STT 810

Homework 2

Aditya Jain

2022-09-28

Contents

Question 1

A fair die is rolled 30 times.

a. What is the probability that exactly half of the rolls are even numbers?

```
xa <- dbinom(x= 15, size = 30, prob = 0.5)
paste(round(xa,3), "is the Probability")
```

```
## [1] "0.144 is the Probability"
```

b. What is the probability that more than 20 of the rolls are even numbers?

```
xb <- 1 - pbinom(q = 20, size = 30,prob = 0.5 )
paste(round(xb,3), "is the Probability")
```

```
## [1] "0.021 is the Probability"
```

c. What is the probability that less than 5 of the rolls are greater than 4?

```
xc <- pbinom(q = 4, size = 30, prob = 1/3)
paste(round(xc,3), "is the Probability")
```

```
## [1] "0.012 is the Probability"
```

Question 2

A web server gets typically pinged according to a Poisson process with rate 30/second.

a. Find the probability that the server gets pinged between 20 and 40 times in a particular second.

```
xa <- ppois(q = 39,lambda = 30) - ppois(q = 20,lambda = 30)
paste(round(xa,3), "is the Probability")
```

```
## [1] "0.918 is the Probability"
```

b. Calculate the number of seconds in a year

```
Secondsinyear <- 60*60*24*365.25
paste(Secondsinyear, "is no of seconds in a year")
```

```
## [1] "31557600 is no of seconds in a year"
```

c. Use (b) to estimate the maximum number of pings in a single second over the course of a year.

```
xc <- max(rpois(n = Secondsinyear,lambda = 30))
paste(xc, "is the maxinum number of pings over an year")
```

```
## [1] "62 is the maxinum number of pings over an year"
```

d. Often a web server creates alerts when the ping rate is alarmingly high (typically, the sign of a Denial of Service attack by a hacker). What would be a good rate to create such an alarm (and why)?

```
# Calculating
xd <- qpois(p = 0.9999999999,lambda = 30)
paste(xd, "is the Alarm rate for dos attack")
```

```
## [1] "71 is the Alarm rate for dos attack"
```

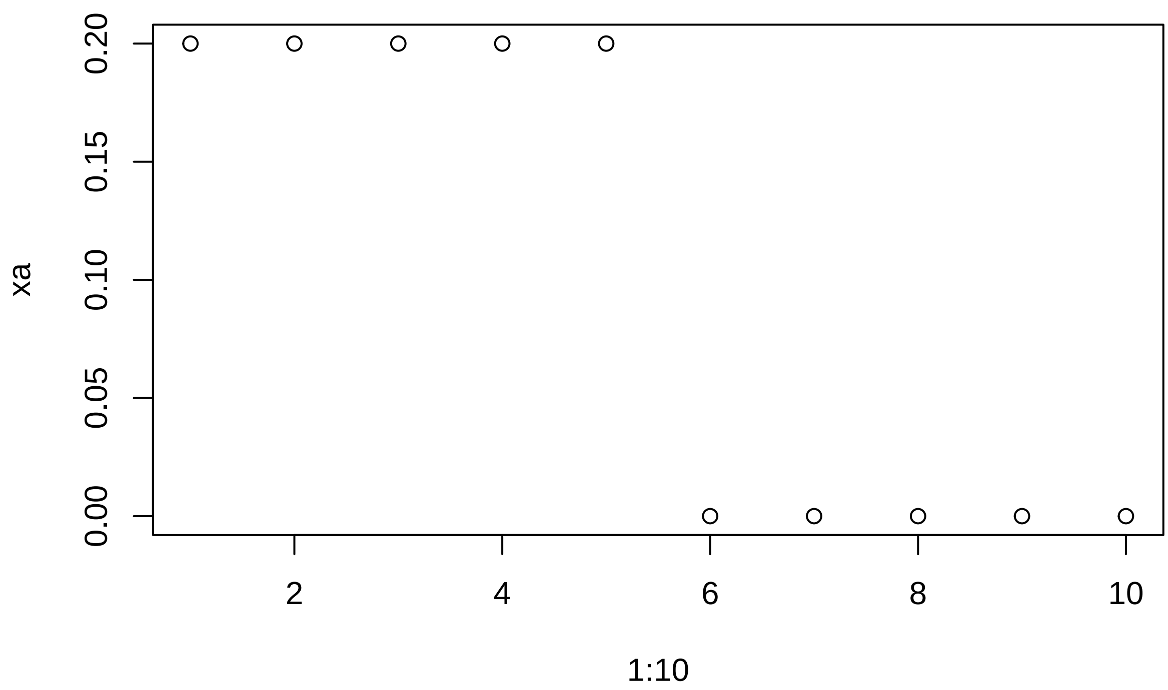
Ans: A good rate must be higher than 99 percent of the ping rate in a poisson distribution. That would surely alarm us for something unusual on the website.

