Problem 1

Let X be a continuous random variable with CDF F and PDF f.

(a) Find the conditional CDF of X given $X > a$ (where a is a constant with $\mathsf{P}(X > a) \neq 0$). That is, find $\mathsf{P}(X \leq x \mid X > a)$ for all a, in terms of F.

(b) Find the conditional PDF of X given $X > a$.

(c) Check that the conditional PDF from (b) is a valid PDF, by showing directly that it is nonnegative and integrates to 1.

Solution: (a) From the definition of conditional probability, we have

$$\mathsf{P}(X \leq x \mid X > a) = \frac{\mathsf{P}(X \leq x, X > a)}{\mathsf{P}(X > a)} = \frac{\mathsf{P}(X \leq x, X > a)}{1 - F(a)}.$$

Note the relationship between the events:

$$\{X \leq x, X > a\} = \{X \leq x\} \cap \{X > a\} = \begin{cases} \{X \in (a, x]\}, & \text{if } x \geq a, \\ \emptyset & \text{if } x < a. \end{cases}$$

Therefore the conditional CDF of X given $X > a$ is

$$\mathsf{P}(X \leq x \mid X > a) = \begin{cases} \frac{F(x) - F(a)}{1 - F(a)}, & \text{if } x \geq a, \\ 0 & \text{if } x < a. \end{cases}$$

because $\mathsf{P}(\emptyset) = 0$ and $\mathsf{P}(X \in (a, x]) = \mathsf{P}(X \leq x) - \mathsf{P}(X \leq a) = F(x) - F(a)$.

(b) We denote by $g_a(x)$ the conditional PDF of X given $X > a$. We determine it by taking the derivative with respect to x of the conditional CDF of X given $X > a$:

$$g_a(x) = \frac{\mathrm{d}}{\mathrm{d}x} \mathsf{P}(X \leq x \mid X > a) = \begin{cases} \frac{f(x)}{1 - F(a)}, & \text{if } x \geq a, \\ 0 & \text{if } x < a. \end{cases}$$

(c) Since

$$\mathsf{P}(X > a) \neq 0 \implies \mathsf{P}(X > a) > 0 \implies F(a) = 1 - \mathsf{P}(X > a) < 1,$$

and $f(x) \geq 0$ for any $x \in \mathbb{R}$, it follows that $g_a(x) \geq 0$ for any $x \in \mathbb{R}$. Next, we show that $g_a(x)$ integrates to 1:

$$\begin{aligned} \int_{-\infty}^{\infty} g_a(x)\,\mathrm{d}x &= \int_a^{\infty} \frac{f(x)}{1 - F(a)}\,\mathrm{d}x, \\ &= \frac{1}{1 - F(a)} \int_a^{\infty} f(x)\,\mathrm{d}x, \\ &= \frac{1}{1 - F(a)} \left(1 - \int_{-\infty}^{a} f(x)\,\mathrm{d}x\right), \\ &= \frac{1}{1 - F(a)} (1 - F(a)), \\ &= 1. \end{aligned}$$

Problem 2

A circle with a random radius $R \sim \mathsf{Unif}(0,1)$ is generated. Let A be its area.

(a) Use simulations in R (the statistical programming language) to numerically estimate the mean and the variance of A.

(b) Find the theoretical mean and the variance of A, without first finding the CDF or PDF of A. Compare with your numerical results from (a).

(b) Find the CDF and PDF of A.

Solution: (a) The code for performing this simulation is in Listing 1. We can make random draws from the distribution of the area A by taking draws from $R \sim \mathsf{Unif}(0,1)$ and noting $A = \pi R^2$. Based on a sample of size $100,000$ from the distribution of R, the Monte Carlo estimate for the mean of A is 1.048, and the Monte Carlo estimate for the variance of A is 0.879.

```r
#set the seed
set.seed(0)

#generate 100000 random samples from uniform (0,1)
R <- runif(100000)
#calculate the area of the circle with the radii in the vector R
A = pi*R^2

#Monte Carlo estimate of the mean of A
mean(A)
#Monte Carlo estimate of the variance of A
var(A)
```

Listing 1: Code implementing the simulations for Problem 2 part (a)

(b) We use the continuous version of LOTUS. We have

$$
\begin{aligned}
\mathsf{E}(A) &= \pi \mathsf{E}\left(R^2\right) = \frac{\pi}{3} \approx 1.047. \\
\mathsf{E}\left(A^2\right) &= \pi^2 \mathsf{E}\left(R^4\right) = \pi^2 \int_0^1 r^4 \, \mathrm{d}r = \pi^2 \left(\frac{r^5}{5}\right)\Big|_0^1 = \frac{\pi^2}{5}, \\
\mathsf{Var}(A) &= \mathsf{E}\left(A^2\right) - (\mathsf{E}(A))^2 = \frac{\pi^2}{5} - \frac{\pi^2}{9} = \frac{4\pi^2}{45} \approx 0.877.
\end{aligned}
$$

Compare how close the Monte Carlo estimates from (a) are to the exact values of the mean and the variance of A.

