POSTS AND TELECOMMUNICATIONS INSTITUTE OF TECHNOLOGY

———————o0o———————

Chapter 1: Introduction to Probability

PROBABILITY AND STATISTICS

Department of Mathematics, Faculty of Fundamental Science 1

Hanoi - 2023

1. 1.1 Random Experiment and Events
2. 1.2 Definitions and Properties of Probability
3. 1.3 Conditional Probability
4. 1.4 Sequence of Bernoulli Trials

1 1.1 Random Experiment and Events

2 1.2 Definitions and Properties of Probability

3 1.3 Conditional Probability

4 1.4 Sequence of Bernoulli Trials

Random Experiment

A random experiment is an action or process that leads to one of several possible outcomes.

Example 1:

- Experiment: Toss a coin.
  Outcomes: Heads and tails.
- Experiment: Measure the time to assemble a computer.
  Outcomes: Positive numbers.

Sample Spaces

- The set of all possible outcomes of a random experiment is called the sample space of the experiment, denoted as S.
- A sample space is often defined based on the objectives of the study.

Example 2: Tossing a die. The sample space is

$$S = \{1, 2, 3, 4, 5, 6\}.$$

Example 3: Tossing a coin three times. If the objective of the study is to consider whether the coin is heads or tails, the sample space is

$$S = \{\text{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}\}.$$

If the objective of the study is to consider the number of heads that appear, the sample space is

$$S = \{0, 1, 2, 3\}.$$

Events

- An event is a subset of the sample space of a random experiment.
- Impossible event: $\varnothing$

Basic Set Operations

- The union of the two events A and B, denoted by $A \cup B$, is the event that consists of all outcomes that are in A or B.
- The intersection of the two events A and B, denoted by $A \cap B$, is the event that consists of all outcomes that are common to A and B.
- The complement of the event A in the sample space S, denoted by A' or $\overline{A}$, is the set of outcomes in S that are not in A.

Example 4: Let $A_i, i = 1, 2, 3$, denote the event that component i is working. Express in terms of A_1, A_2, A_3 the following events:

1) Only component 2 is working.
2) All three components are working.
3) None is working.
4) At least one is working.
5) Exactly two are working.

Mutually Exclusive Events

- Two events A and B are called to be mutually exclusive if

$$A \cap B = \varnothing$$

- A collection of events $E_1, E_2, \ldots, E_k$ is said to be mutually exclusive if for all pairs,

$$E_i \cap E_j = \varnothing$$

1. 1.1 Random Experiment and Events
2. 1.2 Definitions and Properties of Probability
3. 1.3 Conditional Probability
4. 1.4 Sequence of Bernoulli Trials

1.2.1 Classical Definition of Probability

- Probability is used to quantify the chance that an outcome of a random experiment will occur.
- For a sample space consists of N possible outcomes that are equally likely, the probability of an event A, denoted as $P(A)$, is defined by

$$P(A) = \frac{\text{The number of outcomes in } A}{\text{The number of possible outcomes}} = \frac{n(A)}{N}$$

Example 5: A deck of playing cards is thoroughly shuffled and a card is drawn from the deck.

a) What is the probability that the card drawn is the ace of diamonds?
b) What is the probability that the card drawn is a ten?

Example 6: A message can follow different paths through servers on a network. The sender's message can go to one of five servers for the first step, each of them can send to five servers at the second step, each of which can send to four servers at the third step, and then the message goes to the recipient's server.

(a) How many paths are possible?

(b) If all paths are equally likely, what is the probability that a message passes through the first of four servers at the third step?

1.2.2 Relative Frequency Definition of Probability

- Consider a sequence of repetitions of the same experiment under identical conditions. Let n_A denote the number of occurrences of an event A. The ratio $f_n(A) = \dfrac{n_A}{n}$ is called the relative frequency of occurrences of the event A.
- The probability of A is defined by

$$P(A) = \lim_{n\to\infty} f_n(A).$$

1.2.3 Properties of Probability

Basic Properties of Probability

- $0 \leq P(A) \leq 1$
- $P(\varnothing) = 0, P(S) = 1$

Addition Rules

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B)$$
$$- P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Remark:

- If A and B are mutually exclusive events, then

$$P(A \cup B) = P(A) + P(B)$$

- If $E_1, E_2, \dots, E_k$ are mutually exclusive events, then

$$P(E_1 \cup E_2 \cup \dots \cup E_k) = P(E_1) + P(E_2) + \dots + P(E_k)$$

- $P(A) + P(A') = 1$

Example 7: The probability that a person stopping at a gas station will ask to have the tires checked is 0.12, the probability that he or she will ask to have the oil checked is 0.29, and the probability that he or she will ask to have them both checked is 0.07.

a) What is the probability that a person stopping at this gas station will have the tires or the oil checked?

b) What is the probability that a person stopping at this gas station will have neither the tires nor the oil checked?

1. 1.1 Random Experiment and Events
2. 1.2 Definitions and Properties of Probability
3. 1.3 Conditional Probability
4. 1.4 Sequence of Bernoulli Trials

1.3.1 Conditional Probability

Definition

The conditional probability of B given A, denoted as $P(B|A)$, is the probability that the event B occurs given that the event A has already occurred.

Example 8: A lot of 100 semiconductor chips contains 20 that are defective. Two are selected randomly, without replacement, from the lot.

a) What is the probability that the second one selected is defective given that the first one was defective?

b) What is the probability that both are defective?

Formula for Conditional Probability

If $P(A) > 0$, then

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Example 9: It is thought that 30 % of all people in the United States are obese, that 3 % suffer from diabetes, and 31 % are obese or suffer from diabetes. What is the probability that a randomly selected person

a) Have both obese and diabetes?

b) Is diabetic given that he/she is obese?

c) Is diabetic but is not obese?

d) Is diabetic given that he/she is not obese?

Multiplication Rule

- $P(AB) = P(A)P(B|A)$
- $P(A_1A_2\ldots A_n) = P(A_1)P(A_2|A_1)\ldots P(A_n|A_1A_2\ldots A_{n-1})$

Example 10: Suppose that $P(A|B) = 0.4$ and $P(B) = 0.5$. Determine the following:

(a) $P(A \cap B)$

(b) $P(A' \cap B)$

Example 11: Suppose that 29% of Internet users download music files, and 67% of downloaders say they do not care if the music is copyrighted. What is the probability that an Internet user downloads music and does not care about copyright?

Independence (two events)

Two events A and B are independent if any one of the following equivalent statements is true:

(1) $P(A|B) = P(A)$

(2) $P(B|A) = P(B)$

(3) $P(A \cap B) = P(A)P(B)$

Remark:

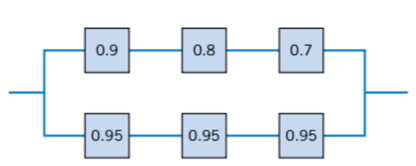
- Two events are independent if the probability of one event is not affected by the occurrence of the other event.
- If A and B are independent events, then so are events A and B', events A' and B, and events A' and B'.

Independence (multiple events)

The events $E_1, E_2, \ldots, E_n$ are independent if and only if for any subset of these events $E_{i_1}, E_{i_2}, \ldots, E_{i_k}$,

$$P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_k}) = P(E_{i_1})P(E_{i_2})\ldots P(E_{i_k})$$

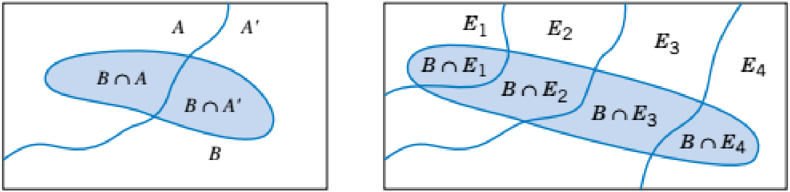
Example 12: The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the probability that a device is functional does not depend on whether or not other devices are functional. What is the probability that the circuit operates?



Total Probability Rule (two event)

For any events A and B,

$$P(B) = P(A)P(B|A) + P(A')P(B|A')$$



Example 13: The probability is 1% that an electrical connector that is kept dry fails during the warranty period of a portable computer. If the connector is ever wet, the probability of a failure during the warranty period is 5%. If 90% of the connectors are kept dry and 10% are wet, what proportion of connectors fail during the warranty period?

Definition

A collection of events $E_1, E_2, \ldots, E_k$ is said to be exhaustive if

$$E_1 \cup E_2 \cup \ldots \cup E_k = S$$

Total Probability Rule (multiple events)

If $E_1, E_2, \ldots, E_k$ are k mutually exclusive and exhaustive events, then

$$P(B) = P(E_1)P(B|E_1) + P(E_2)P(B|E_2) + \ldots + P(E_k)P(B|E_k)$$

Example 14: The edge roughness of slit paper products increases as knife blades wear. Only 1% of products slit with new blades have rough edges, 3% of products slit with blades of average sharpness exhibit roughness, and 5% of products slit with worn blades exhibit roughness. If 25% of the blades in manufacturing are new, 60% are of average sharpness, and 15% are worn, what is the proportion of products that exhibit edge roughness?