Question 3

The number of minutes that a bus is late is modeled by the Uniform density on the interval (0, 5).

a. Draw a picture of the density function.

```
xa <- dunif(x = 1:10,0,5)
plot(1:10,xa)
```



b. What is the probability that the bus is more than 1 minutes late?

```
xb <- 1 - punif(q = 1, min = 0, max = 5)
paste(round(xb,3), "is the Probability")
```

```
## [1] "0.8 is the Probability"
```

c. What is the conditional probability that the bus is more than 4 minutes late, given that it is already 3 minutes late?

```
xc <- punif(q = 3,min = 0,max = 5)/punif(q = 4,min = 0,max = 5)
paste(round(xc,3), "is the Probability")
```

```
## [1] "0.75 is the Probability"
```

Question 4

Using the normal density with mean = 50 cm and sigma = 5 cm as a model for the length of catfish in a lake, answer the following questions. Draw an appropriate picture of a normal density for each question.

a. If a catfish is selected at random, what is the probability that it is more than 60 cm in length?

```
xa = 1 - pnorm(q = 60, mean = 50, sd = 5)
paste(round(xa,3), "is the Probability")
```

```
## [1] "0.023 is the Probability"
```

b. What is the length x such that exactly 10% of catfish are shorter than x?

```
xb = qnorm(p = 0.1, mean = 50, sd = 5 )
paste(round(xb,3),"is the length")
```

```
## [1] "43.592 is the length"
```

c. What is the length y such that exactly 70% of catfish are longer than y?

```
xc = qnorm(p = 0.3, mean = 50, sd = 5)
paste(round(xc,3),"is the length")
```

```
## [1] "47.378 is the length"
```

Question 5

Let X be a random variable with mean = 80 and sigma = 10.

a. Compute P(mean – sigma < X < mean + sigma). Note that this is P(70 < X < 90).

```
xa = pnorm(q = 90, mean = 80, sd = 10) - pnorm(q = 70, mean = 80, sd = 10)
paste(round(xa,3), "is the Probability")
```

```
## [1] "0.683 is the Probability"
```

b. Compute P(60 < X < 100). Note that this can be written as P(mean – 2sigma < X < mean + 2sigma).

```
xb = pnorm(q = 100, mean = 80, sd = 10) - pnorm(q = 60, mean = 80, sd = 10)
paste(round(xb,3), "is the Probability")
```

```
## [1] "0.954 is the Probability"
```

c. Compute P(50 < X < 110). Note that this can be written as P(mean – 3sigma < X < mean + 3sigma).

```
xc = pnorm(q = 110, mean = 80, sd = 10) - pnorm(q = 50, mean = 80, sd = 10)
paste(round(xc,3), "is the Probability")
```

```
## [1] "0.997 is the Probability"
```

d. The relationships among normal densities suggest that you should get the same answers to these three questions no matter what the values of mean and sigma. Verify that this is true by calculating the same three quantities, but this time for a standard normal distribution, i.e., a normal distribution with mean = 0 and sigma = 1

```
xsa = pnorm(q = 1, mean = 0, sd = 1) - pnorm(q = -1, mean = 0, sd = 1)
paste(round(xsa,3), "is the Probability")
```

```
## [1] "0.683 is the Probability"
```

```
xsb = pnorm(q = 2, mean = 0, sd = 1) - pnorm(q = -2, mean = 0, sd = 1)
paste(round(xsb,3), "is the Probability")
```

```
## [1] "0.954 is the Probability"
```

```
xsc = pnorm(q = 3, mean = 0, sd = 1) - pnorm(q = -3, mean = 0, sd = 1)
paste(round(xsc,3), "is the Probability")
```

```
## [1] "0.997 is the Probability"
```

