Solutions to module 3 Homework

1) (5 points)

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

(a) $\int_0^4 2e^{-2(x-1)}dx$;

(b) $\int_1^4 2e^{-2(x-1)}dx$;

(c) $\int_0^4 x2e^{-2(x-1)}dx$;

(d) $\sum_{x=0}^{4} 2e^{-2(x-1)}$;

(e) $\int_1^\infty x2e^{-2(x-1)}dx$.

Solution: As X is a continuous variable defined with x>1 , we compute P(1<X<4) taking integral values within the range. Thus option b) is the answer

(b) $\int_1^4 2e^{-2(x-1)}dx$;

```
> integrate (function(x) 2*exp(-2*(x-1)), lower=1, upper=4)
0.9975212 with absolute error < 1.1e-14
```

2) (10 points)

A random variable X has pdf

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

Find P(X = 1).

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

Give your answers to at least four decimal places.

Solutions to module 3 Homework

Solution:

P(X = 1) = $f(1) = 2^1 / 1! \ * \ e^{-2} = 2e^{-2}$

```
> 2*exp(-2)
[1] 0.2706706
```

Thus, P(X = 1) = 0.2706706

P(-2 < X < 4) = $f(0) + f(1) + f(2) + f(3) = \sum_{x=0}^{3} 2^x / x! \ * \ e^{-2}$

```
> f_x<- function(x) (2^x/factorial(x))*exp(-2)
> sum(f_x(0:3))
[1] 0.8571235
```

Thus, P(-2 < X < 4) = 0.8571235

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 =_____ p = ______

Solution:

If n = number of trials i.e. birth of a child

If p = the probability of success i.e. children having albinism

We get the answer as n = 3 and p = ¼ = 0.25

Solutions to module 3 Homework

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.

Solution:

$P(Y \leq 2) = 0.984375$

```
>  set.x<- c(0,1,2)
> sum(dbinom(set.x, size=3, p=0.25))
[1] 0.984375
```

E(Y) = np for binomial distribution

$E(Y) = (3)(0.25) = 0.75$

```
> 3*0.25
[1] 0.75
```

Var(Y) = np (1 – p) for binomial distribution

$Var(Y) = (3)(0.25)(1 - 0.25) = 0.5625$

```
> 3*0.25*(1-0.25)
[1] 0.5625
```

Solutions to module 3 Homework

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?

Solution:

$P(1 < X < 4)$ is a continuous variable . it can be computed in two ways .

```
> # Method 1: cumulative Distribution Function
> pchisq(4,3)-pchisq(1,3)
[1] 0.5397878
> # Method 2: Integrating the pdf
> integrate( function(x) dchisq(x,3), lower=1, upper=4)
0.5397878 with absolute error < 3.1e-14
```

$P(1 < X < 4) = 0.5397878$

For Chi-square distribution

$E(X) = m = 3$

$Var(X) = 2m = 6$

b) Using monte carlo simulation with size n = 100000 we get

estimate value as 0.53933

```
> # b) use a Monte Carlo simulation with sample size n=100,000 to estimate
> # P(1 < X < 4).
> X<-rchisq( n=100000, df=3)
> mean((1<X)&(X<4))
[1] 0.53933
```

The estimate value we get from monte carlo is approximately same as the value obtained with Chi-square

Solutions to module 3 Homework

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?

Solution:

a) Appy the properties of mean and variance of linear transformation of random variables

$E(Y) = aE(X) + b$

$Var(Y) = a^2Var(X)$

We calculate :

$E(Y) = aE(X) + b = (4)(5) – 10 = 20 – 10 = 10$

$Var(Y) = a^2Var(X) = 4^2(10) = 16(10) = 160$

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

If y followed a Chi-square distribution then,

$E(Y) = m = 10$

$Var(Y) = 2m = (2)(10) = 20$

The value of Var(Y) for linear transformation is = 160

Thus, Y **does not** follow Chi-Square distribution with degree of freedom m = 10

Solutions to module 3 Homework

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?

(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.

Solution:

a) Probability for Zyxin gene expression for values between 1 and 1.6 i.e.

P(1 ≤ X ≤ 1.6) can be computed in two methods

```
> integrate (function(x)dnorm(x, mean=1.6, sd=0.4), lower=1, upper=1.6)$value
[1] 0.4331928
> pnorm(1.6, mean=1.6,sd=0.4)- pnorm(1, mean=1.6, sd=0.4)
[1] 0.4331928
```

Therefore, P(1 ≤ X ≤ 1.6) = 0.4331928

b) Using monte carlo simulation of sample size n=500,000 to estimate the probability

```
> X<-rchisq( n=100000, df=3)
> mean((1<X)&(X<4))
[1] 0.54136
>
```

The value of my estimate is 0.54136

Solutions to module 3 Homework

c) To find the probability that exactly 2 out of 5 patients have the Zyxin gene expression values between 1 and 1.6,

We can define each patient as a trial and gene expression between 1 and 1.6 as success , thus it can be computed as binomial distribution with

n = 5

p = 0.4331928

for P(X = 2) =

```
> dbinom(2, size=5, prob=0.4331928)
[1] 0.3417185
```

Therefore, P(X = 2) = 0.3417185

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.

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

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

Solution:

Solutions to module 3 Homework

```
> # Problem 8
>
> # (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.
>
> EX<- integrate(function(x) x*df(x, df1=2, df2=5), lower=0, upper=Inf)$value
> EX
[1] 1.666667
> EY<- integrate(function(y) y*df(y, df1=10, df2=5), lower=0, upper=Inf)$value
> EY
[1] 1.666667
> VarX<- integrate(function(x) (x-EX)^2*df(x, df1=2, df2=5), lower=0, upper=Inf)$value
> VarX
[1] 13.88889
> VarY<- integrate(function(y) (y-EY)^2*df(y, df1=10, df2=5), lower=0, upper=Inf)$value
> VarY
[1] 7.222222
>
> # (b) Use the formula in Table 3.4.1 to calculate the means and variances
> # directly.
>
> # mean= n/(n-2)
> mean<-5/(5-2)
> mean
[1] 1.666667
> #variance= (2*n^2*(m+n-2))/(m*(n-2)^2*(n-4))
> variance<- function(m,n) (2*n^2*(m+n-2))/(m*(n-2)^2*(n-4))
> variance(2,5)
[1] 13.88889
> variance(10,5)
[1] 7.222222
```

Solutions to module 3 Homework

All the outputs are as follows and the answers from part a) agree
With those of part b)

```
> EX
[1] 1.666667
> EY
[1] 1.666667
> VarX
[1] 13.88889
> VarY
[1] 7.222222
> mean
[1] 1.666667
> variance(2,5)
[1] 13.88889
> variance(10,5)
[1] 7.222222



>
```