QBS120_problem_set_1

Gibran Erlangga

9/15/2021

Import libraries

```
library(gtools)
```

Problems for Rice, Chapter 1

Question 1
Scientists discover the life form in the clouds of Venus that is responsible for unusual levels of phosphine. To their surprise, the genetic material is also encoded using DNA but with five possible bases (A,C,T,G,B) rather than 4 for life on Earth (A,C,T,G). Also, this life form uses 2 base codons rather than 3 base codons. Assume the values of the two independent bases in one codon are measured. Compute answers to the following using R: – How many unique codons are possible? How does this compare the number of unique 3 base codons for Earth DNA?

```
# define list of options
venus_bases <- c('A', 'C', 'T', 'G', 'B')
earth_bases <- c('A', 'C', 'T', 'G')

unique_venus_codons <- permutations(n=5,r=2,v=venus_bases,repeats.allowed=T)
unique_earth_codons <- permutations(n=4,r=3,v=earth_bases,repeats.allowed=T)

num_unique_venus_codons <- nrow(unique_venus_codons)
num_unique_earth_codons <- nrow(unique_earth_codons)

paste("number of unique venus codons is ", num_unique_venus_codons,
      "and the number of unique earth codons is ", num_unique_earth_codons)
```

```
## [1] "number of unique venus codons is  25 and the number of unique earth codons is  64"
```

– What are the sequences of these unique 2 base codons? \

```
unique_venus_codons
```

```
##       [,1] [,2]
##  [1,] "A"  "A"
##  [2,] "A"  "B"
##  [3,] "A"  "C"
```

```
##  [4,] "A"  "G"
##  [5,] "A"  "T"
##  [6,] "B"  "A"
##  [7,] "B"  "B"
##  [8,] "B"  "C"
##  [9,] "B"  "G"
## [10,] "B"  "T"
## [11,] "C"  "A"
## [12,] "C"  "B"
## [13,] "C"  "C"
## [14,] "C"  "G"
## [15,] "C"  "T"
## [16,] "G"  "A"
## [17,] "G"  "B"
## [18,] "G"  "C"
## [19,] "G"  "G"
## [20,] "G"  "T"
## [21,] "T"  "A"
## [22,] "T"  "B"
## [23,] "T"  "C"
## [24,] "T"  "G"
## [25,] "T"  "T"
```

– List the elements of the event E1 that both bases are the same.

```
E1 = c("AA", "CC", "TT", "GG", "BB")
```

– List the elements of the event E2 that both bases are different.

```
E2 = c("AB", "AC", "AT", "AG", "CA", "CT", "CG", "CB", "TA", "TC", "TG", "TB", "GA", "GC" ,"GT", "GB",
```

– List the elements of the event E3 that the first base is B.

```
E3 = c("BB", "BA", "BC", "BT", "BG")
```

– List the elements of the event E1 intersect E3.

```
intersect = c("BB")
```

– List the elements of the event E1 union E3.

```
union = c("AA", "CC", "TT", "GG", "BB", "BA", "BC", "BT", "BG")
```

Question 2
Five people want to play a card game and so they discard the two cards from a 52 card deck and are each dealt 10 cards. How many unique deals are possible? If there were four playing instead and each was dealt 13 cards, how many unique deals are possible? Before computing, which number do you think will be larger? Before computing, I think 52 choose 13 would be larger than 50 choose 10 because it has a larger base number of cards.

```r
five_people <- choose(50, 10) * choose(40, 10) * choose(30, 10) * choose(20, 10) * choose(10, 10)
four_people <- choose(52, 13) * choose(39, 13) * choose(26, 13) * choose(13, 13)

paste("# of combinations in 5-people game: ", five_people)
```

```
## [1] "# of combinations in 5-people game:  4.83347757579012e+31"
```

```r
paste("# of combinations in 4-people game: ", four_people)
```

```
## [1] "# of combinations in 4-people game:  5.36447377654888e+28"
```

Question 3
Rice, 1.18 (changed so that probability that at least one defect turns up is 0.8 instead of 0.9): A lot of n items contains k defectives, and m are selected randomly and inspected. How should the value of m be chosen so that the probability that at least one defective item turns up is 0.80? Apply your answer to (a) n = 1000, k = 10 and (b) n = 1000, k = 100. Notes: solve for m numerically using R; for case b), use n=1000, k=100 to ensure a numerical solution without more complex simplifications/approximations.

