STAT440/840/CM761 Assignment # 1 Winter 2019 Instructor: S. Chenouri

<u>Due</u>: **Feb. 13, 2019**

Instruction: Both graduate and undergraduate student must clearly mention on their submitted solution their level: "Graduate student" or "undergraduate student".

Problem 1. [10 mark] Let $X \sim N(\theta, 1)$. Suppose we take a sample of size n and observe $x_1, \ldots, x_n$, but only $x_{(n)} = \max\{x_1, \ldots, x_n\}$ is reported, while the others are missing.

a) [**4 mark**] Derive the likelihood function, and plot it for $n = 5$ and $x_{(n)} = 3.5$. Find the maximum likelihood estimate of θ.

SOLUTION: Recall that from STAT230 or STAT330, the pdf of the random variable $Y = X_{(n)} = \max\{X_1, \ldots, X_n\}$ is given by

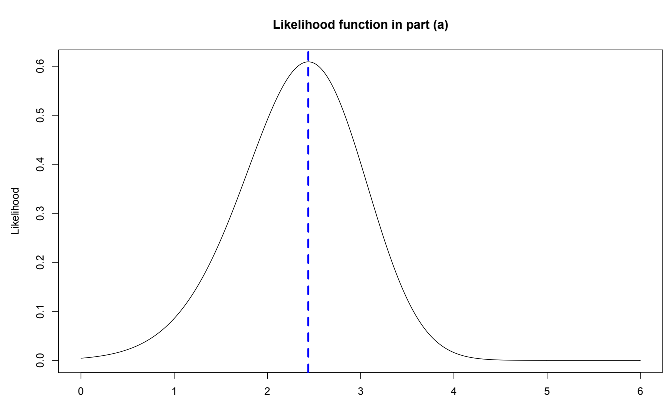
$$g(y \mid \theta) = n\, f(y \mid \theta)\, (F(y \mid \theta))^{n-1}\,, \quad -\infty < y < \infty\,.$$

Let $\phi(z)$ and $\Phi(z)$ denote the pdf and cdf of a standard normal $N(0,1)$ distribution. We can write $Z = X - \theta$ and thus

$$g_1(y \mid \theta) = n\, \phi(y-\theta)\, (\Phi(y-\theta))^{n-1}\,, \quad -\infty < y < \infty\,.$$

The likelihood function for $n = 5$ and $y = x_{(n)} = 3.5$ is $L_1(\theta) = 5\, \phi(3.5-\theta)\, (\Phi(3.5-\theta))^4\,.$

```
lik.Q1.a<-function(theta){
5*dnorm(3.5-theta,mean=0,sd=1)*(pnorm(3.5-theta,mean=0,sd=1,lower.tail=TRUE))^4
}
Theta<-seq(0,6,0.01)
plot(Theta,lik.Q1.a(Theta),type="l",xlab="theta",ylab="Likelihood",main="Likelihood function in part (a)")
> optimize(f=lik.Q1.a,interval=c(2,4),maximum=TRUE)
$maximum
[1] 2.438481
$objective
[1] 0.6090148
abline(v=2.438481,col="blue",lwd=3,lty=2)
```



b) **[6 mark]** Suppose we have two independent samples taken from $N(\theta, 1)$. From the first sample it is reported that the sample size is $n_1 = 5$, and the maximum $x_{(5)} = 3.5$. The second sample has size $n_2 = 3$, and only the sample mean $\overline{y} = 4$ is reported. Plot the three likelihoods, namely based the 1st sample, the 2nd sample and the combined sample.

SOLUTION: Let

$$X_1, X_2, \ldots, X_{n_1} \overset{iid}{\sim} N(\theta, 1)$$
$$Y_1, Y_2, \ldots, Y_{n_2} \overset{iid}{\sim} N(\theta, 1)$$

be two independent samples. Thus $\overline{Y} = n_2^{-1} \sum_{i=1}^{n_2} Y_i \sim N(\theta, 1/n_2)$, with pdf

$$g_2(\overline{y}) = \sqrt{\frac{n_2}{2\pi}}\, e^{-n_2(\overline{y}-\theta)/2} = \phi(\sqrt{n_2}\,(\overline{y}-\theta))$$

Thus the likelihood from the 2nd sample is

$$L_2(\theta) = \phi(\sqrt{3}\,(4-\theta))\,.$$

Therefore the combined likelihood function is

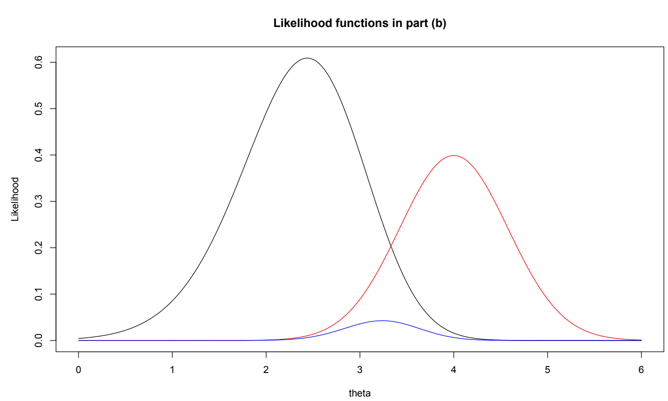
$$L(\theta) = L_1(\theta)\, L_2(\theta)\,.$$

```
Theta<-seq(0,6,0.01)
lik.Q1.a<-function(theta){
5*dnorm(3.5-theta,mean=0,sd=1)*(pnorm(3.5-theta,mean=0,sd=1,lower.tail=TRUE))^4
}

plot(Theta,lik.Q1.a(Theta),type="l",xlab="theta",ylab="Likelihood",main="Likelihood functions in part (b)")

lik.Q1.b<-function(theta){dnorm(sqrt(3)*(4-theta),mean=0,sd=1)}
points(Theta,lik.Q1.b(Theta),lwd=1,type="l",col="red")

lik.Q1.comb<-function(theta){lik.Q1.a(theta)*lik.Q1.b(theta)}
points(Theta,lik.Q1.comb(Theta),lwd=1,type="l",col="blue")
```



Problem 2. **[20 mark]** Over the course of the 1990s the General Social Survey gathered data on the educational attainment and number of children of 155 women who were 40 years of age at the time of their participation in the survey. These women were in their 20s during the 1970s, a period of historically low fertility rates in the United States. In this question, the intention is to compare the women with college degrees to those without in terms of their numbers of children. Let $X_1, \ldots, X_m$ denote the numbers of children for the m women without college degrees and $Y_1, \ldots, Y_n$ be the data for women with degrees. Assume that the sampling models are $X_1, \ldots, X_m \overset{iid}{\sim} \text{Poisson}(\theta_1)$, and $Y_1, \ldots, Y_n \overset{iid}{\sim} \text{Poisson}(\theta_2)$. The observed data are summarized as

$$\text{Less than bachelor's:} \qquad m = 111\,, \quad \sum_{i=1}^{m} X_i = 217\,, \quad \overline{X} = 1.95$$

$$\text{Bachelor's or higher:} \qquad n = 44\,, \quad \sum_{i=1}^{n} Y_i = 66\,, \quad \overline{Y} = 1.50\,.$$

Let the prior distributions be $\theta_1, \theta_2 \overset{iid}{\sim} \text{Gamma}(\alpha = 2, \beta = 1)$.

a) **[3 mark]** What are the marginal and joint posterior distributions of θ_1 and θ_2? Show your steps.
SOLUTION: First note that

$$X_1, \ldots, X_m \overset{iid}{\sim} f_X(x \mid \theta_1) = \frac{\theta_1^x \, e^{-\theta_1}}{x!} \quad \text{for} \quad x = 0, 1, \ldots\,,$$
$$Y_1, \ldots, Y_n \overset{iid}{\sim} f_Y(y \mid \theta_2) = \frac{\theta_2^y \, e^{-\theta_2}}{y!} \quad \text{for} \quad y = 0, 1, \ldots\,.$$

Thus the respective joint distributions of the data $\mathbf{x} = (x_1, \ldots, x_m)$ and $\mathbf{y} = (y_1, \ldots, y_m)$ are

$$f_X(\mathbf{x} \mid \theta_1) = \frac{\theta_1^{m\,\overline{x}} \; e^{-m\,\theta_1}}{\prod_{i=1}^{m} x_i!}\,, \quad \text{and} \quad f_Y(\mathbf{y} \mid \theta_2) = \frac{\theta_2^{n\,\overline{y}} \; e^{-n\,\theta_2}}{\prod_{i=1}^{n} y_i!}$$

In addition, the prior distributions are $\pi(\theta_1) = \theta_1^2 \, e^{-\theta_1}$ and $\pi(\theta_2) = \theta_2^2 \, e^{-\theta_2}$ for $\theta_1, \theta_2 > 0$. The posterior p.d.f of θ_1 given $\mathbf{x}$ is obtained in the following fashion,

$$g_1(\theta_1 \mid \mathbf{x}) \propto \pi(\theta_1) f_X(\mathbf{x} \mid \theta_1) \propto \theta_1 \, e^{-\theta_1} \, \theta_1^{m\,\overline{x}} \; e^{-m\,\theta_1} = e^{-(m+1)\theta_1} \, \theta_1^{m\,\overline{x}+1}$$

which says $\theta_1 \mid \mathbf{x} \overset{d}{\sim} \text{Gamma}(\alpha = m\,\overline{x} + 2, \, \beta = (m+1)^{-1}) = \text{Gamma}(\alpha = 219, \, \beta = (112)^{-1})$. We can similarly show that $\theta_2 \mid \mathbf{y} \overset{d}{\sim} \text{Gamma}(\alpha = n\,\overline{y} + 2, \, \beta = (n+1)^{-1}) = \text{Gamma}(\alpha = 68, \, \beta = (45)^{-1})$. Because of the independence of θ_1 from θ_2 as well as the independence of $\mathbf{X} = (X_1, \ldots, X_m)$ from $\mathbf{Y} = (Y_1, \ldots, Y_n)$, the joint posterior p.d.f. is

$$g(\theta_1, \theta_2 \mid \mathbf{x}, \mathbf{y}) = \frac{(m+1)^{m\,\overline{x}+1} \, (n+1)^{n\,\overline{y}+1}}{\Gamma(m\,\overline{x}+2)\,\Gamma(n\,\overline{y}+1)} \, \theta_1^{m\,\overline{x}+1} \; \theta_2^{n\,\overline{y}+1} \; e^{-(m+1)\theta_1 - (n+1)\theta_2} \qquad \theta_1, \theta_2 > 0\,.$$

b) **[3 mark]** Obtain the posterior means, modes and 95% quantile-based confidence intervals for θ_1 and θ_2. Show your steps, or your `R` codes.

