0 0 3 0 3 0 3 0 1 0 0 2 0 0 0 0
[83] 2 0 5 0 0 0 1 0 2 0 0 4 3 0 0 3 0 0
```

## Part (b)

1. $(\lambda|p, \boldsymbol{r}, \boldsymbol{x})$:

$$\begin{aligned} f(\lambda|p, \boldsymbol{r}, \boldsymbol{x}) &\propto f(\boldsymbol{x}, \boldsymbol{r}, \lambda, p) \\ &= \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \prod_{i=1}^{n} \frac{e^{-\lambda r_i} (\lambda r_i)^{x_i}}{x_i!} p^{r_i} (1-p)^{1-r_i} \\ &\propto \lambda^{a-1} e^{-b\lambda} e^{-\lambda \sum_{i=1}^n r_i} \lambda^{\sum_{i=1}^n x_i} \\ &= \lambda^{a-1+\sum_{i=1}^n x_i} e^{-(b+\sum_{i=1}^n r_i)\lambda} \end{aligned}$$

   Thus $(\lambda|p, \boldsymbol{r}, \boldsymbol{x}) \sim \text{Gamma}(a + \sum_i x_i, b + \sum_i r_i)$.

2. $(p|\lambda, \boldsymbol{r}, \boldsymbol{x})$:

$$\begin{aligned} f(p|\lambda, \boldsymbol{r}, \boldsymbol{x}) &\propto f(\boldsymbol{x}, \boldsymbol{r}, \lambda, p) \\ &= \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \prod_{i=1}^{n} \frac{e^{-\lambda r_i} (\lambda r_i)^{x_i}}{x_i!} p^{r_i} (1-p)^{1-r_i} \\ &\propto p^{\sum_{i=1}^n r_i} (1-p)^{n-\sum_{i=1}^n r_i} \end{aligned}$$

   Thus $(p|\lambda, \boldsymbol{r}, \boldsymbol{x}) \sim \text{Beta}(1 + \sum_i r_i, n + 1 - \sum_i r_i)$.

3. $(r_i|\lambda, p, \boldsymbol{x})$:

$$\begin{aligned} f(r_i|\lambda, \boldsymbol{r}, \boldsymbol{x}) &\propto f(\boldsymbol{x}, \boldsymbol{r}, \lambda, p) \\ &= \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \prod_{j=1}^{n} \frac{e^{-\lambda r_j} (\lambda r_j)^{x_j}}{x_j!} p^{r_j} (1-p)^{1-r_j} \end{aligned}$$

   Note that $r_i$ takes two values 0 and 1. Note that $r_i$ is 0, $x_i$ is 0 with probability 1. Thus, $f(0|\lambda, \boldsymbol{r}, \boldsymbol{x}) \propto (1-p) I_{\{x_i=0\}}$. And $f(1|\lambda, \boldsymbol{r}, \boldsymbol{x}) \propto p e^{-\lambda}$. As a result, $(r_i|\lambda, p, \boldsymbol{x}) \sim \text{Bernoulli}\left(\frac{p e^{-\lambda}}{p e^{-\lambda} + (1-p) I_{\{x_i=0\}}}\right)$.