Question 6

A machine produces nails whose lengths are normally distributed with mean = 2 inches and sigma = 0.05 inches.

a. What proportion of nails are less than 1.9 inches in length?

```
xa = pnorm(q = 1.9, mean = 2, sd = 0.05)
paste(round(xa,3), "is the Proportion")
```

```
## [1] "0.023 is the Proportion"
```

b. What proportion of nails are longer than 2.1 inches in length?

```
xb = 1 - pnorm(q = 2.1, mean = 2, sd = 0.05)
paste(round(xb,3), "is the Proportion")
```

```
## [1] "0.023 is the Proportion"
```

c. What is the length x for which exactly 20% of the nails are longer than x?

```
xc = qnorm(p = 0.8, mean = 2, sd = 0.05)
paste(round(xc,3), "is the length")
```

```
## [1] "2.042 is the length"
```

d. What is the length y for which exactly 20% of the nails are shorter than y?

```
xd = qnorm(p = 0.2, mean = 2, sd = 0.05)
paste(round(xd,3), "is the length")
```

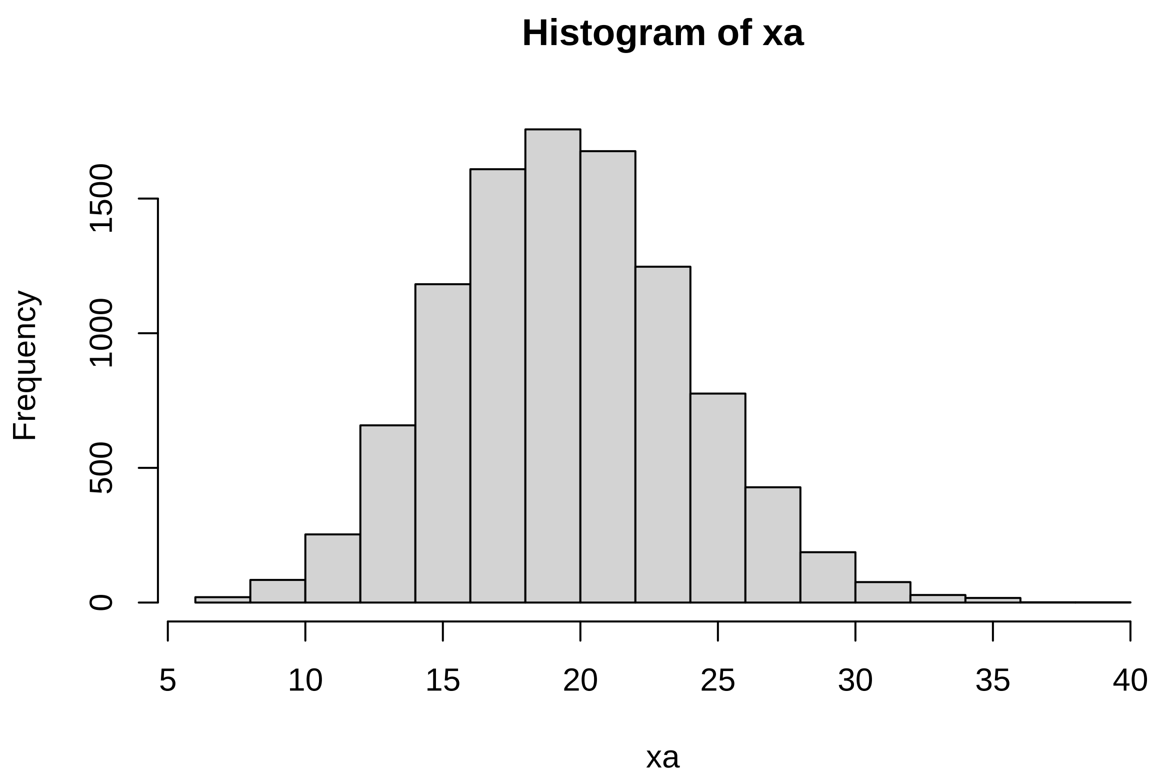
```
## [1] "1.958 is the length"
```

Question 7

You are the data scientist at a pharmaceutical startup. Over the next 4 years the startup is expected to discover new drugs at a random rate which can be fit as a Poisson process with a rate of 5/year. Each drug thus discovered has an 8% chance of obtaining FDA approval.

a. Do 10,000 simulations of the number of drugs that will be discovered (irrespective of FDA approval). Plot a histogram of the results.