(c) Let F and f denote the CDF and PDF of A, respectively. Since $R \in [0,1]$, we have:

$$
F(a) = \mathsf{P}(A \leq a) = \mathsf{P}\left(\pi R^2 \leq a\right) = \mathsf{P}\left(R \leq \sqrt{\frac{a}{\pi}}\right) = \begin{cases} 0, & \text{if } a < 0, \\ \sqrt{\frac{a}{\pi}}, & \text{if } 0 \leq a \leq \pi, \\ 1, & \text{otherwise.} \end{cases}
$$

We can find the PDF f by differentiating the above expression for the CDF F:

$$f(a) = \frac{\mathrm{d}}{\mathrm{d}a}F(a) = \begin{cases} 0, & \text{if } a < 0, \\ \frac{1}{2\sqrt{a\pi}}, & \text{if } 0 \leq a \leq \pi, \\ 0, & \text{otherwise.} \end{cases}$$

Problem 3

A stick of length 1 is broken at a uniformly random point, yielding two pieces. Let X and Y be the lengths of the shorter and longer pieces, respectively, and let $R = X/Y$ be the ratio of the lengths of X and Y.

(a) Use simulations in R (the statistical programming language) to gain some understanding about the distribution of the random variable R. Numerically estimate the expected value of R and $1/R$.

(b) Find the CDF and PDF of R.

(c) Find the expected value of R (if it exists).

(d) Find the expected value of $1/R$ (if it exists).

Solution: (a) The code for performing this simulation is in Listing 2. To sample from the distribution of R, we first sample break points $u_1, u_2, \ldots, u_n$ which are samples from $U \sim \mathsf{Unif}(0,1)$. Then we set $x_i = \min(u_i, 1-u_i)$, and $y_i = 1 - x_i$, $i = 1, 2, \ldots, n$. Finally, we set $r_i = x_i/y_i$, $i = 1, 2, \ldots, n$. These are independent samples from the distribution of R. Similarly, $1/r_i = y_i/x_i$, $i = 1, 2, \ldots, n$, are independent samples from the distribution of $1/R$. Based on $n = 100,000$ samples from the distribution of R, we obtain a Monte Carlo estimate

$$\frac{1}{n}(r_1 + r_2 + \ldots + r_n)$$

for the expected value of R to be 0.387, and a Monte Carlo estimate

$$\frac{1}{n}\left(\frac{1}{r_1} + \frac{1}{r_2} + \ldots + \frac{1}{r_n}\right)$$

for the expected value of $1/R$ to be 24.31.

```
#set the seed
set.seed(0)

#simulate from the distribution of R
U <- runif(100000)
X <- rep(0, length(U))
for(i in 1:length(U)) {
  X[i] = min(U[i], 1-U[i])
}
R <- X/(1-X)

# Examine the distributions of X, R and 1/R
hist(X); hist(R); hist(1/R)
```

```

#Monte Carlo estimates for the means of R and 1/R
mean(R)
mean(1/R)
```

Listing 2: Code implementing the simulations for Problem 3 part (a)

(b) Let F and f denote the CDF and PDF of R, respectively. Note that $R \geq 0$ since $X \geq 0$ and $Y > 0$. Moreover, $R \leq 1$ because $X \leq Y$. Thus $R \in [0,1]$. For $0 \leq r \leq 1$, we have:

$$F(r) = \mathsf{P}\left(\frac{X}{1-X} \leq r\right) = \mathsf{P}(X \leq r - rX) = \mathsf{P}\left(X \leq \frac{r}{1+r}\right)$$

Since $r \geq 0$, we have $0 \leq \frac{r}{1+r} \leq \frac{1}{2}$. But $X = \min(U, 1-U)$. We find the CDF of X for $0 \leq x \leq \frac{1}{2}$:

$$\begin{aligned}
\mathsf{P}(X \leq x) &= \mathsf{P}(\min(U, 1-U) \leq x), \\
&= \mathsf{P}(\{U \leq x\} \text{ or } \{1-U \leq x\}), \\
&= P(U \leq x) + P(1-U \leq x) - P(U \leq x, 1-U \leq x), \\
&= x + \mathsf{P}(U \geq 1-x) - 0, \\
&= x + (1 - \mathsf{P}(U < 1-x)), \\
&= x + (1 - (1-x)), \\
&= 2x.
\end{aligned}$$

Here we used that the CDF of the uniform distribution on $[0,1]$ is the identify function,

$$\mathsf{P}(U \leq u) = u, \text{ for } 0 \leq u \leq 1,$$

and the fact that we cannot have both $U \leq x$ and $1-U \leq x$ for any $0 \leq x < \frac{1}{2}$. Note that this implies X is uniformly distributed $\mathsf{Unif}\left(0, \frac{1}{2}\right)$, which should be roughly in line with our intuition. By substituting this expression into the CDF of R, we obtain

$$F(r) = \mathsf{P}\left(X \leq \frac{r}{1+r}\right) = \begin{cases} 0, & \text{if } r < 0, \\ \frac{2r}{1+r}, & \text{if } 0 \leq r \leq 1, \\ 1, & \text{if } r > 1. \end{cases}$$

