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

Assignment 4

Problem 6.22

(a). For every n, the conditional distribution of X_{n+1}, given $x_n, x_{n-1}, ..., x_0$ is the same as the distribution of X_{n+1} given x_n. Thus (X_n) is a Markov chain.

(b). Let $n_1,\ n_2$ be two states ($n_1,\ n_2$ are two positive integers). We have, if $n_1 < n_2$, $K^{n_2-n_1}(n_1, n_2) = p^{n_2-n_1} > 0$ and $K^{n_2-n_1}(n_2, n_1) = (1-p)^{n_2-n_1} > 0$, if $n_1 = n_2 = n > 0$, $K^2(n, n) = 2p(1-p) > 0$ and $K(0, 0) = 1 - p > 0$. We conclude that every pair of states of (X_n) can be connected in a finite number of steps with positive probability and thus that the chain (X_n) is irreducible.

(c). Let $a = (a_1, a_2, ...)$ be a distribution. a is invariant of the chain if, and only if, $a\mathbb{P} = a$, where $\mathbb{P}$ is the transition matrix of (X_n). This is equivalent to $(1-p)(a_0 + a_1) = a_0$ and for every $k \geq 0$, $pa_k + (1-p)a_{k+2} = a_{k+1}$, or, still equivalently, $a_1 = \frac{p}{1-p}a_0$ and for every $k \geq 0$, $pa_k + (1-p)a_{k+2} = a_{k+1}$.

The recurrence equation $pa_k + (1-p)a_{k+2} = a_{k+1}$ has the characteristic equation

$$(1-p)\rho^2 - \rho + p = 0,$$

which solves into $\rho_1 = \frac{p}{1-p}$ and $\rho_2 = 1$. The recurrence solves into

$$a_k = \alpha\rho_1^k + \beta\rho_2^k$$

Substituting this result into a_0 and a_1 gives $\alpha = a_0$ and $\beta = 0$. Therefore, the invariant distribution of the chain is

$$a = \left\{ \left(\frac{p}{1-p}\right)^k a_0 \right\}_{k\geq 0},$$

where a_0 is arbitrary and a_k is the probability that the chain is at state k. To be a probability, a must satisfy

$$\sum_{k=0}^{\infty} a_k < \infty \quad \text{and} \quad \sum_{k=0}^{\infty} a_k = 1,$$

that is,

$$a_0 = \frac{1-2p}{1-p},\ p < \frac{1}{2}.$$

(d). If $\sum a_k < \infty$ and $X_n \sim a$, we have

$$\begin{aligned}
P(X_{n+1} = k) &= \sum_{j=0}^{\infty} P(X_{n+1} = k | X_n = j) P(X_n = j) \\
&= P(X_{n+1} = k | X_n = k-1) P(X_n = k-1) \\
&+ P(X_{n+1} = k | X_n = k+1) P(X_n = k+1) \\
&= p a_{k-1} + (1-p) a_{k+1} = \\
&= p \left(\frac{p}{1-p} \right)^{k-1} + (1-p) \left(\frac{p}{1-p} \right)^{k+1} = a_k
\end{aligned}$$

Therefore, the invariant distribution is also a stationary distribution of the chain and then the chain is ergodic.

1 Problem 6.35 ab

a) Given $P(X_0 = 1) = 1/2$, it implies $\pi = (1/2 \;\; 1/2)$.

Moreover, since $P(X_{i+1} = 1 | X_i = -1) = 1$ and $P(X_{i+1} = -1 | X_i = 1) = 1$, it follows that

$$\mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Let's consider

$$\pi \mathbb{P} = (1/2 \;\; 1/2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (1/2 \;\; 1/2) = \pi$$

Thus, the chain is stationary.

b) Notice that $\mathbb{P}^{2k} = \mathbb{I}$ and $\mathbb{P}^{2k+1} = \mathbb{P}$. Notice that for the case $2k$, we observe that $X_{2k} = X_0$ which implies $Cov(X_0, X_{2k}) = 1 \neq 0$. Since we just found a subsequence that does not converge to 0, $Cov(X_0, X_k)$ does not go to zero.