SOLUTION: The posterior means are

$$\widehat{\theta}_1^B = \mathrm{E}[\theta_1 \mid \mathbf{x}] = \frac{m\,\overline{x}+2}{m+1} = \frac{m}{m+1}\,\overline{x} + \frac{1}{m+1}\,2 = \frac{219}{112} = 1.96\,,$$
$$\widehat{\theta}_2^B = \mathrm{E}[\theta_2 \mid \mathbf{y}] = \frac{n\,\overline{y}+2}{n+1} = \frac{n}{n+1}\,\overline{y} + \frac{1}{n+1}\,2 = \frac{68}{45} = 1.51\,.$$

To find the modes or the maximum a posteriori (MAP) estimates it is easier to maximize log of the posterior p.d.f.s. For estimating θ_1, we have

$$\log(g_1(\theta_1 \mid \mathbf{x})) = -(m+1)\theta_1 + (m\,\overline{x}+1)\log(\theta_1)$$

and

$$\frac{\partial\,\log(g_1(\theta_1 \mid \mathbf{x}))}{\partial\,\theta_1} = -(m+1) + \frac{(m\,\overline{x}+1)}{\theta_1} = 0\,.$$

Solving this equation we obtain the maximum a posteriori (MAP) estimate of θ_1,

$$\widehat{\theta}_1^{MAP} = \frac{m\,\overline{x}+1}{m+1} = \frac{m}{m+1}\,\overline{x} + \frac{1}{m+1} = \frac{218}{112} = 1.95\,.$$

Similarly we can show that the maximum a posteriori (MAP) estimate θ_2 is

$$\widehat{\theta}_2^{MAP} = \frac{n\,\overline{y}+1}{n+1} = \frac{n}{n+1}\,\overline{y} + \frac{1}{n+1} = \frac{67}{45} = 1.49\,.$$

```
> # for theta1
> # lower confidence bound
> qgamma(0.025,shape=219,rate=112,lower.tail=TRUE)
[1] 1.704943
> # upper confidence bound
> qgamma(0.975,shape=219,rate=112,lower.tail=TRUE)
[1] 2.222679
> # for theta2
> # lower confidence bound
> qgamma(0.025,shape=68,rate=45,lower.tail=TRUE)
[1] 1.173437
> # upper confidence bound
> qgamma(0.975,shape=68,rate=45,lower.tail=TRUE)
[1] 1.890836
```

c) [**3 mark**] Plot the posterior densities for the population means θ_1 and θ_2 of the two groups together and compare them. What do you conclude?

```
theta1<-theta2<-seq(0,3,0.01)
pd1<-dgamma(theta1,shape=219,rate=112)
pd2<-dgamma(theta2,shape=68,rate=45)
plot(theta1,pd1,xlab="theta",ylab="Probability density function",type="l",
main="Probability Density Functions of theta1 & theta2")
points(theta2,pd2,"l",col="red")
```

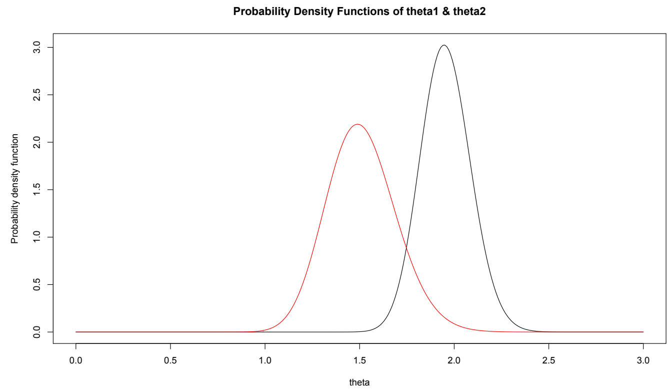


Figure 1: Grey represents the distribution of θ_1 and red represents the distribution of θ_2.

The two distributions are well separated and θ_1 is very likely to be bigger than θ_2, indicating substantial evidence that $\theta_1 > \theta_2$.

d) **[2 mark]** Does the posterior indicates substantial evidence that $\theta_1 > \theta_2$? Why? Compute a suitable quantity and show your steps.
SOLUTION:

```
theta1<-rgamma(100000,shape=219,rate=112)
theta2<-rgamma(100000,shape=68,rate=45)
mean(theta1>theta2)
[1] 0.9734
```

Yes, the posterior indicates substantial evidence that $\theta_1 > \theta_2$, since

$$\Pr\left(\theta_1 > \theta_2 \,\middle|\, \sum_{i=1}^{m} X_i = 217,\, \sum_{i=1}^{n} Y_i = 66\right) = 0.9734\,,$$

which is a very large probability.

e) **[6 mark]** Now consider two randomly sampled individuals X^f and Y^f, one from each of the two groups. What are the predictive distributions of X^f and Y^f? To what extent do we expect the one without the bachelor's degree to have more children than the other? Calculate the relevant probabilities exactly.

SOLUTION: Let X^f be independent of $X_1, \ldots, X_m$ with pdf $f_X(x \mid \theta_1)$, then

$$f(x^f \mid x_1, \ldots, x_n, \theta_1) = f_X(x^f \mid \theta_1)\,.$$

The predictive distribution $p(x^f \mid x_1, \ldots, x_n)$ is obtained by

$$\begin{aligned}
p(x^f \mid x_1, \ldots, x_n) &= \int_0^\infty f(x^f \mid \mathbf{x}, \theta_1)\, g_1(\theta_1 \mid \mathbf{x})\, d\theta_1 \\
&= \int_0^\infty f_X(x^f \mid \theta_1)\, g_1(\theta_1 \mid \mathbf{x})\, d\theta_1 \\
&= \int_0^\infty \frac{e^{-\theta_1}\, \theta_1^{x^f}}{x^f!} \frac{(m+1)^{m\,\overline{x}+2}}{\Gamma(m\,\overline{x}+2)}\, \theta_1^{m\,\overline{x}+1}\, e^{-(m+1)\,\theta_1}\, d\theta_1 \\
&= \int_0^\infty \frac{(m+1)^{m\,\overline{x}+2}}{x^f!\,\Gamma(m\,\overline{x}+2)}\, \theta_1^{m\,\overline{x}+x^f+1}\, e^{-(m+2)\,\theta_1}\, d\theta_1 \\
&= \frac{(m+1)^{m\,\overline{x}+2}\,\Gamma(m\,\overline{x}+x^f+2)}{(m+2)^{m\,\overline{x}+x^f+2}\, x^f!\,\Gamma(m\,\overline{x}+2)} \\
&= \frac{\Gamma(m\,\overline{x}+x^f+2)}{x^f!\,\Gamma(m\,\overline{x}+2)} \cdot \left(\frac{m+1}{m+2}\right)^{m\,\overline{x}+2} \left(\frac{1}{m+2}\right)^{x^f} \\
&= \binom{218+x^f}{x^f} \left(\frac{112}{113}\right)^{219} \left(\frac{1}{113}\right)^{x^f}, \quad x^f = 0,\, 1,\, \ldots
\end{aligned}$$

which is the negative binomial distribution NB$(219,\, \frac{112}{113})$. We can similarly show that the predictive distribution $p(y^f \mid y_1, \ldots, y_n)$ is NB$(68,\, \frac{45}{46})$ that is

$$\begin{aligned}
p(y^f \mid y_1, \ldots, y_n) &= \frac{\Gamma(n\,\overline{y}+y^f+2)}{y^f!\,\Gamma(n\,\overline{y}+2)} \cdot \left(\frac{n+1}{n+2}\right)^{n\,\overline{y}+2} \left(\frac{1}{n+2}\right)^{y^f} \\
&= \binom{67+y^f}{y^f} \left(\frac{45}{46}\right)^{68} \left(\frac{1}{46}\right)^{y^f}, \quad y^f = 0,\, 1,\, \ldots
\end{aligned}$$

Therefore

$$\begin{aligned}
\Pr\left(X^f > Y^f \mid \overline{X} = 1.95, \overline{Y} = 1.50\right) &= \sum_{x^f=1}^{\infty} \sum_{y^f=0}^{x^f-1} \Pr\left(X^f = x^f,\, Y^f = y^f \mid \overline{Y} = 1.50\right) \\
&= \sum_{x^f=1}^{\infty} \sum_{y^f=0}^{x^f-1} \Pr\left(X^f = x^f \mid \overline{X} = 1.95\right) \Pr\left(Y^f = y^f \mid \overline{Y} = 1.50\right) \\
&= \sum_{x^f=1}^{\infty} \Pr\left(X^f = x^f \mid \overline{X} = 1.95\right) \Pr\left(Y^f \le x^f - 1 \mid \overline{Y} = 1.50\right) \\
&= \sum_{x^f=0}^{\infty} \Pr\left(X^f = x^f \mid \overline{X} = 1.95\right) \Pr\left(Y^f \le x^f - 1 \mid \overline{Y} = 1.50\right) \mathbb{I}(x^f \ge 1) \\
&= \mathrm{E}\left[\Pr\left(Y^f \le X^f - 1 \mid \overline{Y} = 1.50\right) \mathbb{I}(X^f \ge 1) \,\Big|\, \overline{X} = 1.95\right],
\end{aligned}$$

in which the expectation is with respect to $X^f \overset{d}{\sim}$ NB$(219,\, \frac{1}{113})$. The `R` function below provides a Monte Carlo approximation to this probability. Based on a sample $N = 10000$, the probability is estimated to be

$$\Pr\left(X^f > Y^f \mid \overline{X} = 1.95, \overline{Y} = 1.50\right) = 0.5584\,,$$

while

$$\Pr\left(\theta_1 > \theta_2 \,\Big|\, \sum_{i=1}^{m} X_i = 217,\, \sum_{i=1}^{n} Y_i = 66\right) = 0.9734\,.$$

```
# Monte Carlo Computation of the probability.
NB.theta<-function(N,alpha){
xf<-rnbinom(N,size=219,prob=112/113)
delta<-pnbinom(xf[xf>=1]-1, size=68, prob=45/46, lower.tail = TRUE)
Theta.hat<-mean(delta)
sd.delta<-sd(delta)
SE.hat<-sd.delta/sqrt(N)
LB<-Theta.hat+qnorm(alpha/2,mean=0,sd=1)*SE.hat
UB<-Theta.hat+qnorm(1-alpha/2,mean=0,sd=1)*SE.hat
return(c(Theta.hat=Theta.hat,Lower.B=LB,Upper.B=UB))
}

round(NB.theta(10000,alpha=0.05),digits=4)
Theta.hat   Lower.B   Upper.B
   0.5584    0.5532    0.5637
```

f) [**3 mark**] Plot these predictive distributions together and compare them. Discuss the distinction between the events $\{\theta_1 > \theta_2\}$ and $\{X^f > Y^f\}$.

