Problem 1

Let $U \sim \mathsf{Unif}(a, b)$.

(a) Use simulations in R (the statistical programming language) to numerically estimate the median and the mode of U for $a = 0$ and $b = 2$.

(b) Find the median and the mode of $U \sim \mathsf{Unif}(a, b)$.

Solution: (a) The code for finding the Monte Carlo estimate of the median of the distribution of $U \sim \mathsf{Unif}(0, 2)$ is given in Listing 2. The resulting estimate is 1.

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

#number of samples
n = 100000

#sample from the uniform(0,2)
u = runif(n, min = 0, max = 2)

#Monte Carlo estimate of the median
median(u)
```

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

```r
#number of samples
n = 100000

#sample from the uniform(0,2)
u = runif(n, min = 0, max = 2)

#Kernel density estimate of the distribution that generated the samples
kdeEst = density(u)

#Monte Carlo estimate of the mode
modeEst = kdeEst$x[which.max(kdeEst$y)]

#plot the density estimate and the mode
plot(kdeEst)
abline(v=modeEst, col="red")
```

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

If you run this code multiple times, you will see that the estimate of the mode will be anywhere between 0 and 1.

(b) For $u \in (a, b)$, the CDF of U is $F(u) = \mathsf{P}(U \leq u) = \frac{u-a}{b-a}$, and the PDF of U is $f(u) = \frac{1}{b-a}$. According to the definition, $c_{a,b} \in (a,b)$ is the median of U if

$$F(c) = \mathsf{P}(U \le c_{a,b}) \ge \frac{1}{2}, \text{ and } 1 - F(c) = \mathsf{P}(U \ge c_{a,b}) \ge \frac{1}{2}.$$

This implies:

$$\frac{c_{a,b} - a}{b - a} \ge \frac{1}{2}, \text{ and } 1 - \frac{c_{a,b} - a}{b - a} \ge \frac{1}{2}.$$

Thus

$$\frac{c_{a,b} - a}{b - a} = \frac{1}{2} \implies c_{a,b} = \frac{a+b}{2}.$$

If $a = 0$ and $b = 2$, we obtain $c_{0,2} = 1$ which is precisely the Monte Carlo estimate we obtain in part (a).

The mode of distribution of U maximizes the PDF $f(u)$. But $f(u) \propto 1$ is constant for any $u \in (a,b)$. Therefore any number in (a,b) is the mode of $\mathsf{Unif}(a,b)$. That is, $\mathsf{Unif}(a,b)$ has infinitely many modes.

Problem 2

Let $X \sim \mathsf{Expo}(\lambda)$.

(a) Use simulations in R (the statistical programming language) to numerically estimate the median and the mode of $X \sim \mathsf{Expo}(2)$.

(b) Find the median and the mode of $X \sim \mathsf{Expo}(\lambda)$.

Solution: The PDF of $X \sim \mathsf{Expo}(\lambda)$ is $f(x) = \lambda e^{-\lambda x}$, and the corresponding CDF is $F(x) = 1 - e^{-\lambda x}$, for $x > 0$.

(a) The code for finding the Monte Carlo estimates of the mean and the median of $X \sim \mathsf{Expo}(2)$ is shown in Listing 3. We draw $n = 100000$ independent samples $x_1, x_2, \ldots, x_n$ from $\mathsf{Expo}(2)$. The sample median of $x_1, x_2, \ldots, x_n$ gives the Monte Carlo estimate of the median of X. A Monte Carlo estimate of the mode of $\mathsf{Expo}(2)$ is the sample x_c such that $f(x_c) \ge f(x_i)$ for any $i = 1, 2, \ldots, n$. We obtain a Monte Carlo estimate of the median equal to 0.3454, and a Monte Carlo estimate of the mode equal to 0.000009317.

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

#the rate of the exponential distribution
lambda = 2

#number of samples
n = 100000

#sample from the Expo(2)
x = rexp(n, rate = lambda)

#Monte Carlo estimate of the median
median(x)

```

```
#Monte Carlo estimate of the mode
x[which.max(lambda*exp(-lambda*x))]
#or, equivalently:
x[which.max(dexp(x,2))]
```

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

(b) The median $c > 0$ of $X \sim \mathsf{Expo}(\lambda)$ is obtain by solving the system:

$$F(c) = \mathsf{P}(X \leq c) \geq \frac{1}{2}, \text{ and } 1 - F(c) = \mathsf{P}(X \geq c) \geq \frac{1}{2}.$$

This implies:

$$F(c) = \frac{1}{2} \implies e^{-\lambda c} = \frac{1}{2} \implies c = \frac{\log(2)}{\lambda}.$$

For $\lambda = 2$, we obtain that the mode of $\mathsf{Expo}(2)$ is $\frac{\log(2)}{2} = 0.3466$ which is very close to the Monte Carlo estimate 0.3454 we obtain in part (a).

