1) (5 points)

Suppose a random variable X has pdf as $f(x) = 2e^{-2(x-1)}, x > 1$. Which of the following represents P(0 < X < 4)? (Note: you do not need to solve for exact number).

1. $\int_0^4 2e^{-2(x-1)}dx;$
2. $\boldsymbol{\int_1^4 2e^{-2(x-1)}dx;}$
3. $\int_0^4 x2e^{-2(x-1)}dx;$
4. $\sum_{x=0}^4 2e^{-2(x-1)};$
5. $\int_0^\infty x2e^{-2(x-1)}dx;$

2) (10 points)

A random variable X has pdf

$$\boldsymbol{f(x) = \frac{2^x}{x!} e^{-2}}, \text{ x= 0,1,2…}$$

Find P(X = 1).

Rscript:

```
((2)/factorial(1))*exp(-2)
```

Answer:

0.2707

Then find $P(-2<X<4)$.

Rscript:

```
sum(((2^(0:3))/factorial(0:3))*exp(-2))
```

or

```
x_range<-c(0:3)
f.x<-function(x) ((2^x)/factorial(x))*exp(-2)
f_x<-function(x) f.x(x)*(x %in% x_range)
sum(f_x(x_range))
```

Note: I don't know which way you prefer I don't want my marks to be cut because I use a longer code instead of one line code. So please let me know in future which one should I use and Please don't cut marks for writing both the scripts here as I don't know what exactly you want here.

Answer:

0.8571

Give your answers to at least four decimal places.

3) (5 points)

If two carriers of the gene for albinism marry and have children, then each of their children has a probability of 1/4 of being albino. Let the random variable Y denote the number of their albino children out of all 3 of their children. Then Y follows a binomial(n, p) distribution. Find the values for n and p.
n =___3__ p = _.25_____

Answer:
n=3
p=1/4=.25

4) (10 points)

For Y following a binomial (n = 3, p = 0.25) distribution, compute the following:
$P(Y \leq 2) =$
$E(Y) =$
$Var(Y) =$
Give your answers to at least four decimal places.

Rscript:
```
sum(dbinom((0:2),size=3,p=0.25))
```

Answer:
$P(Y \leq 2) = 0.9844$

Rscript:
```
y<-c(0:3)
sum(y*dbinom(y,size=3,p=0.25))
```

Answer:
$E(Y) = 0.7500$

Rscript:
```
y<-c(0:3)
EY<-sum(y*dbinom(y,size=3,p=0.25))
sum((y-EY)^2*dbinom(y,size=3,p=0.25))
```

Answer:
$Var(Y) = 0.5625$

5) (20 points)

For X following a Chi-square distribution with degree of freedom m = 3, compute the following:
P(1 < X < 4) =
E(X) =
Var(X) =
Give your answers to at least four decimal places.
Also, use a Monte Carlo simulation with sample size n=100,000 to estimate P(1 < X < 4). What is your Monte Carlo estimate? Does it agrees with the answer above?
Rscript:

```
x<-function c(0:Inf)
integrate(function(x) dchisq(x,df=3), lower=1,upper=4)
```

Answer:
P(1 < X < 4) = .5398

Rscript:

```
x<-function c(0:Inf)
integrate(function(x) x*dchisq(x,df=3), lower=0,upper=Inf)
```

Answer:
E(X) = 3 (mean is same as degree of freedom for chi-square distribution)

Rscript:

```
x<-function c(0:Inf)
EX<-integrate(function(x) x*dchisq(x,df=3), lower=0,upper=Inf)$value
integrate(function(x) ((x - EX)^2)*dchisq(x,df=3), lower=0,upper=Inf)
```

Answer:
Var(x) = 6 (Variance is twice the degree of freedom for chi-square distribution)

Rscript:

```
X<-rchisq(n=100000, df=3)
mean((1<x) & (x<4))
```

Answer:
.5378 Yes the monte carlo estimate agrees with above.