Answer:
with the n and k definition stated above, we get

P(no defect) = number of no defect / number of all possible outcomes
P(no defect) = (n - k)/ n
P(no defect) = 1 - k/n

with P(A) being the probability of at least one item identified as defective, and P(A^C) being the probability of no item identified as defective. With these definitions in mind, we can derive the equation into:

$$
\begin{gathered}
P(A) = 1 - P(A^C) \\
0.8 = 1 - P(A^C) \\
0.8 = 1 - (1 - k/n)^m \\
-0.2 = -(1 - k/n)^m \\
0.2 = (1 - k/n)^m \\
m.ln(1 - k/n) = ln0.2 \\
m = ln0.2/ln(1 - k/n)
\end{gathered}
$$

plug the numbers in r:

```r
get_m <- function(n,k) {
  return(log(0.2)/ log(1 - (k/n)))
}

# a
n_1 <- 1000
k_1 <- 10

paste("value of m when n = 1000 and k = 10 is ", get_m(n_1, k_1))
```

```
## [1] "value of m when n = 1000 and k = 10 is  160.137724340166"
```

```r
# b
n_2 <- 1000
k_2 <- 100

paste("value of m when n = 1000 and k = 100 is ", get_m(n_2, k_2))
```

```
## [1] "value of m when n = 1000 and k = 100 is  15.2755318478223"
```

Question 4
Rice 1.49 (changed for 2 coins and 3 tosses): Two fair coins are simultaneously tossed three times.
– a. What is the probability of two or more heads given that there was at least one head?
– b. What is the probability of two or more heads given that there was at least one tail?

probability of getting two or more heads:
P(A) = 1 - P(getting less than 2 heads)
P(A) = 1 - 7/64
P(A) = 57/64

probability of getting at least one head:
P(B) = 2^6 - 1
P(B) = 63/64

probability of getting at least one tail:
P(C) = 2^6 - 1
P(C) = 63/64

a. P(A|B) = P(A intersect B) / P(B)
   P(A|B) = (57/64) / (63/64)
   P(A|B) = 57/63
b. P(A|C) = P(A intersect C) / P(C)
   P(A|C) = (57/64) / (63/64)
   P(A|C) = 57/63

Question 5
Rice 1.75 (changed so that each individual has either 0, 1 or 2 progeny instead of 0 or 2): A population starts with one member; at a time t = 1, it either has 1 progeny with probability p, 2 progeny with probability 2p, or dies. Also, we know that $3p < 1$. If it successfully reproduces, then its children behave independently with the same alternatives at time t = 2. What is the probability that there are no members in the third generation? For what value people want to play a card game so they discard the two cards from a 52 card deck and 1 of p is this probability equal to 0.5? Note: can solve for p numerically; hint: use law of total probability.

Answer:
option: 1 progeny with probability p; 2 progeny with probability 2p; dies

$$P(n=0) = P(dies|progency)P(progency) + P(dies|2progency)P(2progency) + P(dies|dies)P(dies)$$

$$P(n=0) = (1-3p)(p) + (1-3p)(1-3p) + (1-3p)$$

$$0.5 = 1 - 2p - p^3 - 24p^3 + 36p^4$$

$$p = 0.3$$

Problems for Rice, Chapter 2

Question 1
What is a random variable? Define both discrete and continuous RVs in terms of the sample space.

```
paste("Random variable is a function that maps sample space to real numbers. Discrete RVs are the type o
```

```
## [1] "Random variable is a function that maps sample space to real numbers. Discrete RVs are the type
```

Question 2
One of the observed values in an experiment is 13.4, which is modeled as a RV during analysis. How can a fixed number be modeled as a RV? Where is the random component?

```
paste("Fixed number is an example of the discrete random variable, as it said that 13.4 is one of the ol
```

```
## [1] "Fixed number is an example of the discrete random variable, as it said that 13.4 is one of the o
```