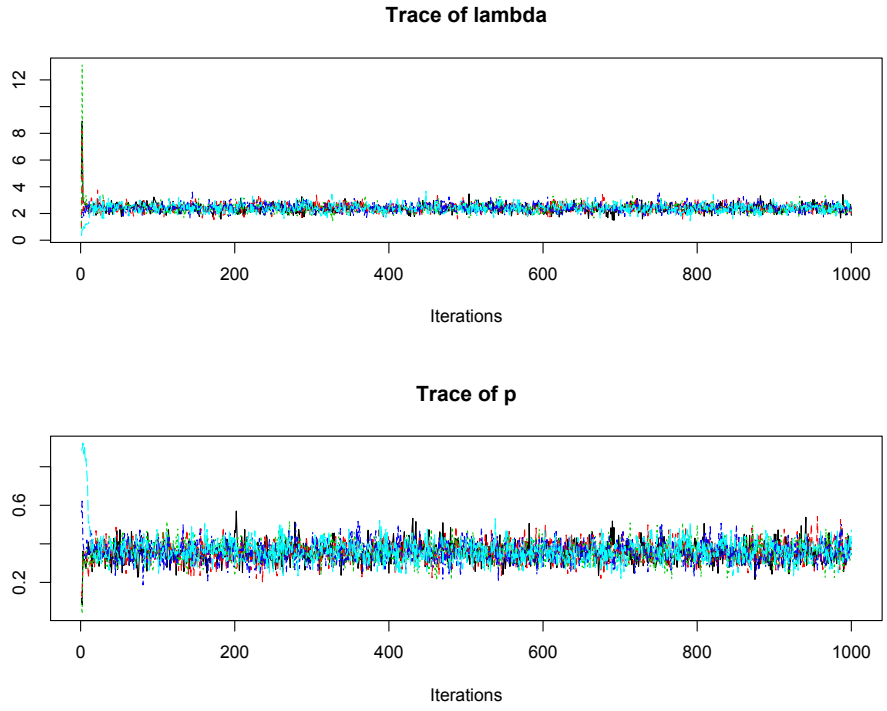
## Part (c)

Here, we focus on the posterior distribution of $\lambda$ and $p$. In order to use the sample generated by the Gibbs sampler, we have to check the convergence of the sample after discarding the burn-in. In order to illustrate the technique, the case of $a = 1$ and $b = 1$ will be discussed in details.

One common method for checking the convergence of multivariate MCMC sequence is Gelman and Rubin's convergence diagnostic. It is available in R package 'coda'. The idea is to check the convergence by multiple MCMC sequences of sampls from different initial values. The details of the diagnostic is shown in the package.

Here, the initial values of $p$ and $\lambda$ are generated from Uniform(0,1) and Gamma(a,b) independently. Then we draw 1000 sample from through Gibbs sampler. In order to produce the sample from the posterior distribution, we discard the first 300 observations. Overall, this sampling procedure is repeated for five sets of randomly drawn initial values to compute the Gelman and Rubin's convergence diagnostic. If there is no problem in convergence, we pool the five sets of samples ($700 \times 5 = 3500$) and use them to compute the confidence intervals.

For the case of $a = 1$ and $b = 1$, we plot the sequence of samples as follows.

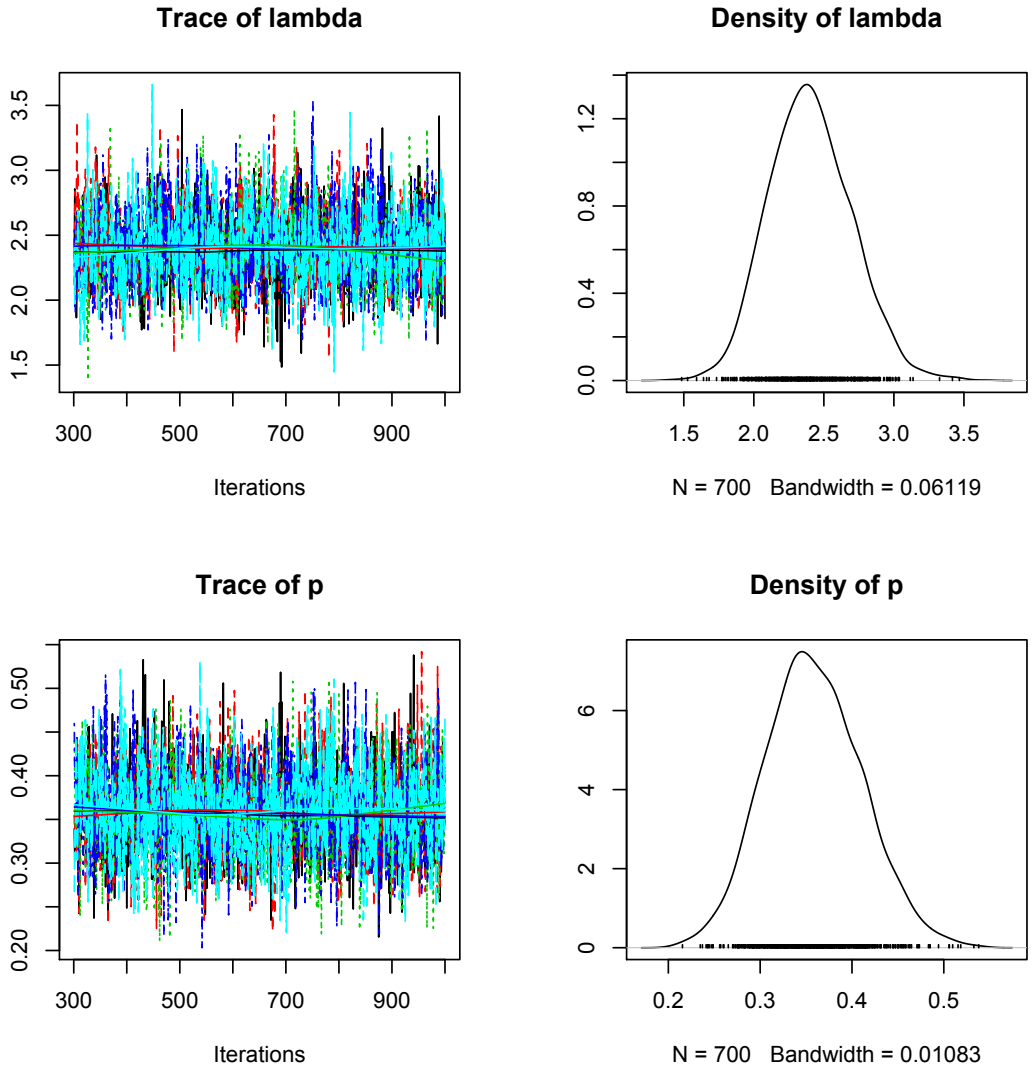