SOLUTION: The following plots show there is only small overlap between two posterior distributions of θ_1 and θ_2 but there is much overlap between the predictive distributions of X^f and Y^f compared to the posteriors. This tells us that the distinction between the two events $\theta_1 > \theta_2$ and $X^f > X^f$, as their probabilities are reported in part (e), is extremely important as the strong evidence of a difference between two populations does not mean that the difference between two randomly selected individuals from these populations itself is large.

```
par(mfrow=c(1,2))
theta<-theta1<-theta2<-seq(0,3,0.01)
pd1<-dgamma(theta1,shape=219,rate=112)
pd2<-dgamma(theta2,shape=68,rate=45)
plot(theta1,pd1,xlab="theta",ylab="Probability density function",type="l",
main="Probability Density Functions of theta1 & theta2",col="grey",lwd=4)
points(theta2,pd2,"l",col="red",lwd=3)
points(theta,dgamma(theta,shape=2,scale=1),"l",col="blue",lwd=3,lty=2)
x<-0:10
dxf<-dnbinom(x,size=219,prob=112/113)
dyf<-dnbinom(x,size=68, prob=45/46)
dxy<-rbind(dxf,dyf)
```

```
barplot(dxy,ylim=c(0,0.35),beside=TRUE,ylab="pdf",col=c("grey","red"),
legend=rownames(dxy))
```

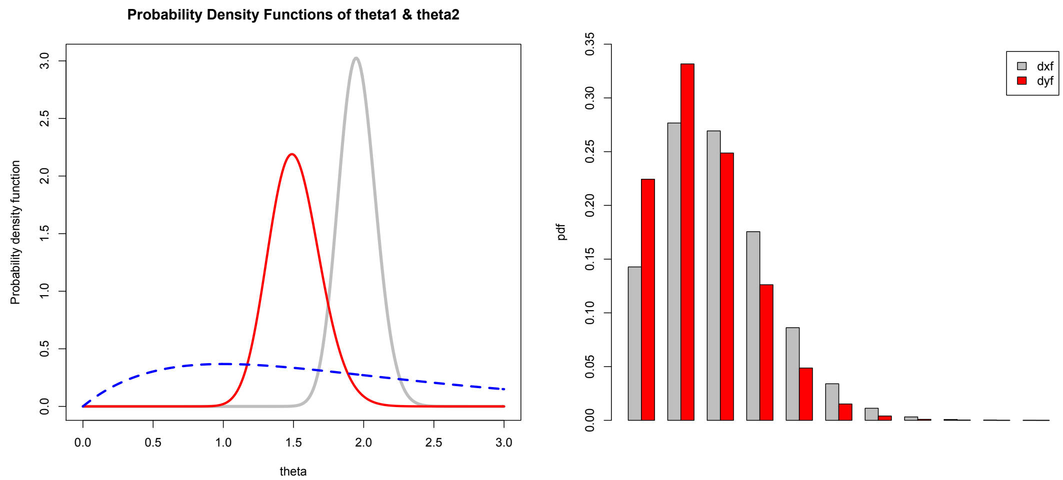


Figure 2: Left: posterior pdf of θ_1 (grey), θ_2 (red) and the common prior (dashed in blue). Right: pdf of X^f (grey) and Y^f (red).

Problem 3. [**12 mark**] Consider the function $f(x) = \exp\left(-\frac{x^2}{2}\right)$ and note that $\theta = \int_{-\infty}^{\infty} f(x)dx = \sqrt{2\pi}$.
SOLUTION: First note that since $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} f(x)dx = 0.9973$, then the integral $\int_{-10}^{10} f(x)\, dx$ is very close to $\theta = \sqrt{2\pi}$. Now let $X \sim U(-10, 10)$, and note that

$$\sqrt{2\pi} = \theta = \mathrm{E}\left[20\, f(X)\right] = \int_{-10}^{10} 20\, f(x) \frac{1}{20}\, dx\,.$$

The simple Monte Carlo estimate of θ is obtained by setting $\delta(X) = 20\, f(X) = 20\, \exp(-X^2/2)$. [**3 mark**]

a) Consider using a control variate $g(x)$, where

$$g(x) = 1 - \frac{|x|}{5}\,.$$

Estimate $\int f(x)dx$ using $g(x)$ as the control variate for 50 times, each time using $n = 500$ samples. What is the average estimate? What is the variance of the estimate?
SOLUTION: [**4 mark**] First note that

$$\theta^* = E\left[g(X)\right] = \int_{-10}^{10} g(x)\, \frac{1}{20}\, dx = \int_{-10}^{10} \left(1 - \frac{|x|}{5}\right) \frac{1}{20}\, dx$$

$$= 1 - \frac{1}{100} \int_{-10}^{10} |x|\, dx = 1 - \frac{2}{100} \int_{0}^{10} x\, dx = 0\,.$$

If we perform the experiment using the simple Monte Carlo estimate of θ by setting $\delta(X) = 20\, f(X) = 20\, \exp(-X^2/2)$, we have

```
  m<-50
  n<-500
  X<-matrix(runif(m*n,min=-10,max=10),nrow=m,ncol=n)
  delta<-20*exp(-X^2/2)
  Theta.hat<-apply(delta,MARGIN=1,FUN=mean)
mean(Theta.hat)  # Average of the estimate for the simple MC
[1] 2.548434
var(Theta.hat)  # Variance of the estimate for the simple MC
[1] 0.05284796
```

For the control variate estimates using g and the same sample as above, we obtain

```
  theta.star.g=0
  delta.g<-1-abs(X)/5
  Theta.hat.g<-apply(delta.g,MARGIN=1,FUN=mean)
  Var.g<-apply(delta.g,MARGIN=1,FUN=var)/n
  Cov.g<-apply((delta-Theta.hat)*(delta.g-Theta.hat.g),MARGIN=1,FUN=mean)/(n-1)
  alpha.g<--Cov.g/Var.g
  Theta.hat.cv.g<-Theta.hat+alpha.g*(Theta.hat.g-theta.star.g)
mean(Theta.hat.cv.g) # Average of the estimate for the control variate g
[1] 2.518592
var(Theta.hat.cv.g)  # Variance of the estimate for the control variate g
[1] 0.02705816
```

b) Now consider using

$$h(x) = 1 - \frac{x^2}{25}$$

as a control variate. Is there a simple way to tell if $h(x)$ will be more or less effective than $g(x)$ as a control variate for this problem? Explain. Verify by actually estimating $\int f(x)dx$ using $h(x)$ as a control variate.

SOLUTION: [5 mark] First note that

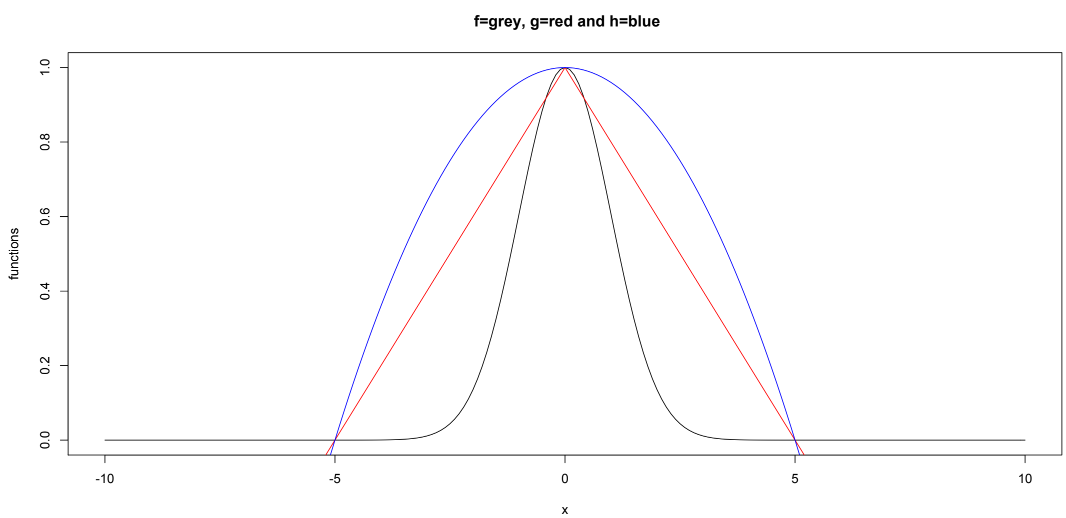
$$\theta^* = E\left[h(X)\right] = \int_{-10}^{10} h(x)\, \frac{1}{20}\, dx = \int_{-10}^{10} \left(1 - \frac{x^2}{25}\right) \frac{1}{20}\, dx$$

$$= 1 - \frac{1}{500} \int_{-10}^{10} x^2\, dx = -0.333\,.$$

```
  theta.star.h=-0.333
```

```
  delta.h<-1-X^2/25
  Theta.hat.h<-apply(delta.h,MARGIN=1,FUN=mean)
  Var.h<-apply(delta.h,MARGIN=1,FUN=var)/n
  Cov.h<-apply((delta-Theta.hat)*(delta.h-Theta.hat.h),MARGIN=1,FUN=mean)/(n-1)
  alpha.h<--Cov.h/Var.h
  Theta.hat.cv.h<-Theta.hat+alpha.h*(Theta.hat.h-theta.star.h)
mean(Theta.hat.cv.h)  # Average of the estimate for the control variate h
[1] 2.527954
var(Theta.hat.cv.h) # Variance of the estimate for the control variate h
[1] 0.03512369

### Plot of the functions f, g and h
x<-seq(-10,10,0.1)
f<-exp(-x^2/2)
g<-1-abs(x)/5
h<-1-x^2/25
plot(x,f,"l",ylab="functions")
points(x,g,"l",col="red")
points(x,h,"l",col="blue")
```



Looking at the plot of the functions f, g and h, it is clear that g provides a better approximation to g compared to h. This implies that g should provide better control variate estimates with smaller variance as it is seen from the variance calculations.

$$\text{Var}(\widehat{\theta}^{g}_{cv}) = 0.02705816 < \ 0.03512369 = \text{Var}(\widehat{\theta}^{h}_{cv}) < \text{Var}(\widehat{\theta}_{MC}) = 0.05284796\,.$$

Problem 4. [15 mark] We would like to estimate the integral

$$\theta = \int_1^\infty \frac{x^2}{\sqrt{2\,\pi}}\, e^{-x^2/2}\, dx$$

using simulations of size 1000.

