Lecture 8

Probability IV

Text: Chapter 4

STAT 8010 Statistical Methods I
September 15, 2020

Whitney Huang
Clemson University

Notes

Agenda

1. Binomial Random Variables
2. Hypergeometric Random Variable
3. Continuous Random Variables

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?

Notes

Example Cont'd

1. Want to find k s.t.
$$\sum_{x=0}^{4} p_X(x) = \sum_{x=0}^{4} k(5-x) = 1$$
$$\Rightarrow 15k = 1 \Rightarrow k = \frac{1}{15}$$

2. $$P(1 \le X \le 3) = P(X=1) + P(X=2) + P(x=3)$$
$$= \frac{4}{15} + \frac{3}{15} + \frac{2}{15} = \frac{9}{15} = 0.6$$

3. $$P(X < 3 | x \neq 0) = \frac{P(X < 3 \& X \neq 0)}{P(X \neq 0)}$$
$$= \frac{P(0 < X < 3)}{1 - P(X=0)} = \frac{4/15 + 3/15}{1 - 5/15} = 0.7$$

Notes

Mean of Discrete Random Variables

Probability IV
CLEMSON UNIVERSITY
Binomial Random Variables
Hypergeometric Random Variable
Continuous Random Variables

The mean of a discrete r.v. X, denoted by $E[X]$, is defined by
$$E[X] = \sum_x x \times p_X(x)$$

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

Example
$$E[X^2] = \sum_x x^2 \times p_X(x)$$

Notes

Properties of Mean

Probability IV
CLEMSON UNIVERSITY
Binomial Random Variables
Hypergeometric Random Variable
Continuous Random Variables

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:

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

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.

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

Notes

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

Probability IV
CLEMSON UNIVERSITY
Binomial Random Variables
Hypergeometric Random Variable
Continuous Random Variables

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

$$\boxed{\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[x])^2]}$$

or

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

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

Notes

Properties of Variance

Probability IV
CLEMSON UNIVERSITY
Binomial Random Variables
Hypergeometric Random Variable
Continuous Random Variables

Let c be a constant. Then the following hold:

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

Notes

Example

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

1. $\mathrm{E}[2X + 1]$
2. $\mathrm{E}[X - Y]$
3. $\mathrm{E}[X^2]$
4. $\mathrm{E}[X^2 - 4]$
5. $\mathrm{E}[(X - 4)^2]$
6. $\mathrm{Var}(2X - 4)$

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: $0, 1, \cdots, n$
- Its parameter(s) and definition(s): p: the probability of success on 1 trial; n is the sample size
- The probability mass function (pmf):

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

- The expected value:

$$\mathrm{E}[X] = np$$

- The variance:

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

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.

$R \sim Binomial(n = 10, p = \frac{1}{4} = .25)$
$\mathrm{E}[R] = n \times p = 2.5$
$\mathrm{P}(X \geq 8) = .000416$

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.

1. What is the probability that X is at least 1?
2. What is the probability that X is at most 3?

Notes

Binomial and Hypergeometric r.v.s

Probability IV
CLEMSON UNIVERSITY
Binomial Random Variables
Hypergeometric Random Variable
Continuous Random Variables

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.

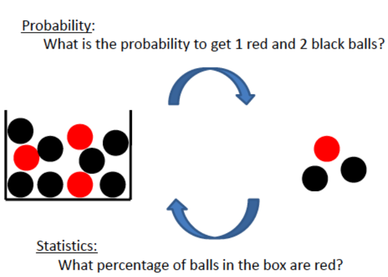
$\Rightarrow$ Hypergeometric Distribution

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 IV
CLEMSON UNIVERSITY
Binomial Random Variables
Hypergeometric Random Variable
Continuous Random Variables



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: $k \in \{\max(0, n+K-N), \cdots, \min(n,K)\}$
- 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: $\mathrm{E}[X] = n\frac{K}{N}$
- The variance: $\mathrm{Var}(X) = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$

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.

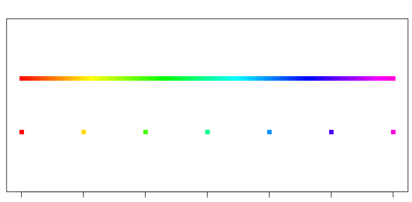
Let D be the number of defective TVs in the sample.

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

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

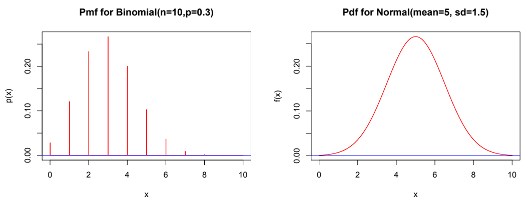
Notes

From Discrete to Continuous Random Variables



Notes

Probability Mass Functions vs. Probability Density Functions



Remarks:

- pmf assigns probabilities to each possible values of a discrete random variable
- pdf describes the relative likelihood for a continuous random variable to take on a given **interval**

Notes

Probability Mass Functions v.s. Probability Density Functions cont'd

Recall the properties of discrete probability mass functions (Pmfs):

- $0 \leq p_X(x) \leq 1$ for all possible values of x
- $\sum_x p_X(x) = 1$
- $\mathbb{P}(a \leq X \leq b) = \sum_{x=a}^{x=b} p_X(x)$

For continuous distributions, the properties for probability density functions (Pdfs) are similar:

- $f_X(x) \geq 0$ for all possible values of x
- $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
- $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\, dx$

Notes

Cumulative Distribution Functions (cdfs) for Continuous r.v.s

- The cdf $F_X(x)$ is defined as $F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(x)\, dx$
- we use cdf to calculate probabilities of a continuous random variable within an interval, i.e. $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\, dx =$ $\int_{-\infty}^{b} f_X(x)\, dx - \int_{-\infty}^{a} f_X(x)\, dx = \boxed{F_X(b) - F_X(a)}$

Remark: $\mathbb{P}(X = x) = \int_x^x f_X(x)\, dx = 0$ for all possible values of x

Notes

Expected Value and Variance

Recall the expected value formula for the discrete random variable: $\mathbb{E}[X] = \sum_x x p_X(x)$
For continuous random variables, we have similar formulas:
Let a, b, and c are constant real numbers

- $\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$
- $\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$
- $\mathrm{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathrm{E}[cX] = c\mathbb{E}[X]$
- $\mathrm{E}[aX+bY] = a\mathrm{E}[X] + b\mathrm{E}[Y]$
- $\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - \left(\int_{-\infty}^{\infty} x f_X(x)\,dx\right)^2$
- $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$
- $\mathrm{Var}(X-c) = \mathrm{Var}(X)$

Notes

Example

Let X represent the diameter in inches of a circular disk cut by a machine. Let $f_X(x) = c(4x - x^2)$ for $1 \le x \le 4$ and be 0 otherwise. Answer the following questions:

1. Find the value of c that makes this a valid pdf
2. Find the expected value and variance of X
3. What is the probability that X is within $.5$ inches of the expected diameter?
4. Find $F_X(x)$

Notes

Notes