Module 3 Homework

Chengbo Gu

Problem 1. (5 points)

Suppose a random variable X has pdf as $f(x) = 2e^{-2(x-1)}, x > 1$. Which of the following represents $P(0 < X < 4)$?

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

Problem 2. (10 points)

A random variable X has pdf

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

Find $P(X = 1)$

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

```
f2 <- function(x) {2^x*exp(-2)/gamma(x+1)}
print(f2(1))
[1] 0.2706706
print(f2(0)+f2(1)+f2(2)+f2(3))
[1] 0.8571235
```

$P(X = 1) = 0.2706706$

$P(-2 < X < 4) = 0.8571235$

Problem 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=1/4
```

Problem 4. (10 points)

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

$P(Y \le 2), E(Y)$ and $Var(Y)$

```
yrange = c(0:2)
```

```
print(sum(dbinom(yrange, size=3, p=0.25)))
[1] 0.984375

Yrange = c(0:3)
EY <- sum(Yrange*dbinom(Yrange, size=3, p=0.25))
print(EY)
[1] 0.75

VarY <- sum( (Yrange-EY)^2 * dbinom(Yrange, size=3, p=0.25))
print(VarY)
[1] 0.5625
```

$P(Y \le 2) = 0.984375$
$E(Y) = 0.7500$
$Var(Y) = 0.5625$

Problem 5. (20 points)
For X following a Chi-square distribution with degree of freedom m=3, compute the following: $P(1 < X < 4), E(X)$ and $Var(X)$.

```
print(integrate(function(x) dchisq(x, df = 3), lower=1, upper=4)$value)
[1] 0.5397878
EX <- integrate(function(x) x*dchisq(x, df = 3), lower=-Inf, upper=Inf)$value
print(EX)
[1] 3
VarX <- integrate(function(x) (x-EX)^2*dchisq(x, df = 3), lower=-Inf, upper=Inf)$value
print(VarX)
[1] 6
```

Also, use a Monte Carlo simulation with sample size n=100,000 to estimate $P(1 < X < 4)$.

```
x <- rchisq(n=100000, df=3)
print( mean( (1<x) & (x<4) ) )
[1] 0.5396
It agrees the answer above.
```

Problem 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?

$E(Y) = 4E(X) - 10 = 10$
$Var(Y) = 4^2 \times Var(X) = 160$
No, Y doesn't follow a Chi-square distribution with degree of freedom m=10.

Problem 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?

```
p <- integrate(function(x) dnorm(x, 1.6, 0.4), lower=1, upper=1.6)$value
print(p)
[1] 0.4331928
```

(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.

```
x <- rnorm(n=500000, 1.6, 0.4)
print(mean((1<x)&(x<1.6)))
[1] 0.433596
```

(c) What is the probability that exactly 2 out of 5 patients have the Zyxin gene expression values between 1 and 1.6?

```
print(dbinom(2, 5, p))
[1] 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).

```
rm(list=ls())
m <- 2
n <- 5
print("mean and variance of X~F(m=2, n=5)")
EX<-integrate(function(x) x*df(x, m, n), lower=0, upper=Inf)$value
print(EX)
[1] 1.666667

VarX<-integrate(function(x) (x-EX)^2*df(x, m, n), lower=0, upper=Inf)$value
print(VarX)
[1] 13.88889

print(n/(n-2))
[1] 1.666667

print( 2*n^2*(m+n-2)/(m*(n-2)^2*(n-4)) )
[1] 13.88889

rm(list=ls())
m <- 10
n <- 5
print("mean and variance of X~F(m=10, n=5)")
EX<-integrate(function(x) x*df(x, m, n), lower=0, upper=Inf)$value
print(EX)
[1] 1.666667

VarX<-integrate(function(x) (x-EX)^2*df(x, m, n), lower=0, upper=Inf)$value
print(VarX)
[1] 7.222222

print(n/(n-2))
[1] 1.666667
```

```
print( 2*n^2*(m+n-2)/(m*(n-2)^2*(n-4)) )
[1] 7.222222
```

Answers from part(a) and part(b) match each other.