Question 3
Rice 2.9 (solve without using R then confirm through visualization): For what values of p is a 2 out of 3 majority decoder better than transmission of the message once? (here p is probability of success on one transmission)

```
print("I tried my best - I have no idea.")
```

```
## [1] "I tried my best - I have no idea."
```

Question 4
Rice 2.40 (changed so that f(x) = cx(x+2) for 0 x 2 and 0 otherwise; also asking to plot the CDF): Suppose X has the density function f(x) = cx(x + 2) for 0 x 2 and f(x) = 0 otherwise. a) Find c
b) Find the CDF
c) What is P(0.1 x 0.5)?
d) Plot the CDF

Answer: a) Find c

$$\int_0^2 f(cx^2 + 2cx)\,dx = \tfrac{1}{3}cx^3 + \tfrac{1}{2}2cx^2\Big|_0^2$$

$$c = \tfrac{3}{20}$$

b) Find the CDF

$$\int_0^2 f(\tfrac{3}{20}x^2 + \tfrac{6}{20}x)\,dx = \tfrac{1}{20}x^3 + \tfrac{3}{20}2x^2\Big|_0^2$$

$$= \tfrac{8}{20} + \tfrac{12}{20}$$

$$= 1$$

c)

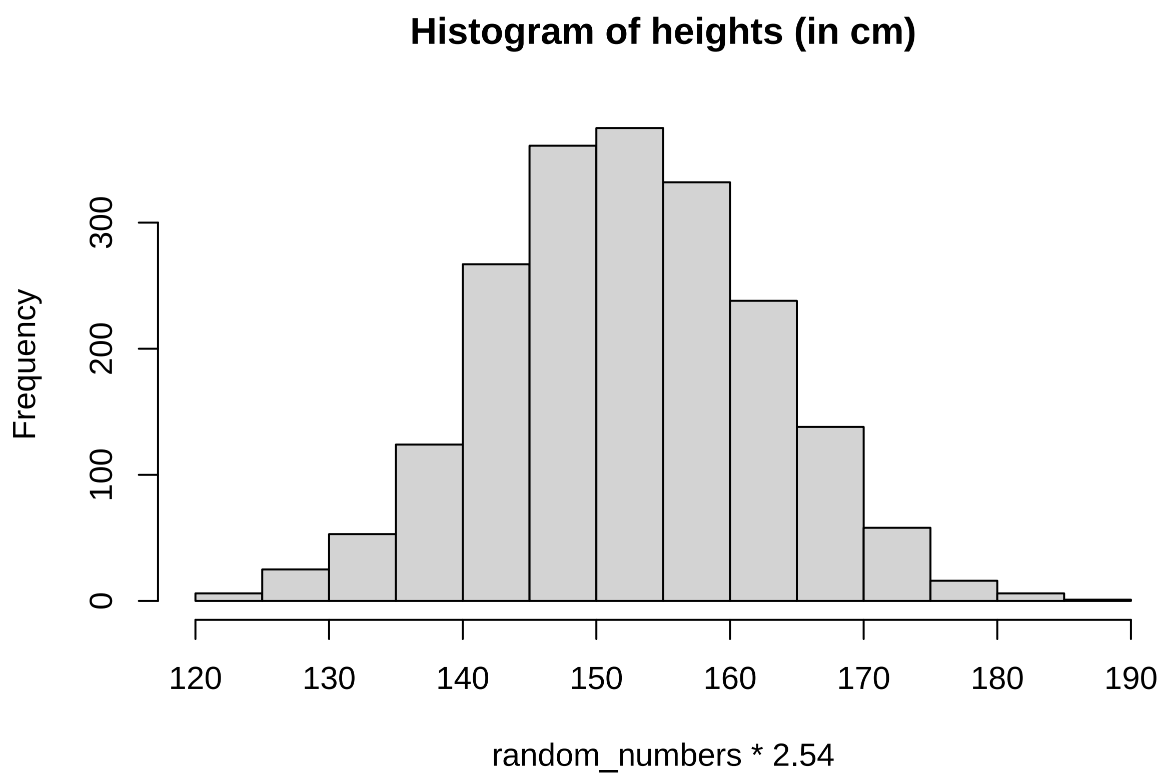
d)

Question 5
Rice 2.52(numbers adjusted; for part a) you may only use R functions for the standard normal, i.e., N(0,1)): Suppose that in a certain population, individual's heights are approximately normally distributed with parameters = 60 inches and = 4 inches. a) What proportion of the population is over 6 ft tall? b) What is the distribution of heights if they are expressed in centimeters? – c) In meters?