The PDF of R is obtained by differentiation of its CDF. By the quotient rule, we obtain:

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{2r}{1+r}\right) = \frac{2(1+r) - 2r}{(1+r)^2} = \frac{2}{(1+r)^2}.$$

Therefore the CDF of R is:

$$f(r) = \begin{cases} \frac{2}{(1+r)^2}, & \text{if } 0 \leq r \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

(c) The expected value of R is obtained by employing the definition of expectation:

$$\begin{aligned}
\mathsf{E}(R) &= \int_0^1 \frac{2r}{(1+r)^2}\,\mathrm{d}r,\\
&= \int_0^1 \frac{2(r+1)-2}{(1+r)^2}\,\mathrm{d}r,\\
&= 2\int_0^1 \frac{1}{1+r}\,\mathrm{d}r - 2\int_0^1 \frac{1}{(1+r)^2}\,\mathrm{d}r,\\
&= (2\log(1+r))\Big|_0^1 + 2\left(\frac{1}{1+r}\right)\Big|_0^1,\\
&= 2\log(2) + 1 - 2,\\
&= 2\log(2) - 1 \approx 0.3862944
\end{aligned}$$

Remark that this is very close to the Monte Carlo estimate of 0.387 we obtain at (a).

(d) We use LOTUS to calculate the mean of $1/R$ based on the distribution of R:

$$\begin{aligned}
\mathsf{E}\left(\frac{1}{R}\right) &= \int_0^1 \frac{2}{r\,(1+r)^2}\,\mathrm{d}r,\\
&= \int_0^1 \frac{2}{r} - \frac{2}{r+1} - \frac{2}{(1+r)^2}\,\mathrm{d}r,\\
&= \left(2\log(r) - 2\log(1+r) + \frac{2}{1+r}\right)\Big|_0^1.
\end{aligned}$$

However, this integral does not converge since $\lim_{r\to 0^+}\log(r) = -\infty$. As such, the expected value of $1/R$ does not exist. When comparing this result with the Monte Carlo estimate for the mean of $1/R$ obtained by simulations in part (a), the non-existence of $\mathsf{E}\left(\frac{1}{R}\right)$ could be confusing. You can convince yourselves that this is indeed the case by rerunning the simulation in part (a) with a different seed. You will see that the Monte Carlo estimate for $\mathsf{E}(R)$ you will obtain will be close to the previous Monte Carlo estimate. On the other hand, the Monte Carlo estimate for $\mathsf{E}\left(\frac{1}{R}\right)$ could be very different when the seed is changed. This is indicative of the lack of convergence (hence the non-existence) of the expectation of $1/R$.

Problem 4

Let $U_1, \ldots, U_n$ be i.i.d. $\mathsf{Unif}(0,1)$, and $X = \max(U_1, \ldots, U_n)$.

(a) What is the PDF of X?

(b) What is $\mathsf{E}X$?

(c) Use simulations in R (the statistical programming language) to numerically estimate $\mathsf{E}X$.

Solution: (a) Since $U_i \in (0,1)$ for $i = 1, 2, \ldots, n$, we must have $X \in (0,1)$. First we find the CDF $F(x)$ of X. Because $U_1, \ldots, U_n$ are independent, for $0 \le x \le 1$, we have:

$$F(x) = \mathsf{P}(X \le x) = \mathsf{P}(\max(U_1, ..., U_n) \le x) = \mathsf{P}\left(\cap_{i=1}^n \{U_i \le x\}\right) = \prod_{i=1}^n \mathsf{P}(U_i \le x) = \prod_{i=1}^n x = x^n.$$

We obtain the PDF $f(x)$ of X by taking the derivative of the CDF of X:

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x) = \begin{cases} nx^{n-1}, & \text{if } 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(b) We calculate the mean of X by using the definition of the expectation:

$$\mathsf{E}(X) = \int_0^1 xf(x)\,\mathrm{d}x = \int_0^1 nx^n\,\mathrm{d}x = \left(\frac{n}{n+1}x^{n+1}\right)\Big|_0^1 = \frac{n}{n+1}.$$

(c) The code for producing Monte Carlo estimates for $\mathsf{E}(X)$ is given in Listing 3. We numerically estimate $\mathsf{E}(X)$ for $n \in \{1, 2, ..., 10\}$. We draw $100,000$ samples for n, calculate the Monte Carlo estimates and compare them with the exact theoretical value of $\mathsf{E}(X)$ we obtain in part (b):

n	1	2	3	4	5	6	7	8	9	10
Monte Carlo Estimate	0.50	0.67	0.75	0.80	0.83	0.86	0.88	0.89	0.9	0.91
Exact value $\frac{n}{n+1}$	0.50	0.67	0.75	0.8	0.83	0.86	0.88	0.89	0.9	0.91