Problem 6.40

(a)

By definition of Total Variation Norm, $\|\mu\|_{TV} = \frac{1}{2} \int_\Omega |\mu(dx)|$. And we use a property,

$$|\mu(dx)| \geq |h(x)\mu(dx)|$$

Where $|h(x)| \leq 1$ for any E. And $|\mu(dx)| = \mu^+(dx) + \mu^-(dx)$ is total variation measure, and $\mu^+(dx) = \max(\mu(dx), 0)$, $\mu^-(dx) = -\min(\mu(dx), 0)$ are signed measures. Then

$$\frac{1}{2}\int_\Omega |\mu(dx)| \geq \frac{1}{2}\int_\Omega |h(x)\mu(dx)| \geq \frac{1}{2}\left|\int_\Omega h(x)\mu(dx)\right|$$

The first inequality is valid by $|\mu(dx)| \geq |h(x)\mu(dx)|$, and the second one is valid by triangle inequality. Equality satisfies where

$$\|\mu\|_{TV} = \frac{1}{2}\int_\Omega |\mu(dx)| = \frac{1}{2}\sup_{|h|\leq 1}\left|\int_\Omega h(x)\mu(dx)\right|_\square$$

Problem 6.54

Let

$$\mathbb{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

be the transition matrix of a two-state ergodic chain. The ergodicity implies that α and β are not null together, that is that, $\alpha + \beta > 0$. The stationary distribution is

$$\pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right),$$

and then we obtain $\tilde{\mathbb{P}} = \mathbb{P}$ and $\mathbb{P}$ is reversible.

For an ergodic chain with symmetric transition matrix and state space equal to $\{1, ..., n\}$, the invariant distribution is

$$\pi = \left(\frac{1}{n}, ..., \frac{1}{n}\right),$$

and so

$$\tilde{p}_{ij} = \frac{\pi_j}{\pi_i} p_{ji} = p_{ji} = p_{ij}.$$

For the given matrix, let $\pi = (0.1, 0.2, 0.4, 0.2, 0.1)$. An easy computation gives $\pi\mathbb{P} = \pi$ and hence, π is the invariant distribution of $\mathbb{P}$. Since, $\tilde{p}_{12} = \frac{\pi_2}{\pi_1} p_{21} = \frac{0.2}{0.1} 0.5 = 1 \neq p_{12} = 0$, $\mathbb{P}$ is not reversible.

2 Problem 6.67

a) Let's consider the distribution of $\bar{X}$ in the multivariate normal case. Notice that $V(\bar{X}) > 0$ requires for all sufficiently large n requires $\rho > 0$. Thus,

$$\bar{X} = \frac{1}{n} JX \Rightarrow V(\bar{X}) = \frac{n\sigma^2}{n^2} + \frac{n-1}{n}\rho \to 0 + \rho > 0$$

Therefore, we can conclude that $\bar{X}$ may not be consistent if $\rho_{j-i} = \rho$ for all $i \neq j$.

b) We will now try to show that $V(\bar{X}) \to 0$ if $|\rho_{j-i}| \leq M\gamma^{j-i}$ with $|\gamma| < 1$. For all $\epsilon > 0$ there exists a k such that $|\gamma|^{k+1} < \epsilon$. Let's consider

$$
\begin{aligned}
|\frac{1}{n^2}(n\sigma^2+\sum_{i<j}\rho_{ij})| &\le \frac{\sigma^2}{n}+\frac{M}{n^2}\sum_{i<j}|\gamma|^{j-i}\\
&= \frac{\sigma^2}{n}+\frac{2M}{n^2}[(n-1)|\gamma|+(n-2)|\gamma|^2+\ldots+|\gamma|^{n-1}]\\
&= \frac{\sigma^2}{n}+\frac{2M}{n^2}[(n-1)|\gamma|+(n-2)|\gamma|^2+\ldots+(n-k)|\gamma|^k+(n-k-1)|\gamma|^{k-1}+\ldots+|\gamma|^n\\
&\le \frac{\sigma^2}{n}+\frac{2M}{n^2}[(n-1)+(n-2)+\ldots+(n-k)+\epsilon*((n-k-1)+\ldots+1)]\\
&= \frac{\sigma^2}{n}+\frac{2M}{n^2}[kn-\frac{k(k+1)}{2}+\epsilon\frac{(n-k-1)(n-k)}{2}]\\
&\le \frac{\sigma^2}{n}+\frac{2M}{n^2}[kn+\epsilon\frac{(n-k-1)(n-k)}{2}]\to_{n\to\infty} M\epsilon
\end{aligned}
$$

