Problem Set 3

Monday, September 11, 2017
9:20 AM

Hickman, Keith

Question 1 Trosset Chapter 4.5, Exercise 1

a. PMF of X:

$$f(x) = \begin{cases} 1/5 \; x \in \{1,6\} \\ 4/5 \; x \in \{3,4\} \\ 0 \; otherwise \end{cases}$$

* The brackets for the functions didn't save correctly in transferring this file type, but they're intended to capture all of the intervals.

b. CDF of X:

$$f(y) = \begin{cases} 0 & y < 1 \\ 1/10 & 1 \le y < 3 \\ 1/2 & 3 \le y < 4 \\ 9/10 & 4 \le y < 6 \\ 1 & y \ge 6 \end{cases}$$

c. Expected Value of X:

i. We can write out the problem as
1, 3, 3, 3, 3, 4, 4, 4, 4, 6
Noting the values weighted by the probabilities:
$1 * \frac{1}{10} + 3 * \frac{4}{10} + 4 * \frac{4}{10} + 6 * \frac{1}{10} = .1 + 1.2 + 1.6 + .6 = \mathbf{3.5, the\ expected\ value\ of\ X}$

d. Variance of X:

i. Var(1, 3, 3, 3, 3, 4, 4, 4, 4, 6)

ii. $Var(X) = \sqrt{\sum_x (x - \mu)^2 \cdot f(x)}$

1) Here, the mean mu is the EX = 3.5. I would take the square root of the sum of the distance of each ∈ of X from the mean, squared times f(x).

iii. In R:

iv.
```
> x <- c(1, 3, 3, 3, 3, 4, 4, 4, 4, 6)
> var(x)
[1] 1.611111
```

e. Standard Deviation of X:

i. The standard deviation of X = $\sqrt{Var(X)} = \sqrt{1.61111} = 1.269$

ii. In R:
```
> sd(x)
[1] 1.269296
```

Question 2 Trosset Chapter 4.5, Exercise 3

X = {1, 1, 1, 1, 2, 5, 5, 10, 10, 10}

a. PMF of X:

$$f(x) = \begin{cases} \frac{4}{10} & x = 1 \\ \frac{1}{10} & x = 2 \\ \frac{2}{10} & x = 5 \\ \frac{3}{10} & x = 10 \\ 0 \; otherwise \end{cases}$$

b. CDF of X:

$$f(y) = \begin{cases} 0 & y < 1 \\ 4/10 & 1 \le y < 2 \\ 1/2 & 2 \le y < 5 \\ 7/10 & 5 \le y < 10 \\ 1 & y \ge 10 \end{cases}$$

c. Expected Value of X:

i. Noting the values weighted by the probabilities:
$1 * \frac{4}{10} + 2 * \frac{1}{10} + 5 * \frac{1}{5} + 10 * \frac{3}{10} = .4 + .2 + 1 + 3 = \mathbf{4.6, the\ expected\ value\ of\ X}$
Or the sum of the values / n = 46/10 = 4.6

d. Variance of X:

i. Var(1, 1, 1, 1, 2, 5, 5, 10, 10, 10)

ii. $Var(X) = \sqrt{\sum_x (x - \mu)^2 \cdot f(x)}$

1) Here, the mean mu is the EX = 3.5. I would take the square root of the sum of the distance of each ∈ of X from the mean, squared times f(x).

iii. In R:

iv.
```
> x <- c(1, 1, 1, 1, 2, 5, 5, 10, 10, 10)
> var(x)
[1] 16.26667
```

e. Standard Deviation of X:
    i. The standard deviation of X = $\sqrt{Var(X)} = \sqrt{16.26667} = 1.269$
    ii. In R:
```
> sd(x)
[1] 4.033196
```

Question 3. Trosset Chapter 4.5 Exercise 10.

a. Arlen has 7 seats and invites 12 people. He knows people will accept with a 50% probability, and actually show up with 80% probability. The probability that a person both accepts an invitation and shows up in a binomial trial is : $.5 * .8 = .4\ or\ 40\%$
    i. $P(Y > 7) = 1 - P(Y \leq 7) = 1 - f(7)$
    ii. In R:
```
> 1 - pbinom(7, 12, .4)
[1] 0.1938477
```
Or 19% probability that Arlen will have more than 7 dinner guests. .

