Course: STA6934- Monte Carlo Statistical Methods

Instructor: Professor Casella

Assignment 1

Problem 1.1

```
mix <- function(x, e, u1, u2, sigma1, sigma2)
(e*dnorm(x, mean=u1, sd=sigma1)+ (1-e)*dnorm(x, mean=u2, sd=sigma2))
min <- function(x, u1, u2, sigma1, sigma2)
((1-pnorm((x-u1)/sigma1))/sigma2*dnorm((x-u2)/sigma2)
+ (1-pnorm((x-u2)/sigma2))/sigma1*dnorm((x-u1/sigma1)))

censplot <- function(e,u1,u2,sigma1,sigma2)
{
lowpoint=pmin(u1,u2)-3*pmax(sigma1,sigma2)
uppoint=pmax(u1,u2)+3*pmax(sigma1,sigma2)
xplot <- seq(from=lowpoint, to=uppoint, length=1000)
mixxplot <- mix(xplot, e, u1, u2, sigma2, sigma2)
minxplot <- min(xplot, u1, u2, sigma1, sigma2)
plot(xplot, mixxplot, xlim=c(lowpoint, uppoint), ylim=c(0,0.8),
type="l", lty=1, ylab="density", col="blue")
lines(xplot, minxplot, lty=2, col="red")
legend(lowpoint, 0.8, c("mixed", "minimum"), lty=c(1,2), col=c("blue", "red"))
mtext(bquote(paste("u=", .(e), ",Normal(", .(u1),",", .(sigma1),"),
Normal(", .(u2),",", .(sigma2), ")")))
}

#library(lattice)
#trellis.device(pdf, file="HW1p1", height=20, width=17)
par(mfrow=c(3,2))
censplot(0.3,1,1,1,1)
censplot(0.3,-1,1,1,1)
censplot(0.3,1,1,2,1)
censplot(0.3,-3,1,3,1)
censplot(0.5,3,1,2,1)
censplot(0.5,-3,1,1,3)
```

Problem 1.4 In order to find an explicit form of the integral

$$\int_{\omega}^{\infty} \alpha\beta x^{\alpha-1} e^{-\beta x^{\alpha}} dx,$$

we use the change of variable $y = x^{\alpha}$. We have $dy = \alpha x^{\alpha-1} dx$ and the integral becomes

$$\int_{\omega}^{\infty} \alpha\beta x^{\alpha-1} e^{-\beta x^{\alpha}} dx = \int_{\omega^{\alpha}}^{\infty} \beta e^{-\beta y} dy = e^{-\beta\omega^{\alpha}}\,.$$

Problem 1.7 The density f of the vector Y_n is

$$f(y_n, \mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \mu}{\sigma}\right)^2\right), \quad \forall y_n \in \mathbb{R}^n, \forall(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$$

This function is strictly positive and the first and second order partial derivatives with respect to μ and σ exist and are positive. The same hypotheses are satisfied for the log-likelihood function

$$\log(L(\mu, \sigma, y_n)) = -n \log\sqrt{2\pi} - n \log\sigma - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \mu}{\sigma}\right)^2$$

thus we can find the ML estimator of μ and σ^2. The gradient of the log-likelihood is

$$\nabla \log(L) = \left\{\begin{array}{l} \frac{\partial \log(L(\mu,\sigma,y_n))}{\partial\mu} \\ \frac{\partial \log(L(\mu,\sigma,y_n))}{\partial\sigma} \end{array}\right. = \left\{\begin{array}{l} \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu) \\ -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{\sigma^3} \end{array}\right.$$

if we equate the gradient to the null vector, $\nabla \log(L) = 0$ and solve the resulting system in μ and σ, we find

$$\begin{aligned} \hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}\,, \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = s^2\,. \end{aligned}$$

Problem 1.13

For $X \sim \mathcal{W}e(\alpha, \beta, \gamma)$, where $\alpha > 0$ is the shape parameter, $\beta > 0$ is the scale parameter, and γ is the translation parameter, the density is given as:

$$f(x; \alpha, \beta, \gamma) = \frac{\alpha}{\beta}\left(\frac{x - \gamma}{\beta}\right)^{\alpha-1} e^{-(\frac{x-\gamma}{\beta})^{\alpha}}, \quad \text{for } x \geq \gamma$$

