Problem Set 1

September 22, 2020

QBS 120 - Problem Set 1

Author: Vala Zeinali

0.1 Problems for Rice, Chapter 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?
    - – What are the sequences of these unique 2 base codons?
    - – List the elements of the event E1 that both bases are the same.
    - – List the elements of the event E2 that both bases are different.
    - – List the elements of the event E3 that the first base is B.
    - – List the elements of the event E1 E3.
    - – List the elements of the event E1 E3.

1.1) How many unique codons are possible? How does this compare the number of unique 3 base codons for Earth DNA?

```
[1]: #Number of unique codons for Earth DNA can be computed using basic principle of␣
     ↪counting
     unqCod = 4 * 4 * 4
     print(unqCod)
```

```
[1] 64
```

```
[2]: # SO the number of unique codons for the five base DNA is show below with the␣
     ↪same logic
     newUnqCod = 5 * 5
     print(newUnqCod)
```

```
[1] 25
```

1.1) Answer: There are 25 possible codon combinations. There are less possible codon combinations than on Earth.

1.2) What are the sequences of these unique 2 base codons?

```
[3]: install.packages('gtools')
     #load library
     library(gtools)
     #DNA with 5 bases
     x <- c('A', 'C', 'T', 'G', 'B')
     #pick 2 bases from the 5 possible bases with replacement
     #get all permutations
     permutations(n=5,r=2,v=x,repeats.allowed=T)
     nrow(permutations(n=5,r=2,v=x,repeats.allowed=T))
```

Installing package into '/Users/valazeinali/Library/R/3.6/library'
(as 'lib' is unspecified)

The downloaded binary packages are in
/var/folders/bx/dp3pc2gx453dz73c35_y9_8h0000gn/T//RtmphjZHu0/downloaded_packages

Warning message:
"package 'gtools' was built under R version 3.6.2"

```
                                       A  A
                                       A  B
                                       A  C
                                       A  G
                                       A  T
                                       B  A
                                       B  B
                                       B  C
                                       B  G
                                       B  T
                                       C  A
                                       C  B
    A matrix: 25 × 2 of type chr       C  C
                                       C  G
                                       C  T
                                       G  A
                                       G  B
                                       G  C
                                       G  G
                                       G  T
                                       T  A
                                       T  B
                                       T  C
                                       T  G
                                       T  T
```

25

1.2) Answer: There are 25 possible codon combinations and they are shown above.

1.3) List the elements of the event E1 that both bases are the same. E1 = {"AA","CC","BB","GG","TT"}

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

```
[5]: x <- c('A', 'C', 'T', 'G', 'B')
     #pick 2 bases from the 5 possible bases with replacement
     #get all permutations
     permutations(n=5,r=2,v=x,repeats.allowed=F)
     E2 <- nrow(permutations(n=5,r=2,v=x,repeats.allowed=F))
     E2
```

3

A matrix: 20 × 2 of type chr

A	B
A	C
A	G
A	T
B	A
B	C
B	G
B	T
C	A
C	B
C	G
C	T
G	A
G	B
G	C
G	T
T	A
T	B
T	C
T	G

20

1.5) List the elements of the event E3 that the first base is B. E3 = {"BB","BA","BC","BT","BG"}

1.6) List the elements of the event E1 ∪ E3. E1 ∪ E3 = {"BB","BA","BC","BT","BG"}

1.7) List the elements of the event E1 ∩ E3. Its just E1 since E3 repeats in E1. Listing below.

```
[77]: x <- c('A', 'C', 'T', 'G', 'B')
      #pick 2 bases from the 5 possible bases with replacement
      #get all permutations
      permutations(n=5,r=2,v=x,repeats.allowed=T)
      nrow(permutations(n=5,r=2,v=x,repeats.allowed=T))
```

```
                                    A   A
                                    A   B
                                    A   C
                                    A   G
                                    A   T
                                    B   A
                                    B   B
                                    B   C
                                    B   G
                                    B   T
                                    C   A
                                    C   B
A matrix: 25 × 2 of type chr        C   C
                                    C   G
                                    C   T
                                    G   A
                                    G   B
                                    G   C
                                    G   G
                                    G   T
                                    T   A
                                    T   B
                                    T   C
                                    T   G
                                    T   T
```

25

1 Problem 2

Five people want to play a card game 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?