When we round to the first two significant digits, the Monte Carlo estimates are the same to the exact theoretical values of $\mathsf{E}(X)$.

```
#set the seed
set.seed(0)

n = seq(from=1,to=10,by=1)

estimate_mean <- function(n)
{
  X <- rep(0, 100000)

  for(i in 1:length(X)) {
    X[i] = max(runif(n))
  }

  #Monte Carlo estimate for the mean of X
  mean(X)
}

sapply(n, estimate_mean)
```

Listing 3: Code implementing the simulations for Problem 4 part (c)

Problem 5

(a) Find $\mathsf{P}(X < Y)$ for $X \sim \mathsf{N}(a, b)$, $Y \sim \mathsf{N}(c, d)$ with X and Y independent. You might need the following result which you should consider known: *if X_1 and X_2 are independent with $X_1 \sim \mathsf{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathsf{N}(\mu_2, \sigma_2^2)$, then $X_1 + X_2 \sim \mathsf{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

(b) Use simulations in R (the statistical programming language) to numerically estimate $\mathsf{P}(X < Y)$ for $X \sim \mathsf{N}(0,1)$, $Y \sim \mathsf{N}(1,5)$ with X and Y independent.

Solution: (a) If $Y \sim \mathsf{N}(c,d)$, then $-Y \sim \mathsf{N}(-c,d)$. Since X and Y are independent, X and $-Y$ are also independent. Using the given result, we obtain:

$$X - Y \sim \mathsf{N}(a-c, b+d).$$

Then

$$\begin{aligned}\mathsf{P}(X<Y) &= \mathsf{P}(X-Y<0), \\ &= \mathsf{P}\left(\frac{X-Y-(a-c)}{\sqrt{b+d}} < -\frac{a-c}{\sqrt{b+d}}\right), \\ &= \Phi\left(\frac{c-a}{\sqrt{b+d}}\right),\end{aligned}$$

where $\Phi(\cdot)$ is the CDF of the standard Normal distribution $\mathsf{N}(0,1)$.

(b) The code for obtaining the Monte Carlo estimate as well as the exact numerical value of $\mathsf{P}(X<Y)$ is given in Listing 4. Based on $100,000$ independent samples from the distributions of X and Y, we obtain a Monte Carlo estimate for $\mathsf{P}(X<Y)$ equal to 0.658. The numerical value of $\mathsf{P}(X<Y)$ based on our derivation from part (a) rounded to the first three digits is also 0.658.

```r
#set the seed
set.seed(0)

X <- rnorm(100000)
Y <- rnorm(100000, mean = 1, sd = sqrt(5))

#Monte Carlo estimate of the probability that X<Y
mean(X < Y)

#Obtaining the numerical value of the (exact) probability that X<Y
pnorm(1/sqrt(1+5))
```

Listing 4: Code implementing the simulations for Problem 5 part (b)

Problem 6

The heights of men in the United States are normally distributed with mean 69.1 inches and standard deviation 2.9 inches. The heights of women are normally distributed with mean 63.7 inches and standard deviation 2.7 inches. Let x be the average height of 100 randomly sampled men, and y be the average height of 100 randomly sampled women.

(a) What is the distribution of $x - y$?

(b) Using simulations in R, calculate the Monte Carlo estimates of the mean and standard deviation of the distribution of $x - y$. Compare these estimates with the exact values of the mean and standard deviation.

(c) What is the probability that a randomly sampled man is taller than a randomly sampled woman? Please do not answer this question by reporting a Monte Carlo estimate of this probability.

Solution: (a) Let X and Y be the heights of a random man and a random woman, respectively. The problem states that $X \sim \mathsf{N}(69.1, 2.9^2)$ and $Y \sim \mathsf{N}(63.7, 2.7^2)$. Let $X_1, \ldots, X_{100}$ be the heights of 100 sampled men, and $Y_1, \ldots, Y_{100}$ be the heights of 100 sampled women. Then

$$\begin{aligned} x = \frac{1}{n}(X_1 + \ldots + X_{100}) &\sim \mathsf{N}\left(69.1, \frac{2.9^2}{100}\right), \\ y = \frac{1}{n}(Y_1 + \ldots + Y_{100}) &\sim \mathsf{N}\left(63.7, \frac{2.7^2}{100}\right). \end{aligned}$$

Thus

$$x - y \sim \mathsf{N}\left(69.1 - 63.7, \frac{2.9^2}{100} + \frac{2.7^2}{100}\right).$$

Therefore the exact mean of $x - y$ is $69.1 - 63.7 = 5.4$, and the exact standard deviation of $x - y$ is $\sqrt{\frac{2.9^2}{100} + \frac{2.7^2}{100}} = 0.396$.

