Course: STA6934- Monte Carlo Statistical Methods
Instructor: Professor Casella

Assignment 5

Problem 8.4

Algorithm: We can find that the integrating factor is $\int_0^\infty \exp(-x^d)dx$. We can transform $x^d = y$, then we can find $\int_0^\infty \exp(-x^d)dx = \frac{1}{d}\Gamma(\frac{1}{d})$, hence $f(x) = \frac{d\exp(-x^d)}{\Gamma(\frac{1}{d})}$. Also, to use in algorithm, we find $f(x) > u$, which results $x < \left[-\log\left(\frac{\Gamma(1/d)}{d}u\right)\right]^{1/d}$.

- Set starting point (x, u) in $\left\{(x, u) : 0 < u < \frac{d\exp(-x^d)}{\Gamma(\frac{1}{d})}\right\}$.
- At iteration t, simulate
  1. $u^{(t+1)} \sim \mathcal{U}\left(0, \frac{d\exp(-x^d)}{\Gamma(\frac{1}{d})}\right)$
  2. $x^{(t+1)} \sim \mathcal{U}\left(0, \left[-\log\left(\frac{\Gamma(1/d)}{d}u\right)\right]^{1/d}\right)$

Simulation: For $(1000, 10000, 100000)$ simulations for each $d = \{0.1, 0.25, 0.4\}$. For each graph, a histogram is the simulation results, and a solid line is $f(x) = \frac{d\exp(-x^d)}{\Gamma(\frac{1}{d})}$. We can check that the d is larger, then histogram is more filling under the area of $f(x)$.

Problem 8.5

Here is the data,

Quartile	0.00000	0.67000	0.84000	1.28000	1.64000	1.96000	2.33000
Probability	0.50000	0.74857	0.79955	0.89973	0.94950	0.97500	0.99010
Slice	0.50099	0.74953	0.80054	0.90061	0.95050	0.97584	0.99034
IID Sampler	0.49887	0.74929	0.80049	0.90141	0.95139	0.97539	0.99011

The first row shows quantiles, and the second row represents corresponding probabilities. The third rows shows the empirical probabilities of corresponding quantiles derived by slice sampler, and the forth represent IID sampler in R system, for 100,000 sample size for both. Both sampler gives similar results, however, slice sampler takes 5.77 seconds, but IID sampler takes only 0.32 second.

Problem 9.2

Let (x_0, y_0) be the starting point, and that suppose we actually use only x_0. From standard properties of the conditional distribution of the bivariate normal, we have

$$Y_0 \sim N(\rho x_0, 1-\rho^2)$$

Now, conditional on Y_0,

$$X_1 \sim N(\rho Y_0, 1-\rho^2)$$

Thus hives us the GS. Combining these two facts, we see that unconditionally,

$$X_1 \sim N(\rho^2 x_0, (1-\rho^2) + \rho^2(1-\rho^2)) = N(\rho^2 x_0, 1-\rho^4)$$

Iterating this kind of calculation, we see that unconditionally,

$$X_n \sim N(\rho^{2n} x_0, 1-\rho^{4n})$$

R-code:

```
nsim <- 50000
X <- array(0, dim=c(nsim,1))
Y <- array(0, dim=c(nsim,1))
Z <- array(0, dim=c(nsim,1))

rho=0.3
Y[1]=runif(1,0,1)
X[1]=runif(1,0,1)*sqrt(1-rho^2)+rho*Y[1]
Z[1]= (X[1])^2+{Y[1]}^2

for (i in 2:nsim) {
   X[i]=  runif(1,0,1)*sqrt(1-rho^2)+rho*Y[i-1]
   Y[i]= runif(1,0,1)*sqrt(1-rho^2)+rho*X[i]
   Z[i]= (X[i])^2+{Y[i]}^2 }

sum((Z>2))/nsim
[1] 0.10266
```

Problem 9.4

We have two ergodic Markov chains $\{X^{(i)}, Y^{(i)}\}$, and corresponding Metropolis-Hastings Algorithm and Gibbs Sampler as following,

- MH Algorithm

  1. Generate $x' \sim g(x'|x)$
  2. Take

  $$x'_{new} = \begin{cases} x', & \text{with probability } \rho(x, x') \\ x, & \text{otherwise.} \end{cases}$$