The problem follows as $\epsilon \to 0$.

Problem 7.9

(a). If an Accept–Reject algorithm is used with a wrong bound M, the probability of acceptance of a given $y \sim g(y)$ is no longer $f(x)/Mg(x)$ but

$$\min\{f(x)/Mg(x), 1\} = \min\{f(x), Mg(x)\}/Mg(x)\,.$$

Therefore, the distribution of the accepted value is given by

$$\mathbb{P}(X \le x) = \int_{-\infty}^{x}\int_{0}^{\min\{f(x)/Mg(x),1\}} g(y)\,\mathrm{d}u\,\mathrm{d}y = \int_{-\infty}^{x} \min\{f(x), Mg(x)\}/M\,\mathrm{d}y\,,$$

i.e. its density is

$$\tilde{f}(x) \propto \min\{f(x), Mg(x)\}\,.$$

(b). Since we are generating from $\tilde{f}$ rather than from f, we can correct for this bias by using $\tilde{f}$ as a proposal and f as a target in an Metropolis–Hastings algorithm. The probability of acceptance of y_t is then

$$
\min\left\{1, \frac{f(y_t)\tilde{f}(x^{(t)})}{f(x^{(t)})\tilde{f}(y_t)}\right\} =
\begin{cases}
\min\left\{1, \frac{f(y_t)g(x^{(t)})}{g(y_t)f(x^{(t)})}\right\} & \text{if } \frac{f(y_t)}{g(y_t)} > M \text{ and } \frac{f(x^{(t)})}{g(x^{(t)})} > M\\
\frac{Mg(x^{(t)})}{f(x^{(t)})} & \text{if } \frac{f(y_t)}{g(y_t)} < M \text{ and } \frac{f(x^{(t)})}{g(x^{(t)})} > M\\
1 & \text{if } \frac{f(y_t)}{g(y_t)} > M \text{ and } \frac{f(x^{(t)})}{g(x^{(t)})} < M\\
1 & \text{otherwise.}
\end{cases}
$$

3 Problem 7.30a

Consider $X_1 \sim f \Rightarrow X_t \sim f \forall t$.

Moreover, $Y_t \sim q(y|x_t) \Rightarrow f_{X_t,Y_t}(x,y) = f(x)q(y|x)$.

Now, let's consider

$$
\begin{aligned}
E\frac{f(y_t)}{q(y_t|x_t)}h(y_t) &= \int \frac{f(y_t)}{q(y_t|x_t)}h(y_t)f(x,y)dxdy \\
&= \int\int h(y)f(y)f(x)dxdy = \int h(y)f(y)dy = E_f h(x)
\end{aligned}
$$