SOLUTION: [2 mark] First note that the integral can be written as the following expectation

$$\theta = \mathrm{E}\left[X^2\, \mathbb{I}(X > 1)\right] \,,$$

where $X \overset{d}{\sim} \mathrm{N}(0,1)$ and $\mathbb{I}(A)$ is the indicator of the event A. So $\delta(x) = x^2\, \mathbb{I}(x > 1)$. A simple Monte Carlo estimate of θ may be obtained by sampling $x_1, \ldots, x_n$ from $N(0,1)$ and computing

$$\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \delta(x_i)\,.$$

```
A1.Q4.SMC<-function(n,c,alpha){
X<-rnorm(n,mean=0,sd=1)
delta<-X^2*(X>c)
m<-length(X[X>c])
Theta.hat<-mean(delta)
sd.delta<-sd(delta)
SE.hat<-sd.delta/sqrt(n)
LB<-Theta.hat+qnorm(alpha/2,mean=0,sd=1)*SE.hat
UB<-Theta.hat+qnorm(1-alpha/2,mean=0,sd=1)*SE.hat
  return(c(n=n,m=m,Theta.hat=Theta.hat,sd.delta=sd.delta,SE.hat=SE.hat,LB=LB,UB=UB))
}

round(A1.Q4.SMC(1000,1,0.05),digits=4)
        n         m  Theta.hat  sd.delta    SE.hat        LB        UB
1000.0000  151.0000     0.3790    1.2351    0.0391    0.3024    0.4555
```

a) **[5 mark]** Estimate the integral by importance sampling using random variables having a truncated normal distribution. That is, the importance function is

$$g(x) = \frac{1}{\sqrt{2\,\pi}\,(1 - \Phi(1))}\, e^{-x^2/2} \qquad \text{for } x > 1\,.$$

SOLUTION: To employ an importance sampling method, we first note that

$$f(x) = \frac{1}{\sqrt{2\,\pi}} e^{-x^2/2} \quad \text{and} \quad \mathrm{supp}(\delta\, f) = \mathrm{supp}(g)\,.$$

If we generate a sample $x_1, \ldots, x_n$ of size n from g, the importance sample estimate of θ is

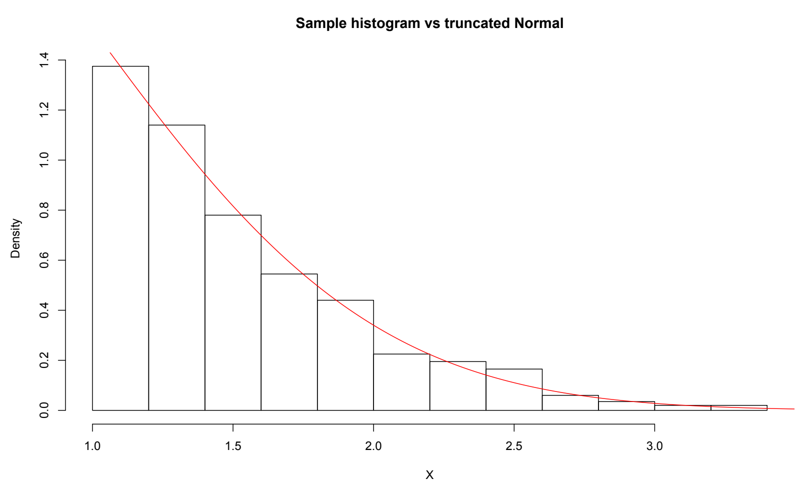
$$\widehat{\theta}^{g}_{IS} = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta(x_i)\, f(x_i)}{g(x_i)}\,.$$

To sample from the truncated normal distribution g, we use the acceptance-rejection sampling method discussed in Example 15 of Chapter 2, where we set $c = 1$, $\lambda = (c+\sqrt{c^2+4})/2 = (1+\sqrt{5})/2 = 1.618034$

and we generate from $\mathrm{Exp}(\lambda)$. Note that in this case the acceptance rate is 0.88. The following `R` codes generate a sample from the truncated normal distribution g using the above acceptance-rejection sampling scheme.

```
A1.Q4.a.AR<-function(n,c){
X<-c()
k<-1
lambda<-(c+sqrt(c^2+4))/2
M<-exp((lambda^2-2*lambda*c)/2)/(sqrt(2*pi)*lambda*(1-pnorm(c,mean=0,sd=1)))
while (k<=n){
y<-rexp(1,rate=lambda)
u<-runif(1,0,1)
f<-dnorm(y+1,mean=0,sd=1)/(1-pnorm(c,mean=0,sd=1))
g<-dexp(y,rate=lambda)
r<-f/(M*g)
if (u<=r){ X<-c(X,y+1)
k<-k+1
}
}
 return(X)
}

X<-A1.Q4.a.AR(1000,1)
hist(X,freq=FALSE,main="Sample histogram vs truncated Normal")
lines(seq(1,5,0.01),dnorm(seq(1,5,0.01),mean=0,sd=1)/(1-pnorm(1,mean=0,sd=1)),col=2)
```




```
A1.Q4.a.IS.g<-function(X,alpha){
n<-length(X)
delta<-X^2
fx<-dnorm(X,mean=0,sd=1)
gx<-dnorm(X,mean=0,sd=1)/(1-pnorm(1,mean=0,sd=1))
```

```
Theta.hat.IS<-mean(delta*fx/gx)
sd.IS<-sd(delta*fx/gx)
SE.hat.IS<-sd.IS/sqrt(n)
LB<-Theta.hat.IS+qnorm(alpha/2,mean=0,sd=1)*SE.hat.IS
UB<-Theta.hat.IS+qnorm(1-alpha/2,mean=0,sd=1)*SE.hat.IS
return(c(n=n,Theta.hat.IS=round(Theta.hat.IS,3),sd.IS=sd.IS,SE.hat.IS=SE.hat.IS,LB=LB,UB=UB))
}

round(A1.Q4.a.IS.g(X,0.05),digits=4)
        n   Theta.hat.IS      sd.IS     SE.hat.IS        LB            UB
   1000.0000        0.4000     0.2602      0.0082      0.3841        0.4164
```

b) [**4 mark**] Estimate the integral by importance sampling using random variables with the p.d.f.

$$h(x) = x \, \exp\left(\frac{1-x^2}{2}\right) \qquad \text{for } x > 1\,.$$

Hint: Prove that such random variables can be obtained as follows. Start with a random variable that has the exponential distribution with parameter 0.5, add 1, then take the square root.

SOLUTION: We first use the method of change of variable to find the p.d.f. of the random variable $Y = X^2 - 1$. Note that for $y = x^2 - 1$ we have $x = \sqrt{y+1}$ and then the Jacobean of the transformation is $J = \frac{dx}{dy} = \frac{1}{2\sqrt{y+1}}$. Thus the p.d.f. of Y, denoted by $k(y)$ is

$$k(y) = h(\sqrt{y+1})\,|J| = \sqrt{y+1}\, e^{-y/2}\, \frac{1}{2\sqrt{y+1}} = \frac{1}{2}\, e^{-y/2} \quad \text{for } y > 0\,.$$

Therefore, to generate a sample x from h one may generate y from Exp(0.5) and set $x = \sqrt{y+1}$. Now generate a sample $x_1, \ldots, x_n$ from h in this fashion and compute the importance sampling estimate of θ as

$$\widehat{\theta}^h_{IS} = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta(x_i)\, f(x_i)}{h(x_i)}\,.$$

The following R codes generate a sample from the h and computes $\widehat{\theta}^h_{IS}$.

```
A1.Q4.a.IS.h<-function(n,alpha){
X<-sqrt(rexp(n,rate=0.5)+1)
delta<-X^2
fx<-dnorm(X,mean=0,sd=1)
hx<-X*exp((1-X^2)/2)
Theta.hat.IS<-mean(delta*fx/hx)
sd.IS<-sd(delta*fx/hx)
SE.hat.IS<-sd.IS/sqrt(n)
LB<-Theta.hat.IS+qnorm(alpha/2,mean=0,sd=1)*SE.hat.IS
UB<-Theta.hat.IS+qnorm(1-alpha/2,mean=0,sd=1)*SE.hat.IS
return(c(n=n,Theta.hat.IS=round(Theta.hat.IS,3),sd.IS=sd.IS,SE.hat.IS=SE.hat.IS,LB=LB,UB=UB))
}
```

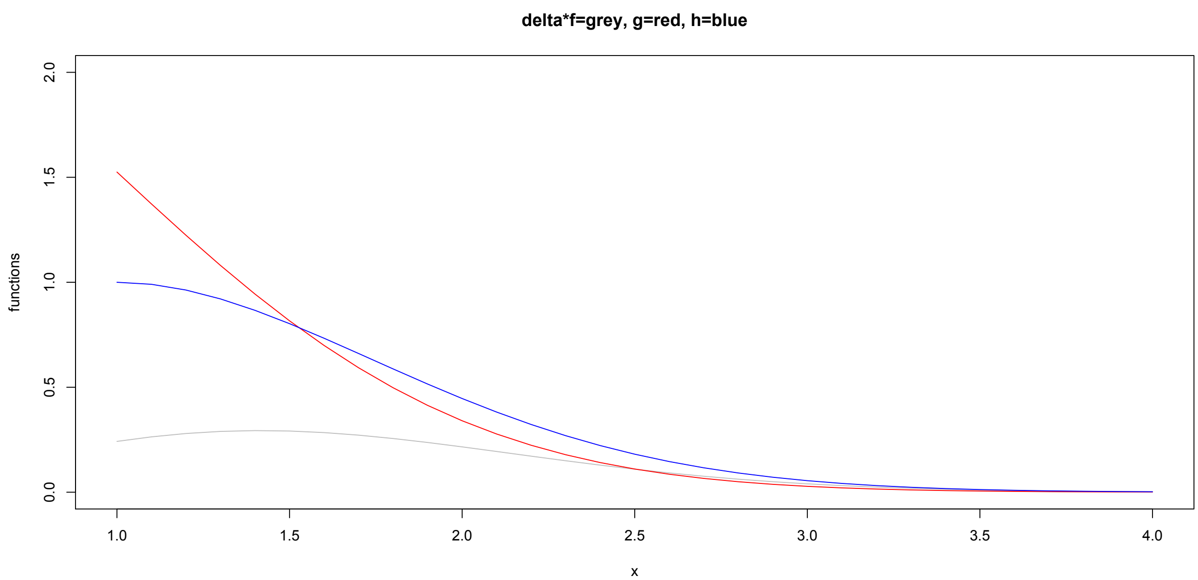
```
round(A1.Q4.a.IS.h(1000,0.05),digits=4)
        n     Theta.hat.IS        sd.IS   SE.hat.IS          LB          UB
1000.0000           0.4000       0.1162      0.0037      0.3927      0.4071
```

c) [**4 mark**] Compute and compare simulation standard errors for the two estimators in parts (a) and (b). Can you explain why one is so much smaller than the other?