Question 4. Trosset Chapter 4.5. Exercise 14.

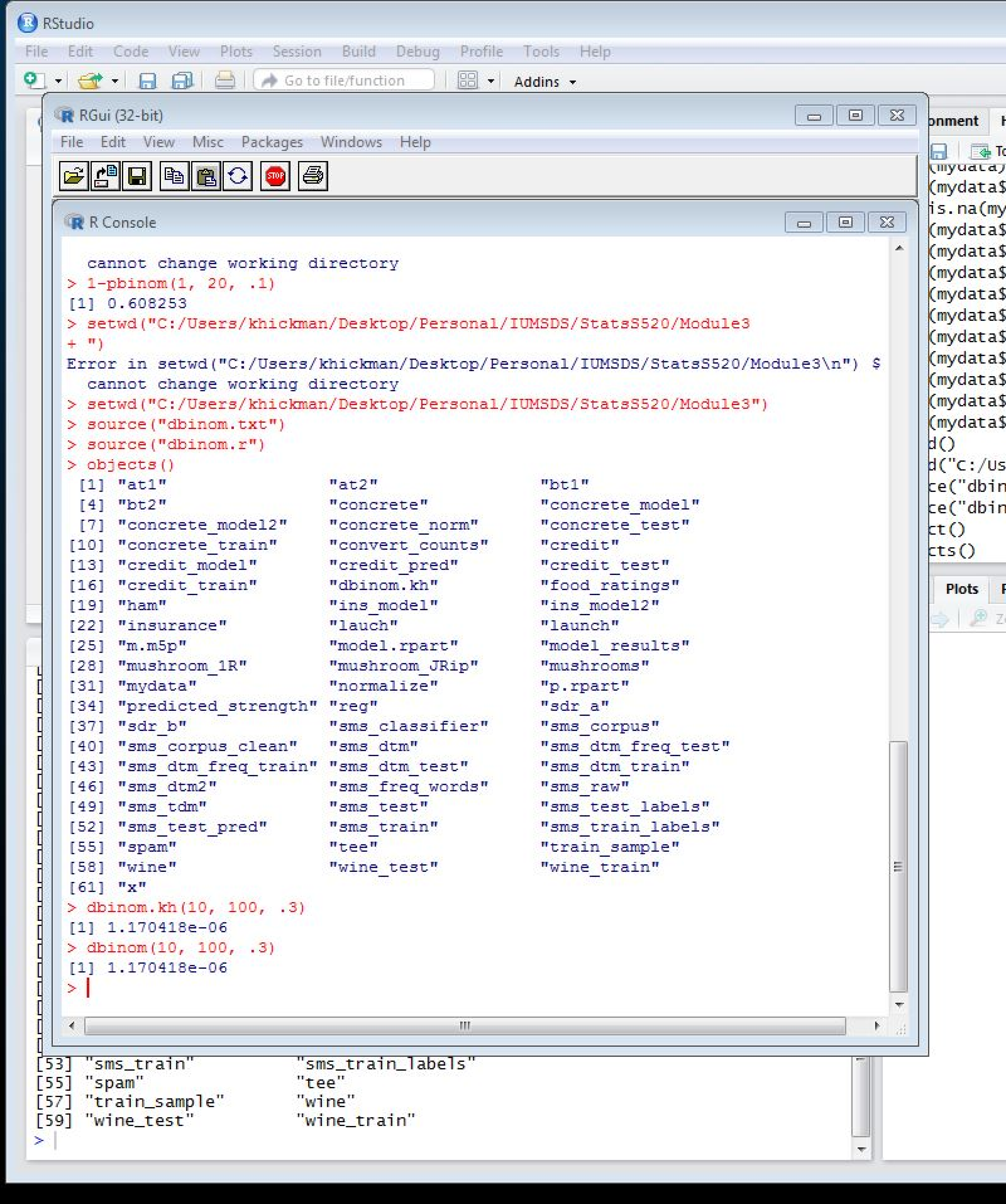
a. How many symbols should we expect the receiver to identify correctly?
    i. The Expected Value for this problem is calculated assuming that the trials are independent and due to chance alone (not ESP). There are 5 symbols, and with each trial we have a 1 out of 5 chance of getting the symbol correct. Therefore, **EX = 0.2**

