Course: STA6934- Monte Carlo Statistical Methods

Instructor: Professor Casella

Assignment 3

Problem 5.1

After the simulation, both numerical maximizer built in R and MC method returns the same objective value 3.832543. However, the $x = 0.3791249$ in numerical maximizer, and $x = 0.563327$ in MC method.

Problem 5.9ab

a) The conditional pdf of $X_i|Z_i = z_i$ is $f_{X_i|Z_i=z_i}(x_i|z_i) = z_i g(x_i) + (1 - z_i)h(x_i)$ and the pdf for $Z_i \sim Ber(\theta)$ is $f_{Z_i}(z_i) = \theta^{z_i}(1-\theta)^{1-z_i}$

Thus, the complete data likelihood is

$$\begin{aligned} L^c(\theta|x,z) &= \Pi_{i=1}^n f_{X_i|Z_i=z_i}(x_i|z_i) f_{Z_i}(z_i) \\ &= \Pi_{i=1}^n [z_i g(x_i) + (1-z_i)h(x_i)]\theta^{z_i}(1-\theta)^{1-z_i} \end{aligned}$$

b) Notice that

$$f_{Z_i|X_i=x_i}(z_i|x_i) = \frac{[z_i g(x_i) + (1-z_i)h(x_i)]\theta^{z_i}(1-\theta)^{1-z_i}}{\theta g(x_i) + (1-\theta)h(x_i)}$$

$$\Rightarrow E[Z_i|\theta, x_i] = 0 * f_{Z_i|X_i=x_i}(0|x_i) + 1 * f_{Z_i|X_i=x_i}(1|x_i) = \frac{\theta g(x_i)}{\theta g(x_i) + (1-\theta)h(x_i)}$$

Now, let's consider

$$\begin{aligned} E_{\hat{\theta}_m}[logL^c(\theta|x,z)] &= \sum_{i=1}^n \int logf(x_i, z_i) * k(z_i|\hat{\theta}_m, x_i) dz_i \\ &= \sum_{i=1}^n \int [log(z_i g(x_i) + (1-z_i)h(x_i)) + z_i log\theta + (1-z_i)log(1-\theta)] * k(z_i|\hat{\theta}_m, x_i) dz_i \\ &= \sum_{i=1}^n [logg(x_i) + log\theta] * k(1|\hat{\theta}_m, x_i) + \sum_{i=1}^n [logh(x_i) + log(1-\theta)] * k(0|\hat{\theta}_m, x_i) \end{aligned}$$

take the first derivative w.r.t θ and set to 0 to solve

$$\begin{aligned}\hat{\theta}_{m+1} &= \frac{1}{n}\sum_{i=1}^{n} k(1|\hat{\theta}_m, x_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} E(Z_i|\hat{\theta}_m, x_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\theta}_m g(x_i)}{\hat{\theta}_m g(x_i) + (1-\hat{\theta}_m)h(x_i)}\end{aligned}$$

Problem 5.10

(a). The likelihood is

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{12} [p\lambda e^{-\lambda x_i} + (1-p)\mu e^{-\mu x_i}],$$

and the complete-data likelihood is

$$L^c(\theta|\mathbf{x}, \mathbf{z}) = \prod_{i=1}^{12} [p\lambda e^{-\lambda x_i}\mathbb{I}_{(z_i=1)} + (1-p)\mu e^{-\mu x_i}\mathbb{I}_{(z_i=2)}],$$

where $\theta = (p, \lambda, \mu)$ denotes the parameter. Let

$$H(x_1, \cdots, x_1 2, Z_1, \cdots, Z_{12}) = \prod_{i=1}^{12} (\lambda e^{-\lambda x_i} I(Z_i = 1) + \mu e^{-\mu x_i} I(Z_i = 2))$$

then it is easy to check the likelihood $h(p, \lambda, \mu)$ can be expressed as $E[H(x_1, \cdots, x_1 2, Z_1, \cdots, Z_{12})]$.