  Where

  $$\rho(x, x') = \min\left\{\frac{f(x')}{f(x)}\frac{g(x|x')}{g(x'|x)},\ 1\right\}$$

- Gibbs Sampler, beginning with $X_0 = x_0$,

  1. Take $Y_t \sim f_{Y|X}(\cdot|x_{t-1})$
  2. Take $X_t \sim f_{X|Y}(\cdot|y_t)$

(a)

Showing $K(x, x') = g(x'|x) = \int f(x'|y)f(y|x)dy$, the first part $K(x, x') = g(x'|x)$ is just matching this and the first step of MH algorithm above. And proving second part is following,

$$\begin{aligned}
K(x, x') &= f(x'|x) = \frac{f(x', x)}{f(x)} \\
&= \frac{1}{f(x)} \int f(x', x|y)f(y)dy \\
&= \int f(x'|y)\frac{f(x|y)}{f(x)}f(y)dy \text{ (by the interleaving property,)} \\
&= \int f(x'|y)\frac{f(x, y)}{f(x)}dy \\
&= \int f(x'|y)f(y|x)dy\square
\end{aligned}$$

(b)

The MH algorithm with ϱ generates $x \sim f$, where $f(\cdot) = \int f(\cdot, y)dy$. What we show is that this marginal distribution is the stationary distribution of $X^{(i)}$, which is following.

$$\begin{aligned} \int f(x)g(x'|x)dx &= \int f(x) \int f(x'|y)f(y|x)dydx \\ &= \iint f(x'|y)f(y|x)f(x)dxdy \\ &= \int f(x'|y)f(y)dy \\ &= f(x') \end{aligned}$$

Hence $f(x)$ is also the stationary distribution, from which we desire to generate x.

(c)

From **(b)**, we know that $f(x)$ is stationary distribution of kernel density $g(x'|x)$, hence it satisfies detailed balance condition,

$$g(x'|x)f(x) = g(x|x')f(x')$$

And we rearrange it,

$$\frac{f(x')}{g(x'|x)} = \frac{f(x)}{g(x|x')}$$

Therefore $\varrho = 1$, and we always accept new x' in MH algorithm.

Problem 9.8

(a)

From Example 9.7, we know the conditional posterior density of a and b, given each other,

$$\pi(a|\mathbf{y}, \mathbf{t}, b) \propto \exp\left(a\sum_i y_i - e^a \sum_i e^{bt_i} - \frac{a^2}{2\sigma^2}\right)$$
$$\pi(b|\mathbf{y}, \mathbf{t}, a) \propto \exp\left(b\sum_i t_i y_i - e^a \sum_i e^{bt_i} - \frac{b^2}{2\tau^2}\right)$$

where observation, (Y_i) given number of passages, (t_i) follows $\mathcal{P}(\exp(a + bt_i))$, and priors are $a \sim \mathcal{N}(0, \sigma^2)$, $b \sim \mathcal{N}(0, \tau^2)$, with known variances.

Algorithm: Since we know $\pi(a|\mathbf{y},\mathbf{t},b)$ and $\pi(b|\mathbf{y},\mathbf{t},a)$, we can use Gibbs sampler, however the normalizing constants of them are problematic. In this case, in generating each of the conditional, we can use random walk Metropolis-Hastings algorithm, in which we can ignore normalizing constants. Also the candidates are $g(a_t|a_{t-1}) = \mathcal{N}(a_{t-1},\sigma^2)$ and $g(b_t|b_{t-1}) = \mathcal{N}(b_{t-1},\tau^2)$, where we can cancel candidate density terms in ρ. The algorithm is, beginning with $a_0 \sim \mathcal{N}(0,\sigma^2)$,