SOLUTION: For part (a) the estimated standard error of the estimator is 0.0082 while for part (b) the standard error is 0.0037, more that 100% reduction in variability. Note that for the simple Monte Carlo this is 0.0391 much larger than those in parts (a) and (b).

```
## c
x<-seq(1,4,0.1)
delta.f<-x^2*dnorm(x,mean=0,sd=1)
gx<-dnorm(x,mean=0,sd=1)/(1-pnorm(1,mean=0,sd=1))
hx<-x*exp((1-x^2)/2)

plot(x,delta.f,ylim=c(0,2),ylab="functions",main="delta*f=grey, g=red, h=blue",
col="grey",type="l")
points(x,gx,type="l",col="red")
points(x,hx,type="l",col="blue")
```



From this plot, we observe that $h(x)$ resembles $\delta(x)\,f(x)$ better that $g(x)$. This can be also observed algebraically in the following way. Note that

$$h(x) = x\,\exp((1-x^2)/2) = e^{0.5}\,x\,\exp((1-x^2)/2) \propto\, x\,\exp(-x^2/2)$$

resembles $\delta(x)\,f(x) \propto x^2\,\exp(-x^2/2)$ beter than $g(x) \propto \exp(-x^2/2)$.

Problem 5. [15 mark] Consider the integral

$$\theta = \int_0^\infty x^2 \sin(\pi x)\, e^{-x/2}\, dx\,.$$

a) **[5 mark]** Use the crude Monte Carlo method to estimate the integral.
SOLUTION: Note that θ can be written as the expectation of the $\delta(X) = 2\, X^2 \sin(\pi\, X)$ for $X \sim \text{Exp}(1/2)$ (exponential) distribution. The estimate based on the simple Monte Carlo is

```
Q5.a<-function(n, alpha){
X<-rexp(n,1/2)
delta<-2*X^2*sin(pi*X)
Theta.hat<-mean(delta)
sd.delta<-sd(delta)
SE.hat<-sd.delta/sqrt(n)
LB<-Theta.hat-+qnorm(alpha/2,mean=0,sd=1)*SE.hat
UB<-Theta.hat+qnorm(1-alpha/2,mean=0,sd=1)*SE.hat
return(c(n=n,Theta.hat=round(Theta.hat,3),sd.delta=sd.delta,SE.hat=SE.hat,LB=LB,UB=UB))
}
> Q5.a(1000,0.05)
 n       Theta.hat      sd.delta       SE.hat             LB           UB
1000     0.1280624    22.5273694      0.7123780        1.5242975    1.5242975
> Q5.a(10000,0.05)
 n        Theta.hat     sd.delta          SE.hat            LB           UB
10000   3.401739e-01 2.606024e+01    2.606024e-01    8.509451e-01 8.509451e-01
> Q5.a(100000,0.05)
 n       Theta.hat       sd.delta      SE.hat            LB             UB
100000  -2.236354e-02  2.774194e+01  8.772770e-02  1.495796e-01  1.495796e-01
>
```

b) **[5 mark]** Use antithetic exponential variate to estimate the integral.
SOLUTION: By the integral transform method, if $U \sim U(0,1)$ then both $-2\,\log(U)$ and $-2\,\log(1-U)$ have Exp(1/2) (exponential) distribution. From this observation we code up an antithetic Monte Carlo estimate as follow:

```
Q5.b<-function(n){
  U<-runif(n,0,1)
  delta1<-2*(-2*log(U))^2*sin(-2*pi*log(U))
  Theta1.hat<-mean(delta1)
  delta2<-2*(-2*log(1-U))^2*sin(-2*pi*log(1-U))
  Theta2.hat<-mean(delta2)
  AS.Theta.hat<-(Theta1.hat+Theta2.hat)/2
```

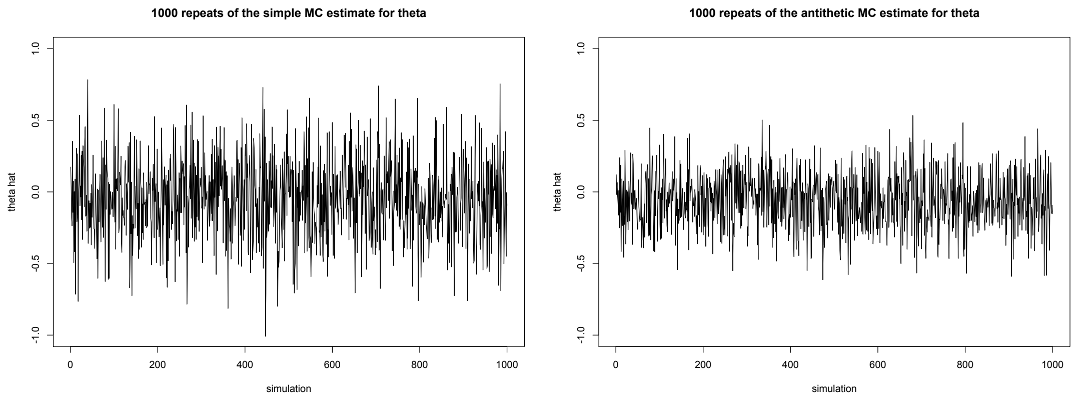
```
  cov.12<-cov(delta1,delta2)
  sd.12<-sd((delta1+delta2)/2)
  AS.SE.12<-sd.12/sqrt(n)
  return(c(AS.Theta.hat=AS.Theta.hat,Cov.deltas=cov.12,sd.12=sd.12,AS.SE=AS.SE.12))
  }
> Q5.b(1000)
AS.Theta.hat   Cov.deltas        sd.12        AS.SE
   0.3418098    2.7172191   19.0931791    0.6037793
> Q5.b(10000)
AS.Theta.hat   Cov.deltas        sd.12        AS.SE
  0.08123443   2.52351126  21.08285527   0.21082855
> Q5.b(100000)
AS.Theta.hat   Cov.deltas        sd.12        AS.SE
 -0.04446168   2.54511951  19.76590423   0.06250528
```

c) [**5 mark**] Compare the variance of the estimator in part (a) with that of the estimator in part (b) by performing the estimation of 1000 times in each case with sample sizes of 10,000.

```
  m<-1000
  n<-10000
  U<-matrix(runif(m*n,min=0,max=1),nrow=m,ncol=n)
  delta1<-2*(-2*log(U))^2*sin(-2*pi*log(U))
  Theta1.hat<-apply(delta1,MARGIN=1,FUN=mean)
> mean(Theta1.hat)  # Average of the estimate for the simple MC
[1] -0.04725437
> var(Theta1.hat)  # Variance of the estimate for the simple MC
[1] 0.07928873
>
  delta2<-2*(-2*log(1-U))^2*sin(-2*pi*log(1-U))
  Theta2.hat<-apply(delta2,MARGIN=1,FUN=mean)
  AS.Theta.hat<-(Theta1.hat+Theta2.hat)/2

> mean(AS.Theta.hat) # Average of the estimate for the antithetic MC
[1] -0.05688856
> var(AS.Theta.hat) # Variance of the estimate for the antithetic MC
[1] 0.03825467
>
```

Note that in the `R` codes above, the random uniform samples for both parts (a) and (b) are kept the same. The variance of the AS is half of that for the simple MC. A 100% reduction in the variance. The following Figure, depicts the time series plot of all 1000 simple MC and antithetic 1000 estimates of θ. Note the higher variability of the simple MC estimates compared to the antithetic estimates.



Problem 6. [**15 mark**] Let

$$\theta = \int_0^{\pi/2} \cos(x)\ \sin(x)\ \exp\{-\cos^2(x)\}\, dx\,.$$

a) [**3 mark**] Use the Monte Carlo method, with $n = 100$, to estimate θ.
SOLUTION

```
Q6<-function(n, alpha){
U<-runif(n,0,pi/2)
delta<-(pi/2)*cos(U)*sin(U)*exp(-(cos(U))^2)
Theta.hat<-mean(delta)
sd.delta<-sd(delta)
SE.hat<-sd.delta/sqrt(n)
LB<-Theta.hat+qnorm(alpha/2,mean=0,sd=1)*SE.hat
UB<-Theta.hat+qnorm(1-alpha/2,mean=0,sd=1)*SE.hat
return(c(n=n,Theta.hat=Theta.hat,SD=sd.delta,SE=SE.hat,LB=LB,UB=UB))
}

round(Q6(100,0.05),digit=6)
         n  Theta.hat         SD         SE         LB         UB
100.000000   0.313457   0.167484   0.016748   0.280630   0.346283
```

From this the Monte Carlo estimate of θ is $\widehat{\theta}_{MC} = 0.313457$.

b) [**3 mark**] Construct a 95% confidence interval for θ based on the Monte Carlo data.
SOLUTION: Based on the calculation above, a 95% confidence interval for θ is (0.280630 0.346283).

c) [**2 mark**] Use the change of variables $y = \cos^2(x)$ to exactly evaluate the integral. Is the true value of θ included in your confidence interval?

SOLUTION: Set $y = \cos^2(x)$ and note that $dy = -2\,\sin(x)\,\cos(x)\,dx$. Thus

$$\begin{aligned}\theta &= \int_0^{\pi/2} \cos(x)\,\sin(x)\,\exp\{-\cos^2(x)\}\,dx \\ &= -\frac{1}{2}\int_1^0 e^{-y}\,dy = \frac{1}{2}\int_0^1 e^{-y}\,dy \\ &= \frac{1}{2}\left(1 - e^{-1}\right) = 0.31606\,.\end{aligned}$$

d) **[3 mark]**Repeat (a)-(c) with $n = 500$ and $n = 1000$.
SOLUTION: For $n = 500$ and $n = 1000$

```
> round(Q6(500,0.05),digit=6)
  n  Theta.hat          SD          SE          LB          UB
500   0.319726    0.167979    0.007512    0.305002    0.334450
> round(Q6(1000,0.05),digit=6)
  n    Theta.hat           SD           SE           LB           UB
1000     0.310741     0.168821     0.005339     0.300278     0.321205
```

e) **[4 mark]** What is the needed sample size if the 95% confidence interval must have total length equal to 0.02?
SOLUTION: First note that from part (c), setting $y = \cos^2(x)$ we obtain

$$\theta = \int_0^{\pi/2} \cos(x)\,\sin(x)\,\exp\{-\cos^2(x)\}\,dx = \int_0^1 0.5\,e^{-y}\,dy = 0.31606\,.$$

This implies that for $Y \sim U(0,\,1)$ and $\eta(Y) = 0.5\,\exp(-Y)$ we have $\theta = \mathrm{E}(\eta(Y)) = 0.31606$ and therefore for a Monte Carlo estimate $\widehat{\theta} = \sum_{i=1}^{n} 0.5\,\exp(-Y_i)$ for $Y_1,\,\ldots,\,Y_n \sim U(0,\,1)$ which is the same as $\widehat{\theta} = \sum_{i=1}^{n} 0.5\,\pi\,\cos(X_i)\,\sin(X_i)\,\exp(-\cos^2(X_i))$, where $X_1,\,\ldots,\,X_n \sim U(0,\,\pi/2)$ and $Y_i = \cos^2(X_i)$ we have

$$\begin{aligned}\mathrm{Var}(\widehat{\theta}) &= \frac{\mathrm{Var}(\delta(Y))}{n} = \frac{1}{n}\left[\int_0^1 0.25\,e^{-2\,y}\,dy - (0.31606)^2\right] \\ &= \frac{1}{n}\left[\frac{1}{8}\left[1 - e^{-2}\right] - (0.31606)^2\right] \\ &= \frac{0.1080831 - 0.09989392}{n} = \frac{0.00818918}{n}\,.\end{aligned}$$

Now for the total length of the 95% confidence interval to be equal to 0.02, the sample size n is obtained by

$$2\,\times\,1.96\sqrt{\frac{\mathrm{Var}(\delta(Y))}{n}} = 0.02 \Rightarrow n = \left(\frac{2\,\times\,1.96}{0.02}\right)^2 \times\,0.00818918 = 314.5955\,.$$

So the required sample size is 315.