(b). Let $g(z)$ be a Bernoulli distribution with parameter q, that is, $P(z = 1) = 1 - P(z = 2) = q$. The approximation (5.26) writes

$$\hat{h}_m(\theta) = \frac{1}{m}\sum_{j=1}^{m} \frac{f(\mathbf{z}^{(j)}|\mathbf{x}, \theta)}{g(\mathbf{z}^{(j)})},$$

where $\mathbf{z}^{(j)} = (z_1^{(j)}, ..., z_{12}^{(j)})$ are simulated from g. Then,

$$\hat{h}_m(\theta) = \frac{1}{m}\sum_{j=1}^{m}\prod_{i=1}^{12} \frac{p^2\lambda^2 e^{-2\lambda x_i}/q\mathbb{I}_{(z_i^{(j)}=1)} + (1-p)^2\mu^2 e^{-2\mu x_i}/(1-q)\mathbb{I}_{(z_i^{(j)}=2)}}{p\lambda e^{-\lambda x_i} + (1-p)\mu e^{-\mu x_i}},$$

and the approximation used in Example (5.25) becomes

$$\hat{h}_\eta(\theta) = \frac{1}{m}\sum_{j=1}^{m}\prod_{i=1}^{12} \left[\frac{p\lambda e^{-\lambda x_i}}{p_0\lambda_0 e^{-\lambda_0 x_i}}\mathbb{I}_{(z_i^{(j)}=1)} + \frac{(1-p)\mu e^{-\mu x_i}}{(1-p_0)\mu_0 e^{-\mu_0 x_i}}\mathbb{I}_{(z_i^{(j)}=2)}\right].$$

For every given value $\eta = (p_0, \lambda_0, \mu_0)$ of the parameter, the variance of the estimator of h is finite for every $\theta = (p, \lambda, \mu)$. In term of computation, the approximation of Example (5.25) is easier than (5.26) because the denominator no longer depends on θ as in (5.26) and it thus leads to faster estimation.

The EM algorithm is based on the optimization of the expected log-likelihood

$$Q(\theta|\hat{\theta}_{(j)}, \mathbf{x}) = \sum_{i=1}^{12} [\log(p\lambda e^{-\lambda x_i}) P_{\hat{\theta}_{(j)}}(z_i = 1) + \log((1-p)\mu e^{-\mu x_i}) P_{\hat{\theta}_{(j)}}(z_i = 2)].$$

The arguments of the maximum are

$$\begin{cases} \hat{p}_{(j+1)} = \hat{P}/12 \\ \hat{\lambda}_{(j+1)} = \hat{S}_1/\hat{P} \\ \hat{\mu}_{(j+1)} = \hat{S}_2/\hat{P}, \end{cases}$$

where

$$\begin{cases} \hat{P} = \sum_{i=1}^{12} P_{\hat{\theta}_{(j)}}(z_i = 1) \\ \hat{S}_1 = \sum_{i=1}^{12} x_i P_{\hat{\theta}_{(j)}}(z_i = 1) \\ \hat{S}_1 = \sum_{i=1}^{12} x_i P_{\hat{\theta}_{(j)}}(z_i = 2), \end{cases}$$

with

$$P_{\hat{\theta}_{(j)}}(z_i = 1) = 1 - P_{\hat{\theta}_{(j)}}(z_i = 2) = \frac{\hat{p}_{(j)}\hat{\lambda}_{(j)} e^{-\hat{\lambda}_{(j)} x_i}}{\hat{p}_{(j)}\hat{\lambda}_{(j)} e^{-\hat{\lambda}_{(j)} x_i} + (1 - \hat{p}_{(j)})\hat{\mu}_{(j)} e^{-\hat{\mu}_{(j)} x_i}}.$$

Problem 5.17

.1 (a)

$$\begin{aligned} L(\beta|\mathbf{x}, y) &= \prod_i^n f(y_i|x_i, \beta) \\ &= \prod_i^n (\Phi(x_i\beta))^{y_i} (1 - \Phi(x_i\beta))^{1-y_i} \quad \square \end{aligned}$$

.2 (b)

The EM algorithm specifies:

$$\beta_{(m)} = arg\max_{\beta} E_{\beta_{m-1}}(\log L(\beta|\mathbf{x}, y, z))$$

Because y_i is a function of z_i, we don't need to care about y_i. Hence,

$$\log L(\beta|\mathbf{x}, y, z)) \propto \sum_i^n \left(-\frac{(z_i - \mathbf{x}_i'\beta)^2}{2}\right) = -\frac{1}{2}(Z - X^T\beta)^T(Z - X^T\beta)$$

Set

$$E_{\beta_{m-1}}\left[\frac{\partial \log L(\beta|\mathbf{x}, y, z))}{\partial \beta}\right] = 0$$