b. The ARE considers a score of more than 7 matches to be indicative of ESP. What is the probability that the receiver will provide such an indication?
    i. Here, we have a binomial random variable Y ~ Binomial(n; p) where n = 25 and p = .2 and need to find the pmf.
    ii. $P(Y > 7) = 1 - P(Y \leq 7) = 1 - f(7)$
    iii. In R:
```
> 1-pbinom(7, 25, .2)
[1] 0.1091228
```

There is a 10% chance of one person correctly guessing on at least 8 attempts.

c. What is the probability that at least one of the 20 receivers will attain a score indicative of ESP?
    i. Now that we've established the probability of attaining a score > 7, we can extrapolate this to the 20 receivers. Any given receiver has a ~10% chance of correctly guessing more than 7 correct symbols. There are 20 receivers - therefore we can compute the likelihood of at least 1 attaining a score of more than 7.
    ii. $P(Y \geq 1) = 1 - P(Y < 2) = 1 - f(2)$
    iii. In R:
```
> 1-pbinom(1, 20, .1)
[1] 0.608253
```

There is a 60% chance that at least 1 of the 20 receivers will attain a score of more than 7 correct guesses.


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
RGui (32-bit)
File Edit View Misc Packages Windows Help
R Console
cannot change working directory
> 1-pbinom(1, 20, .1)
[1] 0.608253
> setwd("C:/Users/khickman/Desktop/Personal/IUMSDS/StatsS520/Module3
+ ")
Error in setwd("C:/Users/khickman/Desktop/Personal/IUMSDS/StatsS520/Module3\n") $
cannot change working directory
> setwd("C:/Users/khickman/Desktop/Personal/IUMSDS/StatsS520/Module3")
> source("dbinom.txt")
> source("dbinom.r")
> objects()
[1] "at1" "at2" "bt1"
[4] "bt2" "concrete" "concrete_model"
[7] "concrete_model2" "concrete_norm" "concrete_test"
[10] "concrete_train" "convert_counts" "credit"
[13] "credit_model" "credit_pred" "credit_test"
[16] "credit_train" "dbinom.kh" "food_ratings"
[19] "ham" "ins_model" "ins_model2"
[22] "insurance" "lauch" "launch"
[25] "m.m5p" "model.rpart" "model_results"
[28] "mushroom_1R" "mushroom_JRip" "mushrooms"
[31] "mydata" "normalize" "p.rpart"
[34] "predicted_strength" "reg" "sdr_a"
[37] "sdr_b" "sms_classifier" "sms_corpus"
[40] "sms_corpus_clean" "sms_dtm" "sms_dtm_freq_test"
[43] "sms_dtm_freq_train" "sms_dtm_test" "sms_dtm_train"
[46] "sms_dtm2" "sms_freq_words" "sms_raw"
[49] "sms_tdm" "sms_test" "sms_test_labels"
[52] "sms_test_pred" "sms_train" "sms_train_labels"
[55] "spam" "tee" "train_sample"
[58] "wine" "wine_test" "wine_train"
[61] "x"
> dbinom.kh(10, 100, .3)
[1] 1.170418e-06
> dbinom(10, 100, .3)
[1] 1.170418e-06
>
[53] "sms_train" "sms_train_labels"
[55] "spam" "tee"
[57] "train_sample" "wine"
[59] "wine_test" "wine_train"
>
Plots