1.3.4 Bayes' Theorem

Bayes' Theorem

If $E_1, E_2, \ldots, E_k$ are mutually exclusive and exhaustive events and B is any event, then

$$P(E_i|B) = \frac{P(E_i)P(B|E_i)}{P(E_1)P(B|E_1) + P(E_2)P(B|E_2) + \ldots + P(E_k)P(B|E_k)}$$

for $P(B) > 0$.

Remark: $P(E_i|B) = \dfrac{P(E_i)P(B|E_i)}{P(B)}$

Example 15: Customers are used to evaluate preliminary product designs. In the past, 95% of highly successful products received good reviews, 60% of moderately successful products received good reviews, and 10% of poor products received good reviews. In addition, 40% of products have been highly successful, 35% have been moderately successful, and 25% have been poor products.

(a) What is the probability that a product attains a good review.

(b) If a new design attains a good review, what is the probability that it will be a highly successful product?

1. 1.1 Random Experiment and Events
2. 1.2 Definitions and Properties of Probability
3. 1.3 Conditional Probability
4. 1.4 Sequence of Bernoulli Trials

Sequence of Bernoulli Trials

A sequence of Bernoulli trials is a sequence of independent trials, repeated under identical conditions, where each trial has two possible outcomes, labeled as "success" and "failure".

Bernoulli Formula

Consider a sequence of n Bernoulli trials. If the probability of a success in each trial is p $(0 < p < 1)$, then the probability of exactly k successes in n trials is

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n,$$

Example 16: The probability that a lab specimen contains high levels of contamination is 0.1. Five samples are checked, and the samples are independent. What is the probability that

(a) none contains high levels of contamination?

(b) exactly one contains high levels of contamination?

(c) at least one contains high levels of contamination?

Most Likely Number of Successes

Consider a sequence of n Bernoulli trials with a success probability p. The most likely number of successes is the integer m satisfying

$$(n+1)p - 1 \leq m \leq (n+1)p.$$

Example 17: A multiple choice test contains 50 questions, each with four answer. Assume a student just guesses at each question. Find the most likely number of questions answered correctly.

POSTS AND TELECOMMUNICATIONS INSTITUTE OF TECHNOLOGY

——————o0o——————

Chapter 2: Random Variables

PROBABILITY AND STATISTICS

Department of Mathematics, Faculty of Fundamental Science 1

Hanoi - 2023

Chapter 2: Random Variables

1. 2.1 Probability Distribution of Random Variables
2. 2.2 Characteristic Parameters of Random Variables
3. 2.3 Some Common Probability Distributions

1. 2.1 Probability Distribution of Random Variables
2. 2.2 Characteristic Parameters of Random Variables
3. 2.3 Some Common Probability Distributions

Definition

A random variable is a function that assigns a real number to each outcome in the sample space of a random experiment.

A random variable is denoted by an uppercase letter such as X.

Discrete and Continuous Random Variables

- A discrete random variable is a random variable with a finite (or countably infinite) range.
- A continuous random variable is a random variable with an interval of real numbers for its range.

Example 1: A group of 10,000 people are tested for a gene called Ifi202 that has been found to increase the risk for lupus. Let X be the number of people who carry the gene $\Rightarrow X$ is a discrete random variable, X can take on the values 0, 1, 2, ..., 10000.

Example 2: Let Y be the number of surface flaws in a large coil of galvanized steel $\Rightarrow Y$ is a discrete random variable, Y can take on the values 0, 1, 2, ...

Example 3: Let Z be the outside diameter of a machined shaft $\Rightarrow Z$ is a continuous random variable.

Independent Random Variables

- Two discrete random variables X and Y are called independent if for all x, y,

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

- Two continuous random variables X and Y are called independent if for all x, y,

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y).$$

2.1.2 Probability Distributions and Probability Mass Function

A probability distribution is a table, formula, or graph that describes the values of a random variable and the probability associated with these values.

Discrete Probability Distribution

For a discrete random variable, the distribution is often specified by just a list of the possible values along with the probability of occurrence of each value.

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

Remark:

- $p_i = P(X = x_i)\ \forall i = 1, \ldots, n.$
- $p_1 + p_2 + \ldots + p_n = 1.$

Example 4: A shipment of 10 similar computers to a retail outlet contains 3 that are defective. A school makes a random purchase of 2 of these computers. Let X be the number of defective computers. Find the probability distribution of the random variable X.

Probability Mass Function

The probability mass function of a discrete random variable X, denoted as $p_X(x)$, is a function defined by

$$p_X(x) = P(X = x), x \in \mathbb{R}.$$

Properties of the Probability Mass Function

If X is a discrete random variable with possible values $x_1, x_2, \ldots, x_n$, then

(1) $p_X(x_i) > 0 \ \forall i = 1, 2, \ldots, n$

(2) $p_X(x_1) + p_X(x_2) + \ldots + p_X(x_n) = 1$

Example 5: An assembly consists of two mechanical components. Suppose that the probabilities that the first and second components meet specifications are 0.95 and 0.98. Assume that the components are independent. Determine the probability mass function of the number of components in the assembly that meet specifications.

Definition

The cumulative distribution function of a random variable X, denoted as $F_X(x)$, is a function defined by

$$F_X(x) = P(X \leq x).$$

Example 6: Determine the cumulative distribution function of the random variable in Example 4.

Remark: If X be a discrete random variable with possible values $x_1, x_2, \ldots, x_n$, then

$$F_X(x) = \begin{cases} 0 & \text{if } x < x_1, \\ p_X(x_1) & \text{if } x_1 \le x < x_2, \\ \ldots & \\ p_X(x_1) + \ldots + p_X(x_{i-1}) & \text{if } x_{i-1} \le x < x_i, \\ \ldots & \\ 1 & \text{if } x \ge x_n \end{cases}$$

Properties of the Cumulative Distribution Function

1) $0 \le F_X(x) \le 1$
2) $F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0,$
   $F_X(+\infty) = \lim_{x \to +\infty} F_X(x) = 1$
3) $P(a < X \le b) = F_X(b) - F_X(a)$
4) $F_X(x)$ is a nondecreasing function.
5) $F_X(x)$ is a continuous function from the right.

2.1.4 Probability Density Functions

Definition

Let X be a continuous random variable with the cumulative distribution function $F_X(x)$. If there exists a nonnegative function $f_X(x)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \ \forall x \in \mathbb{R}$$

then $f_X(x)$ is called the probability density function of X.

Properties of the Probability Density Function

1) $f_X(x) \geq 0, \forall x \in \mathbb{R}$
2) $f_X(x) = F'_X(x)$ if the derivative exists.
3) $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b)$
   $= P(a < X < b) = \int_a^b f_X(x)dx$
4) $\int_{-\infty}^{+\infty} f_X(x)dx = 1$

Example 7: Let X be a random variable with the probability density function

$$f_X(x) = \begin{cases} 0 & \text{if } x \leq 2 \\ \dfrac{k}{x^2} & \text{if } x > 2 \end{cases}$$

a) Determine k.

b) Find the cumulative distribution function $F_X(x)$.

c) Calculate $P(1 < X \leq 10 | X > 4)$.

Example 8: The probability density function of the time to failure of an electronic component in a copier (in hours) is

$$f_X(x) = e^{-x/1000}/1000 \text{ for } x > 0.$$

Determine the probability that

(a) A component lasts more than 3000 hours before failure.

(b) A component fails in the interval from 1000 to 2000 hours.

(c) A component fails before 1000 hours.

(d) Determine the number of hours at which 10% of all components have failed.

1. 2.1 Probability Distribution of Random Variables
2. 2.2 Characteristic Parameters of Random Variables
3. 2.3 Some Common Probability Distributions

2.2.1 Mean of a Random Variable

Definition

The mean or expected value of a random variable X, denoted by $E(X)$, is defined as follows:

- If X a discrete random variable with possible values $x_1, x_2, \ldots, x_n$, then
$$E(X) = \sum_{i=1}^{n} x_i p_X(x_i).$$
- If X a discrete random variable with possible values $x_1, x_2, \ldots$, then
$$E(X) = \sum_{i=1}^{\infty} x_i p_X(x_i),$$
provided that the series $\sum_{i=1}^{\infty} |x_i| p_X(x_i)$ is convergent.

- If X is a continuous random variable with probability density function $f_X(x)$, then

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx,$$

provided that the integral $\int_{-\infty}^{\infty} |x| f_X(x) dx$ is convergent.

Example 9: The number of pizzas delivered to university students each month is a random variable X with the following probability distribution.

X	0	1	2	3
P	0.1	0.3	0.4	0.2

Determine the mean of X.

Properties of the Mean

1) $E(C) = C$, where C is a constant.
2) $E(kX) = kE(X)$, where k is a constant.
3) $E(X + Y) = E(X) + E(Y)$.
4) $E(k_1X_1 + k_2X_2 + \ldots + k_nX_n) = k_1E(X_1) + k_2E(X_2) + \ldots + k_nE(X_n)$, where $k_1, k_2, \ldots, k_n$ are constants.
5) If X, Y are independent random variables, then

$$E(XY) = E(X)E(Y).$$

Mean of a Function of a Random Variable

Let $h(x)$ be a function.

- If X is a discrete random variable with possible values $x_1, x_2, \ldots, x_n$, then

$$E[h(X)] = \sum_{i=1}^{n} h(x_i) p_X(x_i).$$

- If X is a continuous random variable with the probability density function $f_X(x)$, then

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x) dx$$

Example 10: The probability density function of the weight of packages delivered by a post office is

$$f(x) = 70/(69x^2) \text{ for } 1 < x < 70 \text{ pounds}$$

(a) Determine the mean of the weight of packages.

(b) If the shipping cost is \$2.5 per pound, what is the average shipping cost of a package?

2.2.2 Variance of a Random Variable

Definition

The variance of a random variable X, denoted as $V(X)$, is

$$V(X) = E[(X - E(X))^2].$$

Remark:

- The variance of a random variable X is a measure of dispersion or scatter in the possible values for X.
- $V(X) = E(X^2) - [E(X)]^2$.

Example 11: Let X be a random variable with the following probability distribution

X	0	1	2	3
P	0,24	0,46	0,26	0,04

Find the variance of X.

Properties of the Variance

1) $V(C) = 0$, where C is a constant.
2) $V(kX) = k^2 V(X)$, where k is a constant.
3) If X, Y are independent random variables, then

$$V(X + Y) = V(X) + V(Y).$$

4) If $X_1, X_2, \ldots, X_n$ are independent random variables and $k_1, k_2, \ldots, k_n$ are constants, then

$$V(k_1X_1 + k_2X_2 + \ldots + k_nX_n) = k_1^2 V(X_1) + k_2^2 V(X_2) + \ldots + k_n^2 V(X_n).$$

2.2.3 Standard Deviation of a Random Variable

Definition

The standard deviation of a random variable X is

$$\sigma(X) = \sqrt{V(X)}.$$

1 2.1 Probability Distribution of Random Variables

2 2.2 Characteristic Parameters of Random Variables

3 2.3 Some Common Probability Distributions

Definition

A random variable X is said to have a binomial distribution with parameters $n \in \mathbb{N}^*$ and $p \in (0, 1)$, denoted by $X \sim B(n, p)$, if X can take on the values $0, 1, \ldots, n$ and

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \; k = 0, 1, \ldots, n.$$

Remark: If X is the number of successes in a sequence of n Bernoulli trials, then $X \sim B(n, p)$, where p is the probability of a success in each trial.

Example 12:

- Toss a coin 5 times. Let X be the number of heads obtained $\Rightarrow X \sim B(5, 0.5)$.
- A multiple choice test contains 50 questions, each with four choices, and you guess at each question. Let X be the number of questions answered correctly $\Rightarrow X \sim B(50, 1/4)$.
- Each sample of air has a 10% chance of containing a particular rare molecule. Let X be the number of air samples that contain the rare molecule in the next 18 samples analyzed $\Rightarrow X \sim B(18, 0.1)$.

Example 13: The on-line access computer service industry is growing at an extraordinary rate. Current estimates suggest that 20% of people with home-based computers have access to on-line services. Suppose that 15 people with home-based computers were randomly and independently sampled.

a) What is the probability that two of those sampled have access to on-line services at home?

b) What is the probability that at least 1 of those sampled have access to on-line services at home?

Mean and Variance

If X is a binomial random variable with parameters n and p, then

$$E(X) = np$$
$$V(X) = np(1-p)$$

Example 14: A computer system uses passwords that are exactly six characters and each character is one of the 26 letters (a-z) or 10 integers (0-9). Suppose there are 10000 users on the system with unique passwords. A hacker randomly selects (with replacement) one billion passwords from the potential set in the milliseconds before security software closes the unauthorized access.

(a) What is the distribution of the number of user passwords selected by the hacker?

(b) What is the probability that no user passwords are selected?

(c) What is the mean and variance of the number of user passwords selected?

2.3.2 Poisson Distribution

- The Binomial distribution describes the distribution of the number of successes in a sequence of n trials.
- The Poisson distribution focuses on the number of occurrences of an event in an interval of time or specific region of space.
- Here are several examples of Poisson random variables.
  1) The number of cars arriving at a service station in 1 hour.
  2) The number of flaws in a bolt of cloth.
  3) The number of accidents in 1 day at a road intersection.

Definition

A random variable X is said to have a Poisson distribution with parameter λ ($\lambda > 0$), denoted as $X \sim P(\lambda)$, if X can take the values $0, 1, 2, \ldots$ and

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \ldots.$$

Mean and Variance

If X is a Poisson random variable with parameter λ, then

$$E(X) = V(X) = \lambda.$$

Note:

- It is important to use consistent units in the calculation of probabilities involving Poisson random variables.
- Example of unit conversions: If the average number of accidents in 1 day at a road intersection is 0.5, then the average number of accidents in 1 week at that road intersection is 3.5.

Example 15: The number of messages sent to a computer bulletin board is a Poisson random variable with a mean of 5 messages per hour. What is the probability that

(a) 5 messages are received in 1 hour?

(b) 10 messages are received in 1.5 hours?

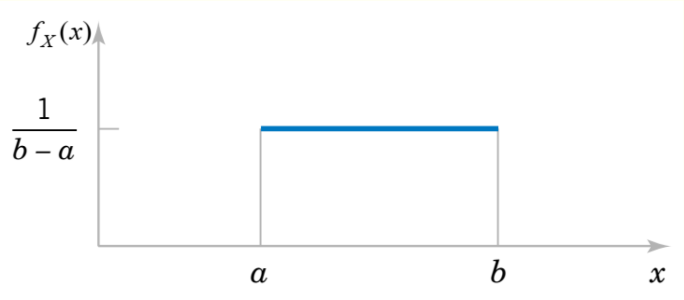
(c) Less than two messages are received in one-half hour?

2.3.3 Continuous Uniform Distribution

Definition

A random variable X is said to have a continuous uniform distribution over $[a;b]$, denoted by $X \sim U[a,b]$, if its probability density function is

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } x \in [a;b] \\ 0 & \text{elsewhere} \end{cases}$$



Mean and Variance

If X is a continuous uniform random variable over $[a, b]$, then

$$E(X) = \frac{a+b}{2},$$
$$V(X) = \frac{(b-a)^2}{12}.$$

Cumulative Distribution Function

If X is a continuous uniform random variable over $[a, b]$, then the cumulative distribution function of X is

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } x > b \end{cases}$$

Example 16: The thickness of a flange on an aircraft component is uniformly distributed between 0.95 and 1.05 millimeters.

(a) Determine the cumulative distribution function of flange thickness.

(b) Determine the proportion of flanges that exceeds 1.02 millimeters.

(c) What thickness is exceeded by 90% of the flanges?

(d) Determine the mean and variance of flange thickness.

Definition

A random variable X is said to have a normal distribution with parameters μ and σ^2 $(\sigma > 0)$, denoted as $X \sim N(\mu, \sigma^2)$, if its probability density function is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

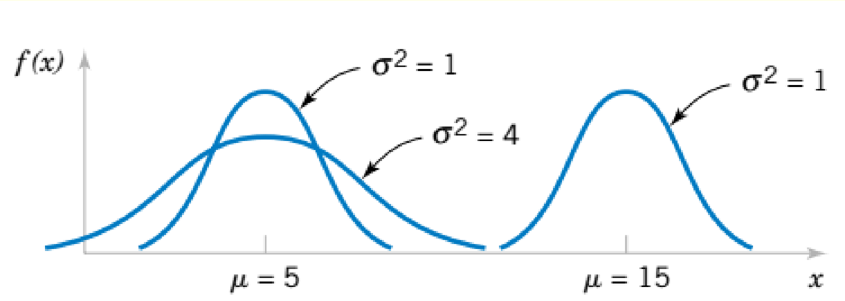


Figure 1: Normal probability density functions for selected values of the parameters μ and σ^2

Mean and Variance

If $X \sim N(\mu, \sigma^2)$, then $E(X) = \mu, V(X) = \sigma^2$.

Standard Normal Distribution

- A normal random variable with

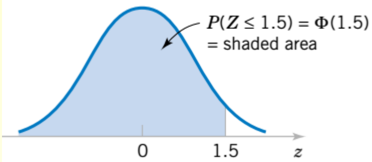
$$\mu = 0, \sigma = 1$$

is called a standard normal random variable, denoted as Z.

- The cumulative distribution function of Z is denoted as $\Phi(z)$,

$$\Phi(z) = P(Z \leq z).$$

- The values of $\Phi(z)$ are given in Appendix Table III.
- In Excel, $\Phi(z) = \text{NORMSDIST}(z)$.



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Properties of the Cumulative Distribution Function $\Phi(z)$

i) $\Phi(-z) = 1 - \Phi(z)$.

ii) If $z \geq 4$, then $\Phi(z) \approx 1$.

Example 17: Find the area under the standard normal curve that lies

a) to the left of $z = -1.3$

b) to the right of $z = 2.5$

c) between $z = -1.48$ and $z = 2$.

Example 18: Find the value of z such that

a) $P(Z > z) = 0.1$

b) $P(-1.24 < Z < z) = 0.8$

c) $P(-z < Z < z) = 0.95$

Note: In Excel, if $\Phi(z) = p$, then $z = \text{NORMSINV}(p)$.

Standardizing a Normal Random Variable

If $X \sim N(\mu, \sigma^2)$ then $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$

Calculating Normal Probabilities

If $X \sim N(\mu, \sigma^2)$, then

a) $P(X < a) = \Phi\left(\dfrac{a - \mu}{\sigma}\right)$

b) $P(a < X < b) = \Phi\left(\dfrac{b - \mu}{\sigma}\right) - \Phi\left(\dfrac{a - \mu}{\sigma}\right)$

Example 19: The compressive strength of samples of cement can be modeled by a normal distribution with a mean of 6000 kilograms per square centimeter and a standard deviation of 100 kilograms per square centimeter.

(a) What is the probability that a sample's strength is less than 6250 Kg/cm^2?

(b) What is the probability that a sample's strength is between 5800 and 5900 Kg/cm^2?

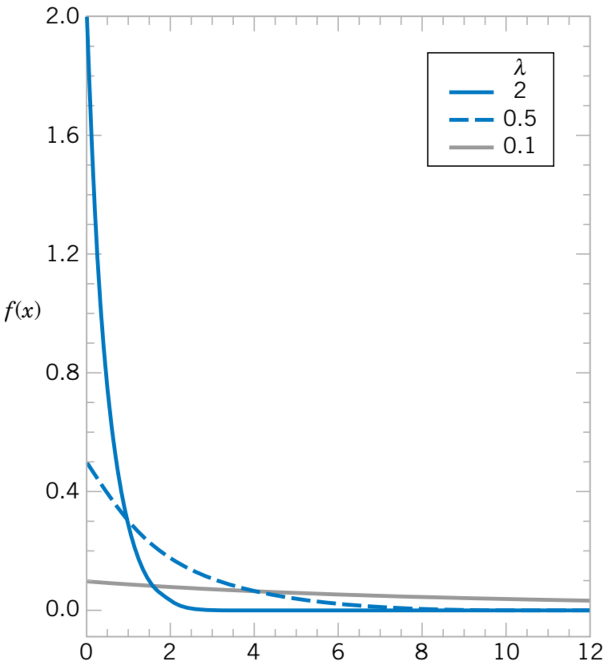
(c) What strength is exceeded by 95% of the samples?

2.3.5 Exponential Distribution

Definition

The random variable X that equals the distance between successive events of a Poisson process with mean $\lambda > 0$ is an exponential random variable with parameter λ. The probability density function of X is

$$f(x) = \lambda e^{-\lambda x},\ x \geq 0$$



2.0
1.6
1.2
0.8
0.4
0.0
f(x)
λ
2
0.5
0.1
0
2
4
6
8
10
12

Mean and Variance

If the random variable X has an exponential distribution with parameter λ, then

$$\mu = E(X) = \frac{1}{\lambda},\ \sigma^2 = V(X) = \frac{1}{\lambda^2}$$

Example 20: The time between calls to a plumbing supply business is exponentially distributed with a mean time between calls of 15 minutes.

(a) What the probability that there are no calls within a 30-minute interval?

(b) What the probability that at least one call arrives within a 10-minute interval?

(c) What the probability that the first call arrives within 5 and 10 minutes after opening?

(d) Determine the length of an interval of time such that the probability of at least one call in the interval is 0.9?

POSTS AND TELECOMMUNICATIONS INSTITUTE OF TECHNOLOGY

———————o0o———————

Chapter 3: Random Vectors

PROBABILITY AND STATISTICS

Department of Mathematics, Faculty of Fundamental Science 1

Hanoi - 2023

1. 3.1 Definition of Random Vectors and Joint Cumulative Distribution Functions
2. 3.2 Discrete Bivariate Random Variables
3. 3.3 Continuous Bivariate Random Variables
4. 3.4 Covariance and Correlation

1. 3.1 Definition of Random Vectors and Joint Cumulative Distribution Functions
2. 3.2 Discrete Bivariate Random Variables
3. 3.3 Continuous Bivariate Random Variables
4. 3.4 Covariance and Correlation

3.1.1 Definition of Random Vectors

Definition

- A random vector of dimension n is an ordered n-tuple $(X_1, X_2, \ldots, X_n)$, where $X_1, X_2, \ldots, X_n$ are random variables.
- A random vector $(X_1, X_2, \ldots, X_n)$ is called discrete if $X_1, X_2, \ldots, X_n$ are discrete random variables.
- A random vector $(X_1, X_2, \ldots, X_n)$ is called continuous if $X_1, X_2, \ldots, X_n$ are continuous random variables.

3.1.2 Joint Cumulative Distribution Functions

Definition

The joint cumulative distribution function of the random variables $X_1, X_2, \ldots, X_n$ is defined as follows.

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n),$$

$\forall x_1, x_2, \ldots, x_n \in \mathbb{R}$.

Properties of the Joint Cumulative Distribution Functions

1) $0 \leq F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) \leq 1,\ \forall x_1,x_2,\ldots,x_n \in \mathbb{R}.$
2) $\lim\limits_{x_k\to-\infty} F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = 0,\ k = 1,2,\ldots,n.$
3) $\lim\limits_{x_1,x_2,\ldots,x_n\to\infty} F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = 1.$
4) The joint cumulative distribution function is nondecreasing in each variable.
5) $P(x_1 < X \leq x_2, y_1 < Y \leq y_2) =$
$F_{X,Y}(x_2,y_2) - F_{X,Y}(x_1,y_2) - F_{X,Y}(x_2,y_1) + F_{X,Y}(x_1,y_1).$
6) $\lim\limits_{y\to\infty} F_{X,Y}(x,y) = F_X(x);\ \lim\limits_{x\to\infty} F_{X,Y}(x,y) = F_Y(y).$
$F_X(x), F_Y(y)$ are called marginal cumulative distribution functions of X and Y, respectively.

Example 1: Let X, Y be random variables with the joint cumulative distribution function

$$F_{X,Y}(x,y) = \begin{cases} (1-e^{-x})(1-e^{-y}) & \text{if } x \geq 0, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

a) Find marginal cumulative distribution functions $F_X(x), F_Y(y)$.

b) Compute $P(X > 1, Y > 2)$.

1 3.1 Definition of Random Vectors and Joint Cumulative Distribution Functions

2 3.2 Discrete Bivariate Random Variables

3 3.3 Continuous Bivariate Random Variables

4 3.4 Covariance and Correlation

3.2.1 Joint Probability Mass Functions and Joint Probability Distributions

Definition

The joint probability mass function of discrete random variables X, Y is defined by

$$p_{X,Y}(x, y) = P(X = x, Y = y),\ \forall x, y \in \mathbb{R}.$$

Let X, Y be discrete random variables. Suppose that the ranges of X and Y are

$$R_X = \{x_1, \ldots, x_n\}, R_Y = \{y_1, \ldots, y_m\},$$

respectively.

Properties of Joint Probability Mass Functions

1) $p_{X,Y}(x_i, y_j) \geq 0,\ \forall i = 1, \ldots, n; \forall j = 1, \ldots, m.$
2) $\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m} p_{X,Y}(x_i, y_j) = 1.$
3) $\displaystyle F_{X,Y}(x, y) = \sum_{x_i \leq x}\sum_{y_j \leq y} p_{X,Y}(x_i, y_j).$

The joint probability distribution of X and Y can be described by the following table.

$X \backslash Y$	y_1	y_2	$\dots$	y_m
x_1	$p_{X,Y}(x_1, y_1)$	$p_{X,Y}(x_1, y_2)$	$\dots$	$p_{X,Y}(x_1, y_m)$
x_2	$p_{X,Y}(x_2, y_1)$	$p_{X,Y}(x_2, y_2)$	$\dots$	$p_{X,Y}(x_2, y_m)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_n	$p_{X,Y}(x_n, y_1)$	$p_{X,Y}(x_n, y_2)$	$\dots$	$p_{X,Y}(x_n, y_m)$

3.2.2 Marginal Probability Distributions

- The marginal probability distribution of X is

X	x_1	x_2	$\ldots$	x_n
P	$p_X(x_1)$	$p_X(x_2)$	$\ldots$	$p_X(x_n)$

where

$$p_X(x_i) = P(X = x_i) = \sum_{j=1}^{m} p_{X,Y}(x_i, y_j), i = 1, \ldots, n.$$

- The marginal probability distribution of Y is

Y	y_1	y_2	$\dots$	y_m
P	$p_Y(y_1)$	$p_Y(y_2)$	$\dots$	$p_Y(y_n)$

where

$$p_Y(y_j) = P(Y = y_j) = \sum_{i=1}^{n} p_{X,Y}(x_i, y_j), j = 1, \dots, m.$$

Remarks: Two discrete random variables X and Y are independent if and only if

$$p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j),\ \forall i = 1, \dots, n; \forall j = 1, \dots, m.$$

Example 2: Let X, Y be discrete random variables with the joint probability distribution

$X \backslash Y$	0	2	3	5
-2	0.1	0.15	0.1	0
1	$5k$	$3k$	0.05	0.07
4	0	$2k$	0	0.13

a) Find k and compute $F_{X,Y}(3,2)$.

b) Find the marginal probability distributions of X and Y. Are X and Y independent?

3.2.3 Conditional Probability Distributions and Conditional Mean

Let X be a discrete random variable with possible values $x_1, \ldots, x_n$ and let B be an event with $P(B) > 0$. Then the conditional probability mass function of X given B is defined by

$$p_{X|B}(x_i) = P(X = x_i|B) = \frac{P((X = x_i) \cap B)}{P(B)}.$$

The conditional probability distribution of X given B is

$X\|B$	x_1	x_2	$\ldots$	x_n
P	$p_{X\|B}(x_1)$	$p_{X\|B}(x_2)$	$\ldots$	$p_{X\|B}(x_n)$

The conditional mean of X given B is defined by

$$E(X|B) = \sum_{i=1}^{n} x_i p_{X|B}(x_i).$$

- The conditional probability distribution of X given $(Y = y_j)$ is

$X\|(Y = y_j)$	x_1	x_2	$\ldots$	x_n
P	$p_{X\|y_j}(x_1)$	$p_{X\|y_j}(x_2)$	$\ldots$	$p_{X\|y_j}(x_n)$

where

$$p_{X|y_j}(x_i) = P(X = x_i|Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}.$$

- The conditional mean of X given $(Y = y_j)$ is

$$E(X|Y = y_j) = \sum_{i=1}^{n} x_i p_{X|y_j}(x_i).$$

- The conditional probability distribution of Y given $(X = x_i)$ is

$Y\|(X = x_i)$	y_1	y_2	$\dots$	y_m
P	$p_{Y\|x_i}(y_1)$	$p_{Y\|x_i}(y_2)$	$\dots$	$p_{Y\|x_i}(y_m)$

where

$$p_{Y|x_i}(y_j) = P(Y = y_j|X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}.$$

- The conditional mean of Y given $(X = x_i)$ is

$$E(Y|X = x_i) = \sum_{j=1}^{m} y_j p_{Y|x_i}(y_j).$$

Remark: Two discrete random variables X and Y are independent if and only if

$$p_{X|y_j}(x_i) = p_X(x_i),\ p_{Y|x_i}(y_j) = p_Y(y_j),$$

$\forall i = 1, \dots, n; \forall j = 1, \dots, m.$

Example 3: Xavier and Yvette are real estate agents. Let X denote the number of houses that Xavier will sell in a month and let Y denote the number of houses Yvette will sell in a month. An analysis of their past monthly performances has the following joint probabilities.

Y \ X	0	1	2
0	0.12	0.42	0.06
1	0.21	0.06	0.03
2	0.07	0.02	0.01

If Xavier sell 1 house in a month, then what is the mean number of houses Yvette will sell in that month?

1. 3.1 Definition of Random Vectors and Joint Cumulative Distribution Functions
2. 3.2 Discrete Bivariate Random Variables
3. 3.3 Continuous Bivariate Random Variables
4. 3.4 Covariance and Correlation

3.3.1 Joint Probability Density Functions

Definition

The joint probability density function of continuous random variables X, Y is the bivariate function $f_{X,Y}(x, y) \geq 0$ satisfying

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) du dv. \quad (1)$$

Properties of Joint Probability Density Functions

1) $f_{X,Y}(x, y) \geq 0 \ \forall x, y \in \mathbb{R}$.
2) If $f_{X,Y}(x, y)$ is continuous on a region $D \subset \mathbb{R}^2$, then

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}, \ \forall (x, y) \in D.$$

3) $P((X, Y) \in D) = \iint\limits_D f_{X,Y}(x, y) dx dy$ if $D \subset \mathbb{R}^2$.
4) $\int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} f_{X,Y}(x, y) dx dy = 1$.

Example 4: Let X, Y be continuous random variables with the joint probability density function

$$f_{X,Y}(x,y) = \frac{c}{(1+x^2)(1+y^2)};\ x,y \in \mathbb{R}.$$

a) Find the constant c.

b) Find the joint cumulative distribution function of X, Y.

c) Calculate $P(1 < X \leq \sqrt{3}, 0 < Y \leq 1)$.

3.3.2 Marginal Probability Density Functions

- The marginal probability density function of X is

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy, \ x \in \mathbb{R}.$$

- The marginal probability density function of Y is

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx, \ y \in \mathbb{R}.$$

Remark: Two continuous random variables X and Y are independent if and only if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y), \ \forall x, y \in \mathbb{R}.$$

Example 5: Let X, Y be continuous random variables with the joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{6\pi} & \text{if } \dfrac{x^2}{9} + \dfrac{y^2}{4} \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

a) Find the marginal probability density functions of X and Y.
b) Are X and Y independent?

3.3.3 Conditional Probability Density Functions and Conditional Mean

- The conditional probability density function of Y given $X = x$ là

$$f_{Y|x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)} \text{ for } f_X(x) > 0.$$

- The conditional mean of Y given $(X = x)$ is

$$E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|x}(y) dy.$$

- The conditional probability density function of X given $Y = y$ is

$$f_{X|y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \text{ for } f_Y(y) > 0.$$

- The conditional mean of X given $Y = y$ is

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|y}(x) dx.$$

Example 6: Let X, Y be continuous random variables with the joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} x+y & \text{if } 0 \le x, y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a) Find the conditional probability density function $f_{Y|x}(y)$ and compute $E(Y|X=x)$.
b) Find the conditional probability density function $f_{X|y}(x)$ and compute $E(X|Y=y)$.

1. 3.1 Definition of Random Vectors and Joint Cumulative Distribution Functions
2. 3.2 Discrete Bivariate Random Variables
3. 3.3 Continuous Bivariate Random Variables
4. 3.4 Covariance and Correlation

3.4.1 Covariance

Definition

The covariance between the random variables X and Y, denoted as $\operatorname{cov}(X, Y)$, is

$$\operatorname{cov}(X, Y) = E[(X - E(X))(Y - E(Y))].$$

The expected value of a function of two random variables is defined as follows.

- If X, Y are discrete random variables and the ranges of X and Y are $R_X = \{x_1, \ldots, x_n\}, R_Y = \{y_1, \ldots, y_m\}$, respectively, then

$$E(g(X,Y)) = \sum_{i=1}^{n} \sum_{j=1}^{m} g(x_i, y_j) p_{X,Y}(x_i, y_j)$$

- If X, Y are continuous random variables, then

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dxdy$$

Properties of the Covariance

1) $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$
2) If X and Y are independent random variables, then $\text{cov}(X, Y) = 0$.
3) If a, b, c, d are constants, then
$$\text{cov}(aX + c, bY + d) = ab\text{cov}(X, Y).$$
4) If a, b are constants, then
$$V(aX + bY) = a^2V(X) + b^2V(Y) + 2ab\text{cov}(X, Y).$$

Covariance Matrix

Let $(X_1, X_2, \ldots, X_n)$ be a random vector. The matrix

$$M = [C_{ij}]_{n\times n}, \text{ where } C_{ij} = \mathrm{cov}(X_i, X_j),$$

is called the covariance matrix of X.

Remark: The covariance matrix is symmetric.

3.4.2 Correlation

Definition

The correlation between random variables X and Y, denoted as $\rho_{X,Y}$, is

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sqrt{V(X)V(Y)}}.$$

Properties of the Correlation

1) If X, Y are independent random variables, then $\rho_{X,Y} = 0$.
2) $-1 \leq \rho_{X,Y} \leq 1$.
3) $\rho_{X,Y} = \pm 1 \Leftrightarrow$ there exist real numbers $a \neq 0$ and b such that $Y = aX + b$.

Remark: The correlation is a measure of the linear relationship between random variables.

Example 7: Let X, Y be discrete random variables with the joint probability distribution

X \ Y	0	2	3	5
-2	0.1	0.15	0.1	0
1	0.2	0.12	0.05	0.07
4	0	0.08	0	0.13

a) Find the correlation between X and Y.

b) Compute $V(2X - 3Y)$.

POSTS AND TELECOMMUNICATIONS INSTITUTE OF TECHNOLOGY

——————o0o——————

Chapter 4: Sampling Theory

PROBABILITY AND STATISTICS

Department of Mathematics, Faculty of Fundamental Science 1

Hanoi - 2023

Chapter 4: Sampling Theory

1. 4.1 Population and Sample
2. 4.2 Sampling Distributions

Chapter 4: Sampling Theory

1. 4.1 Population and Sample

2. 4.2 Sampling Distributions

4.1.1 Population

Definition

A population is a set of all items of interest.

- The number of items in a population is called population size, denoted by N.
- A numerical measurement describing some characteristic of a population is called a parameter.

Example 1: In a study that is trying to determine the mean weight of all 20-year-old males in the United States, the population would be all 20-year-old males in the United States.

Example 2: If we are studying the grade point average (GPA) of students at Harvard, the population is the set of all the students at Harvard.

- Each observation in a population is a value of a random variable X having some probability distribution.
- If X is normally distributed, we say that the population is normally distributed or that we have a normal population.
- Data are the observed values of a variable.
- Why not study the entire population?
  - It is impossible to observe every item in the population.
  - It is too costly.
  - Some testing is inherently destructive.

4.1.2 Sample

Definition

- A sample is a subset of a population.
- The number of items in a sample is called sample size, denoted by n $(n \ll N)$.

4.1.3. Some Methods of Sampling

- Random sampling $\nearrow$ without replacement / $\searrow$ with replacement
- Systematic sampling.
- Stratified sampling.
- Cluster sampling.

4.1.4 Random Sample

Before the data is collected, the observations are considered to be random variables, say $X_1, X_2, \dots, X_n$.

Definition

The random variables $X_1, X_2, \dots, X_n$ constitute a random sample of size n if

i) the X_i's are independent random variables,

ii) every X_i has the same probability distribution as the population.

4.1.5 Describing Data

- Frequency Table:

X	x_1	x_2	$\ldots$	x_k
Frequency	n_1	n_2	$\ldots$	n_k

where

$x_1 < x_2 < \ldots < x_k$.

n_i is the number of observations having the value x_i in the sample.

$n = n_1 + n_2 + \ldots + n_k$.

- Let $f_i = \dfrac{n_i}{n}$ denote the relative frequency of occurrences of the value x_i in the sample, we get the relative frequency table:

X	x_1	x_2	$\dots$	x_k
Relative frequency	f_1	f_2	$\dots$	f_k

- If n is large, it is useful to group the data into intervals or classes. Then data can be arranged in the following table.

Class	$[a_0, a_1)$	$[a_1, a_2)$	$\dots$	$[a_{k-1}, a_k]$
Frequency	n_1	n_2	$\dots$	n_k

where n_i the number of observations falling into the ith class.

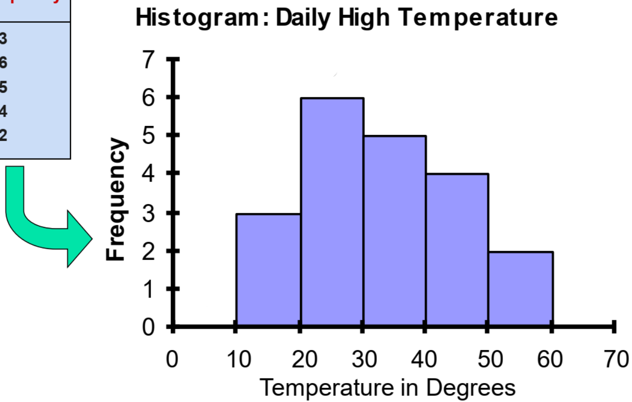
The width of the ith class is $h_i = a_i - a_{i-1}$.

Note:

- The classes should be of equal width.
- Choose the midpoint of each class as a representative value for that class.
- The representative value for the ith class is $x_i = \dfrac{a_{i-1} + a_i}{2}$.
- Grouped data are often represented graphically by histograms.
- A histogram is created by drawing rectangles whose bases are the intervals and whose heights are the frequencies.

Example 3:

Class	Frequency
10 – under 20	3
20 – under 30	6
30 – under 40	5
40 – under 50	4
50 – under 60	2



1 4.1 Population and Sample

2 4.2 Sampling Distributions

Definition

A statistic is a function of the observations in a random sample.

Remark:

- A statistic is a random variable, so it has a probability distribution.
- We use statistics to make inferences about parameters.

Sampling Distribution

The probability distribution of a statistic is called a sampling distribution.

4.2.2 Some common Statistics

Let $(X_1, X_2, \ldots, X_n)$ be a random sample of size n.

- The sample mean is

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

For a specific sample $(x_1, x_2, \ldots, x_n)$, the sample mean is

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

- The sample variance is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

For a specific sample $(x_1, x_2, \ldots, x_n)$, the sample variance is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Remark:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right].$$

- The sample standard deviation, S, is the positive square root of the sample variance.
  For a specific sample $(x_1, x_2, \ldots, x_n)$, the sample standard deviation is $s = \sqrt{s^2}$.

Example 4: A sample of 10 adults was asked to report the number of hours they spent on the Internet the previous month. The results are listed here.

$$0,\ 7,\ 12,\ 5,\ 33,\ 14,\ 8,\ 0,\ 9,\ 22.$$

Calculate the sample mean and sample variance.

4.2.3 How to Compute $\overline{x}, s^2$ for Grouped Data

- Let x_i be the midpoint of the ith class, the data can be represented in the form of a frequency table

x_i	x_1	x_2	$\dots$	x_k
Frequency	n_1	n_2	$\dots$	n_k

Then

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i,$$

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} n_i x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{k} n_i x_i\right)^2\right],$$

where $n = n_1 + n_2 + \dots n_k$.

- If all classes have equal width h, the sample mean and sample variance can be computed by using the transformation $u_i = \dfrac{x_i - a}{h}$. Then

$$\overline{x} = a + h\overline{u};\ \ s^2 = h^2 s_u^2,$$

where

$$\overline{u} = \frac{1}{n}\sum_{i=1}^{k} n_i u_i,$$

$$s_u^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} n_i u_i^2 - \frac{1}{n}\left(\sum_{i=1}^{k} n_i u_i\right)^2\right].$$

Example 5: The data below give the weight in kilograms of 100 college students taken at random in fall 1996.

Weight (kg)	56-58	58-60	60-62	62-64	64-66	66-68	68-70
Frequency	5	8	18	36	27	4	2

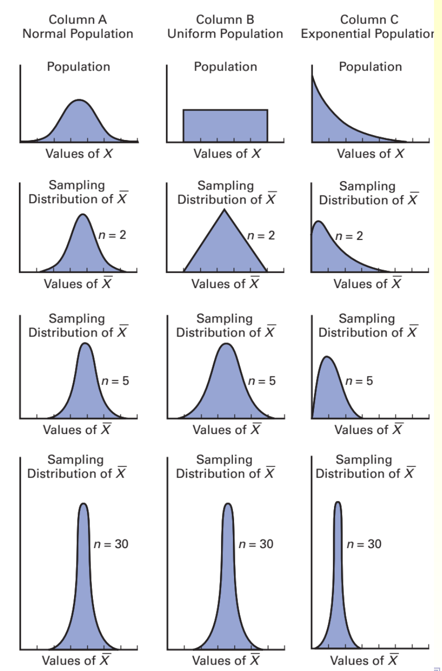
Compute $\overline{x}, s^2$.

4.2.3 Sampling Distribution of the Mean

- If $(X_1, X_2, \ldots, X_n)$ is a random sample of size n taken from a population with mean μ and variance σ^2, then

$$\mu_{\overline{X}} = \mu,\ \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

  ($\sigma_{\overline{X}}$ is called the standard error of the mean)
- If the population is normally distributed, then $\overline{X}$ is also normally distributed.
- If the population is not normally distributed, then $\overline{X}$ is approximately normally distributed for a sufficiently large sample size (The Central Limit Theorem)


Column A
Normal Population
Column B
Uniform Population
Column C
Exponential Populatior
Population
Values of X
Sampling Distribution of $\overline{X}$
$n = 2$
$n = 5$
$n = 30$
Values of $\overline{X}$

POSTS AND TELECOMMUNICATIONS INSTITUTE OF TECHNOLOGY

——————o0o——————

Chapter 5: Estimation and Tests of Statistical Hypotheses

PROBABILITY AND STATISTICS

Department of Mathematics, Faculty of Fundamental Science 1

Hanoi - 2023

Chapter 5: Estimation and Tests of Statistical Hypotheses

1. 5.1 Point Estimation
2. 5.2 Interval Estimation
3. 5.3 Basic Concepts of Testing Statistical Hypotheses
4. 5.4 Tests on Parameters

Chapter 5: Estimation and Tests of Statistical Hypotheses

1. 5.1 Point Estimation
2. 5.2 Interval Estimation
3. 5.3 Basic Concepts of Testing Statistical Hypotheses
4. 5.4 Tests on Parameters

Definition

- Let $(X_1, X_2, \ldots, X_n)$ be a random sample of size n taken from a population with an unknown parameter θ. If a statistic $\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$ is used instead of θ, then $\hat{\Theta}$ is called a point estimator of θ.
- After the sample has been selected, $\hat{\Theta}$ takes on a particular numerical value $\hat{\theta}$ called the point estimate of θ.

Example 1: Suppose that the random variable X is normally distributed with an unknown mean μ.

- The sample mean $\overline{X}$ is a point estimator of μ.
- After the sample has been selected, the numerical value $\overline{x}$ is the point estimate of μ. Thus, if $x_1 = 25, x_2 = 30, x_3 = 29$, and $x_4 = 31$, the point estimate of μ is

$$\overline{x} = \frac{25 + 30 + 29 + 31}{4} = 28.75$$

Definition

A statistic $\hat{\Theta}$ is called an unbiased estimator of the parameter θ if

$$E(\hat{\Theta}) = \theta.$$

Otherwise, $\hat{\Theta}$ is called a biased estimator of θ.

Example 2: Suppose that X is a random variable with mean μ and variance σ^2. Let $(X_1, X_2, \ldots, X_n)$ be a random sample of size n from the population represented by X. Show that the sample mean $\overline{X}$ and sample variance S^2 are unbiased estimators of μ and σ^2, respectively.

5.1.3 Efficient Estimators

Definition

If we consider all unbiased estimators of a parameter θ, the one with the smallest variance is called an efficient estimator of θ.

If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2, the sample mean $\overline{X}$ is the efficient estimator of μ.

5.1.4 Consistent Estimators

Definition

A statistic $\hat{\Theta}$ is called a consistent estimator of the parameter θ if

$$\lim_{n\to\infty} P(|\hat{\Theta} - \theta| < \varepsilon) = 1, \forall \varepsilon > 0.$$

We often need to estimate:

1) The mean μ of a single population.
2) The variance σ^2 (or standard deviation) of a single population.
3) The proportion p of items in a population that belong to a class of interest.

Reasonable point estimates of these parameters are as follow:

1) For μ, the estimate is $\hat{\mu} = \overline{x}$, the sample mean.
2) For σ^2, the estimate is $\hat{\sigma}^2 = s^2$, the sample variance.
3) For p, the estimate is $\hat{p} = x/n$, the sample proportion, where x is the number of items in a random sample of size n that belong to the class of interest.

1 5.1 Point Estimation

2 5.2 Interval Estimation

3 5.3 Basic Concepts of Testing Statistical Hypotheses

4 5.4 Tests on Parameters

5.2.1 Concept of Confidence Interval

- Suppose that $L = L(X_1, X_2, \ldots, X_n), U = U(X_1, X_2, \ldots, X_n)$ are two statistics from a random sample $(X_1, X_2, \ldots, X_n)$, θ is a population parameter, $\alpha \in (0, 1)$.
- The interval $[L, U]$ is called a confidence interval for θ with the $1 - \alpha$ level of confidence if

$$P(L \leq \theta \leq U) = 1 - \alpha$$

- $U - L$ is called the length of the confidence interval.

- If we have selected the sample:

$$X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$$

and computed $L = \ell, U = u$, then the $1 - \alpha$ confidence interval for θ is

$$\ell \leq \theta \leq u$$

ℓ: Lower-confidence limit
u: Upper-confidence limit

5.2.2 Confidence Interval for a Population Mean

Problem

Find a $1-\alpha$ confidence interval for a population mean μ.

Let $(X_1, X_2, \ldots, X_n)$ be a random sample of size n taken from the population.

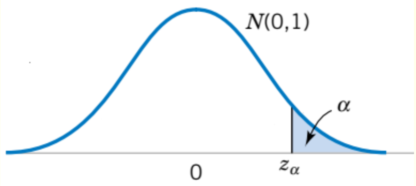
We consider the problem in three cases:

1. The population has a normal distribution with variance σ^2 known.
2. The population has an arbitrary distribution, large-sample.
3. The population has a normal distribution with variance σ^2 unknown.

Case 1: The population has a normal distribution with variance σ^2 known

- Use the notation z_α to represent the value of Z such that the area to its right under the standard normal curve is α; that is,

$$P(Z > z_\alpha) = \alpha.$$

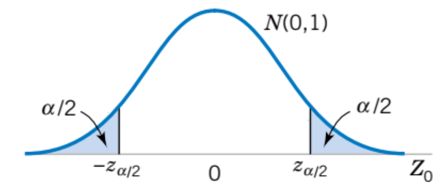


- $z_{0.05} = 1.64,\ z_{0.025} = 1.96,\ z_{0.01} = 2.33.$
- In Excel, $z_\alpha = \text{NORM.S.INV}(1 - \alpha)$.

If the population is normally distributed with mean μ and standard deviation σ, then

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

is standard normally distributed.



$$P\left(-z_{\alpha/2} \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) = 1-\alpha$$

$$\Leftrightarrow P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Case 1: The population has a normal distribution with variance σ^2 known

A 1–α confidence interval for μ is

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Example 3: The life in hours of a 75-watt light bulb is known to be normally distributed with $\sigma = 25$ hours. A random sample 20 bulbs has a mean life of 1014 hours. Find a 95% confidence interval on the mean life.

Interpreting a Confidence Interval

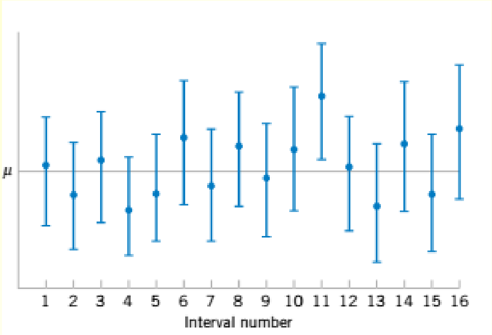


Figure 1: Repeated construction of a confidence interval for μ

If an infinite number of random samples are collected and a $1-\alpha$ confidence interval for μ is computed from each sample, $100(1-\alpha)\%$ of these intervals will contain the true value of μ.

The length of a confidence interval is a measure of the precision of estimation.

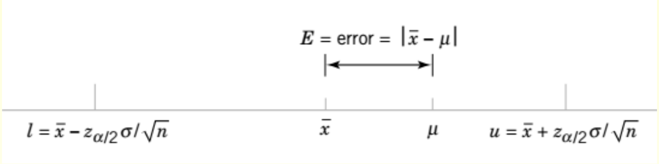


Figure 2: Error in estimating μ with $\overline{x}$

Determining the Sample Size

If $\overline{x}$ is used as an estimate of μ, we can be $(1-\alpha)$ confident that the error $|\overline{x}-\mu|$ will not exceed a specified amount E when the sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 \qquad (1)$$

If the right-hand side of Equation (1) is not an integer, it must be rounded up.

Example 4: How large a sample is required in Example 3 if we want to be 95% confident that the error in estimating the mean life is less than 5 hours.

Case 2: The population has an arbitrary distribution, large-sample ($n > 30$)

A $1–\alpha$ confidence interval for μ is

$$\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

Example 5: A random sample of 100 students from a large college showed an average IQ score of 112 with a standard deviation of 10. Find a 99% confidence interval for the mean IQ score of all students in this college.

t Distribution

Let $(X_1, X_2, \ldots, X_n)$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. The random variable

$$T = \frac{\overline{X} - \mu}{S}\sqrt{n}$$

has a t distribution with $n-1$ degrees of freedom.

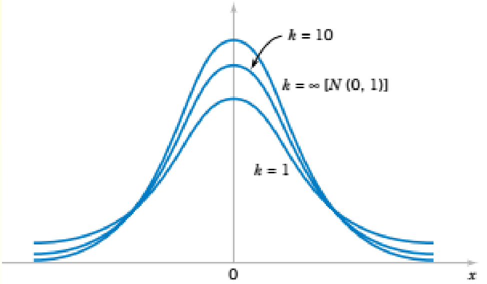
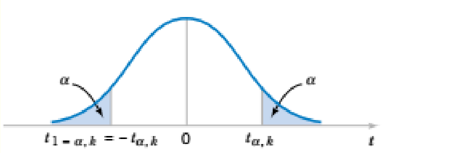


Figure 3: Probability density functions of several t distributions, k is the number of degrees of freedom

Note:

- As the number of degrees of freedom $k \to \infty$, the limiting form of the t distribution is the standard normal distribution.
- The Appendix Table V lists values of $t_{\alpha,k}$, which are the values of the random variable T with k degrees of freedom such that $P(T > t_{\alpha,k}) = \alpha$.



- In Excel, $t_{\alpha,k} = \text{T.INV.2T}(2\alpha, k)$.

Case 3: The population has a normal distribution with variance σ^2 unknown

A 1–α confidence interval for μ is

$$\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

Example 6: The nicotine contents of five cigarettes of a certain brand, measured in milligrams, are 21, 19, 23, 19, 23. Find a 99% confidence interval for the average nicotine content of this brand of cigarette. Assume the population is normally distributed.

5.2.3 Large-Sample Confidence Interval for a Population Proportion

Approximate Confidence Interval for a Population Proportion

A $1-\alpha$ confidence interval for the proportion p of items in a population that belong to a class of interest is

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where

$$\hat{p} = \frac{x}{n}$$

x is the number of items in the sample that belong to a class of interest, n is the sample size.

Example 7: Of 1000 randomly selected cases of lung cancer, 823 resulted in death within 10 years. Calculate a 95% confidence interval on the death rate from lung cancer.

Choice of Sample Size

If we want to be $1-\alpha$ confident that the error in estimating p by $\hat{p}$ is less than E, the appropriate sample size is

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p)$$

When there is no information concerning the value of p, the sample size is

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 0.25$$

Example 8: Of 1000 randomly selected cases of lung cancer, 823 resulted in death within 10 years.

(a) Using the point estimate of p obtained from the preliminary sample, what sample size is needed to be 95% confident that the error in estimating the true value of p is less than 0.03.

(b) How large must the sample be if we wish to be at least 95% confident that the error is less than 0.03, regardless of the true value of p.

Chapter 5: Estimation and Tests of Statistical Hypotheses

1. 5.1 Point Estimation
2. 5.2 Interval Estimation
3. 5.3 Basic Concepts of Testing Statistical Hypotheses
4. 5.4 Tests on Parameters

5.3.1 Statistical Hypotheses

Definition

A statistical hypothesis is a statement about the parameters of one or more populations.

Example 9: The manager of a fast-food restaurant wants to determine whether the waiting time to place an order has changed in the past month from its previous population mean value of 4.5 minutes. We may express this formally as

- $H_0 : \mu = 4.5 \longleftarrow$ null hypothesis
- $H_1 : \mu \neq 4.5 \longleftarrow$ alternative hypothesis

In some situations, we may wish to formulate a one-sided alternative hypothesis as in

$$\begin{aligned} H_0 &: \quad \mu = 4.5 \\ H_1 &: \quad \mu > 4.5 \end{aligned}$$

or

$$\begin{aligned} H_0 &: \quad \mu = 4.5 \\ H_1 &: \quad \mu < 4.5 \end{aligned}$$

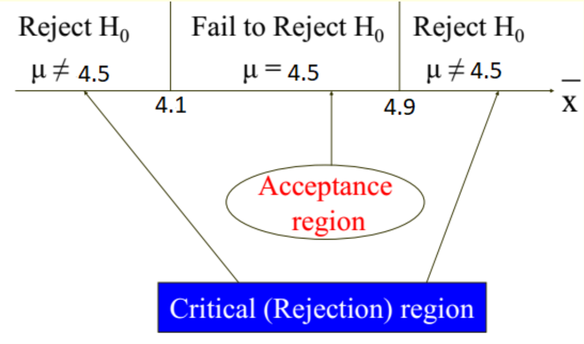
A procedure for deciding whether to accept or reject the null hypothesis, based on sample data, is called a test of a statistical hypothesis.

- If reject the null hypothesis, we have strong statistical evidence that the alternative hypothesis is correct.
- If do not reject the null hypothesis, we have not proven the null hypothesis.

- Consider Example 9, we wish to test

$$H_0: \quad \mu = 4.5$$
$$H_1: \quad \mu \neq 4.5$$

- Suppose that a sample of $n = 36$ orders is selected and that the sample mean waiting time $\overline{x}$ is computed.



Two possible Errors

Type I Error

Rejecting the null hypothesis H_0 when it is true.

Probability of type I error:

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$$

α is called the significance level of the test.

Type II error

Failing to reject the null hypothesis H_0 when it is false.

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false})$$

$1-\beta$ is called the power of the test.

5.3.3 General Procedure for Hypothesis Tests

1) State H_0, H_1.
2) Determine an appropriate test statistic.
3) Compute the value of the test statistic from the sample data.
4) State the critical region.
5) Decision: Reject H_0 if the test statistic has a value in the critical region, otherwise fail to reject H_0.

Chapter 5: Estimation and Tests of Statistical Hypotheses

1. 5.1 Point Estimation
2. 5.2 Interval Estimation
3. 5.3 Basic Concepts of Testing Statistical Hypotheses
4. 5.4 Tests on Parameters

Problem

Suppose a population has a normal distribution with mean μ and variance σ^2. Test the hypotheses

$$\begin{aligned} H_0 &: \quad \mu = \mu_0 \\ H_1 &: \quad \mu \neq \mu_0 \text{ (or } H_1 : \mu > \mu_0, \text{ or } H_1 : \mu < \mu_0) \end{aligned}$$

Let $(X_1, X_2, \ldots, X_n)$ be a random sample taken from the population.

Case 1: σ^2 is known

- Test statistic

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n}$$

- If the null hypothesis is true, $Z_0 \sim N(0, 1)$

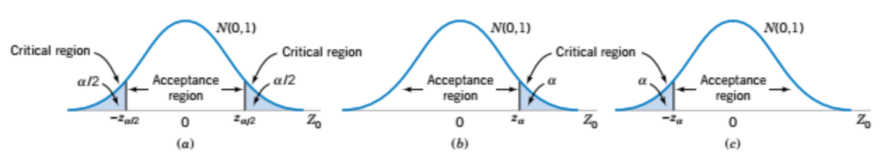


Figure 4: The distribution of Z_0 when $H_0 : \mu = \mu_0$ is true, with critical region for

(a) the two-sided alternative $H_1 : \mu \neq \mu_0$,

(b) the one-sided alternative $H_1 : \mu > \mu_0$,

(c) the one-sided alternative $H_1 : \mu < \mu_0$.

For a sample $x_1, x_2, \ldots, x_n$, we test the hypothesis $H_0 : \mu = \mu_0$ as follows

1) Compute $z_0 = \dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$
2) Conclusion

Alternative hypothesis	Rejection criteria
$H_1 : \mu \neq \mu_0$	$\lvert z_0 \rvert > z_{\alpha/2}$
$H_1 : \mu > \mu_0$	$z_0 > z_\alpha$
$H_1 : \mu < \mu_0$	$z_0 < -z_\alpha$

Example 10: Spam e-mail has become a serious and costly nuisance. An office manager believes that the average amount of time spent by office workers reading and deleting spam exceeds 25 minutes per day. To test this belief, he takes a random sample of 18 workers and measures the amount of time each spends reading and deleting spam. The results are listed here.

35, 48, 29, 44, 17, 21, 32, 28, 34, 23, 13, 9, 11, 30, 42, 37, 43, 48

If the population of times is normal with a standard deviation of 12 minutes, can the manager infer at the 1% significance level that he is correct?

Case 2: σ^2 is unknown, $n > 30$.

- Test statistic

$$Z_0 = \frac{\overline{X} - \mu_0}{S}\sqrt{n}$$

- If the null hypothesis is true, Z_0 has approximately the standard normal distribution.

Remark: In the case of large sample sizes ($n > 30$), it is not necessary to make the assumption that the population has a normal distribution.

For a specific sample $x_1, x_2, \ldots, x_n$, we test the hypothesis $H_0 : \mu = \mu_0$ as follows

1) Compute $z_0 = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n}$
2) Conclusion: the same as Case 1.

Example 11: The breaking strengths of cables produced by a manufacturer have mean 1800 lb. By a new technique in the manufacturing process, it is claimed that the breaking strength can be increased. To test this claim, a sample of 50 cables is tested, and it is found that the mean breaking strength is 1850 lb with the standard deviation 100 lb. Can we support the claim at a 0.01 level of significance?

Case 3: σ^2 is unknown, $n < 30$.

- Test statistic: $T_0 = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$
- If the null hypothesis is true, T_0 has a t distribution with $n-1$ degrees of freedom.

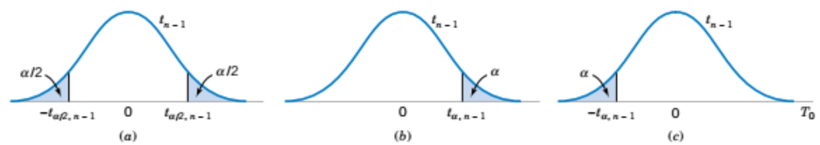


Figure 5: The distribution of T_0 when $H_0 : \mu = \mu_0$ is true, with critical region for (a) $H_1 : \mu \neq \mu_0$, (b) $H_1 : \mu > \mu_0$, and (c) $H_1 : \mu < \mu_0$

For a specific sample $x_1, x_2, \ldots, x_n$, we test the hypothesis $H_0 : \mu = \mu_0$ as follows

1) Compute $t_0 = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n}$
2) Conclusion

Alternative hypothesis	Rejection criteria
$H_1 : \mu \neq \mu_0$	$\lvert t_0 \rvert > t_{\alpha/2,n-1}$
$H_1 : \mu > \mu_0$	$t_0 > t_{\alpha,n-1}$
$H_1 : \mu < \mu_0$	$t_0 < -t_{\alpha,n-1}$

Example 12: The mean time for mice to die when injected with 1000 leukemia cells is known to be 12.5 days. When the injection was doubled in a sample of 10 mice, the survival times were

$$10.5,\ 11.2,\ 12.9,\ 12.7,\ 10.3,\ 10.4,\ 10.9,\ 11.3,\ 10.6,\ 11.7$$

If the survival times are normally distributed, do the results suggest that the increased dosage has decreased survivorship at the 5 % significance level?

5.4.2 Tests on a Population Proportion

Problem

Let p be the proportion of items in a population that belong to a class of interest. Test the hypotheses

$$H_0: \quad p = p_0$$
$$H_1: \quad p \neq p_0 \text{ (or } H_1: p > p_0, \text{ or } H_1: p < p_0)$$

- Suppose that a random sample of size n has been taken from the population and that X observations in this sample belong to the class of interest. Then $\hat{P} = X/n$ is a point estimator of p.
- Test statistic: $Z_0 = \dfrac{\hat{P} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$
- If the null hypothesis is true, $Z_0 \approx N(0, 1)$.

For a specific sample, we test the hypothesis $H_0 : p = p_0$ as follows

1) Compute $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$

2) Conclusion

Alternative hypothesis	Rejection criteria
$H_1 : p \neq p_0$	$\lvert z_0 \rvert > z_{\alpha/2}$
$H_1 : p > p_0$	$z_0 > z_\alpha$
$H_1 : p < p_0$	$z_0 < -z_\alpha$

Example 13: An article in Fortune (September 21, 1992) claimed that nearly one-half of all engineers continue academic studies beyond the B.S. degree, ultimately receiving either an M.S. or a Ph.D. degree. Data from an article in Engineering Horizons (Spring 1990) indicated that 117 of 484 new engineering graduates were planning graduate study. Are the data from Engineering Horizons consistent with the claim reported by Fortune? Use $\alpha = 0.05$ in reaching your conclusions.