In practice though we don't have the value of $\mathrm{Var}(\delta(Y))$ and it has to be estimated. From $n = 1000$ the estimate is given by $\widehat{\mathrm{Var}}(\delta(Y)) = \mathrm{SD}^2 = 0.168821^2 = 0.02850053$. Thus

$$2 \times 1.96\sqrt{\frac{\mathrm{Var}(\delta(Y))}{n}} = 0.02 \Rightarrow n \approx \left(\frac{2 \times 1.96}{0.02}\right)^2 \times 0.02850053 \approx 1095\,,$$

which is roughly 3.48 times larger than the exact value 3.15. This due to extra uncertainty from estimating $\mathrm{Var}(\delta(Y))$. This value also can change from a Monte Carlo estimate to another.

Problem 7. [**15 mark**]

a) Generate 100 samples of size $n = 10$ from the following distributions:
SOLUTION: Generate 100 samples of size $n = 10$ from $U \sim U(0, 1)$

```
m<-100 # number of samples
n<-10  # sample size
U<-matrix(runif(n*m,0,1),nrow=m,ncol=n)
```

i. [**2 mark**] Exponential with mean 1.
SOLUTION: For Exponential random variables with mean 1, set $X = -\log(U) \sim \mathrm{Exp}(1)$.

```
Xi<--log(U)
```

ii. [**3 mark**] Discrete distribution with

$$f(x) = \begin{cases} 1/8 & x = 1,\, 3 \\ 3/4 & x = 9 \\ 0 & \text{o.w.} \end{cases}$$

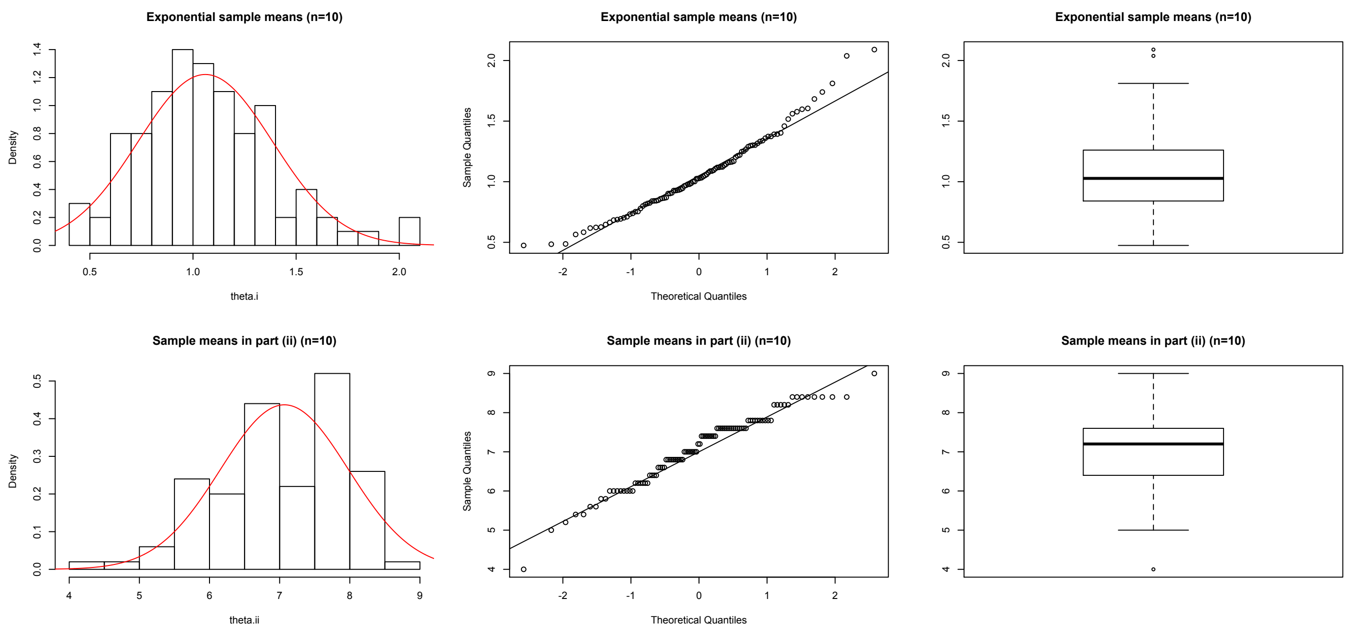
SOLUTION: For this distribution, use the method discussed in Chapter 2. That is define $F_1 = 1/8$, $F_2 = 2/8$ and generate from X in the following fashion

$$X = \begin{cases} 1 & \text{if } U \leq F_1 = \frac{1}{8} \\ 3 & \text{if } \frac{1}{8} = F_1 < U \leq F_2 = \frac{2}{8} \\ 9 & \text{if } U > F_2 = \frac{2}{8} \end{cases}$$

```
Xii<-matrix(NA,nrow=m,ncol=n)
Xii[U<=1/8]<-1
Xii[U<=2/8 & U>1/8 ]<-3
Xii[U>2/8]<-9
```

b) [**5 mark**] For each distribution calculate the corresponding sample means and discuss the merits of the CLT approximation to the distribution of the sample mean in each case. You can use histogram, Q-Q plots, box plots, etc. for your analysis.
SOLUTION:

```
theta.i<-apply(Xi,MARGIN=1,FUN=mean)
theta.ii<-apply(Xii,MARGIN=1,FUN=mean)
par(mfrow=c(2,3))
hist(theta.i,15,freq=FALSE,main="Exponential sample means (n=10)")
lines(seq(0,3,0.01),dnorm(seq(0,3,0.01),mean=mean(theta.i),sd=sd(theta.i)),col=2)
qqnorm(theta.i, main="Exponential sample means (n=10)")
qqline(theta.i)
boxplot(theta.i,main="Exponential sample means (n=10)")
hist(theta.ii,15,freq=FALSE,main="Sample means in part (ii) (n=10)")
lines(seq(0,10,0.01),dnorm(seq(0,10,0.01),mean=mean(theta.ii),sd=sd(theta.ii)),col=2)
qqnorm(theta.ii, main="Sample means in part (ii) (n=10)")
qqline(theta.ii)
boxplot(theta.ii,main="Sample means in part (ii) (n=10)")
```



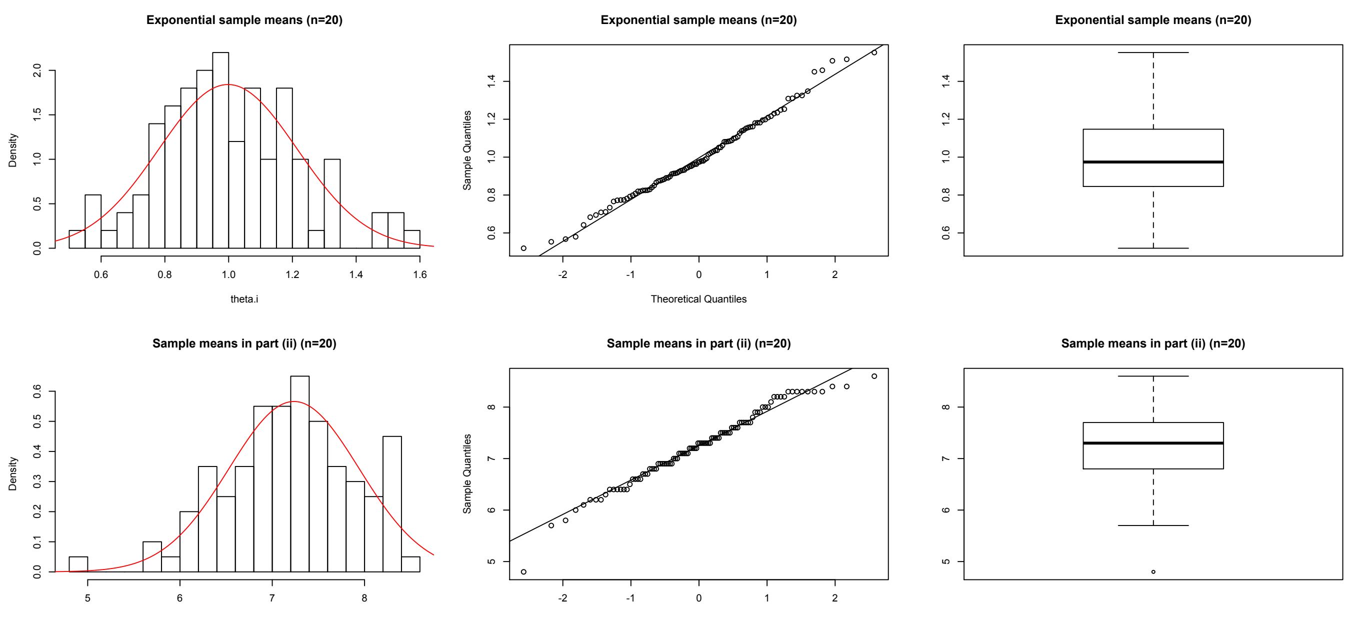
c) [**3 mark**] Repeat (a) and (b) with $n = 20$ and 50.
SOLUTION:

```
m<-100 # number of samples
n<-20  # sample size
U<-matrix(runif(n*m,0,1),nrow=m,ncol=n)
# Part (a) (i)
Xi<--log(U)

# Part (a) (ii)
```

```
Xii<-matrix(NA,nrow=m,ncol=n)
Xii[U<=1/8]<-1
Xii[U<=2/8 & U>1/8 ]<-3
Xii[U>2/8]<-9

# (b)
theta.i<-apply(Xi,MARGIN=1,FUN=mean)
theta.ii<-apply(Xii,MARGIN=1,FUN=mean)
par(mfrow=c(2,3))
hist(theta.i,15,freq=FALSE,main="Exponential sample means (n=20)")
lines(seq(0,3,0.01),dnorm(seq(0,3,0.01),mean=mean(theta.i),sd=sd(theta.i)),col=2)
qqnorm(theta.i, main="Exponential sample means (n=20)")
qqline(theta.i)
boxplot(theta.i,main="Exponential sample means (n=20)")
hist(theta.ii,15,freq=FALSE, main="Sample means in part (ii) (n=20)")
lines(seq(0,10,0.01),dnorm(seq(0,10,0.01),mean=mean(theta.ii),sd=sd(theta.ii)),col=2)
qqnorm(theta.ii, main="Sample means in part (ii) (n=20)")
qqline(theta.ii)
boxplot(theta.ii,main="Sample means in part (ii) (n=20)")
```