For $X_1, \ldots, X_n$ are *iid* as $\mathcal{W}e(\alpha, \beta, \gamma)$, the likelihood function is given as:

$$
\begin{aligned}
L(\alpha, \beta, \gamma | x_1, \ldots, x_n) &= \prod_{i=1}^{n} f(x_i | \alpha, \beta, \gamma) \\
&= (\frac{\alpha}{\beta})^n \cdot \prod_{i=1}^{n} (\frac{x_i - r}{\beta})^{\alpha - 1} \cdot exp\{-\sum (\frac{x_i - r}{\beta})^{\alpha}\} \\
l(\alpha, \beta, \gamma | x_1, \ldots, x_n) &= \log L(\cdot) \\
&= n[log\alpha - log\beta] + (\alpha - 1) \cdot \sum [log(x_i - r) - log\beta] - \sum (\frac{x_i - r}{\beta})^{\alpha} \\
&= nlog\alpha - n\alpha log\beta + (\alpha - 1) \cdot \sum log(x_i - r) - \sum (\frac{x_i - r}{\beta})^{\alpha} \\
\frac{\partial}{\partial \alpha} l(\cdot) &= \frac{n}{\alpha} - nlog\beta + \sum log(x_i - r) - \sum [log \frac{x_i - r}{\beta} \cdot (\frac{x_i - r}{\beta})^{\alpha}] \\
\frac{\partial}{\partial \beta} l(\cdot) &= -\frac{n\alpha}{\beta} + \sum \alpha \beta^{-(\alpha+1)} \cdot (x_i - r)^{\alpha} \\
\frac{\partial}{\partial \gamma} l(\cdot) &= (\alpha - 1) \cdot \sum (\frac{-1}{x_i - r}) + \sum \frac{\alpha}{\beta^{\alpha}} (x_i - r)^{\alpha - 1}
\end{aligned}
$$

1. $\gamma = 100, \ \alpha = 3$

   n=19, in this case

   $$\frac{\partial}{\partial \beta} l(\cdot) = -\frac{19 * 3}{\beta} + \sum 3(x_i - 100)^3 \cdot \beta^{-4}$$

   and we get $\hat{\beta} = 125.6846$

   Also we could use the *nlm* function in R to solve this problem. *nlm* function carries out a minimization of the function using a Newton-type algorithm. Thus to find the maximum likelihood estimator for the parameter is equivalent to get the value minimizing the negative log-likelihood. The R code and the output is given as below:

```
>logweib1 <- function (beta)
 { alpha=3; size=19;
   x <- c(143, 164, 188, 188, 190, 192, 206, 209, 213, 216,
          220, 227, 230, 234, 246, 265, 304, 216, 244)
   -(size*log(alpha)-size*alpha*log(beta)+(alpha-1)*sum(log(x-100))
     -sum(((x-100)/beta)^alpha))
 }

>nlm(logweib1,100)
```

```
$minimum
[1] 95.13106

$estimate
[1] 125.6845
```

As we can see from the above, the two beta values are quite close, hence the fitted model is $Weibull(3, 125.6845, 100)$. We will use this method for the following steps.

2. $\gamma = 100,\ \alpha$ unknown;

```
>logweib2 <- function (p)
{  size=19;
   x <- c(143, 164, 188, 188, 190, 192, 206, 209, 213, 216,
          220, 227, 230, 234, 246, 265, 304, 216, 244)
   -(size*log(p[1])-size*p[1]*log(p[2])+(p[1]-1)*sum(log(x-100))
     -sum(((x-100)/p[2])^p[1]))
}
>nlm(logweib2, c(3, 125.6845))    # use the result from (a) as the initial v
$minimum
[1] 94.75059

$estimate
[1] 3.504232 128.104290
```

Hence the fitted model is $Weibull(3.504232, 128.104290, 100)$.

3. γ and α both unknown;

```
>logweib3 <- function (p)
{  size=19;
   x <- c(143, 164, 188, 188, 190, 192, 206, 209, 213, 216,
          220, 227, 230, 234, 246, 265, 304, 216, 244)
   -(size*log(p[1])-size* p[1]*log(p[2])+(p[1]-1)*sum(log(x-p[3]))
```

```
    -sum(((x-p[3])/p[2])^p[1]))
}

>nlm(logweib3,c(3.504232, 128.104290, 100))
# use the result from (b) as the initial value
$minimum
[1] 94.59973

$estimate
[1]   2.849366 105.345531 121.425922
```

Hence the fitted model is $Weibull(2.849366, 105.345531, 121.425922)$.

