**Module 3 Home Work**

**Problem 1:**

Suppose a random variable X has a pdf of f(x)=2e^{-2(x-1)}$, $x>1$. Which of the

following represents P(0<X<4)?

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

**Answer:**

Here x is a Continuous variable defined as x>1 so we must compute P(0<X<4) by taking the integral values within the allowable range as defined by pdf. So choice b) should be used.

p <- function (x) 2\*exp(-2\*(x-1))

p.x<-function (x) p(x)\*(1<x & x<4)

integrate(p,lower=1,upper=4)

Output: 0.9975212 with absolute error < 1.1e-14

b) int\_{1}^{4} 2e^{-2(x-1)}dx

**Problem 2:**

A random variable X has pdf :

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

a)Find P(X=1)

p<- function(x) (2^x/factorial(x))\*exp(-2)

p(1)

[1] 0.2706706

b) Then find P(-2<X<4)

p<- function(x) (2^x/factorial(x))\*exp(-2)

sum(p(0:3))

[1] 0.8571235

**Problem 3:**

If two carriers of the gene for albinism marry and have children, then each

of their children has a probability of frac{1}{4} of being albino. Let the random

variable Y denote the number of their children having the gene for albinism out of all

3 of their children. Then Y follows a binomial(n, p) distribution.

Find the values for n and p.

Answer:

Let n= the number of trials, where a trial is defined as the birth of a child.

And p= the probability of success, where success is defined as a child having albinism

Therefore n = 3 and p = frac {1} {4} = 0.25

**Problem 4:**

For Y following a binomial (n = 3, p = 0.25) distribution,

calculate the following:

a)P(Y <= 2)

(a)

dbinom(0,3,0.25)+dbinom(1,3,0.25)+dbinom(2,3,0.25)

[1] 0.984375

b) E(Y) = np

Therefore, E(Y) = (3)(0.25) = 0.75

range<-c(0:3)

EY<-sum(range\*dbinom(range,size=3,p=0.25))

EY

[1] 0.75

c)Var(Y) = np(1-p)

Therefore, Var(Y)= (3)(0.25)-(1-0.25) = 0.5625

range<-c(0:3)

EY<-sum(range\*dbinom(range,size=3,p=0.25))

VARY<-sum((range-EY)^2\*dbinom(range,3,0.25))

VARY [1] 0.5625

**Problem 5:**

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

Answers:

a) P(1 < X < 4) = P(X < 4) - P(X < 1)

pchisq(4,3)-pchisq(1,3)

Output: 0.5397878

b)&c)

for the Chi-Square Distribution following properties hold true

b) E(X) = m = 3

c) Var(X) = 2m = 6

d)

x<-rchisq(100000,3)

mean((1<x)&(x<4))

[1] 0.53945

The output does agree with the answer above as the Monte carlo output is not exactly

same as it is just the estimation but it is similar to the output of P(1 < X < 4)

**Problem 6:**

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?

a) E(Y) =

b) Var(Y) =

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

Answers:

a)E(Y) = aE(X)+b = (4)(5)-10 = 20-10 = 10

EX<-5

VARX<-10

EY<-4\*EX-10

EY [1] 10

b) Var(Y) = a^{2}Var(X) = (4)(4)(10) = 160

VX<-10

VY<-4^2\*VX

VY [1] 160

If Y were to follow a Chi-square distribution with degrees of freedom m = 10, then

the following would be true.

E(Y) = m = 10

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

But however, in our transformation the Var(Y) = 160. So, Y {does not} follow a Chisquare

distribution with degrees of freedom m = 10.

**Problem 7:**

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.

Answers:

a)The probability that Zyxin gene expression falls within the range P(1<=X<=1.6) for the continuous random variable X distributed according to the normal distribution

N(mu = 1.6, sigma = 0.4) P(1<=X<=1.6)= P(X <= 1.6) - P(X <= 1)

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

[1] 0.4331928

b) x<-rnorm(500000,1.6,0.4)

mean((1<x)&(x<1.6))

[1] 0.43321 As this is the monte carlo simulation the output is similar.

c) As we try to calculate the probability, that exactly 2 out of 5 patients have the Zyxin gene expression values between 1 and 1.6. We define each patient as a trial, and success is defined as gene expression between the values 1 and 1.6. Then, it follows that this can be computed by examining the binomial distribution, binomial (n=5, p=0.4331928).Now, calculating the probability density function when the discrete random variable X for this distribution is equal to 2, P(X=2).

dbinom(2,size=5,prob=0.4331928)

[1] 0.3417185

**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.

(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).

Answers:

a)

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

print(EX)

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

print(EY)

VarX <- integrate(function(x) (x - EX)^2\*df(x, df1=2, df2=5), lower=0,

upper=Inf)$value

print(VarX)

VarY <- integrate(function(y) (y - EY)^2\*df(y, df1=10, df2=5), lower=0,

upper=Inf)$value print(VarY)

Output: E(X)= 1.666667

E(Y)= 1.666667

Var(X)= 13.88889

Var(Y)= 7.222222

b) Here for X~F(m=2,n=5) and Y~F(m=10,n=5)

mean and variances:

(I) Mean for F(df(x,m,n))= n/n − 2

E(X)=5/(5-2)=5/3 = 1.666666667

E(Y)=5/(5-2)=5/3 = 1.666666667

(II) Variance for F(df(x,m,n)) = (2 \*n^2 \*( m + n − 2)) / (m \*( n − 2)^ 2 \* ( n − 4))

VarX= 2\*25\*5 / 2\*9\*1 = 250/18 = 13.888888889

VarY= 2\*25\*13 / 10\*9\*1 = 650/90 = 7.222222222

# mean = n/(n-2)

print(5/(5-2))

# variance = (2\*n^2\*(m+n-2))/(m\*(n-2)^2\*(n-4))

Fvar <- function(m,n) (2\*n^2\*(m+n-2))/(m\*(n-2)^2\*(n-4))

print(Fvar(2,5))

print(Fvar(10,5))

Outputs according to the formulas:

E(X) = 1.666667

E(Y) = 1.666667

(VarX) = 13.88889

(VarY) = 7.222222

c) By using the formula in Table 3.4.1 and even the R-Script in (a) gives the same values for the means and variances of X, Y. Checking the answers from (a) and (b) it can be said that (b) agrees with the answers from (a).