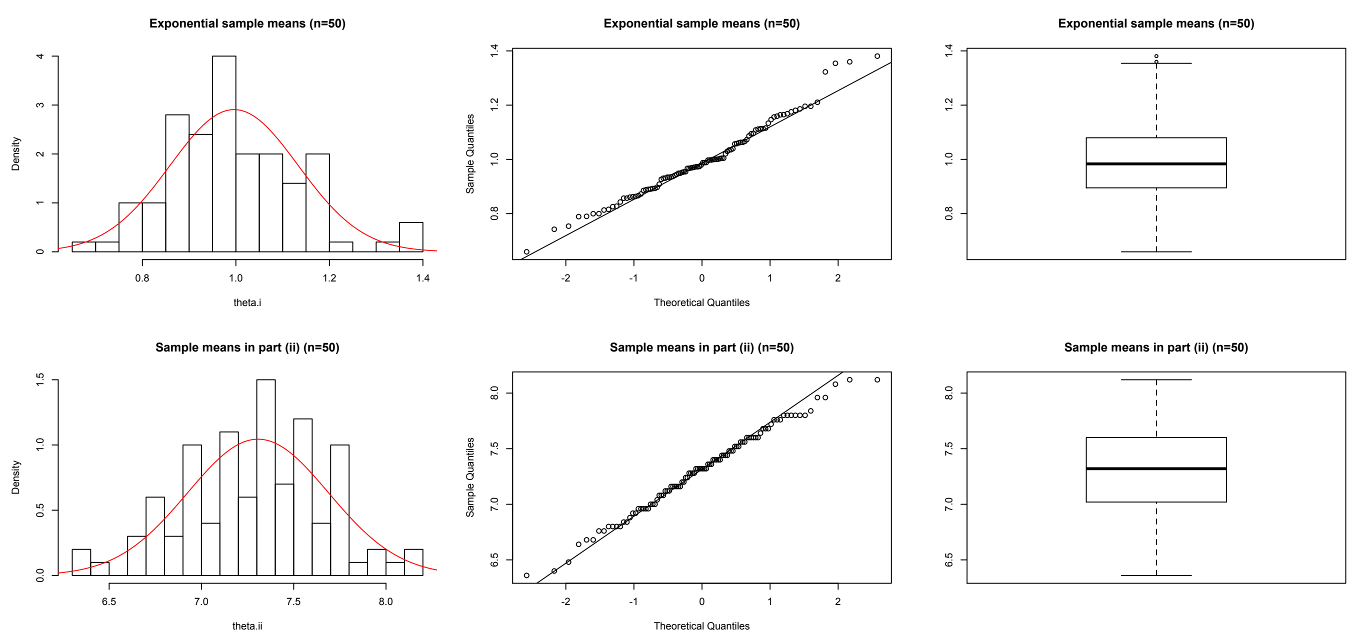



```
m<-100 # number of samples
n<-50  # sample size
U<-matrix(runif(n*m,0,1),nrow=m,ncol=n)
# Part (a) (i)
```

```
Xi<--log(U)

# Part (a) (ii)
Xii<-matrix(NA,nrow=m,ncol=n)
Xii[U<=1/8]<-1
Xii[U<=2/8 & U>1/8 ]<-3
Xii[U>2/8]<-9

# (b)
theta.i<-apply(Xi,MARGIN=1,FUN=mean)
theta.ii<-apply(Xii,MARGIN=1,FUN=mean)
par(mfrow=c(2,3))
hist(theta.i,15,freq=FALSE, main="Exponential sample means (n=50)")
lines(seq(0,3,0.01),dnorm(seq(0,3,0.01),mean=mean(theta.i),sd=sd(theta.i)),col=2)
qqnorm(theta.i, main="Exponential sample means (n=50)")
qqline(theta.i)
boxplot(theta.i,main="Exponential sample means (n=50)")
hist(theta.ii,15,freq=FALSE,main="Sample means in part (ii) (n=50)")
lines(seq(0,10,0.01),dnorm(seq(0,10,0.01),mean=mean(theta.ii),sd=sd(theta.ii)),col=2)
qqnorm(theta.ii, main="Sample means in part (ii) (n=50)")
qqline(theta.ii)
boxplot(theta.ii,main="Sample means in part (ii) (n=50)")
```



d) **[2 mark]** Concisely state your conclusions.

SOLUTION: For both distributions, the normal approximation as stated by the CLT becomes better and better as the sample size n increases. Both distributions are asymmetric and therefore it is expected that the approximation on the tails may not be good enough for the small samples as it is observed.

Problem 8. [10 mark] (Graduate students only) If $X \sim N(0,1)$, the c.d.f. of X is given by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\Big(-\frac{z^2}{2}\Big)\, dz.$$

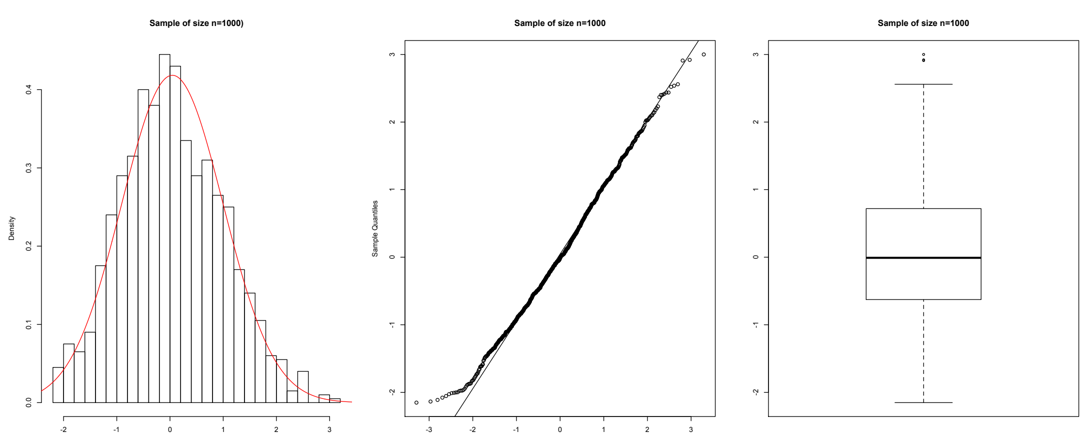
It can be shown that

$$\Phi^{-1}(x) \approx t - \frac{a_0 + a_1 t}{1 + b_1 t + b_2 t^2}$$

for constants $a_0 = 2.30753, a_1 = 0.27061, b_1 = 0.99229, b_2 = 0.04481$ ($t^2 = -2\log x$) (Abramowitz and Stegun, Handbook of mathematical functions, 1964). Write a function that generates the standard normal random numbers using the integral transform method. Is your random number generator sufficiently good enough?

SOLUTION:

```
normR<-function(n){
a0<-2.30753; a1 <- 0.27061
b1 <- 0.99229; b2<- 0.04481
u<-runif(n,0,1)
t<-sqrt(-2*log(u))
x<-t-(a0+a1*t)/(1+b1*t+b2*t^2)
return(x)
}
par(mfrow=c(1,3))
X<-normR(1000)
hist(X,20,freq=FALSE,main="Sample of size n=1000)")
lines(seq(-4,4,0.01),dnorm(seq(-4,4,0.01),mean=mean(X),sd=sd(X)),col=2)
qqnorm(X, main="Sample of size n=1000")
qqline(X)
boxplot(X,main="Sample of size n=1000")
```



The above random number generator seems to behave well except on the tails of the distribution. We need to have more terms in the formula to improve the approximations.

Problem 9. [10 mark] (Graduate students only) Suppose that the random variable X has mean μ and can be simulated. Further, suppose that δ is non-linear function, and that we wish to estimate $\theta = \mathrm{E}\left[\,\delta(X)\,\right]$ using simulation.

Using $h(x) = \delta(\mu) + (x-\mu)\,\delta'(\mu)$ and tuning parameter $\alpha = 1$, estimate θ using control variates. That is, if $X_1, \ldots, X_n$ are iid sample distributed as X, show that for $\alpha = 1$, the controlled estimate of θ is

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\delta(X_i) + (\overline{X}-\mu)\,\delta'(\mu)\,.$$

SOLUTION: The Monte Carlo estimate of θ is

$$\widehat{\theta}_{MC} = \frac{1}{n}\sum_{i=1}^{n}\delta(X_i)\,.$$

If we define $\eta(X) = h(X) = \delta(\mu) + (X-\mu)\,\delta'(\mu)$ and thus $\theta^* = \mathrm{E}(\eta(X)) = \delta(\mu)$ with its Monte Carlo estimate being

$$\widehat{\theta}^*_{MC} = \frac{1}{n}\sum_{i=1}^{n}\eta(X_i) = \frac{1}{n}\sum_{i=1}^{n}\left[\delta(\mu) + (X_i-\mu)\,\delta'(\mu)\right] = \delta(\mu) + (\overline{X}-\mu)\delta'(\mu)\,.$$

Thus for $\alpha = 1$, the control variates estimate is

$$\begin{aligned}\widehat{\theta}_{cv} &= \widehat{\theta}_{MC} + (\widehat{\theta}^*_{MC} - \theta^*)\\ &= \frac{1}{n}\sum_{i=1}^{n}\delta(X_i) + \big(\delta(\mu) + (\overline{X}-\mu)\,\delta'(\mu) - \delta(\mu)\big)\\ &= \frac{1}{n}\sum_{i=1}^{n}\delta(X_i) + (\overline{X}-\mu)\,\delta'(\mu) = \widehat{\theta}\,.\end{aligned} \tag{1}$$

Furthermore, using the fact that for x close to μ, $h(x) \approx \delta(x)$, show that the controlled estimate can be written approximately as

$$\frac{1}{n}\sum_{i=1}^{n}\delta(X_i) + \delta(\overline{X}) - \delta(\mu)\,.$$

SOLUTION: Since by the SLLN $\overline{X}$ is a consistent estimator μ, i.e $\overline{X} \to \mu$ as $n \to \infty$ thus $\overline{X}$ is close to μ and therefore $\eta(\overline{X}) = h(\overline{X}) = \delta(\mu) + (\overline{X}-\mu)\,\delta'(\mu) \approx \delta(\overline{X})$, thus from (1) we have

$$\begin{aligned}\widehat{\theta}_{cv} &= \frac{1}{n}\sum_{i=1}^{n}\delta(X_i) + \big(\delta(\mu) + (\overline{X}-\mu)\,\delta'(\mu) - \delta(\mu)\big)\\ &= \frac{1}{n}\sum_{i=1}^{n}\delta(X_i) + \delta(\overline{X}) - \delta(\mu)\end{aligned} \tag{2}$$

Finally, derive the optimal (theoretical) value of α.

SOLUTION: Recall that the optimal α in $\widehat{\theta}_{cv} = \widehat{\theta}_{MC} + (\widehat{\theta}^*_{MC} - \theta^*)$ is given by

$$\alpha = -\frac{\mathrm{Cov}(\widehat{\theta}_{MC}, \widehat{\theta}^*_{MC})}{\mathrm{Var}(\widehat{\theta}^*_{MC})}\,,$$

where

$$\begin{aligned}
\mathrm{Cov}(\widehat{\theta}_{MC},\, \widehat{\theta}^*_{MC}) &= \mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n}\delta(X_i),\, \delta(\mu) + (\overline{X}-\mu)\,\delta'(\mu)\right)\\
&= \mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n}\delta(X_i),\, \frac{1}{n}\sum_{j=1}^{n}X_j\,\delta'(\mu)\right)\\
&= \frac{\delta'(\mu)}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(\delta(X_i),\, X_j)\\
&= \frac{\delta'(\mu)}{n^2}\sum_{i=1}^{n}\mathrm{Cov}(\delta(X_i),\, X_i) + 0\\
&= \frac{\delta'(\mu)}{n}\mathrm{Cov}(\delta(X_1),\, X_1)
\end{aligned}$$

and

$$\mathrm{Var}(\widehat{\theta}^*_{MC}) = \frac{\delta'^2(\mu)}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\delta'^2(\mu)}{n}\mathrm{Var}(X_1)\,.$$

Thus

$$\alpha = -\frac{\mathrm{Cov}(\widehat{\theta}_{MC},\, \widehat{\theta}^*_{MC})}{\mathrm{Var}(\widehat{\theta}^*_{MC})} = -\frac{\frac{\delta'(\mu)}{n}\mathrm{Cov}(\delta(X_1),\, X_1)}{\frac{\delta'^2(\mu)}{n}\mathrm{Var}(X_1)} = -\frac{\mathrm{Cov}(\delta(X_1),\, X_1)}{\delta'(\mu)\,\mathrm{Var}(X_1)}\,.$$

Problem 10. [10 mark] (Graduate students only) Let $f(\mathbf{x}_1,\mathbf{x}_2)$ and $g(\mathbf{x}_1,\mathbf{x}_2)$ be two p.d.fs, where support of f is a subset of the support of g.

a) Show that,

$$\mathrm{Var}_g\left[\frac{f(\mathbf{X}_1,\mathbf{X}_2)}{g(\mathbf{X}_1,\mathbf{X}_2)}\right] \geq \mathrm{Var}_g\left[\frac{f_1(\mathbf{X}_1)}{g_1(\mathbf{X}_1)}\right]\,,$$

where $f_1(\mathbf{x}_1) = \int f(\mathbf{x}_1,\mathbf{x}_2)d\mathbf{x}_2$ and $f_2(\mathbf{x}_2) = \int f(\mathbf{x}_1,\mathbf{x}_2)d\mathbf{x}_1$ are the marginal densities. Note that variances are taken with respect to g.

SOLUTION: First note

$$\begin{aligned}
\frac{f_1(\mathbf{x}_1)}{g_1(\mathbf{x}_1)} &= \frac{\int f(\mathbf{x}_1,\mathbf{x}_2)d\mathbf{x}_2}{g_1(\mathbf{x}_1)}\\
&= \int \frac{f(\mathbf{x}_1,\mathbf{x}_2)}{g_1(\mathbf{x}_1)\,g_{2|1}(\mathbf{x}_2\mid\mathbf{x}_1)}\,g_{2|1}(\mathbf{x}_2\mid\mathbf{x}_1)d\mathbf{x}_2\\
&= \int \frac{f(\mathbf{x}_1,\mathbf{x}_2)}{g(\mathbf{x}_1,\mathbf{x}_2)}\,g_{2|1}(\mathbf{x}_2\mid\mathbf{x}_1)d\mathbf{x}_2\\
&= \mathrm{E}_g\left[\frac{f(\mathbf{X}_1,\mathbf{X}_2)}{g(\mathbf{X}_1,\mathbf{X}_2)}\,\middle|\,\mathbf{X}_1=\mathbf{x}_1\right]\,.
\end{aligned} \tag{3}$$

Thus

$$\mathrm{Var}_g\left[\frac{f(\mathbf{X}_1,\mathbf{X}_2)}{g(\mathbf{X}_1,\mathbf{X}_2)}\right] = \mathrm{Var}_g\left[\mathrm{E}_g\left[\frac{f(\mathbf{X}_1,\mathbf{X}_2)}{g(\mathbf{X}_1,\mathbf{X}_2)}\,\middle|\,\mathbf{X}_1=\mathbf{x}_1\right]\right] + \mathrm{E}_g\left[\mathrm{Var}_g\left[\frac{f(\mathbf{X}_1,\mathbf{X}_2)}{g(\mathbf{X}_1,\mathbf{X}_2)}\,\middle|\,\mathbf{X}_1=\mathbf{x}_1\right]\right]\,.$$

Sine $\mathrm{E}_g \left[\mathrm{Var}_{_g} \left[\frac{f(\mathbf{X}_1,\mathbf{X}_2)}{g(\mathbf{X}_1,\mathbf{X}_2)} \,\Big|\, \mathbf{X}_1 = \mathbf{x}_1\right]\right] \geq 0$, then from equation (3)

$$\mathrm{Var}_g \left[\frac{f(\mathbf{X}_1,\mathbf{X}_2)}{g(\mathbf{X}_1,\mathbf{X}_2)}\right] \geq \mathrm{Var}_g \left[\mathrm{E}_{_g} \left[\frac{f(\mathbf{X}_1,\mathbf{X}_2)}{g(\mathbf{X}_1,\mathbf{X}_2)} \,\Big|\, \mathbf{X}_1 = \mathbf{x}_1\right]\right] = \mathrm{Var}_g \left[\frac{f_1(\mathbf{X}_1)}{g_1(\mathbf{X}_1)}\right] .$$

The amount of reduction in the variance is given by

$$\mathrm{Var}_g \left[\frac{f(\mathbf{X}_1,\mathbf{X}_2)}{g(\mathbf{X}_1,\mathbf{X}_2)}\right] - \mathrm{Var}_g \left[\frac{f_1(\mathbf{X}_1)}{g_1(\mathbf{X}_1)}\right] = \mathrm{E}_g \left[\mathrm{Var}_{_g} \left[\frac{f(\mathbf{X}_1,\mathbf{X}_2)}{g(\mathbf{X}_1,\mathbf{X}_2)} \,\Big|\, \mathbf{X}_1 = \mathbf{x}_1\right]\right] .$$

Using the analysis of variance (ANOVA) terminology, this is the average “within-group” variation with the group is identified by the variable $\mathbf{X}_1$.

b) Explain what is the message of this result in general, and also in connection to the Rao-Blackwellization and importance sampling.

SOLUTION: The results says that in Monte Carlo computations, it is encouraged to do as much analytical work as possible to reduce the dimensionality of the problem before using a Monte Carlo approximation, which in this process the variance is also reduced. This process could be regarded as Rao-Blackwellization of importance sampling estimate

$$\widehat{\theta}_{IS} = \frac{1}{n} \sum_{i=1}^{n} \delta(\mathbf{X}_{1\,i},\, \mathbf{X}_{2\,i})\, \frac{f(\mathbf{X}_{1\,i},\, \mathbf{X}_{2\,i})}{g(\mathbf{X}_{1\,i},\, \mathbf{X}_{2\,i})} ,$$

of the quantity $\theta = \mathrm{E}_f\left[\delta(\mathbf{X}_1,\mathbf{X}_2)\right]$ is given by

$$\begin{aligned}
\mathrm{E}_{_g} \left[\delta(\mathbf{X}_1,\mathbf{X}_2)\, \frac{f(\mathbf{X}_1,\mathbf{X}_2)}{g(\mathbf{X}_1,\mathbf{X}_2)} \,\Big|\, \mathbf{X}_1 = \mathbf{x}_1\right] &= \int \delta(\mathbf{x}_1,\mathbf{x}_2)\, \frac{f(\mathbf{x}_1,\mathbf{x}_2)}{g(\mathbf{x}_1,\mathbf{x}_2)}\, g_{2\,|\,1}(\mathbf{x}_2 \mid \mathbf{x}_1)\, d\mathbf{x}_2 \\
&= \int \delta(\mathbf{x}_1,\mathbf{x}_2)\, \frac{f(\mathbf{x}_1,\mathbf{x}_2)}{g(\mathbf{x}_1,\mathbf{x}_2)}\, \frac{g(\mathbf{x}_1,\mathbf{x}_2)}{g_1(\mathbf{x}_1)}\, d\mathbf{x}_2 \\
&= \frac{f_1(\mathbf{x}_1)}{g_1(\mathbf{x}_1)} \int \delta(\mathbf{x}_1,\mathbf{x}_2)\, f_{2\,|\,1}(\mathbf{x}_2 \mid \mathbf{x}_1)\, d\mathbf{x}_2 \\
&= \frac{f_1(\mathbf{x}_1)}{g_1(\mathbf{x}_1)}\, \mathrm{E}_{_f} \left[\delta(\mathbf{X}_1,\mathbf{X}_2) \,\Big|\, \mathbf{X}_1 = \mathbf{x}_1\right] .
\end{aligned}$$

Hence, an importance sampling estimate using the ratio $f(\mathbf{X}_1,\mathbf{X}_2)/g(\mathbf{X}_1,\mathbf{X}_2)$ and function $\delta(\mathbf{X}_1,\mathbf{X}_2)$ is always less efficient that the one using the ratio $f_1(\mathbf{x}_1)/g_1(\mathbf{x}_1)$ and $\mathrm{E}_{_f}\left[\delta(\mathbf{X}_1,\mathbf{X}_2) \,\Big|\, \mathbf{X}_1 = \mathbf{x}_1\right]$.