Problem 1.22

(a). Since $L(\delta, h(\theta)) \geq 0$ by using Fubini's theorem, we get

$$
\begin{aligned}
r(\pi, \delta) &= \int_\Theta \int_\mathcal{X} L(\delta, h(\theta)) f(x|\theta)\pi(\theta) dx d\theta \\
&= \int_\mathcal{X} \int_\Theta L(\delta, h(\theta)) f(x|\theta)\pi(\theta) d\theta dx \\
&= \int_\mathcal{X} \int_\Theta L(\delta, h(\theta)) m(x)\pi(\theta|x) d\theta dx \\
&= \int_\mathcal{X} \varphi(\pi, \delta|x) m(x) dx \,,
\end{aligned}
$$

where m is the marginal distribution of X and $\varphi(\pi, \delta|x)$ is the posterior average cost.

The estimator that minimizes the integrated risk r is therefore, for each x, the one that minmizes the posterior average cost and it is given by

$$
\delta^\pi(x) = \arg\min_\delta \varphi(\pi, \delta|x) \,.
$$

(b). The average posterior loss is given by :

$$
\begin{aligned}
\varphi(\pi, \delta|x) &= \mathbb{E}^\pi \left[L(\delta, \theta)|x \right] \\
&= \mathbb{E}^\pi \left[||h(\theta) - \delta||^2 |x \right] \\
&= \mathbb{E}^\pi \left[||h(\theta)||^2 |x \right] + \delta^2 - 2 < \delta, \mathbb{E}^\pi \left[h(\theta)|x \right] >
\end{aligned}
$$

A simple derivation shows that the minimum is attained for

$$
\delta^\pi(x) = \mathbb{E}^\pi \left[h(\theta)|x \right] \,.
$$

(c). Take m to be the posterior median and consider the auxiliary function of θ, $s(\theta)$, defined as

$$s(\theta) = \begin{cases} -1 & \text{if} \quad h(\theta) < m \\ +1 & \text{if} \quad h(\theta) > m \end{cases}$$

Then s satisfies the propriety

$$\begin{aligned} \mathbb{E}^{\pi}\left[s(\theta)|x\right] &= -\int_{-\infty}^{m} \pi(\theta|x)d\theta + \int_{m}^{\infty} \pi(\theta|x)d\theta \\ &= -\mathbb{P}(h(\theta) < m|x) + \mathbb{P}(h(\theta) > m|x) = 0 \end{aligned}$$

For $\delta < m$, we have $L(\delta,\theta) - L(m,\theta) = |h(\theta) - \delta| - |h(\theta) - m|$ from which it follows that

$$L(\delta,\theta) - L(m,\theta) = \begin{cases} \delta - m = (m-\delta)s(\theta) & \text{if} \quad \delta > h(\theta) \\ m - \delta = m - \delta & \text{if} \quad m < \delta \\ 2h(\theta) - (\delta + m) > (m-\delta)s(\theta) & \text{if} \quad \delta < h(\theta) < m \end{cases}$$

It turns out that $L(\delta,\theta) - L(m,\theta) > (m-\delta)s(\theta)$ which implies that

$$\mathbb{E}^{\pi}\left[L(\delta,\theta) - L(m,\theta)|x\right] > (m-\delta)\mathbb{E}^{\pi}\left[s(\theta)|x\right] = 0\,.$$

This still holds, using similar argument when $\delta > m$, so the minimum of $\mathbb{E}^{\pi}\left[L(\delta,\theta)|x\right]$ is reached at $\delta = m$.

Problem 1.23

(a). When $X|\sigma \sim \mathcal{N}(0,\ \sigma^2)$, $\frac{1}{\sigma^2} \sim \mathcal{G}a(1,\ 2)$, the posterior distribution is

$$\begin{aligned} \pi\left(\sigma^{-2}|X\right) &\propto f(x|\sigma)\pi(\sigma^{-2}) \\ &\propto \frac{1}{\sigma} e^{-\frac{(x^2/2+2)}{\sigma^2}} \\ &= (\sigma^2)^{\frac{3}{2}-1} e^{-\frac{(x^2/2+2)}{\sigma^2}}\,, \end{aligned}$$

which means that $1/\sigma^2 \sim \mathcal{G}a(\frac{3}{2},\ 2 + \frac{x^2}{2})$. The marginal distribution is

$$m(x) = \int f(x|\sigma)\pi(\sigma^{-2})d(\sigma^{-2}) \propto \left(\frac{x^2}{2} + 2\right)^{-\frac{3}{2}}\,,$$

that is, $X \sim \mathcal{T}(2, 0, 2)$.

