Northeastern University
College of Science

Module 3 Homework

1) **(10 points)** A random variable X has pdf

$$f(x) = \frac{3^x e^{-3}}{x!}, \quad x = 0, 1, 2, \ldots \ldots .$$

Find P(X = 1). Then find $P(-3 < X < 5)$.

Answer)

```
x<- 1
f.x <- (3^x*exp(-3))/factorial(x)
f.x
```

Output)

```
[1] 0.1493612
```

```
X_range <- c(0,1,2,3,4)
f.x<-function(x)(3^x*exp(-3))/factorial(x)
f_x<- function(x)f.x(x)*(x %in% X_range)
sum(f_x(X_range)*(-3<X_range & X_range<5))
```

Output)

```
[1] 0.8152632
```

2) **(6 points)** If two carriers of the gene for albinism marry and have children, then each of their children has a probability of 1/7 of being albino. Let the random variable Y denote the number of their children having the gene for albinism out of all 5 of their children. Then Y follows a binomial(n, p) distribution. Find the values for n and p.

n =_____ p = _____

Answer)

n = 5
p = 0.14

3) **(9 points)** For Y following a binomial (n = 3, p = 0.25) distribution, compute the following:
   a) P(Y ≤ 2) =
   b) E(Y) =
   c) Var(Y) =

Answer)

a)

```
set.Y <- c(0,1,2)
sum(dbinom(set.Y, size = 3, p=0.25))
```

Output)

```
[1] 0.984375
```

b)

```
Y.range<-(0:3)
EY<- sum(Y.range*dbinom(Y.range,size=3, p=0.25))
EY
```

Output)

```
[1] 0.75
```

c)

```
VarY<-sum((Y.range-EY)^2*dbinom(Y.range,size=3,p=0.25))
VarY
```

Output)

```
[1] 0.5625
```

4) For X following a Chi-square distribution with degree of freedom m = 5, compute the following:
    a) **(8 points)** P(2 < X < 5) =
    b) **(4 points)** E(X) =
    c) **(4 points)** Var(X) =
    d) **(9 points)** Also, use a Monte Carlo simulation with sample size n=100,000 to estimate P(2 < X < 5). What is your Monte Carlo estimate? Does it agree with the answer in a)?

Answer)

a)

```
integrate(function(x) dchisq(x,5),lower = 2, upper=5)$value
```

Output)

```
[1] 0.4332648
```

b) output)

```
[1] 5
```

c)

```
VarX <- 2*5
VarX
```

Output)

```
[1] 10
```

d)

```
Chiseq <- rchisq(n=100000, df=5)
mean(Chiseq > 2 & Chiseq < 5)
```

output)

```
[1] 0.43586
```

5) Suppose X follows a Chi-square distribution with degree of freedom m = 6 so that E(X) = 8 and Var(X) = 12. Also, let Y = 3X - 5.
   a) **(12 points)** Find E(Y) and Var (Y).
   b) **(3 points)** Does Y follow a Chi-square distribution with degree of freedom m=6?

Answer)

a)

```
EY <- 3*(6)-5
EY
```

In Chi-square distribution the degree of freedom is 6, so here mean is taken as 6 because here chi-square distribution and mean should be equal.

Output)

```
[1] 13
```

```
VarY <- (3^2)*12
VarY
```

Output)

```
[1] 108
```

b)

It will not follow. variance is 108 so it doesn't follow because variance always should be twice of mean.

6) **(35 points)** The distribution of the expression values of the patients with the Zyxin gene are distributed according to $N(\mu = 1.6, \sigma = 0.4)$.
   a. What is the probability that a randomly chosen patient have the Zyxin gene expression values between 1 and 1.6?
   b. Use a Monte Carlo simulation of sample size n=500,000 to estimate the probability in part (a). Give your R code, and show the value of your estimate.
   c. What is the probability that exactly 2 out of 5 patients have the Zyxin gene expression values between 1 and 1.6? Please show your work on how to arrive at the answer. Give your answer to at least four decimal places.

Answer)

a)

```
pnorm(1.6, mean=1.6, sd=0.4) - pnorm(1, mean=1.6, sd=0.4)
```

output)

```
[1] 0.4331928
```

b)

```
MCM<-rnorm(500000, mean = 1.6, sd = 0.4)
mean(MCM > 1 & MCM < 1.6)
```

output)

```
[1] 0.433222
```

c)

```
dbinom(2, size=5, p=0.4331928)
```

output)

```
[1] 0.3417185
```

Note: Make sure that you clearly understand the difference between discrete and continuous random variables and use the appropriate functions to find the expected value, variance etc. It is conceptually wrong to use “integrate” for discrete random variables and “sum” for continuous random variables and very few points will be given if you get this wrong.