Problem 7.39

Write

$$
\begin{aligned}
\lim_{N\longrightarrow\infty}\frac{1}{N}\text{var}\left[\sum_{t=1}^{N}h(X_1^{(t)})\right] &- \lim_{N\longrightarrow\infty}\frac{1}{N}\text{var}\left[\sum_{t=1}^{N}h(X_2^{(t)})\right] \\
&= \sum_{i=1}^{m}\sum_{j=1}^{m}h_i c_{ij}^1 h_j - \sum_{i=1}^{m}\sum_{j=1}^{m}h_i c_{ij}^2 h_j \\
&= \sum_{i=1}^{m}\sum_{j=1}^{m}h_i(\pi_i(z_{ij}^1 - z_{ij}^2) + \pi_j(z_{ji}^1 - z_{ji}^2))h_j \\
&= \sum_{i=1}^{m}\sum_{j=1}^{m}h_i\pi_i(z_{ij}^1 - z_{ij}^2)h_j + \sum_{i=1}^{m}\sum_{j=1}^{m}h_i\pi_j(z_{ji}^1 - z_{ji}^2)h_i \\
&\quad \text{(Exchanging i and j of the second term,)} \\
&= 2\sum_{i=1}^{m}\sum_{j=1}^{m}h_i\pi_i(z_{ij}^1 - z_{ij}^2)h_j = 2\mathbf{h}^T\Pi(\mathbb{Z}^1 - \mathbb{Z}^2)\mathbf{h}
\end{aligned}
$$

From exchanging summation index i and j, we know $\Pi\mathbb{Z}$ is symmetric, hence $\Pi\mathbb{Z} = \mathbb{Z}^T\Pi$, and Π is a diagonal matrix for each diagonal term is $\Pi_{ij} = \pi_i$. Also, from Kenemy and Snell 1969, $\mathbb{Z} = (\mathbb{I} - (\mathbb{P} - \mathbb{A}))^{-1}$, where $\mathbb{A}$ is from problem 6.10.

From this, we consider $\mathbf{h}^T\Pi\mathbb{Z}\mathbf{h}$. If we take derivative w.r.t P_{kl}, where $(k \neq l)$, and if it is negative, then it proves $2\mathbf{h}^T\Pi(\mathbb{Z}^1 - \mathbb{Z}^2)\mathbf{h} \geq 0$. We write,

$$
\begin{aligned}
\frac{\partial \mathbf{h}^T \Pi \mathbb{Z} \mathbf{h}}{\partial P_{kl}} &= \mathbf{h}^T \Pi \frac{\partial \mathbb{Z}}{\partial P_{kl}} \mathbf{h} \\
& \quad \text{From} \left(\frac{\partial \mathbb{Z}}{\partial P_{kl}} = -\mathbb{Z} \frac{\partial \mathbb{Z}^{-1}}{\partial P_{kl}} \mathbb{Z} \right), \\
&= -\mathbf{h}^T \Pi \mathbb{Z} \frac{\partial (\mathbb{I} - \mathbb{P} + \mathbb{A})}{\partial P_{kl}} \mathbb{Z} \mathbf{h} \\
&= -\mathbf{h}^T \Pi \mathbb{Z} \frac{\partial (-\mathbb{P})}{\partial P_{kl}} \mathbb{Z} \mathbf{h} \\
&= -(\mathbb{Z}\mathbf{h})^T \Pi \frac{\partial (-\mathbb{P})}{\partial P_{kl}} \mathbb{Z} \mathbf{h}
\end{aligned}
$$

Let $-\Pi \frac{\partial (\mathbb{P})}{\partial P_{kl}} = \mathbb{Q}$. Then by theorem 6.46, $\mathbb{P}$ is reversible, so that we can remove dependencies in the transition matrix with P_{kl}, by setting $P_{kk} = 1 - \sum_{j=1, j\neq k}^{m} P_{kj}$, and by reversibility, $P_{lk} = \frac{\pi_k}{\pi_l} P_{kl}$. Then $\mathbb{Q}$ has all zeros except following elements with the values, $\mathbb{Q}_{(k,k)} = \pi_k$, $\mathbb{Q}_{(k,l)} = -\pi_k$, $\mathbb{Q}_{(l,k)} = -\pi_k$, $\mathbb{Q}_{(l,l)} = \pi_k$. Finally, we check whether $\mathbb{Q}$ is positive definite, by calculating $\mathbf{x}^T \mathbb{Q} \mathbf{x}$, since it is $\pi_k (x_k - x_l)^2 \geq 0$, hence true. Therefore,

$$
\frac{\partial \lim_{N \longrightarrow \infty} \frac{1}{N} \text{var} \left[\sum_{t=1}^{N} h(X^{(t)}) \right]}{\partial P_{kl}} = \frac{\partial \mathbf{h}^T \Pi \mathbb{Z} \mathbf{h}}{\partial P_{kl}} = -(\mathbb{Z}\mathbf{h})^T \Pi \frac{\partial (-\mathbb{P})}{\partial P_{kl}} \mathbb{Z} \mathbf{h} \leq 0
$$