(b). When $X|\lambda \sim \mathcal{P}(\lambda)$, $\lambda \sim \mathcal{G}a(2,\ 1)$, the posterior distribution is

$$\pi(\lambda) \propto f(x|\lambda)\pi(\lambda) \propto \lambda^{x+1} e^{-2\lambda}$$

which means that $\lambda \sim \mathcal{G}a(x+2,\ 2)$. The marginal distribution is

$$m(x) = \int f(x|\lambda)\pi(\lambda)d\lambda \propto \frac{\Gamma(x+2)}{\sqrt{\pi}2^{x+2}x!} = \frac{(x+1)}{\sqrt{\pi}2^{x+2}}\,.$$

Problem 1.24

(a). Let the interval $[a, b]$ satisfy $\int_a^b f(x)dx = 1 - \alpha$ and $f(a) = f(b) > 0$. Also let $x^* \in [a, b]$ be the mode of $f(x)$. We will show that for any interval $[a', b']$ such that $b' - a' < b - a$, $\int_{a'}^{b'} f(x)dx < 1 - \alpha$, thus proving that $[a, b]$ is the shortest interval. WLOG, assume $a' \leq a$ and split the problem into two cases.

Case 1. Suppose $b' \leq a$. Then $a' \leq b' \leq a \leq x$ and

$$\int_{a'}^{b'} f(x)dx \leq f(b')(b' - a') < f(a)(b - a) \leq \int_a^b f(x)dx = 1 - \alpha.$$

Case 2. Otherwise, assume $b' > a$. Then $b' < b$ and

$$\int_{a'}^{b'} f(x)dx = \int_a^b f(x)dx + \int_{a'}^{a} f(x)dx - \int_b^{b'} f(x)dx$$

Hence we only need to show that $\int_{a'}^{a} f(x)dx - \int_b^{b'} f(x)dx < 0$. Note that $a' \leq a \leq b' \leq b$, which implies $\int_{a'}^{a} f(x)dx \leq f(a)(a - a')$ and $\int_b^{b'} f(x)dx \geq f(b)(b - b')$. Hence

$$\begin{aligned}\int_{a'}^{a} f(x)dx - \int_b^{b'} f(x)dx &\leq f(a)(a - a') - f(b)(b - b') \\ &= f(a)(a - a' - b + b') \\ &= f(a)[(b' - a') - (b - a)] < 0.\end{aligned}$$

(b). If f is strictly monotone on either side of its mode, which we take to be 0, then $f(x) = f(-x)$ for any $x \in \mathbb{R}$. If $[a, b]$ is the shortest intercal such that $\int_a^b f(x)dx = 1 - \alpha$, then

$$f(a) = f(b) = f(-a) = f(-b) \quad \text{where } a < 0 < b.$$

Now that $f(a) = f(-b)$ for $a, -b < 0$ and f is strictly monotone, $a = -b$ must hold.

Problem 1.28

(a). If $X \sim \mathcal{G}(\theta, \beta)$, then

$$\pi(\theta|x, \beta) \propto \pi(\theta) \times (\beta x)^\theta / \Gamma(\theta)$$

and a family of functions $\pi(\theta)$ that are similar to the likelihood is given by

$$\pi(\theta) \propto \xi^\theta / \Gamma(\theta)^\alpha\,,$$

where $\xi > 0$ and $\alpha > 0$ (in fact, α could even be restricted to be an integer). This distribution is integrable when $\alpha > 0$ thanks to the Stirling approximation,

$$\Gamma(\theta) \approx \theta^{\theta - 1/2}\, e^{-\theta}\,.$$

(b). When $X \sim \mathcal{B}e(1,\theta)$, $\theta \in \mathbb{N}$, we have

$$f(x|\theta) = \frac{(1-x)^{\theta-1}}{B(1,\theta)} = \frac{\Gamma(1+\theta)\,(1-x)^{\theta-1}}{\Gamma(\theta)} = \theta\,(1-x)^{\theta-1}$$

and this suggest using a gamma-like distribution on θ,

$$\pi(\theta) \propto \theta^m e^{-\alpha\theta}\,,$$

where $m \in \mathbb{N}$ and $\alpha > 0$. This function is clearly summable, due to the integrability of the gamma density, and conjugate.