

Lecture 5

Discrete Random Variables

Today's Updates / Reminders

- Homework 2 is due today
- Lab 2 is due Friday and requires the use of Excel. The tech guide suggests using filters. I used logical operators... =OR, =AND, =COUNTIF
 - Let A = ankle, B = wrist, and C = bicep
 - For part b, $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$
 - Hint: what is P(C)?
- Homework 3 is due a week from today (since we finish ch. 3 today)
- Stat Hub is now open for business. Free help/tutoring. SAS1101, M-Th, 5:30-8:00pm

Chapter 3.1 – 3.2

Random Variables and Probability
Distributions of Discrete R.V.

Random Variable

- A **random variable** is a variable whose value is a numerical representation of the outcomes in an experiment.
- Events are represented by A, B, C...
- Variables are represented by X, Y, Z...

Types of Random Variables

- **Discrete** random variables – countable; integers on a number line; discrete variables cannot take on a value between the integers
 - Ex., Number of manatees in Florida
- **Continuous** random variables – decimals and fractions; can slide anywhere along a number line
 - Ex., Number of picocuries per liter of radon as measured by a radon detector



Probability of Discrete Random Variables

- Possible values of the random variable (X), denoted by lower case, (x).
- The probability statement is written: $P(X = x)$. This is also called the probability distribution or probability mass function (pmf).
- For example, we flip a coin two times. The sample space is $S = \{HH, HT, TH, TT\}$. Let X be the number of heads.
 - $P(X = 0) =$
 - $P(X = 1) =$
 - $P(X = 2) =$

Distribution of Discrete Variables

- The following table shows the probability of the number of roommates that an NC State student has:

X	0	1	2	3
$P(X = x)$	0.1	0.4	0.3	0.2

Rules

$$1) 0 \leq P(x) \leq 1$$

$$2) \sum P(x) = 1$$

- What is the probability of having 3 roommates?
- $P(X < 2) =$
- $P(X \leq 2) =$ (think of 2 ways!)
- What is the probability of having 0-3 roommates?

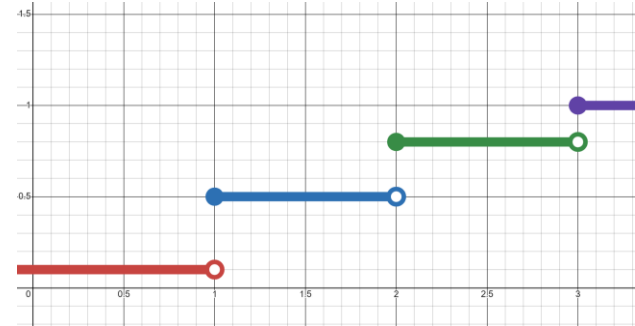
Cumulative Distribution Function

- The cumulative distribution function (cdf) of a discrete R.V. X with probability distribution (or pmf), $p(x)$ is defined by:

$$F(x) = P(X \leq x) = \sum_{y:y \leq x} p(y)$$

- We can create a plot of the previous situation and it looks like a step function.

X	0	1	2	3
$P(X = x)$	0.1	0.4	0.3	0.2



Review

- This table shows the probability of an affirmative response, based on the number of women asked out by the undergrad in the upper right.
- Is this a valid probability distribution?
- Think about the nature of this experiment as well. How could I change the experiment so that it would be valid?



x	$P(x)$
0	0
1	0.001
2	0.020
3	0.105
4	0.233
5	0.242

Geometric Distribution

- There is a distribution for the scenario I just described.
- Let's say the probability someone says "yes" is 0.2. What is the probability that I have to ask out 4 women before one says yes?
- We would take $P(4) = (0.8)*(0.8)*(0.8)*(0.2) = 0.1024$
- The general form is then:

$$P(x) = \begin{cases} (1-p)^{x-1}p, & x = 1, 2, 3, \dots \\ 0 & \textit{otherwise} \end{cases}$$

- The cdf is: $P(x) = 1 - (1-p)^x$
- This situation is generalizable to various manufacturing scenarios, such as quality control with a known defect rate or time to failure.

Let's try this in
MATLAB...
geopdf
geocdf

Chapter 3.3

Expected Value
Variance of a Linear Function

Expected Value

- Let's continue with the roommate distribution:

X	0	1	2	3
$P(X = x)$	0.1	0.4	0.3	0.2

- What to do if someone asks what is the average number of roommates of an NC State student?
- Expected value, also known as mean, of a discrete R.V. Let X be a discrete R.V. with set of possible values D .

$$E(X) = \mu_x = \sum_{x \in D} x \cdot P(x)$$

- [TI83](#) video showing how to do this. Let's do it by hand and in Excel.