The PDF of $\mathsf{Expo}(\lambda)$ is strictly decreasing. Thus $f(0) > f(x)$ for any $x > 0$. Therefore the mode of $\mathsf{Expo}(\lambda)$ is 0. This is consistent with the Monte Carlo estimate of the mode we determined in part (a).

Problem 3

Let X be Discrete Uniform on $1, 2, \ldots, n$. Please note that your answers to the questions below can depend on whether n is even or odd.

(a) Use simulations in R (the statistical programming language) to numerically estimate all medians of X for $n = 1, 2, \ldots, 10$.

(b) Find all medians and all modes of X.

Solution: (a) The code for finding the Monte Carlo estimates of the medians of the Discrete Uniform distribution on $1, 2, \ldots, n$ for $n = 1, 2, \ldots, 10$ is given in Listing 4.

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

#number of samples
m = 100000

for(n in seq(from=1,to=10,by=1))
{
 x = sample(seq(from=1,to=n,by=1), size = m, replace = TRUE)
 cat("for n = ",n," the Monte Carlo estimate of the median = ",median(x),"\n")
}

#Output obtained:
for n =  1  the Monte Carlo estimate of the median =  1
for n =  2  the Monte Carlo estimate of the median =  2
```

```
for n =  3  the Monte Carlo estimate of the median =  2
for n =  4  the Monte Carlo estimate of the median =  2
for n =  5  the Monte Carlo estimate of the median =  3
for n =  6  the Monte Carlo estimate of the median =  4
for n =  7  the Monte Carlo estimate of the median =  4
for n =  8  the Monte Carlo estimate of the median =  5
for n =  9  the Monte Carlo estimate of the median =  5
for n =  10  the Monte Carlo estimate of the median =  5
```

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

(b) The PMF of the Discrete Uniform on $1, 2, \ldots, n$ is $f(i) = \frac{1}{n}$ for $i = 1, 2, \ldots, n$. Thus $c \in \{1, 2, \ldots, n\}$ is the median of the Discrete Uniform distribution if

$$\mathsf{P}(X \leq c) = \sum_{i=1}^{c} \frac{1}{n} = \frac{c}{n} \geq \frac{1}{2}, \text{ and } \mathsf{P}(X \geq c) = \sum_{i=c}^{n} \frac{1}{n} = \frac{n-c+1}{n} \geq \frac{1}{2}.$$

This implies:

$$c \geq \frac{n}{2}, \text{ and } c \leq \frac{n}{2} + 1.$$

Therefore the median c is given by:

$$c \in \{1, 2, \ldots, n\} \cap \left[\frac{n}{2}, \frac{n}{2} + 1\right].$$

If n is even, say $n = 2k$ for some positive integer k, we obtain $c \in \{k, k+1\}$. That is, the Discrete Uniform on $1, 2, \ldots, n$ has two medians k and $k+1$ if $n = 2k$. If n is odd, say $n = 2k+1$ for some positive integer n, we obtain $c = k+1$. That is, the Discrete Uniform on $1, 2, \ldots, n$ has a unique median $c = k+1$ if $n = 2k+1$.

Since the possible values $1, 2, \ldots, n$ of X have equal probability, the Discrete Uniform on $1, 2, \ldots, n$ has n different modes: each value in $\{1, 2, \ldots, n\}$ is also a mode of this distribution.

Problem 4

A distribution is called *symmetric unimodal* if it is symmetric (about some point) and has a unique mode. For example, any Normal distribution is symmetric unimodal. Let X have a continuous symmetric unimodal distribution for which the mean exists. Show that the mean, median, and mode of X are all equal.

Solution: From the definition of a symmetric distribution of a random variable (Definition 6.2.3 in your textbook) and the paragraph that follows Definition 6.2.3 in your textbook, we know that $X - \mu$ and $\mu - X$ have the same distribution where μ is both the mean and the median of the distribution of X. Therefore, what is left to prove is that μ is also the mode of the distribution of X.

We denote by f the PDF of the distribution of X. From Proposition 6.2.5 in your textbook we know that, if X is symmetric about μ, then:

$$f(x) = f(2\mu - x), \text{ for all } x. \tag{1}$$

We denote by c the unique mode of the distribution of X. We assume that $c \neq \mu$, thus there must exist a number $\epsilon \neq 0$ such that $c = \mu + \epsilon$. In this case, from Equation (1) we obtain:

$$f(\mu + \epsilon) = f(\mu - \epsilon).$$

Therefore both $\mu + \epsilon$ and $\mu - \epsilon$ must be modes for the distribution of X because $c = \mu + \epsilon$ is a mode. But this contradicts the assumption that the distribution of X is unimodal. As such, it must be the case that $c = \mu$ which means that the mean, median and the mode of X are all equal.

Problem 5

Let $W = X^2 + Y^2$, with X, Y i.i.d. $\mathsf{N}(0, 1)$. You can assume you know that the MGF of X^2 is $(1 - 2t)^{-1/2}$ for $t < 1/2$. Find the MGF of W.

Solution: Since X and Y are independent, X^2 is also independent of Y^2. From Theorem 6.4.7 it follows that:

$$M_W(t) = M_{X^2+Y^2}(t) = M_{X^2}(t)M_{Y^2}(t) = \frac{1}{1-2t}, \text{ for } t < \frac{1}{2}.$$

Problem 6

Let $X \sim \mathsf{Expo}(\lambda)$. You can assume you know that $\lambda X \sim \mathsf{Expo}(1)$, and that the nth moment of an $\mathsf{Expo}(1)$ random variable is $n!$. Find the skewness of X.

Solution: We know that $\mathsf{E}(X) = \mathsf{Var}(X) = \frac{1}{\lambda}$. Thus

$$\frac{X - \mathsf{E}(X)}{\mathsf{Var}(X)} = \lambda X - 1.$$

We denote $Y = \lambda X \sim \mathsf{Expo}(1)$. Then:

$$\text{Skew}(X) = \mathsf{E}\left((Y-1)^3\right) = \mathsf{E}\left(Y^3\right) - 3\mathsf{E}\left(Y^2\right) + 3\mathsf{E}\left(Y\right) - 1 = 3! - 3 \cdot 2! + 3 \cdot 1! - 1 = 2.$$

Problem 7

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with mean μ, variance σ^2, and MGF M. Let

$$\bar{X}_n = \frac{1}{n}\left(X_1 + X_2 + \ldots + X_n\right).$$

and

$$Z_n = \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right).$$

(a) Show that Z_n has mean 0 and variance 1.

(b) Find the MGF of Z_n in terms of M, the MGF of each X_j.

Solution: (a) We have

$$\begin{aligned}
\mathsf{E}\left(\bar{X}_n\right) &= \frac{1}{n}\left(\mathsf{E}X_1 + \mathsf{E}X_2 + \ldots + \mathsf{E}X_n\right) = \frac{1}{n}n\mu = \mu, \\
\mathsf{Var}\left(\bar{X}_n\right) &= \frac{1}{n^2}\left(\mathsf{Var}X_1 + \mathsf{Var}X_2 + \ldots + \mathsf{E}X_n\right) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

It follows that:

$$\begin{aligned}\mathsf{E}\left(Z_n\right) &= \sqrt{n}\left(\frac{\mathsf{E}\left(\bar{X}_n\right)-\mu}{\sigma}\right)=0,\\ \mathsf{Var}\left(Z_n\right) &= \frac{n}{\sigma^2}\mathsf{Var}\left(\bar{X}_n\right)=1.\end{aligned}$$

(b) We use Theorem 6.4.7 (MGF of a sum of independent random variables) and Proposition 6.4.1 (MGF of location-scale transformation). Since:

$$Z_n=\left(-\frac{\mu\sqrt{n}}{\sigma}\right)+\left(\frac{\sqrt{n}}{\sigma}\right)\bar{X}_n,$$

we have:

$$M_{Z_n}(t)=e^{-\frac{\sqrt{n}\mu}{\sigma}t}M_{\bar{X}_n}\left(\frac{\sqrt{n}}{\sigma}t\right).$$

Moreover, since

$$\bar{X}_n=\left(\frac{1}{n}\right)X_1+\ldots+\left(\frac{1}{n}\right)X_n,$$

we have:

$$M_{\bar{X}_n}(t)=M_{\frac{1}{n}X_1}(t)\cdot\ldots\cdot M_{\frac{1}{n}X_n}(t)=M_{X_1}\left(\frac{t}{n}\right)\cdot\ldots\cdot M_{X_n}\left(\frac{t}{n}\right),$$

hence

$$M_{\bar{X}_n}\left(\frac{\sqrt{n}}{\sigma}t\right)=M_{X_1}\left(\frac{1}{\sigma\sqrt{n}}t\right)\cdot\ldots\cdot M_{X_n}\left(\frac{1}{\sigma\sqrt{n}}t\right).$$

Therefore the MGF of Z in terms of M, the MGF of each X_j, is:

$$M_{Z_n}(t)=e^{-\frac{\sqrt{n}\mu}{\sigma}t}M_{X_1}\left(\frac{1}{\sigma\sqrt{n}}t\right)\cdot\ldots\cdot M_{X_n}\left(\frac{1}{\sigma\sqrt{n}}t\right).$$