This result implies that $\lim_{N \longrightarrow \infty} \frac{1}{N} \text{var} \left[\sum_{t=1}^{N} h(X^{(t)}) \right]$ is decreasing function of P_{kl}, hence if $\mathbb{P}_1 \leq \mathbb{P}_2$, then

$$
\lim_{N \longrightarrow \infty} \frac{1}{N} \text{var} \left[\sum_{t=1}^{N} h(X_1^{(t)}) \right] - \lim_{N \longrightarrow \infty} \frac{1}{N} \text{var} \left[\sum_{t=1}^{N} h(X_2^{(t)}) \right] \geq 0 \square
$$

Problem 7.44

(a)

To show

$$
L(\boldsymbol{\beta}, \phi, D | \mathbf{y}) = \int \prod_{i=1}^{n} f(y_i | \mathbf{b}, \beta_i, \phi) f_{\mathbf{b}}(\mathbf{b} | D) d\mathbf{b}
$$

, where $\mathbf{b}$ is unobserved, and $\mathbf{y}$ is observed. By the link function $h(\xi_i) = \mathbf{x}_i' \beta_i + z_i' \mathbf{b}$, $\mathbf{y}$ depends on $\mathbf{b}$ through D. Therefore,

$$\begin{aligned}
L(\boldsymbol{\beta}, \phi, D|\mathbf{y}) &= \int f(\mathbf{y}, \mathbf{b}|\boldsymbol{\beta}, \phi, D)d\mathbf{b} \\
&= \int f(\mathbf{y}, |\boldsymbol{\beta}, \phi)f(\mathbf{b}|D)d\mathbf{b} \\
&= \int \prod_{i=1}^{n} f(y_i|\mathbf{b}, \beta_i, \phi)f_{\mathbf{b}}(\mathbf{b}|D)d\mathbf{b}\square
\end{aligned}$$

(b)

$$\begin{aligned}
L_w &= L(\boldsymbol{\beta}, \phi, D|\mathbf{y}, \mathbf{b}) = \prod_{i=1}^{n} f(y_i|\mathbf{b}, \beta_i, \phi)f_{\mathbf{b}}(\mathbf{b}|D) \\
\log L_w &= \sum_{i=1}^{n} \log f(y_i|\mathbf{b}, \beta_i, \phi) + \log f_{\mathbf{b}}(\mathbf{b}|D)\square
\end{aligned}$$

(c)

At each step, EM algorithm maximizes:

$$E_{\hat{\theta}_{(m)}}(\log L(\boldsymbol{\beta}, \phi, D|\mathbf{y}, \mathbf{b}))$$

,where $\hat{\theta}_{(m)} = \{\hat{\boldsymbol{\beta}}_{(m)}, \hat{\phi}_{(m)}, \hat{D}_{(m)}\}$, and by theorem 5.16 and 5.17, $\lim_{m\to\infty} \hat{\theta}_{(m)} = \hat{\theta}^{MLE}_{(m)}$. Then,

$$E_{\hat{\theta}_{(m)}}(\sum_{i=1}^{n} \log f(y_i|\mathbf{b}, \beta_i, \phi) + \log f_{\mathbf{b}}(\mathbf{b}|D)) = \sum_{i=1}^{n} E_{\hat{\theta}_{(m)}}(\log f(y_i|\mathbf{b}, \beta_i, \phi)) + E_{\hat{\theta}_{(m)}}(\log f_{\mathbf{b}}(\mathbf{b}|D))$$

Since independent variables in the first$(\boldsymbol{\beta}, \phi)$, and the second$(D)$ terms do not depend on each other, hence we can maximize them separately, then they are the step 2 and 3 in the given algorithm, hence proved.