Another Example – Geometric

- Using the geometric distribution, let's say we know the proportion of hydraulic cylinders that fail inspection is 0.002.

$$E(X) = \mu_x = \sum_{x \in D} x \cdot P(x) = \sum_{x=1}^{\infty} x \cdot p \cdot (1-p)^{x-1} = p \sum_{x=1}^{\infty} \left[-\frac{d}{dp} (1-p)^x \right]$$

- Interchanging the order of taking the derivative and the summation, the sum is that of a geometric series. After the sum is computed, the derivative is taken, resulting in $E(x) = 1/p$.
- So what is the average number of hydraulic cylinders you would inspect before you find a failure?

$$E(X) = \frac{1}{0.002} = 500$$

Variance

- Let X have pmf $p(x)$ and expected value, μ . Then the variance of X , denoted by $V(X)$, or just σ^2 , is:

$$V(X) = \sum_D (x - \mu)^2 \cdot p(x)$$

$$V(X) = \left[\sum_D x^2 \cdot p(x) \right] - \mu^2$$

- The second version of the formula is often called the “shortcut” method.
- And standard deviation is the square root of the variance.

Variance Example

- Continuing with the roommate example, we found $E(X) = 1.6$

X	0	1	2	3
$P(X = x)$	0.1	0.4	0.3	0.2

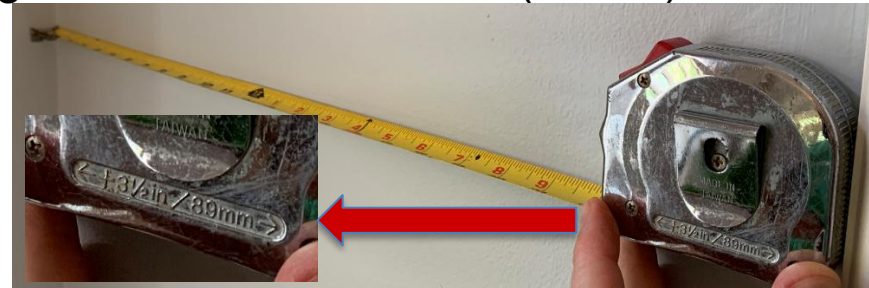
- Now let's use the shortcut method...

Expected Value / Variance – Linear $f(x)$

$$E(aX + b) = a \cdot E(X) + b$$
$$V(aX + b) = \sigma_{aX+b}^2 = a^2 \cdot \sigma_X^2$$

- These rules apply to discrete or continuous.
- These are the formulas to find expected value and variance of a linear function. Why does this work?
- Suppose I take measurements with a tape measure. I write numbers in inches but I need centimeters. I also forgot to add the case size (8.9cm) to my measurements.

Write a mathematical function that would convert my data.



General Rules for Mean and Variance

- Suppose my original data was 20, 30, and 40 inches.
 - What is the average in inches?
 - Now add the case width (let's say it is 3 inches). What does this do to the average in inches? *Think: move the distribution... 23, 33, 43*
 - What is the average in centimeters? *Think: scale the distribution... 58.42, 83.82, 109.22*
- The rule:
 - $E(aX + b) = a \cdot E(X) + b$

General Rules for Mean and Variance

- Suppose my original data was 20, 30, and 40 inches.
 - The variance in inches is 100 inches²
 - What do you think will happen to the variance when I add the 3 inches? *Think: move the distribution... 23, 33, 43*
 - What do you think will happen to the variance when I convert to centimeters? *Think: scale the distribution... 58.42, 83.82, 109.22*
- The rule:
 - $V(aX + b) = \sigma_{aX+b}^2 = a^2 \cdot \sigma_X^2$
 - $\sigma_{aX+b} = |a| \cdot \sigma_X$

Example

X	16	18	20
P(x)	0.2	0.5	0.3

- A certain brand of upright freezer is available in three different rated capacities: 16 ft³, 18 ft³, and 20 ft³. Let X = the rated capacity of a freezer of this brand sold at a certain store. Suppose X has pmf
- $E(X) =$ _____ , $V(X) =$ _____
- If the price of the freezer having capacity X is $70X - 650$, what is the expected price paid by the next customer to buy a freezer?

Chapter 3.4

Binomial Probability Distribution

Requirements for Binomial

1. The experiment consists of a fixed number of trials, n .
2. Each trial can result in only two outcomes, success and failure.
3. The trials are independent.
4. The probability of success is constant from one trial to the next.

Not Binomial – Why?

1. I flip a coin until I get 10 heads and count how many of each I got.
2. I roll a die 20 times and count the frequency of each number rolled.
3. I pull 8 cards from a deck of cards without replacement and count whether it is a heart or not.
4. I throw 100 bags at a cornhole board, counting the number of bags that land on the board... while tailgating at the football game and maybe sipping an “oat soda.”

Binomial Probability Function

- The binomial probability function (or binomial pmf) is given by:

$$b(x; n, p) = \binom{n}{x} p^x (1 - p)^{n-x}$$

- The formula seems scary but it's not that bad. Let's look at 4 coin flips. What is the probability I get three heads?
- How many ways can I get 3 heads? Order does not matter, right?
- What is the probability of three heads?
- What is the probability of one "not heads" or tails?
- Put it all together...

Example

- If I roll a die 10 times, what is the probability of getting the number 4 (or really any single number) three times?
- [Google Sheets Example](#)
- [Galton Board](#) aka quincunx
- FYI, there are tables in the appendix of textbooks for common cases of n , p , and x .

How Does It Work?

- Oddly enough, the probability of flipping a coin and:
 - getting 4 heads out of 4 (0.0625) **IS GREATER THAN**
 - getting 59 heads or more out of 100 (0.0443)
- The former is more likely than the latter.
- Why is this true?
- Let's go back to [Google Sheets](#).

Cumulative Binomial

- When the problem says, “What is the probability of **exactly** x successes?” we use the previous formula.
- What if we want to know the probability of **greater than or equal to** x successes?
- We add up all the probabilities from x to n .
- If it said **less than** x , we would count from 0 to $x - 1$.
- Recall most calculators do the probability of $\leq x$.

Practice

- You are taking a 10-question test with just True/False answers. You have no idea what any of the right answers are so you guess. You need at least a 60% to pass. What is the probability that you pass?
- This year's Super Bowl had a 43.5 Nielsen rating, meaning 43.5% of households watched it. There are 12 houses on your street. What is the probability that at least one household had the game on?
- The sun shines 300 days a year in Phoenix. What is the probability that there are 52 non-sunny days in Phoenix this year?

Example

- Airlines oversell flights. If the daily 3:15pm flight from RDU to DFW has a historic track record where 95% of passengers actually show up to the flight, and the plane seats 88, what is the probability that they will oversell the flight if they sell 91 tickets?

Mean and Standard Deviation

- Recall we use the terms “mean,” μ , and “expected value,” $E(X)$, interchangeably.

$$E(X) = \mu = np$$

- We also use $V(X)$ or σ^2 for variance.

$$V(X) = \sigma^2 = npq$$

- And standard deviation is the square root of variance.

$$\sigma_X = \sqrt{npq}$$

Example

- One in 20 people have a food allergy. Suppose I have a classroom with 90 students in it. What is the expected value of students with allergies? What is the standard deviation?
- If you took the mean \pm standard deviation, what is the range of the number of students?

Example

- When you donate blood, they draw samples to test for various diseases. [Pooling samples](#) is a way to shorten screening time and reduce the number of tests needed and was used during the height of the COVID pandemic. Let's say the incidence of COVID in the population is 5%. If we pool samples in groups of 8, what is the probability that the pooled sample would test positive?

Chapter 3.6

Poisson Probability Distribution

Definition

- A discrete random variable X is said to have a Poisson distribution with parameter μ ($\mu > 0$) if the pmf of X is:

$$p(x; \mu) = \frac{e^{-\mu} \cdot \mu^x}{x!}, \text{ where } x = 0, 1, 2, 3, \dots$$

- As we saw, if we flip a coin 100 times, the probability of any single event is small. Imagine we flip a coin 100,000 times. The probability of getting 50,000 heads is 0.0025, according to binom.dist. That's less than the probability of getting 4 heads out of 4 flips! So as $n \rightarrow \infty$ and $p \rightarrow 0$, the binomial can be approximated to the Poisson.

Explaining Terms

- Other versions of the formula use λ instead of μ . And it is often called the rate.
- Also binomial doesn't model events that occur at the same time. The successes occur one after the other.
- When you have a large number of events with a low probability, then the number of events that occur in a fixed time interval follows a Poisson distribution.
- Poisson can be used to describe the probability of a number of events that occur at a specific rate, λ , and within a specific time interval.

Example

- Suppose a restaurant receives an average of 10 customers per hour. We wish to know the probability of exactly 15 customers coming.
 - Note why binomial won't apply here.
- So this means $\mu = 10$ and $x = 15$.
- What about the probability of greater than or equal to 15? We must use tables. Let's go to the textbook in [WebAssign](#).

Let's try [Matlab](#)...

poisspdf

poisscdf