Then

$$XE_{\beta_{m-1}}[Z|X] = X^TX\beta$$

So

$$\beta_m = (X^TX)^{-1}E_{\beta_{m-1}}[Z|X]$$

.3 (c)

If $y_i = 1$, then

$$\begin{aligned} E_\beta(Z_i|x_i, y_i = 1) &= \frac{1}{\Phi(x_i^T\beta)}\left[\int_0^{+\infty}(Z_i - x_i^T\beta)\phi(Z_i - x_i^T\beta)dZ_i + x_i^T\beta\int_0^{+\infty}\phi(Z_i - x_i^T\beta)dZ_i\right] \\ &= \frac{1}{\Phi(x_i^T\beta)}\left(\int_0^{+\infty} -\phi'(Z_i - x_i^T\beta)dZ_i + x_i^T\beta(1 - \Phi(-x_i^T\beta))\right) \\ &= \frac{\phi(x_i^T\beta))}{\Phi(x_i^T\beta))} + x_i^T\beta \end{aligned}$$

Similarly, if $y_i = 0$, then

$$E_\beta(Z_i|x_i, y_i = 0) = \frac{-\phi(x_i^T\beta))}{1 - \Phi(x_i^T\beta))} + x_i^T\beta$$

Problem 5.24

The observed data likelihood is $L(\theta|x) = [1 - F(216)][1 - F(244)]\Pi_{i=1}^m f(x_i; \alpha, \beta, \gamma)$.

The complete data likelihood is $L^c(\theta|x, z) = f(z_1)f(z_2)\Pi_{i=1}^m f(x_i)$ where $z = (z_1, z_2)$ is the missing data.Thus, we obtain

$$k(z|\theta, x) = \frac{L^c(\theta|x, z)}{L(\theta|x)} = \frac{f(z_1)f(z_2)}{[1 - F(216)][1 - F(244)]}$$

and

$$\begin{aligned} E_{\theta_n} log L^c(\theta|x, z) &= \int [log f(z_1) + log f(z_2) + \sum_{i=1}^m og f(x_i)] * k(z|\theta, x)dz \\ &= \sum_{i=1}^m log f(x_i) + E_{\theta_n}[log f(z_1)] + E_{\theta_n}[log f(z_2)] \qquad (*) \end{aligned}$$

a) For $\gamma = 100$ and $\alpha = 3$

Consider $(*)$, take the first derivative w.r.t β and set to 0, i.e

$$0 = -\frac{(m+2)\alpha}{\beta} + \alpha\beta^{-\alpha-1}[\sum_{i=1}^{m}(x_i-\gamma)^\alpha + E_{\theta_n}[(z_1-\gamma)^\alpha] + E_{\theta_n}[(z_2-\gamma)^\alpha]]$$

```
#a=3,c=100
> x=c(143,164,188,188,190,192,206,209,213,216,220,227,230,234,246,265,304);
> f=function(z,a,b,c){((z-c)^(a+2))*a*exp(-((z-c)/b)^a)*b^(-a)};
> g=function(z,a,b,c){((z-c)^(a-1))*a*exp(-((z-c)/b)^a)*b^(-a)};
> beta=c(1:8);
> beta[1]=100;
> for(i in 1:7){
+ beta[i+1]=((sum((x-100)^3)+integrate(f,216,Inf,a=3,b=beta[i],c=100)$value/inte
+ }
> plot(beta,type='l',main="EM sequence for alpha=3,gamma=100");
> #the EM estimator is
> beta[8];
[1] 130.4318
```

b) For $\gamma = 100$ and α unknown

```
#c=100 and a unknown
> x=c(143,164,188,188,190,192,206,209,213,216,220,227,230,234,246,265,304);
> theta=array(0:0,c(20,2));
> theta[1,]=c(3,130);
> for(i in 1:19){
+ #generate z1 and z2
+ we=theta[i,2]*(-log(runif(1000)))^(1/theta[i,1])+100;
+ size1=sum(as.numeric(we>216));
+ z1=c(1:size1);
+ k=1;
+ for(j in 1:1000){if(we[j]>216){z1[k]=we[j]; k=k+1} }
+ size2=sum(as.numeric(we>244));
+ z2=c(1:size2);
+ k=1;
+ for(l in 1:1000){if(we[l]>244){z2[k]=we[l]; k=k+1} }
+ #EM step
```

```
+ ex=function(p){-(19*log(p[1])-19*p[1]*log(p[2])+(p[1]-1)*sum(log(x-100))-sum((
+ theta[i+1,]=nlm(ex,c(theta[i,]))$estimate;
+ }
> par(mfrow=c(1,2));
> plot(theta[,1],type='l',ylim=c(3,4),main="EM sequence for alpha(gamma=100)");
> plot(theta[,2],type='l',ylim=c(130,135),main="EM sequence for beta(gamma=100)"
> #the MCEM estimator is
> theta[20,];
[1]   3.362795 131.786271
```