(b) You need to repeatedly sample the heights of 100 men and 100 women, and calculate the difference of their average heights. This gives you samples from the distribution of $x - y$. The code below generates 1000 such samples.

```
n <- 1000
x <- rep(0,n)
y <- rep(0,n)
for(i in 1:n)
{
    men <- rnorm(100,mean=69.1,sd=2.9)
    women <- rnorm(100,mean=63.7,sd=2.7)
    x[i] <- mean(men)
    y[i] <- mean(women)
}
z <- x-y
#calculate mean and std of the difference
cat(mean(z),sd(z))
```

Listing 5: Code for Problem 6 (b)

The simulated mean of the difference in heights between men and women is 5.41, while its standard deviation is 0.387. The exact values are 5.4 and 0.396. Figure 1 shows the histogram of the samples from the distribution of $x - y$ together with the true normal distribution of $x - y$.

This is the code you need to reproduce Figure 1:

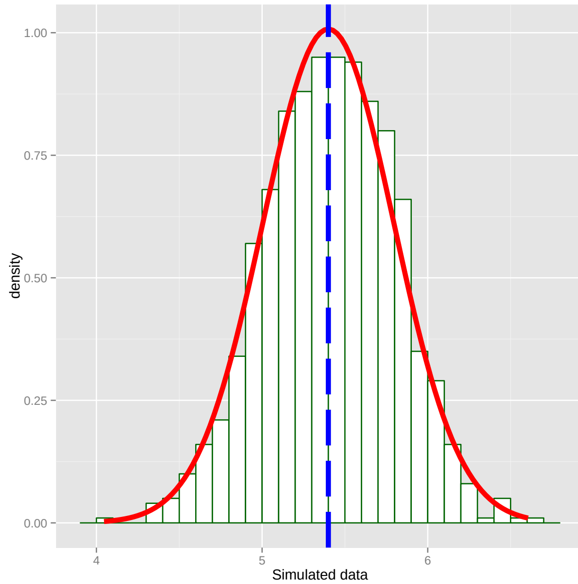

Figure 1: Histogram of the simulated difference in heights between men and women.

```
require(ggplot2)
pdf("simulated.pdf")
m=qplot(z, geom="blank",xlab="Simulated data")
m1 = m + geom_histogram(aes(y = ..density..),binwidth=0.1,colour = "darkgreen", fill =
     "white")
m2 = m1 + stat_function(geom="line", size=2,color="red",fun=dnorm, arg=list(mean=5.4,
     sd=0.396))
m3 = m2 + geom_vline(xintercept = 5.4,color="blue",linetype="longdash",size=2)
print(m3)
dev.off()
```

Listing 6: Code for Figure 1

(c) You need to calculate $\mathsf{P}(X-Y>0)$. Since X and Y are two independent Normally distributed random variables, $X-Y$ also follows a Normal distribution:

$$X - Y \sim \mathsf{N}(69.1 - 63.7, 2.9^2 + 2.7^2) = \mathsf{N}(5.4, 3.962323^2).$$

This implies:

$$\frac{(X-Y)-5.4}{3.962323} \sim \mathsf{N}(0,1).$$

Thus

$$\mathsf{P}(X-Y>0) = \mathsf{P}\left(\frac{(X-Y)-5.4}{3.962323} > -\frac{5.4}{3.962323}\right) = \Phi\left(\frac{5.4}{3.962323}\right) = \Phi(1.362837) = 0.9135331.$$

Problem 7

Suppose that Y is binomially distributed $Y \sim \mathsf{Bin}(n = 5, \theta)$ where θ is unknown. Furthermore, assume that

$$\theta \in \Theta = \{0.0, 0.1, \ldots, 0.9, 1.0\},$$

and that *a priori* we view each of these 11 possibilities as equally likely, i.e.

$$\mathsf{P}(\theta) = \frac{1}{11}, \text{ for each } \theta \in \Theta.$$

(a) Using the Bayes' rule, write down a formula for the posterior distribution $\mathsf{P}(\theta \mid y)$ in terms of θ_i, and simplify as much as possible.

(b) For $y = 0$, make a plot for $\mathsf{P}(\theta \mid y)$ for each $\theta \in \Theta = \{0.0, 0.1, \ldots, 0.9, 1.0\}$. In other words, make a plot with the horizontal axis representing the 11 values of θ and the vertical axis representing the corresponding values of $\mathsf{P}(\theta \mid y)$.

(c) Repeat (b) for each $y \in \{1, 2, 3, 4, 5\}$, so in the end you have six plots (including the one in (b)). Describe what you see in your plots and discuss whether or not they make sense.