[34]:
```
#(50 choose 10) * (40 choose 10) * (30 choose 10) * (20 choose 10) * (10 choose
 ↪10) <- FIRST scenario
#(52 choose 13) * (39 choose 13) * (26 choose 13) * (13 choose 13) ??

exp1 <- (choose(52,50)*choose(50, 10) * choose(40,10) * choose(30, 10) *
 ↪choose(20,10) *choose(10,10))
exp2 <- (choose(52, 13) * choose(39,13) * choose(26, 13) * choose(13,13))
print(exp1)
print(exp2)
```

```
[1] 6.409191e+34
[1] 5.364474e+28
```

2.1a) Answer: 6.409191e+34

5

2.1b) Answer: 5.364474e+28

2 Problem 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; hint: remember that $P(A) = 1 - P(A^C)$; for case b), use n=1000, k=100 to ensure a numerical solution without more complex simplifications/approximations.

```
[43]: # p(1 or more defects) >= .80
      # p(no defects) <= .2
      # (n-k/n)^m <= .2
      #((1000-10)/1000)^m <= .2
      m <- ((log(.2))/log(990/1000))
      m
```

160.137724340166

```
[44]: # p(1 or more defects) >= .80
      # p(no defects) <= .2
      # (n-k/n)^m <= .2
      #((1000-100)/1000)^m <= .2
      m <- ((log(.2))/log(900/1000))
      m
```

15.2755318478223

3.1a) Answer: 160.137724340166

3.1b) Answer: 15.2755318478223

3 Problem 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?

Let A be the event of getting two or more heads for one coin with 3 tosses

```
A = {HHT, THH, HTH, HHH}
P(A) = 4/8
```

Let B be the event of getting at least one head for one coin with 3 tosses

```
B = {THH, HTT, TTH, HHT, THT, HTH, HHH}
P(B) = 7/8
P(A   B) = {HHT, THH, HTH, HHH} = 4/8
```

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{4/8}{7/8} = \frac{4}{7} * \frac{4}{7} = \frac{16}{49}$$

Since we have two experiments and these two experiments are independent we can multiply 4/7 * 4/7 = 16/49 to get the probability for 2 coins.

– b. What is the probability of two or more heads given that there was at least one tail?

Let A be the event of getting two or more heads for one coin with 3 tosses

```
A = {HHT, THH, HTH, HHH}
P(A) = 4/8
```

Let B be the event of getting at least one tail for one coin with 3 tosses

```
B = {HTT, HHT, HTH, THT, TTH, THH, TTT}
P(B) = 7/8
P(A  B) = {HHT, THH, HTH} = 3/8
```

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3/8}{7/8} = \frac{3}{7} * \frac{3}{7} = \frac{9}{49}$$

Since we have two experiments and these two experiments are independent we can multiply 3/7 * 3/7 = 9/49 to get the probability for 2 coins.

4   Problem 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 1 of p is this probability equal to 0.5? Note: can solve for p numerically; hint: use law of total probability.

$$P(1 progeny) = p = \frac{1}{3}$$

$$P(2 progeny) = 2p = \frac{1}{3} * 2 = \frac{2}{3}$$

$$P(dies) = 1 - P(A\ B) = 1 - [\frac{1}{3} + \frac{2}{3} - (\frac{1}{3} * \frac{2}{3})] = \frac{2}{9}$$

$$P(NO.Third.Gen) = [(\frac{1}{3} * \frac{1}{3} * \frac{2}{9}) + (\frac{1}{3} * \frac{2}{3} * \frac{2}{9}) + ((\frac{1}{3} * \frac{2}{9} * \frac{2}{9}) + (\frac{2}{3} * \frac{1}{3} * \frac{2}{9}) + (\frac{2}{3} * \frac{2}{3} *$$

$$\frac{2}{9}) + (\frac{2}{3} * \frac{2}{9} * \frac{2}{9}) + (\frac{2}{9} * \frac{1}{3} * \frac{2}{9}) + (\frac{2}{9} * \frac{2}{3} * \frac{2}{9}) + (\frac{2}{9} * \frac{2}{9} * \frac{2}{9})] = .332$$

5 Problem 6

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

A "random variable" is a set of possible values from a random experiment.

A discrete random variable is a variable that can take on a finite or countably infinite set of possible values x with $P(X = x) > 0$ for each of these values.