6) (10 points)

Suppose X follows a Chi-square distribution with degree of freedom m = 5 so that E(X) = 5 and Var(X) = 10. Also, let Y = 4X - 10. Find E(Y) and Var (Y). Does Y follow a Chi-square distribution with degree of freedom m=10?

E(Y) =

Var(Y) =

Does Y follow a Chi-square distribution with degree of freedom m = 10?

Answer:
Y = 4X-10
Using the formula E(aX+b)= aE(X) + b
E(Y) = 4*E(X) – 10
E(Y) = 20-10 = 10

Using the formula Var (aX+b) = a^2Var(X)
Var(Y) = 4^2Var(X)
Var(Y) = 16*10= 160

For Y to follow chi-square distribution with degree of freedom m=10 then it should follow 2 conditions namely;

1. E(Y) = 10 as mean for chi square distribution is same as degree of freedom
2. Var(Y) = 20 as variance for chi square distribution is twice the degree of freedom.

E(Y) is 10 but Var(Y) is 160 so Y doesn't follow a chi square distribution.

7) (20 points)

The Zyxin gene expression values 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?

Rscript:

```
x<-c(-Inf,Inf)
integrate(function(x) dnorm(x,mean=1.6,sd=0.4), lower=1,upper=1.6)$value
```

Answer:

0.4332

(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.

Rscript:

```
x<-rnorm(n=500000, mean=1.6, sd=.4)
mean((1<x) & (x<1.6))
```

Answer:

0.4334

(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.

Rscript:

```
dbinom(2,size =5, prob=.4332)
```

Answer:

Now this normal distribution become binomial distribution with set =5 and probability = 0.4332.
We have to find P(X=2) which we can calculate by dbinom(2,size=5, prob=.4332)
Which gives us the answer - .3417
.3417 is the probability of exactly 2 patients out of 5 having the zyxin gen expression values between 1 and 1.6.

8) (20 points)

(a) Hand in a R script that calculates the mean and variance of two random variables X~F(m=2,n=5) and Y~F(m=10,n=5) from their density functions.

Rscript:

Mean of X:

```
x<-c(0,Inf)
integrate(function(x) x*df(x,df1=2,df2=5), lower=0,upper=Inf)$value
```

Answer:

1.6667

Variance of X:

```
x<-c(0,Inf)
EX<-integrate(function(x) x*df(x,df1=2,df2=5), lower=0,upper=Inf)$value
integrate(function(x) (x-EX)^2*df(x,df1=2,df2=5), lower=0,upper=Inf)
```

Answer:

13.8889

Mean of Y:

```
y<-c(0,Inf)
integrate(function(y) y*df(y,df1=10,df2=5), lower=0,upper=Inf)$value
```

Answer:

1.6667

Variance of Y:

```
y<-c(0,Inf)
EY<-integrate(function(y) y*df(y,df1=10,df2=5), lower=0,upper=Inf)$value
integrate(function(y) (y-EY)^2*df(y,df1=10,df2=5), lower=0,upper=Inf)
```

Answer:

7.2222

(b) Use the formula in Table 3.4.1 to calculate the means and variances directly.

Mean of X:

$\frac{n}{n-2}$ = 5/3 = 1.6667

Variance of X:

$\frac{\boldsymbol{2n^2(m+n-2)}}{\boldsymbol{m(n-2)^2(n-4)}}$ = 50*5/18 = 13.8889

Mean of Y:

$\frac{\boldsymbol{n}}{\boldsymbol{n-2}}$ = 5/3 = 1.6667

Variance of Y:

$\frac{\boldsymbol{2n^2(m+n-2)}}{\boldsymbol{m(n-2)^2(n-4)}}$ = 50*13/90 = 7.2222

(c) Run your script in (a), and check that your answers agree with those from part (b).

Yes, my scripts agrees with the answers from part (b)