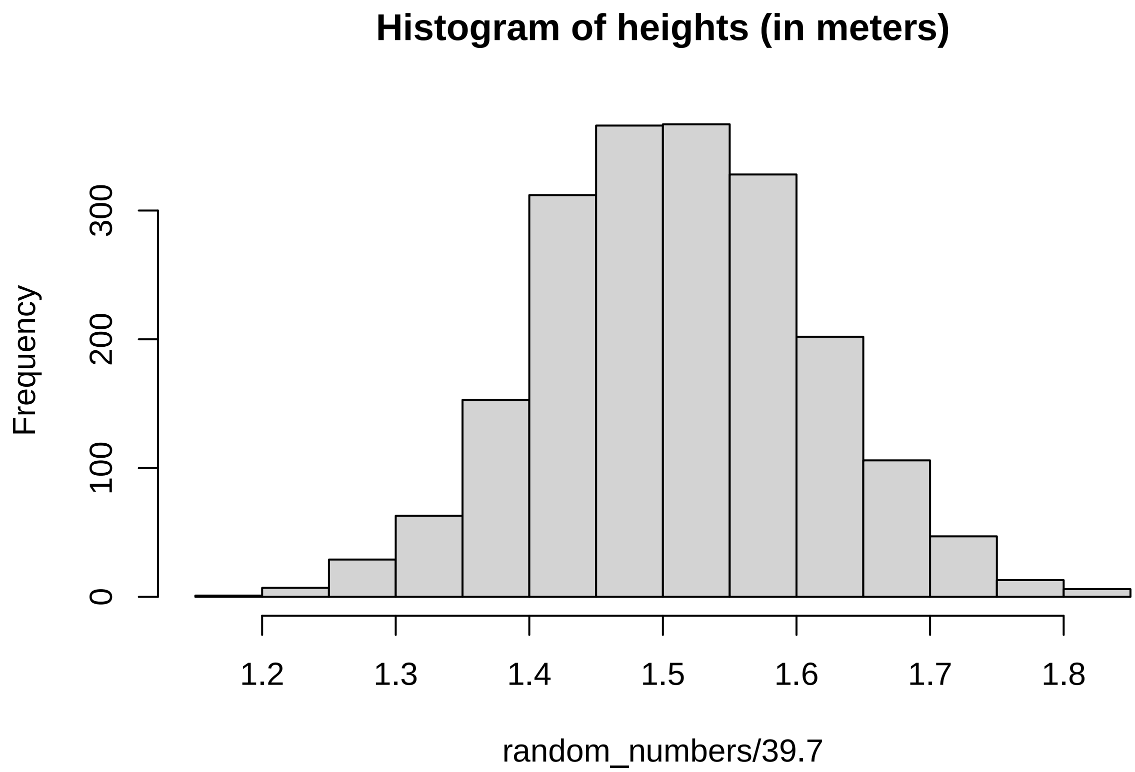
```
mean = 60
sd = 4

# a
# 6 ft = 72 inches, so:
print(pnorm(72, mean, sd, lower.tail=FALSE))
```

```
## [1] 0.001349898
```

```
# b
random_numbers <- rnorm(2000, mean, sd)
# multiply the numbers with 2.54 for conversion to cm
hist(random_numbers * 2.54, main="Histogram of heights (in cm)")
```




```
# c
# divide the numbers with 39.7 for conversion to meters
hist(random_numbers / 39.7, main="Histogram of heights (in meters)")
```


Histogram of heights (in meters)
Frequency
0
100
200
300
1.2
1.3
1.4
1.5
1.6
1.7
1.8
random_numbers/39.7