A random variable X is called continuous if it satisfies $P(X = x) = 0$ for each x.

6 Problem 7

- 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?

It is a mapping of $P(X = 13.4)$. It is the mapping of some event on Omega.

7 Problem 8

- 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)

$$\sum_{k=0}^{1} \binom{3}{k} p^k (1-p)^3 + \binom{3}{1} p^1 (1-p)^2$$

$$= (1 - 3p + 3p^2 - p^3) + 3p(1 - 2p + p^2)$$

$$= (2p^3 - 3p^2 + 1)$$

8 Problem 9

- 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.

9.a)

$$\int_0^2 cx(x + 2)dx = 1$$

evaluate at 2 and 0

$$= \frac{cx^2(x + 3)}{3} = 1$$

$$= \frac{20c)}{3} = 1$$

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

9.b)

$$\int_0^2 \frac{3}{20} x(x+2) dx$$

evaluate at 2 and 0

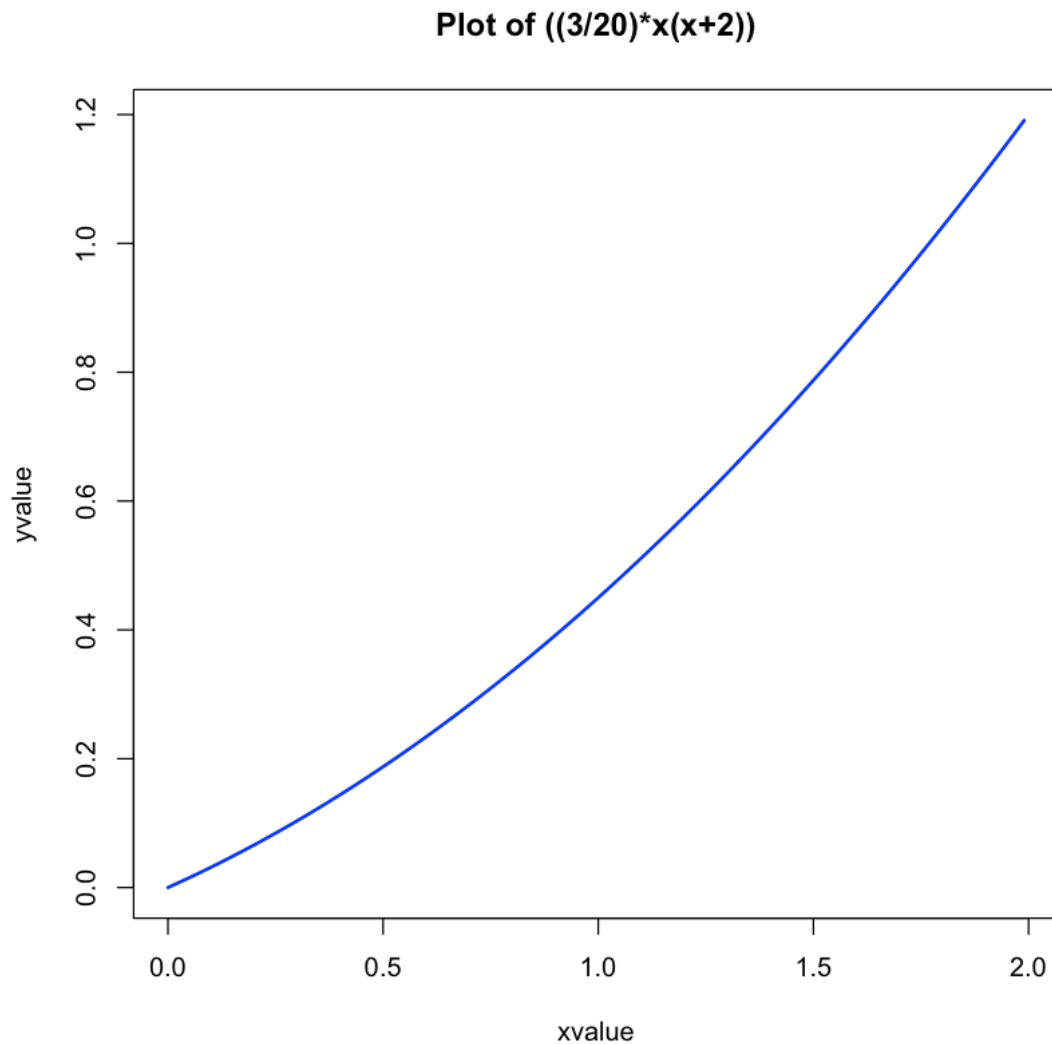$$= \frac{3}{20}[(\int_0^2 x^2 dx + \int_0^2 2x dx)] = \frac{3}{20}(\frac{8}{3} + 4) = 1$$

