Lecture 7

Binomial, Hypergeometric R.V.s & Continuous Random Variables

Text: Chapter 4

STAT 8010 Statistical Methods I
January 30, 2020

Whitney Huang
Clemson University

Notes

Agenda

1. Binomial Random Variables
2. Hypergeometric Random Variable
3. Continuous Random Variables
4. Normal Distributions

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:

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

- The variance:

$$Var(X) = np(1-p)$$

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.

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

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

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

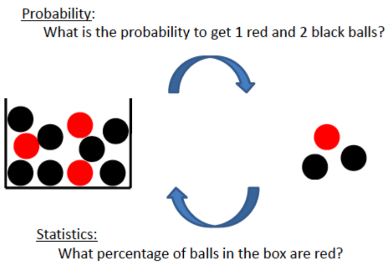
We want a distribution to describe the probability of k successes in n trials without replacement from a finite population of size N containing exactly K successes.

$\Rightarrow$ 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.



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: $\mathbb{E}[X] = n\frac{K}{N}$
- The variance: $Var(X) = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$

Binomial, Hypergeometric R.V.s & Continuous Random Variables
CLEMSON UNIVERSITY
Binomial Random Variables
Hypergeometric Random Variable
Continuous Random Variables
Normal Distributions

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.

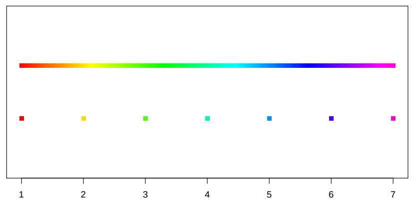
$D \sim Hyp(N = 100, n = 8, K = 10)$

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

Binomial, Hypergeometric R.V.s & Continuous Random Variables
CLEMSON UNIVERSITY
Binomial Random Variables
Hypergeometric Random Variable
Continuous Random Variables
Normal Distributions

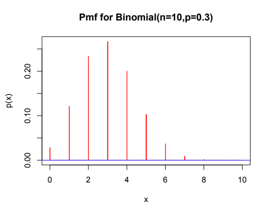
Notes

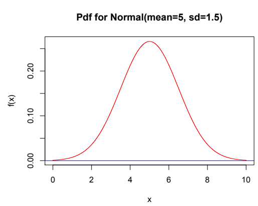
From Discrete to Continuous Random Variables



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):

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

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

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

Notes

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

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

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

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

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

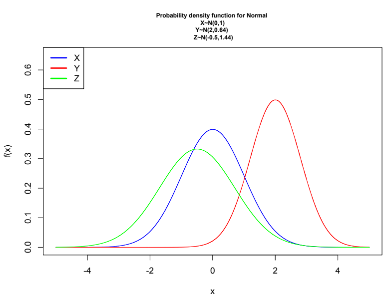
Normal Distribution

Characteristics of the Normal random variable:
Let X be a Normal r.v.

- The support for X: $(-\infty, \infty)$
- Its parameter(s) and definition(s): μ : mean and σ^2 : variance
- The probability density function (pdf): $\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ for $-\infty < x < \infty$
- The cumulative distribution function (cdf): No explicit form, look at the value $\Phi(\frac{x-\mu}{\sigma})$ for $-\infty < x < \infty$ from standard normal table
- The expected value: $\mathbb{E}[X] = \mu$
- The variance: $Var(X) = \sigma^2$

Notes

Normal Density Curves



- The parameter μ determines the center of the distribution
- The parameter σ^2 determines the spread of the distribution
- Also called bell-shaped distribution

Notes

Standard Normal $Z \sim N(\mu = 0, \sigma^2 = 1)$

- Normal random variable X with mean μ and standard deviation σ can convert to standard normal Z by the following :
$$Z = \frac{X-\mu}{\sigma}$$
- The cdf of the standard normal, denoted by $\Phi(z)$, can be found from the standard normal table
- The probability $\mathbb{P}(a \le X \le b)$ where $X \sim N(\mu, \sigma^2)$ can be compute
$$\mathbb{P}(a \le X \le b) = \mathbb{P}(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma})$$
$$= \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$$

Notes

Properties of Φ

- $\Phi(0) = .50 \Rightarrow$ Mean and Median (50_{th} percentile) for standard normal are both 0
- $\Phi(-z) = 1 - \Phi(z)$
- $\mathbb{P}(Z > z) = 1 - \Phi(z) = \Phi(-z)$

Notes

Example

Let us examine Z. Find the following probabilities with respect to Z:

1. Z is at most -1.75
2. Z is between -2 and 2 inclusive
3. Z is less than $.5$

Notes

Example Cont'd

Solution.

1. $\mathbb{P}(Z \leq -1.75) = \Phi(-1.75) = .0401$
2. $\mathbb{P}(-2 \leq Z \leq 2) = \Phi(2) - \Phi(-2) = .9772 - .0228 = .9544$
3. $\mathbb{P}(Z < .5) = \Phi(.5) = .6915$

Notes

Sums of Normal Random Variables

If X_i $1 \leq i \leq n$ are independent normal random variables with mean μ_i are variance σ_i^2, respectively.

- Let $S_n = \sum_{i=1}^n X_i$ then $S_n \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$
- This can be applied for any integer n

Notes

Example

Let X_1, X_2, and X_3 be mutually independent, Normal random variables. Let their means and standard deviations be $3k$ and k for $k = 1$, 2, and 3 respectively. Find the following distributions:

1. $\sum_{i=1}^3 X_i$
2. $X_1 + 2X_2 - 3X_3$
3. $X_1 + 5X_3$

Notes

Example Cont'd

Solution.

1. $\sum_{i=1}^3 X_i \sim N(\mu = 3+6+9 = 15, \sigma^2 = 1^2+2^2+3^2 = 14)$
2. $X_1 + 2X_2 - 3X_3 \sim N(\mu = 3 + 12 - 27 = -12, \sigma^2 = 1^2 + 4 \times 2^2 + 9 \times 3^2 = 98)$
3. $X_1 + 5X_3 \sim N(\mu = 3+45 = 48, \sigma^2 = 1^2+25 \times 3^2 = 226)$

Notes