1. Generate $b_t \sim \pi(b_t|\mathbf{y},\mathbf{t},a_{t-1})$
   (a) Generate $b' \sim \mathcal{N}(b_{t-1},\tau^2) = g(b'|b_{t-1})$
   (b) Take
   $$b_t = \begin{cases} b', & \text{with probability } \rho(b_{t-1},b') \\ b_{t-1}, & \text{otherwise.} \end{cases}$$
   Where
   $$\rho(b_{t-1},b') = \min\left\{\frac{\pi(b'|\mathbf{y},\mathbf{t},a_{t-1})}{\pi(b_{t-1}|\mathbf{y},\mathbf{t},a_{t-1})},1\right\}$$
2. Generate $a_t \sim \pi(a_t|\mathbf{y},\mathbf{t},b_t)$
   (a) Generate $a' \sim \mathcal{N}(a_{t-1},\sigma^2) = g(a'|a_{t-1})$
   (b) Take
   $$a_t = \begin{cases} a', & \text{with probability } \rho(a_{t-1},a') \\ a_{t-1}, & \text{otherwise.} \end{cases}$$
   Where
   $$\rho(a_{t-1},a') = \min\left\{\frac{\pi(a'|\mathbf{y},\mathbf{t},b_t)}{\pi(a_{t-1}|\mathbf{y},\mathbf{t},b_t)},1\right\}$$

(b)

Algorithm: Now, we have a full specification that the last number of passages, t_4 is not 4, but "4 or more" with observations $y_4 = 13$. In this case we need to generate t_4 as well as $\{a,b\}$, So we have three-stage Gibbs sampler, with posterior for t_4 as

$$\pi(t_{4,t}|y_4,a_t,b_t) \propto \exp[-\exp(a+bt_4)+y_4(a+bt_4)]$$

However, since we cannot ascertain $\sum_{t_4=4}^{\infty}\pi(t_{4,t}|y_4,a_t,b_t) < \infty$, we generate a candidate $x \sim$ Truncated Poisson$(\mu) = g(x|\mu), x \geq 4$ and we can generate t_4 through subsequent independent MH algorithm. In this way we can avoid calculating normalizing constant of $\pi(t_{4,t}|y_4,a_t,b_t)$. For the first two stages, we use previous t_4, so that we can use the same posteriors $\pi(a|\mathbf{y},\mathbf{t},b), \pi(b|\mathbf{y},\mathbf{t},a)$

as in **(a)**. The algorithm is following.

1. Generate $b_t \sim \pi(b_t|\mathbf{y}, \mathbf{t}, a_{t-1})$
   (a) Generate $b' \sim \mathcal{N}(b_{t-1}, \tau^2) = g(b'|b_{t-1})$
   (b) Take
   $$b_t = \begin{cases} b', & \text{with probability } \rho(b_{t-1}, b') \\ b_{t-1}, & \text{otherwise.} \end{cases}$$
   Where
   $$\rho(b_{t-1}, b') = \min\left\{\frac{\pi(b'|\mathbf{y}, \mathbf{t}, a_{t-1})}{\pi(b_{t-1}|\mathbf{y}, \mathbf{t}, a_{t-1})}, 1\right\}$$
2. Generate $a_t \sim \pi(a_t|\mathbf{y}, \mathbf{t}, b_t)$
   (a) Generate $a' \sim \mathcal{N}(a_{t-1}, \sigma^2) = g(a'|a_{t-1})$
   (b) Take
   $$a_t = \begin{cases} a', & \text{with probability } \rho(a_{t-1}, a') \\ a_{t-1}, & \text{otherwise.} \end{cases}$$
   Where
   $$\rho(a_{t-1}, a') = \min\left\{\frac{\pi(a'|\mathbf{y}, \mathbf{t}, b_t)}{\pi(a_{t-1}|\mathbf{y}, \mathbf{t}, b_t)}, 1\right\}$$
3. Generate $t_{4,t} \sim \pi(t_{4,t}|y_4, a_t, b_t)$
   (a) Generate $x' \sim$ Truncated $\mathcal{P}(\mu) = g(x'|\mu), x' \geq 4$
   (b) Take
   $$t_{4,t} = \begin{cases} x', & \text{with probability } \rho(t_{4,t-1}, x') \\ t_{4,t-1}, & \text{otherwise.} \end{cases}$$
   Where
   $$\rho(t_{4,t-1}, x') = \min\left\{\frac{\pi(x'|y_4, a_t, b_t)g(t_{4,t-1}|\mu)}{\pi(t_{4,t-1}|y_4, a_t, b_t)g(x'|\mu)}, 1\right\}$$

Problem 9.14

(a)

The distribution of the missing data is $Z_i \sim \frac{\Phi(z-\theta)}{1-\Phi(a-\theta)}$ and the complete data likelihood is

$$\begin{aligned}
L(\theta|x,z) &= \Pi_{i=1}^{m} e^{-(x_i-\theta)^2/2}\left[\Pi_{i=m+1}^{n} e^{-(z_i-\theta)^2/2}\right](2\pi)^{-n/2}(1-\Pi(a-\theta))^{n-m} \\
&\propto exp\{-\frac{1}{2}\sum_{i=1}^{m}(x_i-\theta)^2 - \frac{1}{2}\sum_{i=m+1}^{n}(z_i-\theta)^2\} \\
&\propto exp\{\frac{1}{2}[n\theta^2 - 2\theta(\bar{x}+(n-m)\bar{z}) + (m\bar{x}^2+(n-m)\bar{z}^2)]\} \\
&\propto exp\{-\frac{n}{2}[\theta - \frac{m\bar{x}+(n+m)\bar{z}}{n}]^2 + \frac{1}{2}[\frac{(m\bar{x}+(n-m)\bar{z})^2}{n} - m\bar{x}^2 + (n-m)\bar{z}^2]\} \\
&\propto exp\{-\frac{n}{2}[\theta - \frac{m\bar{x}+(n+m)\bar{z}}{n}]^2
\end{aligned}$$

Thus, as a function of θ, the normalized complete data likelihood is $N(\frac{m\bar{x}+(n+m)\bar{z}}{n}, \frac{1}{n})$.

Problem 9.17

(1)

We assume X_{i1} and X_{i2} are conditionally independent given $\{\theta_1, \theta_2\}$. Then,

$$\begin{aligned}
L(\theta_1,\theta_2|Y_i,n_{i1},n_{i2}) &= L(\theta_1,\theta_2|Y_i = X_{i1}+X_{i2}, n_{i1}, n_{i2}) \\
&= \sum_{j_i=0}^{Y_i} L(\theta_1,\theta_2|X_{i1}=j_i, X_{i2}=Y_i-j_i, n_{i1}, n_{i2}) \\
&= \sum_{j_i=0}^{Y_i} L(\theta_1|X_{i1}=j_i, n_{i1}) L(\theta_2|X_{i2}=Y_i-j_i, n_{i2})
\end{aligned}$$

Since $X_{i1} \sim \mathcal{B}(n_{i1}, \theta_1)$, and $X_{i2} \sim \mathcal{B}(n_{i2}, \theta_2)$, we have,

$$L(\theta_1,\theta_2|\mathbf{Y},\mathbf{n}_1,\mathbf{n}_2) = \prod_{i=1}^{3}\left[\sum_{j_i=0}^{Y_i}\binom{n_{i1}}{j_i}\binom{n_{i2}}{Y_i-j_i}\theta_1^{j_i}(1-\theta_1)^{n_{i1}-j_i}\theta_2^{Y_i-j_i}(1-\theta_1)^{n_{i2}-Y_i+j_i}\right]\Box$$

Problem 9.26

(a)

We want to derive $\pi(\mu, \tau|\mathbf{x})$ where $\mathbf{x} = \{x_1, x_2, \cdots x_n\}$, with prior of $\tau_j^2 \sim \mathcal{IG}(\nu, A)$,

$$\pi(\tau_j|\nu, A) = \frac{A^\nu}{\Gamma(\nu)}(\tau_i^2)^{-(\nu+1)} \exp\left(-\frac{A}{\tau_j^2}\right)$$

independent with known ν and $\pi(A) \propto \frac{1}{A}$. For the prior of μ_j,

$$\pi(\mu_j|\mu_{j-1}, \boldsymbol{\tau}) = \mathcal{N}^+_{\mu_{j-1}}(\mu_{j-1}, B(\tau_j^{-2} + \tau_{j-1}^{-2})^{-1})$$

Where $\mathcal{N}_a^+(\mu, \sigma^2)$ is a density of truncated normal greater than a, except $\pi(\mu_1) = \mathbb{I}_{-\infty<\mu_1<\infty}$, improper uniform. Then the prior of $\boldsymbol{\mu}$,

$$\pi(\boldsymbol{\mu}|\boldsymbol{\tau}) \propto \prod_{j=2}^k (\tau_j^{-2} + \tau_{j-1}^{-2})^{1/2} \exp\left[-\frac{B}{2}\sum_{j=2}^k (\mu_j - \mu_{j-1})^2(\tau_j^{-2} + \tau_{j-1}^{-2})\right] \mathbb{I}_{\mu_1<\mu_2<\cdots<\mu_k} \qquad (*)$$

Where $\mathbf{T}_j = \begin{cases} \tau_j^{-2} + \tau_{j-1}^{-2}, & j > 1 \\ 1, & j = 1 \end{cases}$, $a_j = \begin{cases} B(\mu_j - \mu_{j-1})^2, & j > 1 \\ 0, & j = 1 \end{cases}$, and $b_j = \begin{cases} a_{j+1}, & j < k \\ 0, & j = k \end{cases}$, then the prior of $\boldsymbol{\tau}$

$$\pi(\boldsymbol{\tau}|\nu, A) \propto A^{k\nu-1}\left[\prod_{j=1}^k \tau_j^2\right]^{-(\nu+1)} \exp\left(-A\sum_{j=1}^k \tau_j^{-2}\right), \qquad (**)$$

Assuming B, ν known, and $\pi(A) \propto A^{-1}$.
For the sample distribution, normal mixture, we introduce latent variables $z_i \sim \mathcal{M}_k(1|p_1, p_2, \cdots, p_k)$, and for the prior of $\mathbf{p}$ is $\pi(\mathbf{p}|\boldsymbol{\gamma}) \propto \prod_{j=1}^k p_j^{\gamma_j - 1}$, assuming all $\gamma_j, 1 \le j \le k$ are known. Therefore the likelihood function is,

$$L(\boldsymbol{\mu}, \boldsymbol{\tau}, \mathbf{p}|\mathbf{x}, \mathbf{z}) \propto \prod_{j=1}^k p_j^{n_j+\gamma_j-1}\tau_j^{-n_j} \exp\left[-\frac{1}{2\tau_j^2}[n_j(\bar{x}_j - \mu_j)^2 + (n_j-1)s_j^2]\right] \qquad (***)$$

by square completion, where $\bar{x}_j = \sum_{j=1}^k x_i \mathbb{I}_{z_i=j}$, and $s_j^2 = \sum_{j=1}^k (x_i - \bar{x}_j)^2 \mathbb{I}_{z_i=j}$. Finally the posterior distribution $\pi(\boldsymbol{\mu}, \boldsymbol{\tau}, \mathbf{p}, A|\mathbf{x}, \mathbf{z})$ that is derived by multiplying $(*) \cdot (**) \cdot (***)$ and reorganizing, is proportional to

$$A^{k\nu-1}\prod_{j=1}^k \mathbf{T}_j^{1/2}\frac{p_j^{n_j+\gamma_j-1}}{(\tau_j^2)^{\nu+n_j/2+1}} \exp\left[-\frac{1}{2\tau_j^2}\sum_{j=1}^k (2A + a_j + b_j + n_j(\bar{x}_j - \mu_j)^2 + (n_j-1)s_j^2)\right] \mathbb{I}_{\mu_1<\mu_2<\cdots<\mu_k} \square$$

(b)

Since $\pi(\mathbf{p}|\mathbf{x},\mathbf{z}) \propto \prod_{j=1}^{k} p_j^{n_j+\gamma_j-1}$ is independent to $\boldsymbol{\mu}, \boldsymbol{\tau}, A$, we consider posterior of $\boldsymbol{\mu}, \boldsymbol{\tau}, A$ only. We use two identities,

$$\int_{-\infty}^{\infty} \exp\left[-\frac{1}{\tau^2}(C_1(\mu-a)^2 + C_2(\mu-b)^2)\right] d\mu = \frac{\sqrt{2\pi}\tau}{\sqrt{C_1+C_2}} \exp\left\{-\frac{C_1C_2(a-b)^2}{\tau^2(C_1+C_2)}\right\}$$

$$\int_{0}^{\infty} A^{-a-1} \exp\left(-\frac{C_3}{2A^2}\right) dA = \frac{2^{a/2-1}\Gamma(a/2)}{C_3^{a/2}}$$