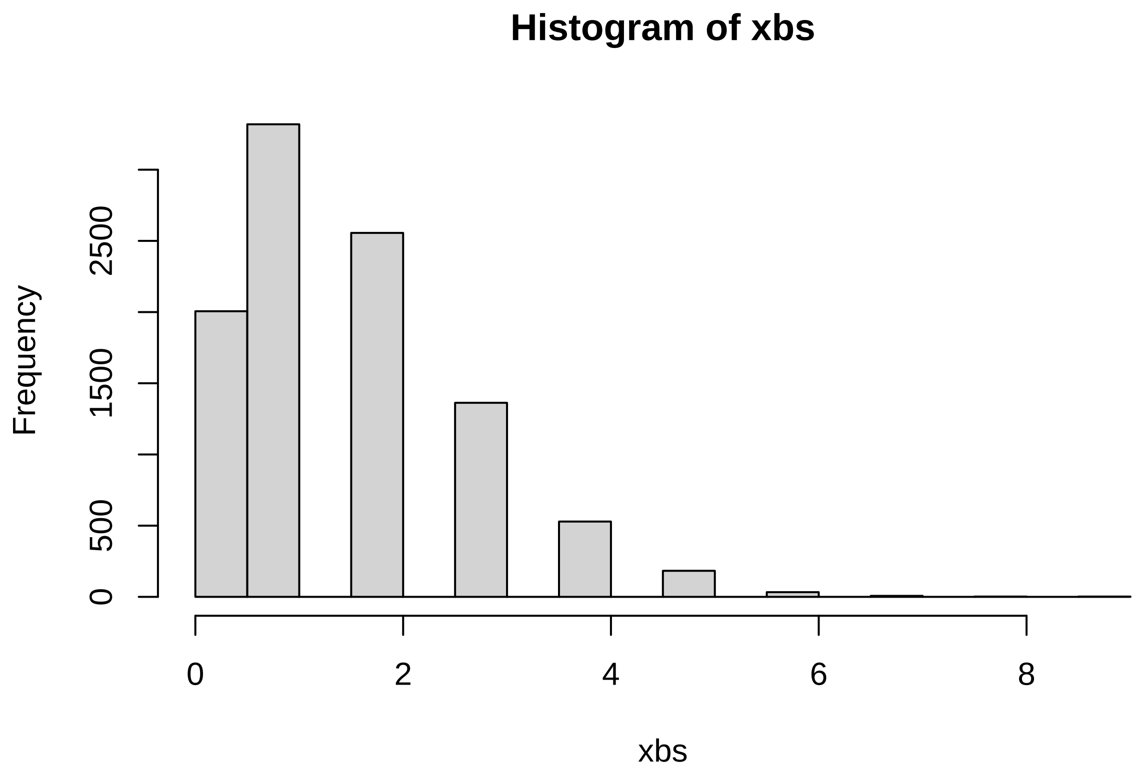
```
xa <- rpois(n = 10000, 5*4)
hist(xa)
```



b. Next, use those 10,000 simulations to determine the number of drugs which will survive FDA approval (1 simulation for each of the 10,000 simulations in (a)). Plot a histogram of those results.