(Here, the five sets of samples are overlaid in one plot with different color. Please refer to the electronic version of this assignment.) You can see that the five sets mixed together quite well, which suggests convergence. Further, we perform the Gelman and Rubin's convergence diagnostic and the program output is as follows.

```
Potential scale reduction factors:

       Point est. 97.5% quantile
lambda          1           1.00
```

```
p                1            1.00

Multivariate psrf

1.00
```

All of the statistics are close to 1, which implies convergence. We discard the burn-in. The trace plot and plot of estimated density are given as follows.




Similar analyses are performed for the other settings of $a$ and $b$. Results show that samples from different settings converged. Here we summarize the 95% Bayesian confidence intervals for different settings as follows.

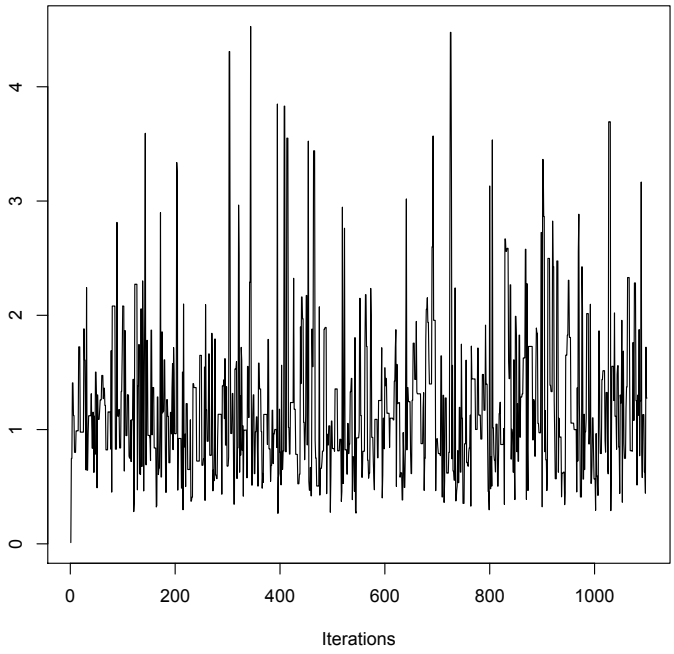
| $\lambda$ | $a = 1$ | $a = 5$ |
|---|---|---|
| $b = 1$ | (1.8684354, 2.9943104) | (1.9686129, 3.1813159) |
| $b = 5$ | (1.8348044, 2.9731457) | (1.9801840, 3.128568) |
| $p$ | $a = 1$ | $a = 5$ |
| $b = 1$ | (0.2567675, 0.4601536) | (0.2547732, 0.4511376) |
| $b = 5$ | (0.2530773, 0.4567579) | (0.2533904, 0.452584) |

Here, you can see that the true value $p = 0.3$ and $\lambda = 2$ are included in all the confidence intervals. The length of the confidence intervals basically depends on three things, the mean and variance of the prior, and the data. Here, our discussion focuses on $\lambda$. As the variance of Gamma$(a, b)$ is $a/b^2$, we may expect that the confidence intervals become wider for larger value of $a/b^2$. However, it is not the case in our situation. For example the interval of $a = 1$ and $b = 1$ is shorter than that of $a = 1$ and $b = 5$. It is basically because the mean of Gamma$(a, b)$ is $a/b$ and thus the case of $a = 1$ and $b = 5$ actually gives a prior with mean far away from the true $\lambda$ and small variance. That leads to conflicting information between the prior and the data and thus resulting a wider confidence intervals. In general, it is difficult to judge the length of the confidence intervals solely from the prior since realization of the data plays an important role.

## Problem 6

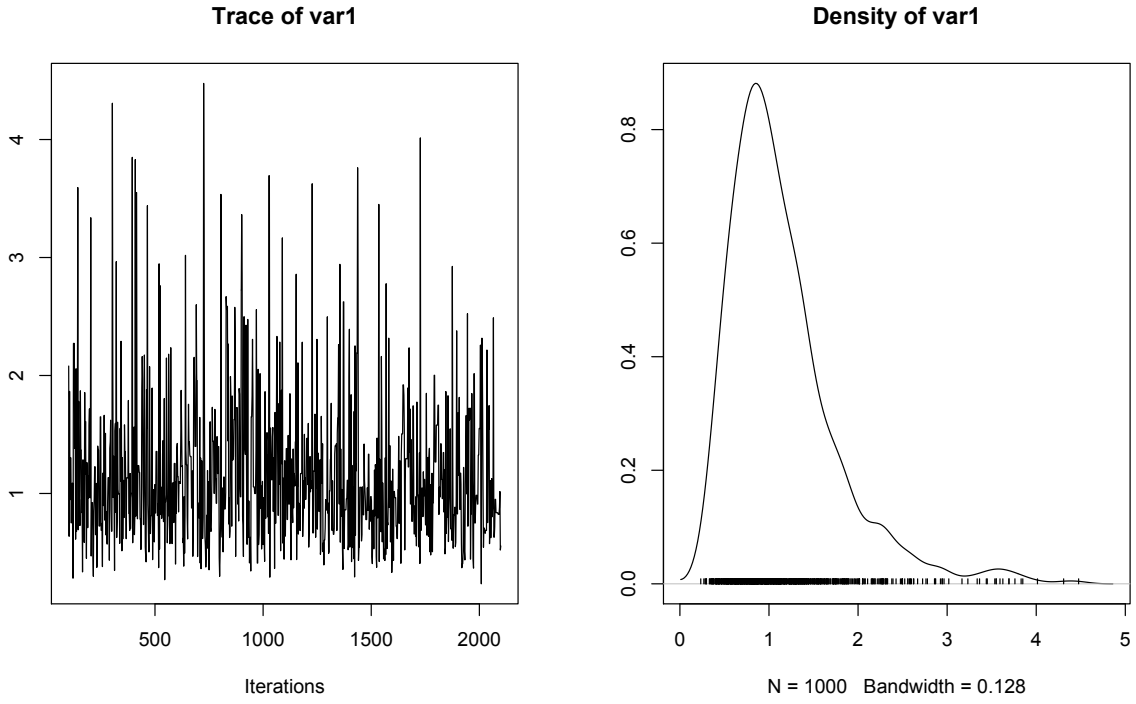
In this question, we draw a sample of size 1000 using the independence Metropolis-hastings algorithm. First, we try Gamma(1,1) as the proposal. Note that in this case, the posterior distribution is univariate and thus we can use a simpler technique to check convergence. Here, we use the Geweke's convergence diagnostic, which is available in R package 'coda'. The basic idea of this diagnostic is to check if the mean of the first portion of the sequence is the same as the that of the last portion of the sequence. In my first trial, first, a sample of 1100 is sampled and the first 100 is regarded as burn-in. The Geweke's statistic is -0.7865, suggesting the convergence. The plot of

the sample is shown as follows.