Solution: (a) Since Y follows a Binomial distribution and all possible values of θ are a priori equally likely, we obtain from the Bayes' rule:

$$\mathsf{P}(\theta \mid y) \propto \theta^y (1-\theta)^{n-y}$$

We denote the i-th value θ can take by $\theta_i = 0.1 * i$ for $i = 0, \ldots, 10$. It follows that the posterior distribution of θ has the form:

$$\mathsf{P}(\theta_j \mid y) = \frac{\theta_j^y (1-\theta_j)^{n-y}}{\sum_{i=0}^{10} \theta_i^y (1-\theta_i)^{n-y}}, \text{ for } j = 0, \ldots, 10.$$

(b) and (c). The six plots of $\mathsf{P}(\theta \mid y)$ are shown in Figures 2, 3, 4, 5, 6 and 7. In each case, the MLE of θ (i.e., the value of $\theta \in \Theta$ which maximizes the binomial likelihood

$$L(\theta) = \mathsf{P}(y \mid \theta) \propto \theta^y (1-\theta)^{n-y}$$

is equal with the mode of the posterior (i.e., the value of $\theta \in \Theta$ which maximizes the posterior distribution $\mathsf{P}(\theta \mid y)$). This makes perfect sense since, a priori, all the possible values of $\theta \in \Theta$ are equally likely, i.e.

$$\mathsf{P}(\theta \mid y) \propto L(\theta)$$

```
n=5
theta = seq(0,1,0.1)
postdistTheta = function(theta,n,y)
{
  return (theta^y*(1-theta)^(n-y))
```

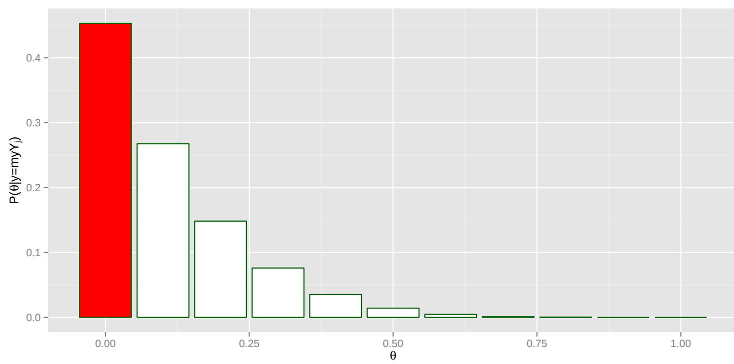

Figure 2: Problem 7 (b). Posterior of θ for $y = 0$. The MLE of θ is $\hat{\theta}_{MLE} = 0$ which coincides with the posterior mode which is shown in red.

```
}
myY = seq(0,5,1)

require(ggplot2)
j=6
pdf(paste('plotProb2Y',myY[j],'.pdf',sep=""),width=10,height=5)
z = postdistTheta(theta,n,myY[j])
z = z/sum(z)
f <- ggplot(data.frame(x=factor(theta),y=z), aes(x=theta,y=z))+
  geom_bar(stat = "identity",fill=c(rep("white",10),"red"), colour="darkgreen")+
  xlab(expression(theta))+
  ylab(expression(paste('P(',theta,'|y=',myY[j],')',sep="")))
plot(f)
dev.off()
```

Listing 7: R code for Problem 7

Problem 8

(a) Starting from independent uniform random variables ($U \sim \mathsf{Uni}(0, 1)$), devise an algorithm to generate independent samples from a Logistic distribution, having density

$$f(x) = \frac{e^{-x}}{(1 + e^{-x})^2}, \qquad \text{for } x \in \mathbb{R}$$

(b) Implement your sampling algorithm in R, and use your code to produce a Monte Carlo estimate of $\mathsf{P}(X \in (2, 3))$ where X is a random variable that has a Logistic distribution. [*Please note that using a function that already exists in an R library that samples from the Logistic distribution will not help you obtain any points for this question*.]

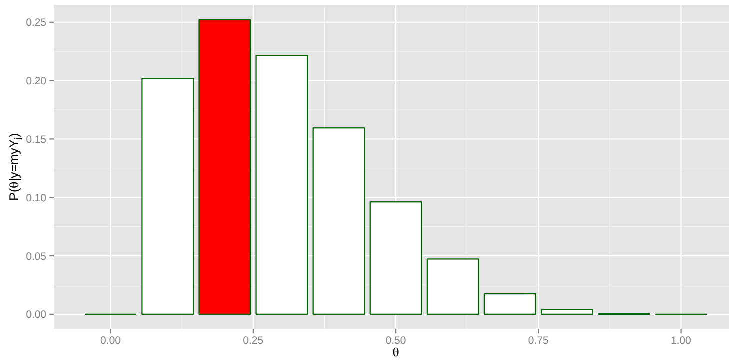

Figure 3: Problem 7 (c). Posterior of θ for $y = 1$. The MLE of θ is $\hat{\theta}_{MLE} = \frac{1}{5} = 0.2$ which coincides with the posterior mode which is shown in red.

