

# Midterm - S520

Keith Hickman

September 26, 2017

## R Markdown

**Problem 1** I toss six fair coins.

a. What is the probability I get exactly five heads?

We're looking for the probability of choosing five outcomes from 6 possibilities, or  $\binom{6}{5}$  because we are choosing exactly five outcomes from 6 possibilities divided by  $2^6$  possible outcomes.

We can solve with a fraction, or with the `dbinom` function in R.

$\binom{6}{5} = 6$  and  $2^6 = 64$  thus  $\frac{6}{64} = .0938$  or slightly less than 10%

To solve/confirm in R:

```
A <- dbinom(5,6,.5)
A
```

```
## [1] 0.09375
```

b. If I get at least five heads(B), what is the conditional probability I get exactly five heads(A)??

The formula for conditional probability is  $P(A|B) = P(A \cap B)/P(B)$ . We know the probability of obtaining exactly five heads (.0938), which is event (B)

To solve for the probability of obtaining at least 5 heads or event (A), which is the same as the probability of obtaining 5 or 6 heads,  $P(A) = P(X = 5) + P(X = 6)$ , we can use A from problem 1 and add to it  $P(X = 6)$  or  $1 - \text{pbinom}(4, 6, .5) = 0.109375$ . which evaluates to .109.

```
B <- 1-pbinom(4, 6, .5)
B
```

```
## [1] 0.109375
```

Because set A is completely contained within Set B (A intersect B is the same as A), we can solve for the conditional probability by dividing  $P(A)/P(B) = .857$

```
A/B
```

```
## [1] 0.8571429
```

Thus there is an 85% chance that I get exactly five heads if I get at least five heads.

c. Are the events "I get exactly five heads" and "I get at least five heads" independent? No. Here the events are independent because getting exactly five heads (A) does not affect getting at least five heads (B).

**Problem 2:** 48% percent of U.S. adults have visited a library in the past year. I select two U.S. adults at random. Let X be a discrete random variable representing the number of the adults I selected who have visited a library in the last year.

a) Find  $P(X = 0)$ ,  $P(X = 1)$ , and  $P(X = 2)$  This is another binomial variable problem in that there are two outcomes in the underlying sample space, either 1:has visited, or 0:has not visited. Our sample space  $S = 0, 2$ , and  $P = .48$ . Thus, we can test for  $P(X)$  by using the `dbinom` function in R for each X value:

```
dbinom(0,2,.48)
```

```
## [1] 0.2704
```

```
dbinom(1,2,.48)
```

```
## [1] 0.4992
```

```
dbinom(2,2,.48)
```

```
## [1] 0.2304
```

b) Write down an expression for  $F(y)$ , the cumulative distribution function (CDF) of  $X$ , for all  $y$ -values from  $-\infty$  to  $\infty$ .

$$f(y) = \begin{cases} 0 & x < 0 \\ .27 & 0 \leq x < 1 \\ .77 & 1 \leq x < 2 \\ 1 & x > 2 \end{cases}$$

c) Find the expected value and variance of  $X$ .

$EX = np$ ,  $2 * .48 = .96$  or we can use the sum of the weighted values:  $(0 * .27) + (1 * 0.49) + (2 * .23) = .96$

The Variance of  $X = p * (1 - p) * n$  or  $.48 * .52 * 2 = .499$

### Problem 3

Let  $X$  be a continuous random variable with the probability density function

$$f(x) = \begin{cases} .1 & 0 \leq x < 8 \\ .05 & 8 \leq x < 12 \\ 0 & \text{Otherwise} \end{cases}$$

a) Find the cdf  $f(y)$  for all values from negative infinity to infinity: First, we can verify that  $X$  is a pdf because all values  $f(x) > 0$  for every  $x \in \mathbb{R}$  and the area under the graph  $= 1$ .  $.1 * 8 = .8 + .05 * (12 - 8) = .2$  and  $.2 + .8 = 1$ .

