

Lecture 7

Probability III

Readings: IntroStat Chapter 4; OpenIntro Chapter 3

STAT 8010 Statistical Methods I
May 24, 2023

Whitney Huang
Clemson University

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

7.1

Notes

Agenda

- 1 Random Variables
- 2 Bernoulli and Binomial Random Variables
- 3 Hypergeometric Random Variable

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

7.2

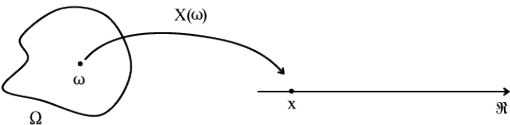
Notes

Random Variables

A **random variable** is a real-valued function whose domain is the sample space of a random experiment. In other words, a random variable is a function

$$X : \Omega \mapsto \mathbb{R}$$

where Ω is the sample space of the random experiment under consideration and \mathbb{R} represents the set of all real numbers.



Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

7.3

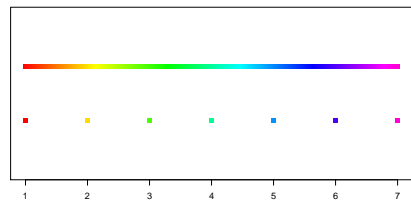
Notes

Discrete and Continuous Random Variables

There are two main types of quantitative random variables (r.v.s): **discrete** and **continuous**. A discrete r.v. often involves a count of something.

Discrete random variable

A random variable X is called a discrete random variable if the outcome of the random variable is limited to a countable set of real numbers (usually integers).



Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

74

Notes

Example

The following is a chart describing the number of siblings each student in a particular class has.

Siblings	Frequency	Relative Frequency
0	8	.200
1	17	.425
2	11	.275
3	3	.075
4	1	.025
Total	40	1

Let's define the event A as the event that a randomly chosen student has 2 or more siblings. What is $\mathbb{P}(A)$?

Solution.

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(X \geq 2) = \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4) \\ &= .275 + .075 + .025 = .375\end{aligned}$$

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

75

Notes

Probability Mass Function

Let X be a discrete random variable. Then the probability mass function (pmf) of X is the real-valued function defined on \mathbb{R} by

$$p_X(x) = \mathbb{P}(X = x)$$

The capital letter, X , is used to denote random variable. Lowercase letter, x , is used to denote possible values of the random variable.

$p_X(x)$: The probability that the discrete random variable X is exactly equal to x .

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

76

Notes

Probability Mass Function Example

Flip a fair coin 3 times. Let X denote the number of heads tossed in the 3 flips. Create a pmf for X

Solution.

The random variable X maps any outcome to an integer (e.g. $X((T, T, T)) = 0, X((H, H, T)) = 2$)

x	0	1	2	3
$p_X(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

7.7

Notes

Properties of a PMF

- $0 \leq p_X(x) \leq 1, x \in \{0, 1, 2, \dots\}$
- $\sum_x p_X(x) = 1$

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

7.8

Notes

Example

Let X be a random variable with pmf defined as follows:

$$p_X(x) = \begin{cases} k(5 - x) & \text{if } x = 0, 1, 2, 3, 4 \\ 0 & \text{otherwise} \end{cases}$$

- 1 Find the value of k that makes $p_X(x)$ a legitimate pmf.
- 2 What is the probability that X is between 1 and 3 inclusive?
- 3 If X is not 0, what is the probability that X is less than 3?

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

7.9

Notes

Mean of Discrete Random Variables

The mean of a discrete r.v. X , denoted by $\mathbb{E}[X]$, is defined by

$$\mathbb{E}[X] = \sum_x x \times p_X(x)$$

Remark:

The mean of a discrete r.v. is a weighted average of its possible values, and the weight used is its probability. Sometimes we refer to the expected value as the **expectation (expected value)**, or the **first moment**.

For any function, say $g(X)$, we can also find an expectation of that function. It is

$$\mathbb{E}[g(X)] = \sum_x g(x) \times p_X(x)$$

Example

$$\mathbb{E}[X^2] = \sum_x x^2 \times p_X(x)$$

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

7.10

Notes

Properties of Mean

Let X and Y be discrete r.v.s defined on the same sample space and having finite expectation (i.e. $\mathbb{E}[X], \mathbb{E}[Y] < \infty$). Let a and b be constants. Then the following hold:

- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[aX + b] = a \times \mathbb{E}[X] + b$

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

7.11

Notes

Number of Siblings Example Revisited

Siblings (X)	Frequency	Relative Frequency
0	8	.200
1	17	.425
2	11	.275
3	3	.075
4	1	.025
Total	40	1

Find the expected value of the number of siblings

Solution.

$$\mathbb{E}[X] = \sum_x x p_X(x) = 0 \times .200 + 1 \times .425 + 2 \times .275 + 3 \times .075 + 4 \times .025 = 1.3$$

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

7.12

Notes

Variance/Standard Deviation of Discrete r.v.'s

The **variance** of a (discrete) r.v., denoted by $Var(X)$, is a measure of the spread, or variability, in the r.v. $Var(X)$ is defined by

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

or

$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

The **standard deviation**, denoted by $sd(X)$, is the square root of its variance

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

7.13

Notes

Properties of Variance

Let c be a constant. Then the following hold:

- $Var(cX) = c^2 \times Var(X)$
- $Var(X + c) = Var(X)$

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

7.14

Notes

Example

Suppose X and Y are random variables with $\mathbb{E}[X] = 3$, $\mathbb{E}[Y] = 4$ and $Var(X) = 4$. Find:

- $\mathbb{E}[2X + 1]$
- $\mathbb{E}[X - Y]$
- $\mathbb{E}[X^2]$
- $\mathbb{E}[X^2 - 4]$
- $\mathbb{E}[(X - 4)^2]$
- $Var(2X - 4)$

Probability III

CLEMSON UNIVERSITY

Random Variables

Bernoulli and Binomial Random Variables

Hypergeometric Random Variable

7.15

Notes

Bernoulli Trials

Many problems in probability and its applications involve **independently** repeating a random experiment and observing at each repetition whether a specified event occurs. We label the occurrence of the specified event a **success** and the nonoccurrence of the specified event a **failure**.