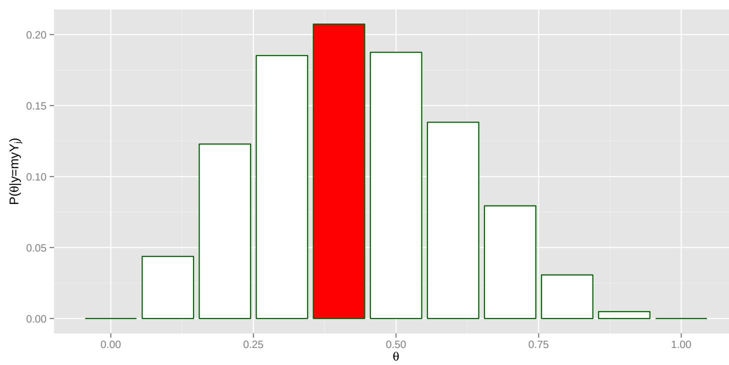

Figure 4: Problem 7 (c). Posterior of θ for $y = 2$. The MLE of θ is $\hat{\theta}_{MLE} = \frac{2}{5} = 0.4$ which coincides with the posterior mode which is shown in red.

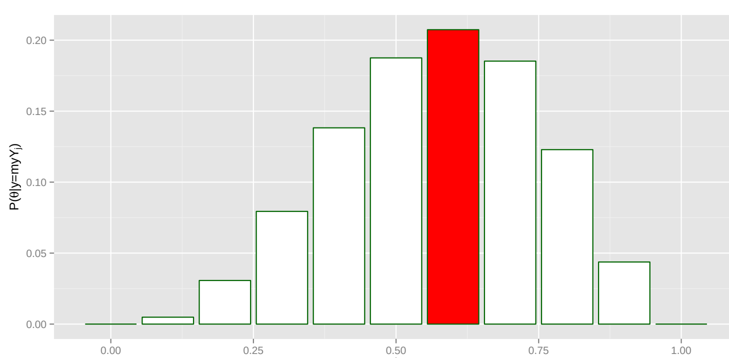

Figure 5: Problem 7 (c). Posterior of θ for $y = 3$. The MLE of θ is $\hat{\theta}_{MLE} = \frac{3}{5} = 0.6$ which coincides with the posterior mode which is shown in red.

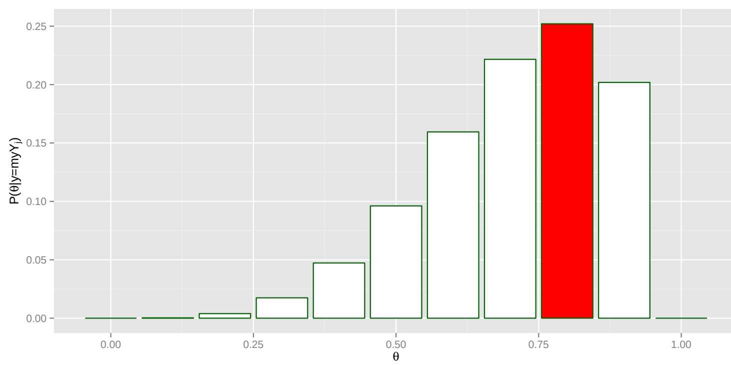

Figure 6: Problem 7 (c). Posterior of θ for $y = 4$. The MLE of θ is $\hat{\theta}_{MLE} = \frac{4}{5} = 0.8$ which coincides with the posterior mode which is shown in red.

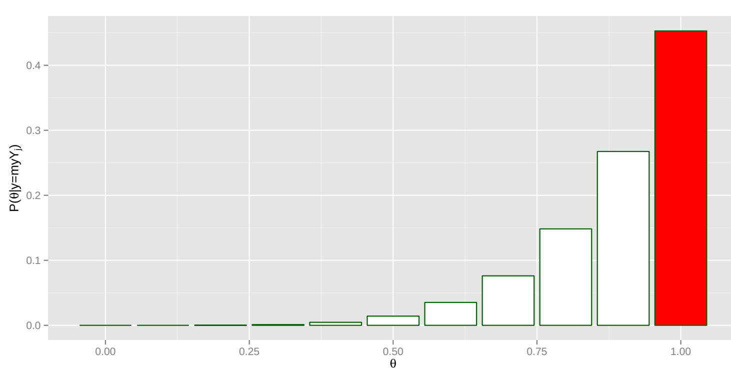

Figure 7: Problem 7 (c). Posterior of θ for $y = 5$. The MLE of θ is $\hat{\theta}_{MLE} = \frac{5}{5} = 1$ which coincides with the posterior mode which is shown in red.