The cdf of  $X$ :

$$f(y) = \begin{cases} 0 & x < 0 \\ .8 & 0 \leq x < 8 \\ 1 & x \geq 8 \end{cases}$$

b) Find the median. The Median of  $X$  is 5. The median is the  $q_2$  of  $X$ , or the .5 quantile of  $X$ . We can use geometry to solve by using the area under the graph to find the .5 quantile. The area of a region between 0 and  $q_2$  of function  $x$  and above the horizontal axis is .5. Since the region between  $[0,8]$  is a rectangle with a base of 8 and ht of .1, the area is .8, thus  $q_2$  lies on the interval  $[0,8]$ . The length of the rectangle's base that gives us .5 is 5 as,  $5 * .1 = .5$ .

c) Find the expected value The Expected Value of  $X$  is 4.8.

We can solve using the weighted averages of the centers of the two blocks from the pdf:  $.1 * 8 = .8$  and the center of  $[0,7] = 3.5$   $.05 * 4 = .2$  and the center of  $[8,12] = 10$  Thus, we have  $(.8 * 3.5) + (.2 * 10) = 2.8 + 2 = 4.8$

### Problem 4

$$f(y) = \begin{cases} 0 & y < 0 \\ .25y^2 & 0 \leq y < 1 \\ .25y & 1 \leq y < 4 \\ 1 & y \geq 4 \end{cases}$$

a) Is  $X$  discrete, continuous, or neither? Explain.  $X$  is a continuous variable as the Set of values  $X$  is not countable, and the values lie on an interval.

b) Find the standard deviation of  $X$ . The standard deviation is equal to  $\sqrt{Var(X)}$ .  $Var(x) = E(X^2) - EX$ . Plugging in  $E(X)^2$  and  $E(X^2) = \frac{48}{8} - \frac{49^2}{24} = \frac{49}{24}$  or 4.16  $\sqrt{Var(X)} = \sqrt{3.95} = 1.989$

I know the notation on  $E(X)^2$  isn't correct

in R:

```
EX2 <- (49/24)^2
EXsq <- 48/8
VarX <- EXsq - EX2
StdDev <- sqrt(VarX)
EX2
```

```
## [1] 4.168403
```

```
VarX
```

```
## [1] 1.831597
```

```
StdDev
```

```
## [1] 1.353365
```

c) Let  $A$  and  $B$  be independent random variables with the same distribution as  $X$ . Let  $Y = A - B$ . What are the expected value and standard deviation of  $Y$ ?

Here, both  $EX$  and Standard Deviation of  $Y$  are 0. Since the distributions are independent, we can subtract the expected values and standard deviations of  $A - B$  as follows:

$$E(Y) = E(X_1 - X_2) = E(X_1) - E(X_2) = \frac{49}{24} - \frac{49}{24} = 0$$

The formula for standard deviation is similar:

$$Var(Y) = Var(X_1) - Var(X_2) = 1.831 - 1.831 = 0$$

### Problem 5

a) The median of  $X$  is .81.  $q_2$  of  $X$  equals the .5 quantile of  $X$ . Looking at the cdf of  $X$ , the .5 quantile will fall obviously somewhere between 0 and 1, or we can say that  $.5 = 1 - \frac{1}{y^2}$ . This problem can be solved using algebra.

$$.5 = 1 - \frac{1}{y^2} \quad -1.5y^2 = -\frac{1}{y^2} \quad y^2 = .667 \quad y = .81$$

b) The IQR of  $X$  is 1.5. The IQR is  $q_3 - q_1$  or  $.75 - .25$  of  $X$ .