Since Metropolis-Hastings algorithm is used, there is chance for rejection of new moves. Thus, you can see (with good eyesight or magnifying glass) that there are consecutive points having the same values in the above plot. In fact, the acceptance rate (after burn-in) is just 0.545. It suggests that thinning may help. We choose to take every second points in the sample as the final sample. And thus in order to get 1000 sample, we will take have to generate a sample of size 2100. Then we use the first 100 as burn-in, and sample every second point and form the final sample. The plot of the

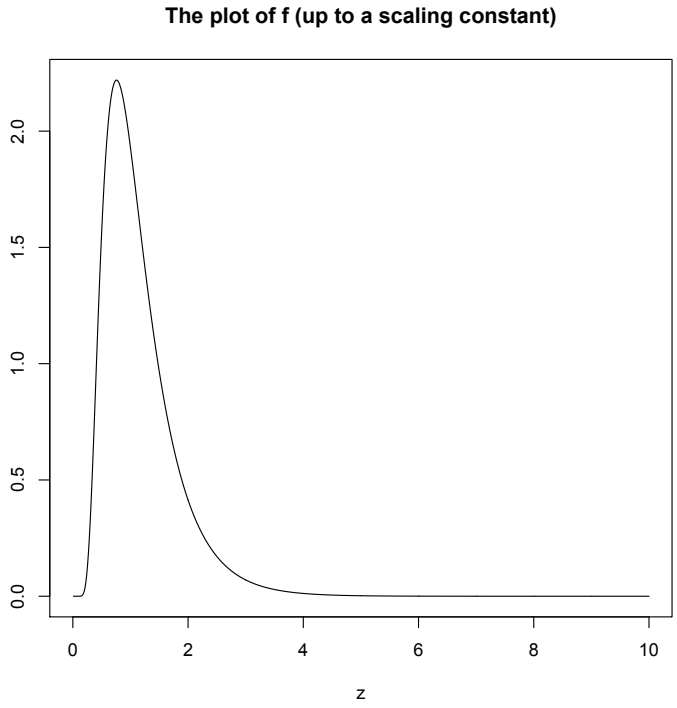
final sample and the plot of its estimated density are shown as follows.




The Geweke's statistic is 0.4595, suggesting the convergence. In order to assess the accuracy, we estimate the mean of $Z$ and $1/Z$ from the sample. The estimates are 1.156345 and 1.104967 respectively. Note that the true values are 1.154701 and 1.116025.

In order to get a better sample, we try different Gamma distribution. To get some clues, we

plot $f$ (up to a scaling constant) as follows.




As for trying different proposals, it seems that there are too many combinations of the two parameters of Gamma distributions to try. However, we see that there is a peak at around 1 to 2. Also, we know that the mean of $Z$ is around 1. One particular choice is fixing the two parameters (shape and rate) the same so that the Gamma distribution will always has mean 1. Here are the five combinations that I tried. (The standard errors are listed in the brackets.)

| Value of the two parameters | Estimated mean of $Z$ | Estimated mean of $1/Z$ |
|---|---|---|
| 1 | 1.156345 (0.01977445) | 1.104967 (0.01782304) |
| 2 | 1.121914 (0.01799992) | 1.131602 (0.01839537) |
| 3 | 1.136918 (0.01898390) | 1.136645 (0.01927243) |
| 4 | 1.150230 (0.02062173) | 1.139721 (0.01985531) |

To me, it seems that all of them provide reasonably good estimates. Also, note that, here, we can only get a brief idea of how well the proposal distribution is since only one sample is used.