And we consider a case $k=2, n_1=2, n_2=0$, the full proof can be done by induction. Then,

$$\begin{aligned} f(\boldsymbol{\mu},\boldsymbol{\tau},A|\mathbf{x},\mathbf{z}) &\propto A^{2\nu-1}(\tau_1^2)^{-(\nu+2)} \exp\left[-\frac{1}{2\tau_1^2}(2A + B(\mu_2-\mu_1)^2 + 2(\bar{x}_1-\mu_1)^2 + s_1^2)\right] \\ &\cdot (\tau_2^{-2}+\tau_1^{-2})^{1/2}(\tau_2^2)^{-(\nu+1)} \exp\left[-\frac{1}{2\tau_2^2}(2A + B(\mu_2-\mu_1)^2)\right] \mathbb{I}_{\mu_1<\mu_2} \end{aligned}$$

From the above posterior, we can extract posterior for μ_1, μ_2,

$$f(\mu_1,\mu_2|\mathbf{x},\mathbf{z}) = \exp\left[-\frac{1}{2}B(\tau_2^{-2}+\tau_1^{-2})(\mu_2-\mu_1)^2 + 2(\bar{x}_1-\mu_1)^2\right] \mathbb{I}_{\mu_1<\mu_2}$$

We have upper bound of $\int_{-\infty}^{\infty}\int_{-\infty}^{\mu_2} f(\mu_1,\mu_2|\mathbf{x},\mathbf{z})d\mu_1 d\mu_2$ when we integrate over $-\infty < \mu_1, \mu_2 < \infty$, by using the first identity,

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\mu_2} f(\mu_1,\mu_2|\mathbf{x},\mathbf{z})d\mu_1 d\mu_2 &\leq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\mu_1,\mu_2|\mathbf{x},\mathbf{z})d\mu_1 d\mu_2 \\ &= \int_{-\infty}^{\infty} \sqrt{\frac{2\pi}{B(\tau_2^{-2}+\tau_1^{-2})+2}} \exp\left[-\frac{B(\tau_2^{-2}+\tau_1^{-2})(\bar{x}_1-\mu_2)^2}{B(\tau_2^{-2}+\tau_1^{-2})+2}\right] d\mu_2 \\ &= \sqrt{\frac{\pi}{B}} \frac{\tau_1^2\tau_2^2}{\sqrt{(\tau_2^2+\tau_1^2)}} = \frac{C\tau_1^2\tau_2^2}{\sqrt{(\tau_2^2+\tau_1^2)}} \end{aligned}$$

Where $C = \sqrt{\frac{\pi}{B}}$. Then we plug-in the above to find,

$$f(A,\boldsymbol{\tau}|\mathbf{x},\mathbf{z}) \leq CA^{2\nu-1}(\tau_1^2)^{-(\nu+5/2)}(\tau_2^2)^{-(\nu+3/2)} \exp\left[-\frac{1}{2\tau_1^2}(2A+s_1^2) - \frac{A}{\tau_2^2}\right].$$

We integrate over $f(A,\boldsymbol{\tau}|\mathbf{x},\mathbf{z})$ over τ_1,τ_2, referring the second identities. Then we have,

$$\begin{aligned} f(A|\mathbf{x},\mathbf{z}) &= \int_0^{\infty}\int_0^{\infty} f(A,\boldsymbol{\tau}|\mathbf{x},\mathbf{z})d\tau_1 d\tau_2 \\ &\leq C\frac{2^{2\nu+1}\Gamma(\nu+2)\Gamma(\nu+1)A^{2\nu-1}}{(2A+s_1^2)^{\nu+2}(2A)^{\nu+1}} \\ &= C'\frac{A^{\nu-2}}{(A+s_1^2/2)^{\nu+2}} \end{aligned}$$

Where $C' = C\frac{\Gamma(\nu+2)\Gamma(\nu+1)}{4}$. Finally, we get total posterior upper bound by integrating $f(A|\mathbf{x},\mathbf{z})$ over A (referring a known identity),

$$\begin{aligned}\int_0^\infty f(A|\mathbf{x},\mathbf{z})dA &= \int_0^\infty C' \frac{A^{\nu-2}}{(A+s_1^2/2)^{\nu+2}}dA \\ &= C''\left(\frac{2}{s_1^2}\right)^3 < \infty\end{aligned}$$

Where $C'' = C'\Gamma(\nu-1)\Gamma(3)$. Since the upper bound is finite, the posterior is proper.