c) For all parameters unknown

```
> #all parameters unknown
> x=c(143,164,188,188,190,192,206,209,213,216,220,227,230,234,246,265,304);
> theta=array(0:0,c(20,3));
> theta[1,]=c(3,131,100);
> for(i in 1:19){
+ #generate z1 and z2
+ we=theta[i,2]*(-log(runif(1000)))^(1/theta[i,1])+100;
+ size1=sum(as.numeric(we>216));
+ z1=c(1:size1);
+ k=1;
+ for(j in 1:1000){if(we[j]>216){z1[k]=we[j]; k=k+1} }
+ size2=sum(as.numeric(we>244));
+ z2=c(1:size2);
+ k=1;
+ for(l in 1:1000){if(we[l]>244){z2[k]=we[l]; k=k+1} }
+ #EM step
+ ex=function(p){-(19*log(p[1])-19*p[1]*log(p[2])+(p[1]-1)*sum(log(x-p[3]))-sum(
+ theta[i+1,]=nlm(ex,c(theta[i,]))$estimate;
+ }
> par(mfrow=c(2,2));
> plot(theta[,1],type='l',main="EM sequence for alpha");
> plot(theta[,2],type='l',main="EM sequence for beta");
> plot(theta[,3],type='l',main="EM sequence for gamma");
>
> #the MCEM estimator is
> theta[20,];
```

```
[1]   2.792445 108.797566 120.883778
```

Problem 5.29

a) Let's consider the complete data log likelihood

$$\begin{aligned} logL(\theta|x,z) &= (z_2+x_4)log\theta + (x_2+x_3)log(1-\theta) \\ \Rightarrow \frac{\delta}{\delta\theta}logL(\theta|x,z) &= \frac{z_2+x_4}{\theta} - \frac{x_2+x_3}{1-\theta} \\ \Rightarrow \frac{\delta^2}{\delta\theta^2}logL(\theta|x,z) &= -\frac{z_2+x_4}{\theta^2} - \frac{x_2+x_3}{(1-\theta)^2} \end{aligned}$$

It follows that

$$\begin{aligned} E\frac{\delta}{\delta\theta}logL(\theta|x,z) &= \frac{E(z_2)+x_4}{\theta} - \frac{x_2+x_3}{1-\theta} \\ \Rightarrow E[\frac{\delta}{\delta\theta}logL(\theta|x,z)]^2 &= \frac{Ez_2^2+2x_4Ez_2+x_4^2}{\theta^2} + (\frac{x_2+x_3}{1-\theta})^2 - \frac{2(x_2+x_3)(Ez_2+x_4)}{\theta(1-\theta)} \\ E\frac{\delta^2}{\delta\theta^2}logL(\theta|x,z) &= -\frac{Ez_2+x_4}{\theta^2} - \frac{x_2+x_3}{(1-\theta)^2} \end{aligned}$$

Therefore, applying (5.22), we obtain

$$\begin{aligned} \frac{\delta^2}{\delta\theta^2}logL(\theta|x) &= -\frac{x_4}{\theta^2} - \frac{x_2+x_3}{(1-\theta)^2} - \frac{Ez_2}{\theta^2} + \frac{Ez_2^2}{\theta^2} + (\frac{2x_4}{\theta^2} - 2\frac{x_2+x_3}{\theta(1-\theta)})Ez_2 + \frac{x_4^2}{\theta} + \frac{(x_2+x_3)^2}{(1-\theta)^2} \\ &\quad -2\frac{(x_2+x_3)x_4}{\theta(1-\theta)} - \frac{(x_2+x_3)^2}{(1-\theta)^2} - \frac{x_4^2}{\theta^2} - \frac{(Ez_2)^2}{\theta^2} + \frac{2x_4(x_2+x_3)}{\theta(1-\theta)} - \frac{2x_4Ez_2}{\theta^2} + \frac{2(x_2+x_3)Ez_2}{\theta(1-\theta)} \\ &= -\frac{Ez_2(1-\theta)^2 + x_4(1-\theta)^2 + (x_2+x_3)\theta^2}{\theta^2(1-\theta)^2} + \frac{1}{\theta^2}Varz_2 \end{aligned}$$