Solve for .75 or  $q_3$ :

$$.75 = 1 - \frac{1}{y^2}$$

$$-.25 = \frac{1}{y^2}$$

$$-.25y^2 = 1$$

$$y^2 = 4$$

$$y = 2$$

$$\text{Solve for } .25 \text{ or } q_1 \quad .25 = 1 - \frac{1}{y^2} \quad -.75 = \frac{1}{y^2} \quad -.75y^2 = 1 \quad y^2 = .25 \quad y = .5$$

$$\text{Solve for } q_3 - q_1 \quad 2 - .5 = 1.5$$

The IQR for  $X = 1.5$

## Problem 6

Let  $X$  be standard normal random variable. Let  $Y = |X|$ .

a) What is the median of  $Y$ ?

b) What is  $P(1 < Y < 2)$ ? This is the difference of the distribution CDF at 2 and at 1.

```
pnorm(2) - pnorm(1)
```

```
## [1] 0.1359051
```

Or 13.5%

c) Find the .95 quantile of  $Y$ .  $P(Y < y)$   $P(|X| < .95)$   $(P - y < X < y)$  Since  $-y$  goes to infinity, we can drop that as impossible to calculate. We're looking then for  $P(X < y) = .975$  We can use `qnorm` which is the inverse of the cdf.

```
qnorm(.975)
```

```
## [1] 1.959964
```

**Problem 7** 1) I select an adult U.S. woman at random. What is the probability she is under five feet (152.4 cm) tall or solve for  $P(X < 152.4)$  We can solve with the `pnorm` function in R by passing in the value, mean, and standard deviation of the variable. `pnorm` is used for finding the quantile of values from negative infinity up to  $X$ .

```
pnorm(152.4, 162.1, 7.3)
```

```
## [1] 0.09196241
```

There is a 9.2% chance of selecting a woman under five feet tall.

2) I select an adult U.S. man at random. What is the probability he is over six feet (182.88 cm) tall? Here, since we're interested in  $P(X > 182.88)$ , we need to subtract  $P(X \leq 182.88)$  from 1.

```
prob6 <- 1 - pnorm(182.88, 175.9, 7.5)
prob6
```

```
## [1] 0.2965343
```

Thus, there is a 29.6% chance that a given male will be taller than 6 feet.

3) I select four adult U.S. men at random. What is the probability at least one of them is over six feet tall? Thus, we have a binomial distribution (either is taller than 6' or is not). Here, we can pass in the probability that a given man is taller than 6' obtained using `pnorm` above (`prob6`), as the probability that one out of the four males selected will be taller than six feet using the `pbinom` function in R. If  $P(A)$  is "at least one of four males is taller than six feet" then  $A^c$  is the probability that none of the four men is taller than six feet. The probability of  $A^c$  is `pbinom(0,4,prob6)`. Thus we can subtract that value from 1 to obtain the  $P(A)$ :

```
1- pbinom(0, 4, prob6)
```

```
## [1] 0.7551096
```

There is a 75.5% chance that one of the four men selected will be taller than 6 feet.

4) I select at random an adult U.S. man and, independently, an adult U.S. woman. What is the probability the woman is taller?

Another way of wording this problem is to say, what's the probability I select a woman from a quantile of  $X_1$  and that the quantile I select the female from is greater than the quantile of  $X_2$  for the man I select.

The difference between the expected value of the two variables is  $E(X_1 - X_2)$  or  $E(X_1) - E(X_2)$ , or  $175.9 - 162.1 = 13.8$

```
1 - pnorm(175.9, 162.7, 7.3)
```

```
## [1] 0.0352862
```

There is a 3% chance that the woman I select will be taller than the man I select.

**Problem 8** 1) Suppose this test is taken by a student who does not have psychic powers. What is the probability she gets 15 or more correct? Here, we have a binomial distribution, and are seeking the  $P(X \geq 15)$ . For this we can use `dbinom` in R:

```
d <- dbinom(15, 20, .5)
```

```
d
```

```
## [1] 0.01478577
```

2) Suppose this test is taken by eighty students who do NOT have psychic powers. What is the probability at least one student gets 15 or more correct? We can use the previously-established probability that one person gets at least 15 questions corrected (1%) as the probability for our second `dbinom` function. Here, we want to know the probability that at least one student out of 80 will get at least 15 answers correct, or  $P(X \geq 1)$

```
dbinom(1, 80, d)
```

```
## [1] 0.3646342
```

There is a 36.4% probability that at least one student out of 80 will get 15 or more answers correct.