Lecture 9

Probability V

Text: Chapter 4

STAT 8010 Statistical Methods I
September 17, 2020

Whitney Huang
Clemson University

Notes

Agenda

1. **Binomial and Hypergeometric Random Variables**
2. **Continuous Random Variables**
3. **Normal Distributions**

Notes

Review: 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 a 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.

$$R \sim Binomial(n = 10, p = \frac{1}{4} = .25; \mathrm{E}[R] = n \times p = 2.5$$

$$\begin{aligned}
\mathrm{P}(X \geq 8) &= \mathrm{P}(X = 8) + \mathrm{P}(X = 9) + \mathrm{P}(X = 10) \\
&= \binom{10}{8}(0.25)^8(1-0.25)^2 + \binom{10}{9}(0.25)^9(1-0.25)^1 + \binom{10}{10}(0.25)^{10}(1-0.25)^0 \\
&= .000416
\end{aligned}$$

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.

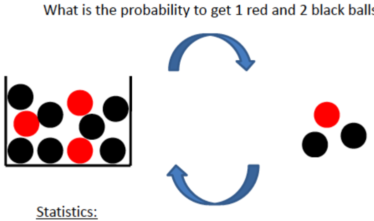
$\Rightarrow$ Hypergeometric Distribution

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.

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



Statistics:
What percentage of balls in the box are red?

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.

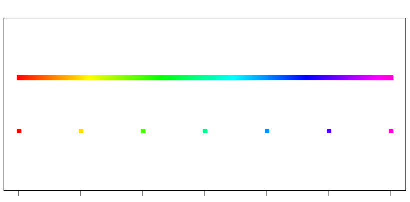
Let D be the number of defective TVs in the sample.

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

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

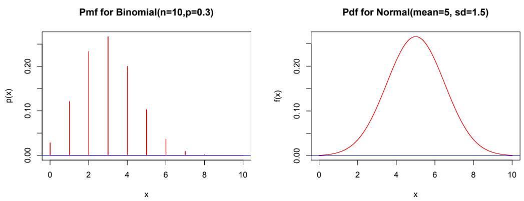
Notes

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):

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

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

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

Notes

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

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

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

Notes

Expected Value and Variance

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

- $\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$
- $\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$
- $\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - \left(\int_{-\infty}^{\infty} x f_X(x)\, dx\right)^2$

Notes

Example

Probability V
CLEMSON UNIVERSITY
Binomial and Hypergeometric Random Variables
Continuous Random Variables
Normal Distributions

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

Example Cont'd

Probability V
CLEMSON UNIVERSITY
Binomial and Hypergeometric Random Variables
Continuous Random Variables
Normal Distributions

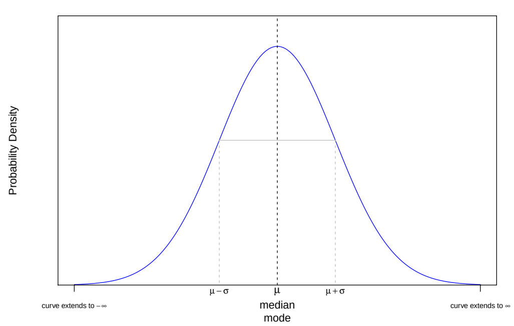
1. $\int_1^4 f_X(x)\, dx = \int_1^4 c(4x - x^2)\, dx = 9c = 1 \Rightarrow c = \frac{1}{9}$
2. $\mathrm{E}[X] = \int_1^4 x f_X(x)\, dx = 2.25$
   $\mathrm{E}[X^2] = \int_1^4 x^2 f_X(x)\, dx = 5.6$
   $\Rightarrow Var(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = .5375$
3. $\mathrm{P}(1.75 \le X \le 2.75) = \int_{1.75}^{2.75} f_X(x)\, dx = .4282$
4. $F_X(x) = \int f_X(x)\, dx = \frac{1}{9}(2x^2 - \frac{x^3}{3})$ for $1 \le X \le 4$ and 0 if $x < 1$ and 1 if $x > 4$

Notes

Normal Distributions

Notes

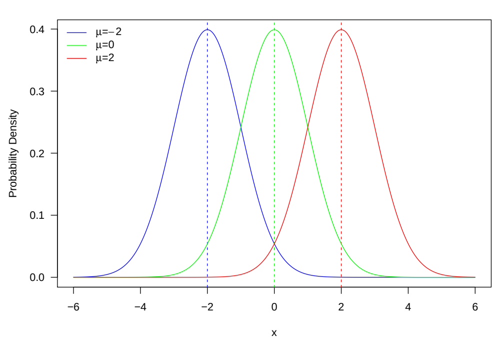
Probability Density Curve for Normal Random Variable



Notes

Normal Density Curves

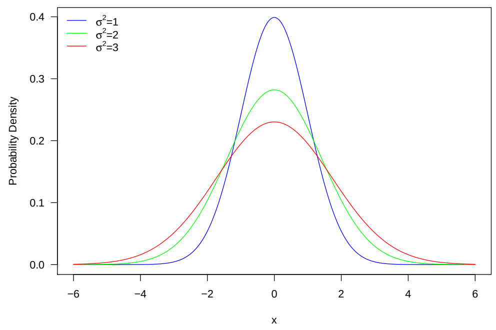
Different μ but same σ^2



Notes

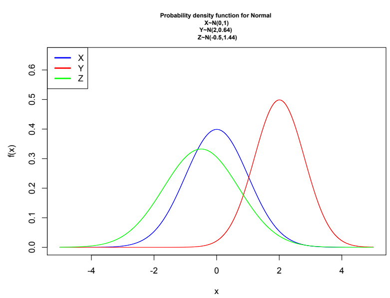
Normal Density Curves Cont'd

Same μ but different σ^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

Characteristics of Normal Random Variables

Let X be a Normal r.v.

- The support for X: $(-\infty, \infty)$
- Parameters: 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

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

- Normal random variable X with mean μ and standard deviation σ can be converted to standard normal Z by the following :

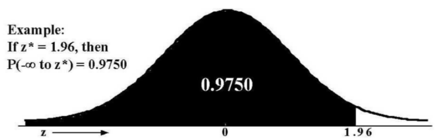
$$Z = \frac{X - \mu}{\sigma}$$

- The cdf of the standard normal, denoted by $\Phi(z)$, can be found from the standard normal table
- The probability $\mathbb{P}(a \leq X \leq b)$ where $X \sim N(\mu, \sigma^2)$ can be computed

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma})$$
$$= \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$$

Notes

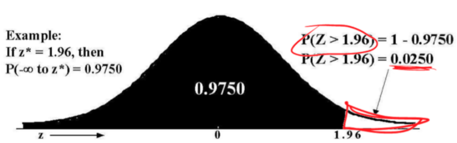
Standard Normal (Z) Table



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	**0.9750**	0.9756	0.9761	0.9767

Notes

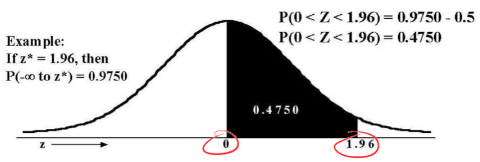
Standard Normal (Z) Table Cont'd



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	**0.9750**	0.9756	0.9761	0.9767

Notes

Standard Normal (Z) Table Cont'd



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	**0.9750**	0.9756	0.9761	0.9767

Notes

Properties of Φ

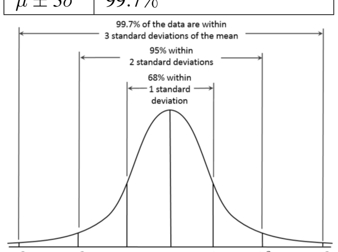
- $\Phi(0) = .50 \Rightarrow$ Mean and Median (50_{th} percentile) for standard normal are both 0
- $\Phi(-z) = 1 - \Phi(z)$
- $\mathbb{P}(Z > z) = 1 - \Phi(z) = \Phi(-z)$

Notes

The Empirical Rules

The Empirical Rules provide a quick way to approximate certain probabilities for the Normal Distribution as the following table:

Interval	Percentage with interval
$\mu \pm \sigma$	68%
$\mu \pm 2\sigma$	95%
$\mu \pm 3\sigma$	99.7%



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. $\mathbb{P}(Z < .5) = \Phi(.5) = .6915$

Probability V
CLEMSON UNIVERSITY
Binomial and Hypergeometric Random Variables
Continuous Random Variables
Normal Distributions

Notes

Example

Suppose a STAT-8020 exam score follows a normal distribution with mean 78 and variance 36. Let X to denote the exam score, answer the following questions:

1. What is the probability that a randomly chosen test taker got a score greater than 84?
2. Suppose the passing score for this exam is 75. What is the probability that a randomly chosen test taker got a score greater than 84 given that she/he pass the exam?
3. Using the empirical rule to find the 84_{th} percentile.

Probability V
CLEMSON UNIVERSITY
Binomial and Hypergeometric Random Variables
Continuous Random Variables
Normal Distributions

Notes

Example

Find the following percentile with respect to Z

1. 10_{th} percentile
2. 55_{th} percentile
3. 90_{th} percentile

Notes

Example Cont'd

Solution.

1. $Z_{10} = -1.28$
2. $Z_{55} = 0.13$
3. $Z_{90} = 1.28$

Notes

Example

Let X be Normal with a mean of 20 and a variance of 49. Find the following probabilities and percentile:

1. X is between 15 and 23
2. X is more than 30
3. X is more than 12 knowing it is less than 20
4. What is the value that is smaller than 20% of the distribution?

Notes

Example Cont'd

Solution.

1. $\mathbb{P}(15 \le X \le 23) = \Phi(\frac{15-20}{7}) - \Phi(\frac{23-20}{7}) = \Phi(0.43) - \Phi(-0.71) = .6664 - .2389 = .4275$
2. $\mathbb{P}(X > 30) = 1 - \mathbb{P}(X \le 30) = 1 - \Phi(\frac{30-20}{7}) = 1 - .9236 = .0764$
3. $\mathbb{P}(X > 12 | X < 20) = \frac{\mathbb{P}(12<X<20)}{\mathbb{P}(X<20)} = \frac{\Phi(0)-\Phi(-1.14)}{\Phi(0)} = .7458$
4. $Z_{80} = 0.84 \Rightarrow X_{80} = \mu + Z_{80} \times \sigma = 20 + 0.84 \times \sqrt{49} = 25.88$

Notes

Notes

Notes