Solution: The CDF of the logistic distribution is

$$F(x) = \frac{1}{1 + e^{-x}}.$$

It is easy to check that $F'(x) = f(x)$ where $f(x)$ is the density function given in the text of the problem. This CDF is strictly increasing, hence it is injective, which implies it is invertible. Its inverse is determined as follows:

$$F(x) = y \Leftrightarrow 1 + e^{-x} = \frac{1}{y} \Leftrightarrow x = -\log\left(\frac{1}{y} - 1\right).$$

Thus the inverse of the CDF $F(x)$ is

$$F^{-1}(y) = -\log\left(\frac{1}{y} - 1\right)$$

The method of inversion says that independent samples θ from the Logistic distribution are generated by sampling $u \sim \mathsf{Uni}(0, 1)$ and applying the inverse CDF $\theta = F^{-1}(u) = -\log\left(\frac{1}{u} - 1\right)$. Please note that, by using the method of inversion, we satisfied the requirement of using only uniform $\mathsf{Uni}(0, 1)$ random variables. The R code implementing the method of inversion is given here:

```r
#problem 8
library(stats)
#first use the function "rlogis"
nSimulations = 10000
simRlogis = rlogis(nSimulations, location=0, scale=1)
plot(density(simRlogis), main="rlogis samples")
curve(dlogis, from=min(simRlogis), to=max(simRlogis), add=TRUE, col="red")
length(which((simRlogis >2) & (simRlogis <3)))/nSimulations
#0.0701
#MC estimate of P(x in (2,3))
#Method of inversion sampling
simInversion = sapply(runif(nSimulations), function(x) {-log(1/x-1)})
plot(density(simInversion), main="Method of inversion samples")
curve(dlogis, from=min(simInversion), to=max(simInversion), add=TRUE, col="red")
length(which((simInversion >2) & (simInversion <3)))/nSimulations
#0.0738
```

Listing 8: R code for Problem 8

We used the function `rlogis` which is available in R to simulate from the Logistic distribution – see Figure 8. The Monte Carlo estimate of $\mathsf{P}(X \in (2, 3))$ based on the samples from `rlogis` is 0.07. We also used the method of inversion sampler developed above to sample from the Logistic distribution – see Figure 9. We see that the results obtained with the method of inversion are quite similar to the results obtained with the `rlogis` function. In particular, the Monte Carlo estimate of $\mathsf{P}(X \in (2, 3))$ is also 0.07.

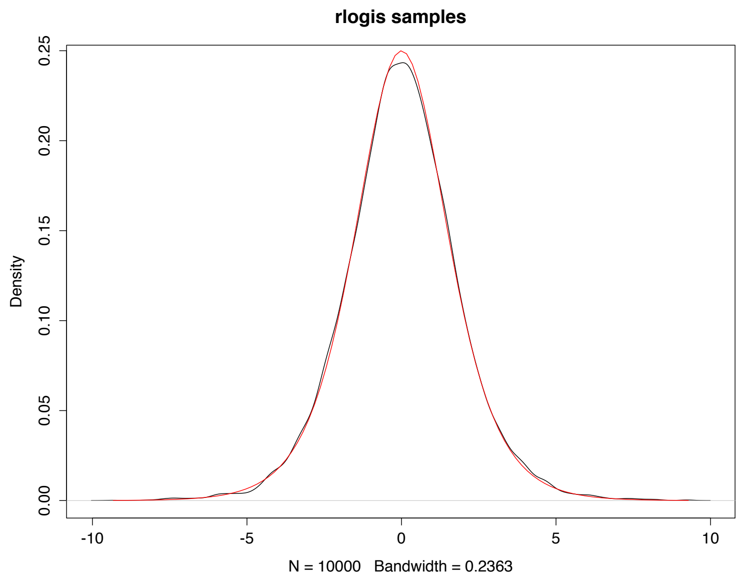

Figure 8: Monte Carlo estimate of the Logistic density based on samples obtained with the `rlogis` function (black curve). The red curve represents the true density function.

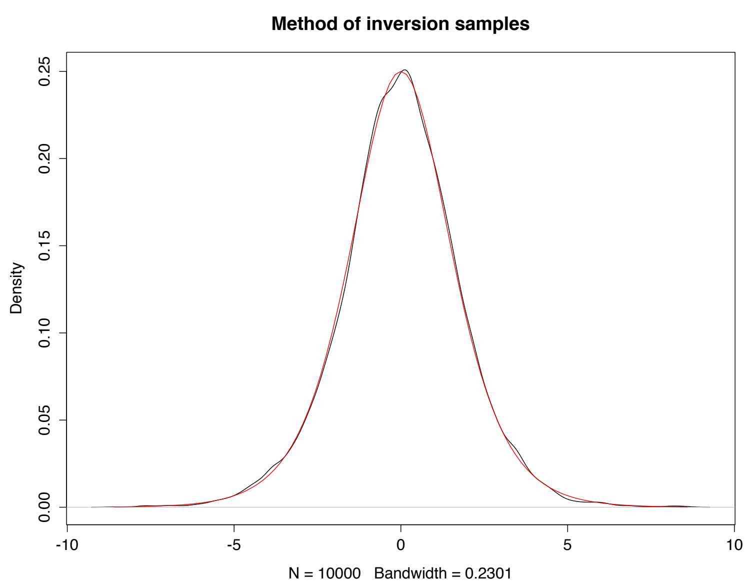

Figure 9: Monte Carlo estimate of the Logistic density based on samples obtained with the method of inversion (black curve). The red curve represents the true density function.