b)

```
> #Reproduce the EM estimator and its standard error
> theta=c(1:25);
> theta[1]=0.6;
> sd=c(1:25);
> x1=125;
> x2=18;
> x3=20;
> x4=34;
```

```
> #do the iteration
> for(i in 1:24){
+ ex=(theta[i]*x1)/(2+theta[i])
+ theta[i+1]=(ex+x4)/(ex+x2+x3+x4)
+ sd[i]=sqrt((ex/theta[i+1]^2+x4/theta[i+1]^2+(x2+x3)/(1-theta[i+1])^2)^(-1))
+ }
> sd[25]=sd[24];
> #plot the EM sequence +- one standard deviation
> plot(theta,ylim=c(0.55,0.70),type='l',xlab="iteration",main="Problem 5.29(b)")
> lines(theta+sd,lty=2);
> lines(theta-sd,lty=2);
> #the EM estimator is
> theta[25];
[1] 0.6268215
> #the standard error is
> sd[25];
[1] 0.04792882
```

c)

```
> #Produce the MCEM estimator and its standard error
> nsim=100;
> theta=c(1:25);
> theta[1]=0.6;
> sd=c(1:25);
> x1=125;
> x2=18;
> x3=20;
> x4=34;
> #do the iteration
> for(i in 1:24){
+ ex=mean(rbinom(nsim,x1,theta[i]/(2+theta[i])))
+ theta[i+1]=(ex+x4)/(ex+x2+x3+x4)
+ sd[i]=sqrt((ex/theta[i+1]^2+x4/theta[i+1]^2+(x2+x3)/(1-theta[i+1])^2)^(-1))
+ }
> sd[25]=sd[24];
> #plot the MCEM sequence +- one standard deviation
```

```
> plot(theta,ylim=c(0.55,0.70),type='l',xlab="iteration",main="Problem 5.29(c)")
> lines(theta+sd,lty=2);
> lines(theta-sd,lty=2);
> #the MCEM estimator is
> theta[25];
[1] 0.6277794
> #the standard error is
> sd[25];
[1] 0.04784231
```

Problem 6.7

The transition matrix

$$\mathbb{P} = \begin{pmatrix} 0.0 & 0.4 & 0.6 & 0.0 & 0.0 \\ 0.65 & 0.0 & 0.35 & 0.0 & 0.0 \\ 0.32 & 0.68 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.12 & 0.88 \\ 0.0 & 0.0 & 0.0 & 0.56 & 0.44 \end{pmatrix},$$

has two diagonal blocks that show two irreducible classes $\{1, 2, 3\}$ and $\{4, 5\}$. The chain is therefore reducible. On each class, it is aperiodic.

Problem 6.21

$$\begin{aligned} P(Z_{n+1} = y | Z_n = x) &= \frac{P(Z_{n+1} = y, Z_n = x)}{P(Z_n = x)} \\ &= \frac{P(Y_{n+1} = y - x, Z_n = x)}{P(Z_n = x)} \\ &= \frac{P(Y_{n+1} = y - x)P(Z_n = x)}{P(Z_n = x)} \\ &= P(Y_{n+1} = y - x) \end{aligned}$$

Thus, Z_n is a Markov chain. However, it's NOT irreducible.

Problem 6.12a

Let's consider

$$\begin{aligned} diagP^1 &= (0,0,0,0,0) \\ diagP^2 &= (0.432, 0.493, 0.43, 0.4928, 0.5087) \\ diagP^3 &= (0.2924, 0.2924, 0.2896, 0.00, 0.0028) \\ diagP^4 &= (0.2958, 0.3473, 0.2968, 0.2525, 0.270) \\ diagP^5 &= (0.3284, 0.34625, 0.3245, 0.00181, 0.0094) \end{aligned}$$

Thus, the periods of each state are $d_1 = d_2 = d_3 = d_4 = d_5 = 1$ which implies that the chain is aperiodic.

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

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

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