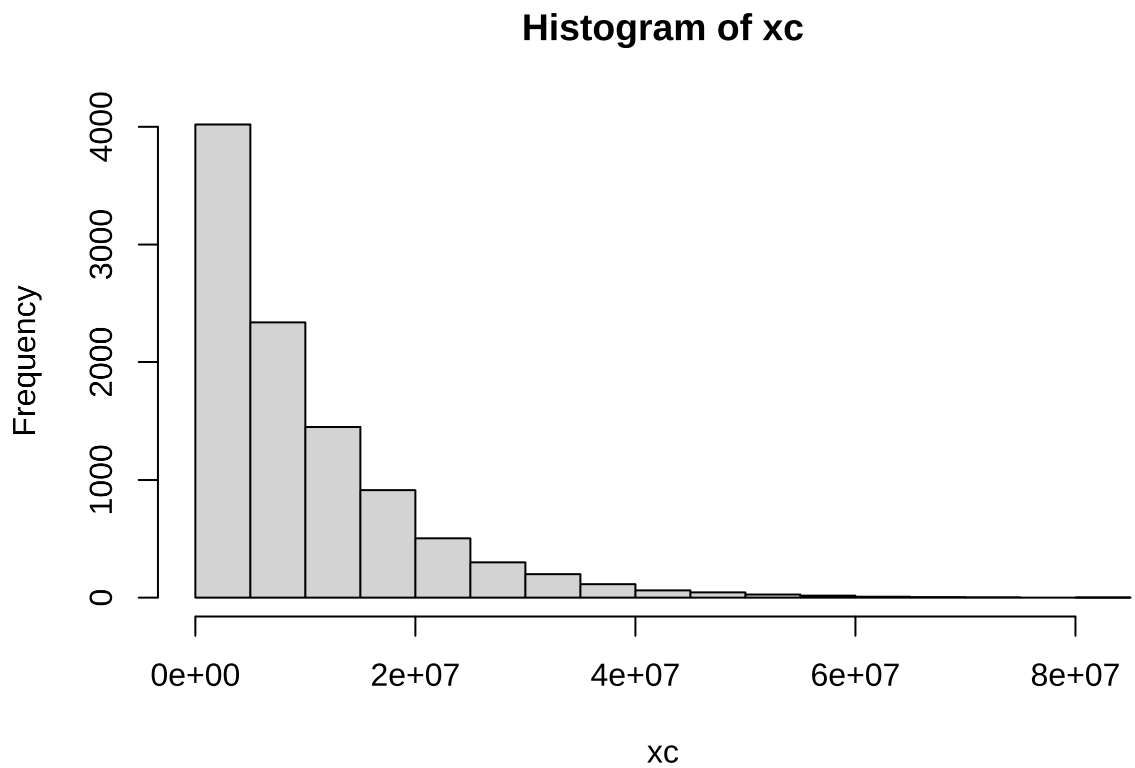
```
xb <- lapply(1:10000 ,function(x) sum(sample(x = 0:1, size = xa[x]
                                    ,replace = TRUE, prob = c(0.92,0.08))))

xbs <- c()
for (i in 1:10000){
  xbs[i] <- xb[[i]]}
hist(xbs)
```



c. Each drug obtaining FDA approval will give the startup revenue which fits an exponential distribution with mean $10,000,000. Use the simulations to model the amount of revenue, and then plot a histogram of the revenue results.

```
# Expected mean = 1/lamda for exponential distribution so
xc <- rexp(xbs,1/10000000)
hist(xc)
```



d. The R&D cost for the startup is $26,000,000. Find the percent of simulation results for which the startup makes a profit.

```
xd <- length(xc[xc > 26000000])/100
paste(round(xd,3), "is the Percentage")
```

```
## [1] "6.96 is the Percentage"
```

e. You are asked to determine business strategy to determine whether more money should be spent on R&D or marketing. If the money is spent on R&D, 20% more drugs will be discovered, so that drugs will be discovered at a rate of 6 per year instead of 5. If the money is spent on marketing, the revenue for each drug will be increased by 20%, so that each drug's revenue will fit an exponential distribution with parameter $12,000,000. Which alternative will yield higher revenue on average for the startup?

```
#case 1 - Money spent on R&D
xa1 <- rpois(n = 10000, 6*4)
xb1 <- lapply(1:10000 ,function(x)sum(sample(x = 0:1, size = xa1[x]
                                      ,replace = TRUE, prob = c(0.92,0.08))))

xbs1 <- c()
for (i in 1:10000){
  xbs1[i] <- xb1[[i]]}

xc1 <- rexp(xbs1,1/10000000)
```

```r
xd1 <- mean(xc1)

#case 2 - Money spent on marketing
xa2 <- rpois(n = 10000, 5*4)
xb2 <- lapply(1:10000 ,function(x) sum(sample(x = 0:1, size = xa2[x]
                                       ,replace = TRUE, prob = c(0.92,0.08))))

xbs2 <- c()
for (i in 1:10000){
  xbs2[i] <- xb2[[i]]}

xc2 <- rexp(xbs2,1/12000000)
xd2 <- mean(xc2)

round(xd1)
```

```
## [1] 9972579
```

```r
round(xd2)
```

```
## [1] 12113713
```

```r
paste("Spending money on marketing would yeild more revenue")
```

```
## [1] "Spending money on marketing would yeild more revenue"
```