9.C)

$$\int_0^2 \frac{3}{20} x(x+2) dx$$

evaluate at .1 and .5

$$= \frac{3}{20}[(\int_{.1}^{.5} x^2 dx + \int_{.1}^{.5} 2x dx)] = \frac{3}{20}(0.04133 + 0.24) = 0.0421995$$

```
[57]: curve(((3/20)*x*(x+2)), from=0, to=1.99, n=300, xlab="xvalue", ylab="yvalue",
             col="blue", lwd=2, main="Plot of ((3/20)*x(x+2))")
```

Plot of ((3/20)*x(x+2))

[ ]: `## inverse cdf can produce the probability for every distribution`

9 Problem 10

Rice2.52(numbersadjusted;forparta)youmayonlyuseRfunctionsforthestandardnormal, 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?

```
[62]:  #a) What proportion of the population is over 6 ft tall?
       sixfoot <- (1 - pnorm(72, mean = 60, sd = 4))
       sixfoot
```
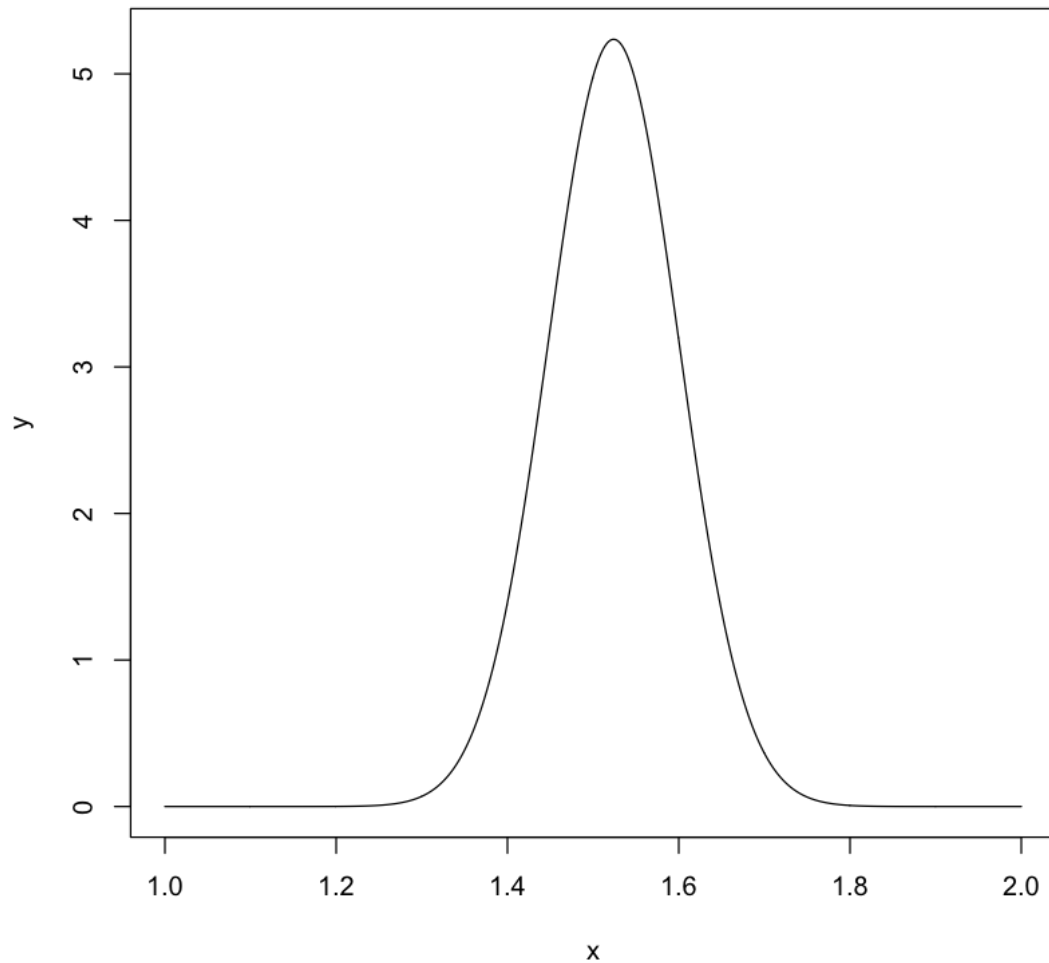
0.0013498980316301

```
[73]:  # b) What is the distribution of heights if they are expressed in centimeters?
       x <- seq(100, 200, length=1000)
       y <- dnorm(x, mean= 60 * 2.54, sd= 3 * 2.54)
       plot(x, y, type="l", lwd=1)
```
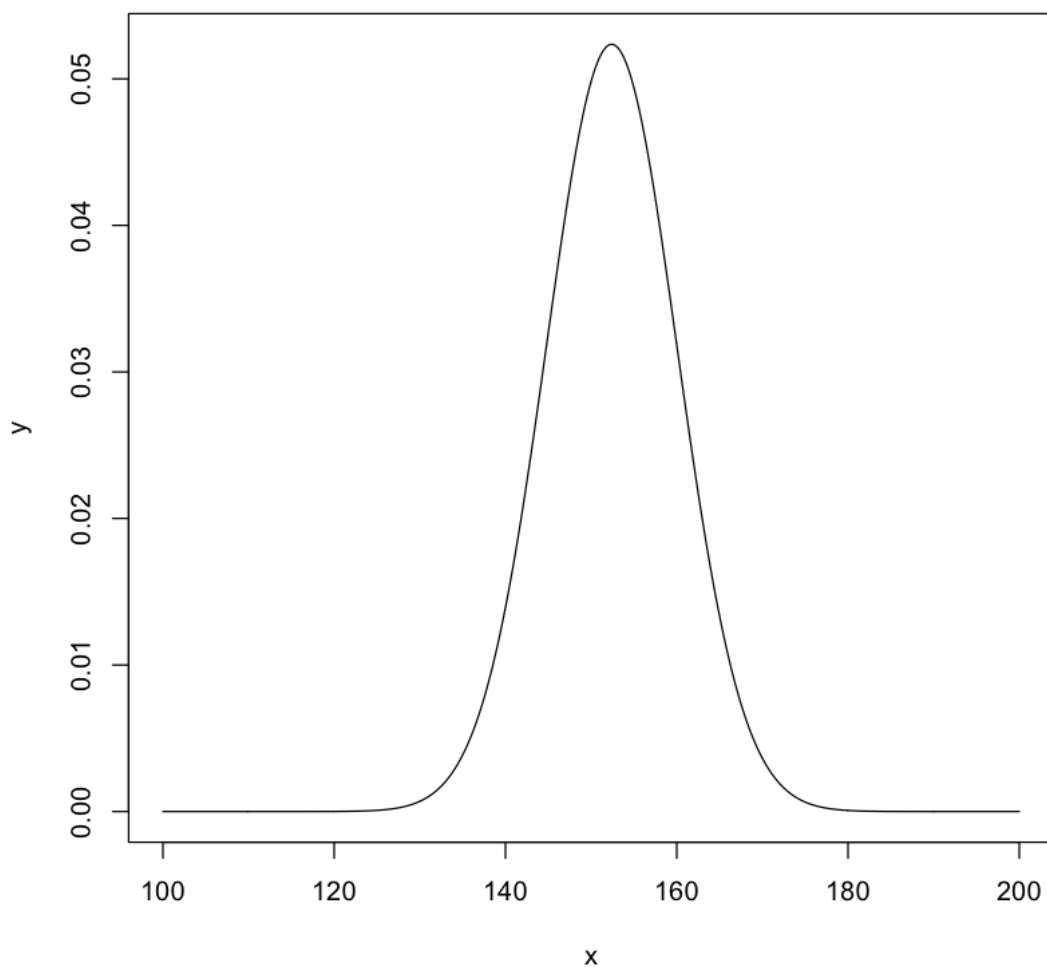


```
[76]:  # c) What is the distribution of heights if they are expressed in meters?
       x <- seq(1, 2, length=1000)
```

```
y <- dnorm(x, mean= 60 / 39.37, sd= 3 / 39.37)
plot(x, y, type="l", lwd=1)
```


10   Problem 11 (could not finish this problem)

[ ]: