Module 4 Lecture - Discrete Random Variables

Introduction to Statistical Methods

Quinton Quagliano, M.S., C.S.P

Department of Educational Psychology

Table of Contents

1 Overview and Introduction

1.1 Textbook Learning Objectives

- Recognize and understand discrete probability distribution functions, in general.
- Calculate and interpret expected values.
- Recognize the binomial probability distribution and apply it appropriately.
- Classify discrete word problems by their distributions.

1.2 Instructor Learning Objectives

- Refresh on understanding of behavior and characteristics of discrete variables
- Understand the how to write probability distribution function tables and use them in calculation
- Be able to continue to write coherent and readable notation for probability problems

1.3 Introduction

- We will be expanding on what we spoke about with __________ in the last lecture, but now explicitly applying those ideas to variables

Discuss: Try describing discrete variables in your own words, based on what you know from Module 1

- We previously described __________ data, or that which is numeric and "count-able", that has no intervals between integers.
  - Those will be the focus of this lecture, before we focus on __________ variables in the next lecture
- In this lecture, we will introduce the concept of **random variables**, or those that can vary in subsequent __________ (used in the sense as how it was introduced during the probability lecture)
- The __________ of random variables is as follows:
  - Uppercase alphabetical character, e.g. X, Y, Z, etc. $\rightarrow$ written/verbose description of variable

- Lowercase alphabetical equivalent, e.g. x, y, z, etc. $\rightarrow$ the possible values X can take on

- Example from book:
  - X = the number of heads you get when you toss three ________ coins (concrete description of what the variable is)
  - x = 0, 1, 2, 3 (what values it could be)
  - This is a discrete, random variable, because x has ________ different values (random) and because those different values follow the discrete rules, i.e. "countable" in the sense that it is a number of times and no intervals between integers

? Review: What is a synonym for 'fair' when discussing probability?

- (●) A) Equally similar
- ( ) B) Similar
- ( ) C) Equivalent
- ( ) D) Equally likely

Explanation:

2 Probability Distribution Function (PDF) for a Discrete Random Variable

2.1 Introduction

- A **probability distribution function** or PDF is a very scary way to say the list of possible x values and their respective ________
  - The ________ they follow are simple:
    1. All probabilities must sum up to 1, or $\sum P(x) = 1$
    2. Each probability is between 0 and 1, inclusive, or $0 \leq P(x) \leq 1$

Discuss: Review: try writing the mathematical notation for probability of event A equaling a 30% chance of occurring

- In a probability distribution function, we can think of each ____________ value of X, i.e., the xs, having some specified probability of occurring

2.2 PDF Tables

- Example:
  - A survey is being done on how many vehicles families own. Assume the following probability distribution function:
  - X = the number of vehicles owned by a family
  - x = 0, 1, 2, 3, 4

x	P(X = x)
0	0.10
1	0.20
2	0.30
3	0.25
4	0.15

or…

x	P(x)
0	P(x = 0) = 10/100
1	P(x = 1) = 20/100
2	P(x = 2) = 30/100
3	P(x = 3) = 25/100
4	P(x = 4) = 15/100

- In this example, If I were to randomly select a single family from this distribution, I would have a 30% chance of picking a family with 2 cars

Discuss: On 30% of days I bring one umbrella with me, and on 70% of days I bring no umbrella with me. Write out a PDF table as above to demonstrate this probability distribution function

- There are 4 types of distributions for discrete random variables:
  - ________ Distributions
  - ________ Distributions
  - ________ Distributions
  - ________ Distributions
  - Each of these has their place, but the last 3 come into play more on more ________ statistical techniques - we'll focus just on the binomial distribution for this lecture

Important

Just because I leave these distribution out in this lecture, doesn't mean that they aren't important! It's mostly just that the others won't be as useful immediately to beginners in statistics.

3 Mean or Expected Value and Standard Deviation

3.1 Introduction

- With PDFs, we sometimes may wish to find the **expected value**, or the **"long-term" average or mean**. Thus, doing running this experiment over and over again, we'd expect to ________ on this expected mean
- This is based in the **Law of Large Numbers** a topic alluded to several times in the previous modules
  - Somewhat review from the ________ module: this law states that relative observed frequency approaches the theoretical probability as the number of experiments or trials increases
- For a discrete probability function:

$$\mu = \sum (x \cdot P(x))$$

$$\sigma = \sqrt{\sum [(x-\mu)^2 \cdot P(x)]}$$

- If outcomes for the experiment are equally likely then these formulas work to find the ________________ mean and standard deviation for each of the outcomes

> Discuss: Why would we use the mu and sigma notation here, instead of x bar and lowercase s? Do they represent statistics or parameters?

3.2 Mean Calculation Example

Car example from earlier:

x	P(x)	x * P(x)
0	0.10	0.00
1	0.20	0.20
2	0.30	0.60
3	0.25	0.75
4	0.15	0.60

$$\mu = \sum (x * P(x)) = 2.15$$

3.3 Standard Deviation Calculation Example

Car example from earlier:

x	P(x)	x * P(x)	(x - mu)^2 * P(x)
0	0.10	0.00	0.462250
1	0.20	0.20	0.264500

x	P(x)	x * P(x)	(x - mu)^2 * P(x)
2	0.30	0.60	0.006750
3	0.25	0.75	0.180625
4	0.15	0.60	0.513375

$$\sigma = \sqrt{\sum (x-\mu)^2 * P(x)} = 1.1948$$

3.4 Section Conclusion

- Probability distribution functions serve a purpose - that is - they can describe a particular ________________ in the probability of outcomes in the data.
- Once we understand what pattern the variable fits, we can use this information for other ________________, as there are the 4 specialty distributions mentioned earlier: geometric, hypergeometric, poisson, and binomial

4 Binomial Distribution

4.1 Introduction

- The binomial distribution has a certain, ________________ number of trials represented as n
  - Each trial is ________________ and does not affect subsequent trials
- There are two possible ________________ in the binomial distribution, “success” and “failure”
  - $P(success) = p$
  - $P(failure) = q$
  - $p + q = 1$
- The experiment described above fits the **binomial probability distribution**. Where the random discrete variable X represents the number of successes obtain in n trials
- For the binomial probability distribution:

$$\mu = n * p$$

$$\sigma^2 = n * p * q$$

$$\sigma = \sqrt{n * p * q}$$

4.2 Notation

$$X \sim B(n, p)$$

- This formula is read as “X is a random variable with a binomial distribution”, where n and p represent the same things as before

5 Conclusion

5.1 Recap

- In this shorter lecture, we introduced the concept of discrete random variables, and how they can be represented with probability distribution functions
- We played around with calculating expected mean and standard deviations for outcomes from the probability distribution function table
- We introduced the binomial distribution as a special, specific pattern for a probability distribution represented via “successes” and “failures”

5.2 Lecture Check-in

- Make sure to complete and submit the lecture check-in