Example:

Tossing a coin several times



Notes

Bernoulli Trials Cont'd

Bernoulli trials:

- Each repetition of the random experiment is called a trial
- We use p to denote the probability of a success on a single trial

Properties of Bernoulli trials:

- Exactly two possible outcomes **success** and **failure**
- The outcomes of trials are **independent** of one another
- The success probability, p , and therefore the failure probability, $(1 - p)$, remains the same from trial to trial



Notes

Binomial Random Variable

We define the **Binomial** r.v. as the number of successes in n Bernoulli trials, where the probability of success in one trial is p . Let X be a Binomial r.v.

- The definition of X : **# of successes in n trials of Bernoulli trials.**
- The support: $0, 1, \dots, n$
- Its parameter(s) and definition(s): **p : the probability of success on 1 trial; n is the sample size**
- The probability mass function (pmf):

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$$

- The expected value:

$$E[X] = np$$

- The variance:

$$\text{Var}(X) = np(1-p)$$



Notes

Example

To test for Extrasensory perception (ESP), we have 4 cards. They will be shuffled and one randomly selected each time, and you are to guess which card is selected. This is repeated 10 times. Suppose you do not have ESP. Let R be the number of times you guess a card correctly. What are the distribution and parameter(s) of R ? What is the expected value of R ? Furthermore, suppose that you get certified as having ESP if you score at least an 8 on the test. What is the probability that you get certified as having ESP?

Solution.

$$R \sim \text{Binomial}(n = 10, p = 1/4 = .25)$$

$$E[R] = n \times p = 2.5$$

$$P(X \geq 8) = .000416$$



Notes

Example

Suppose that 95% of consumers can recognize Coke in a blind taste test. Assume consumers are independent of one another. The company randomly selects 4 consumers for a taste test. Let X be the number of consumers who recognize Coke.

- ❶ What is the probability that X is at least 1?
- ❷ What is the probability that X is at most 3?



Notes

Binomial and Hypergeometric r.v.s

The binomial distribution describes the probability of k successes in n trials **with replacement**.

We want a distribution to describe the probability of k successes in n trials **without replacement** from a finite population of size N containing exactly K successes.

⇒ **Hypergeometric Distribution**

Important applications are **quality control** and statistical **estimation of population proportions**. The hypergeometric r.v. is the equivalent of a Binomial r.v. except that sampling is done **without replacement**.

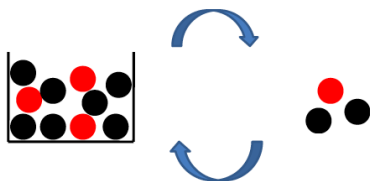


Notes

An Example of Hypergeometric r.v.

Probability:

What is the probability to get 1 red and 2 black balls?



Statistics:

What percentage of balls in the box are red?

Probability III
CLEMSON UNIVERSITY

Random Variables
Bernoulli and Binomial Random Variables
Hypergeometric Random Variable

7.22

Notes

Hypergeometric r.v.s

Let X be a hypergeometric r.v.

- The definition of X : # of successes in n trials of a random experiment, where sampling is done without replacement (or trials are dependent)
- The support: $k \in \{\max(0, n + K - N), \dots, \min(n, K)\}$
- Its parameter(s) and definition(s): N : the population size, n : the sample size, and K : number of success in the population
- The probability mass function (pmf):

$$p_X(k) = \frac{\binom{K}{k} \times \binom{N-K}{n-k}}{\binom{N}{n}}$$
- The expected value: $E[X] = n \frac{K}{N}$
- The variance: $\text{Var}(X) = n \frac{K}{N} \frac{N-K}{N} \frac{N-n}{N-1}$

Probability III
CLEMSON UNIVERSITY

Random Variables
Bernoulli and Binomial Random Variables
Hypergeometric Random Variable

7.23

Notes

Example

There are 100 identical looking 52" TVs at Best Buy in Anderson, SC. Let 10 of them be defective. Suppose we want to buy 8 of the aforementioned TVs (at random). What is the probability that we don't get any defective TVs?

Solution.

Let D be the number of defective TVs in the sample.

$D \sim \text{Hyp}(N = 100, n = 8, K = 10)$

$$P(D = 0) = \frac{\binom{10}{0} \binom{90}{8}}{\binom{100}{8}} = 0.4166$$

Probability III
CLEMSON UNIVERSITY

Random Variables
Bernoulli and Binomial Random Variables
Hypergeometric Random Variable

7.24

Notes

Summary

In this lecture, we learned

- Random Variables
- The probability mass function, mean, and variance of a discrete random variable
- Examples of discrete random variables: Bernoulli/Binomial, Hyper-geometric

Probability III

CLEMSON
UNIVERSITY

Random Variables
Bernoulli and
Binomial Random
Variables
Hypergeometric
Random Variable

7.25

Notes

Notes

Notes
