CHAPTER 1: BASIC CONCEPTS OF PROBABILITY

- Basic probability concepts and definitions
- Operations of events
    - Complement events, mutually exclusive events
    - Union of events, intersection of events
- Counting methods
    - Multiplication principle
    - Addition principle
- Permutation
- Combination
- Approaches to probability
    - Classical, relative frequency, subjective
- Axioms of probability
- Basic properties of probability
- Conditional probability
- Multiplicative rule of probability
- The Law of Total Probability
- Bayes' Theorem
- Independent events

CHAPTER 2: CONCEPTS OF RANDOM VARIABLES

- Random variables
- Discrete probability distributions (probability function)
- Continuous probability distributions (probability density function)
- Cumulative distribution function
- Expectation
    - Mean, variance
    - Expectation of functions of random variables
    - Properties of expectation
- Chebychev's Inequality

CHAPTER 3: 2-D RANDOM VARIABLES AND CONDITIONAL PROBABILITY DISTRIBUTIONS

- 2 dimensional random variables
- Joint probability functions for discrete random variables
- Joint probability density functions for continuous random variables
- Marginal distributions
- Conditional distributions
- Independent random variables
- Expectation

CHAPTER 4: SPECIAL PROBABILITY DISTRIBUTIONS

- Discrete distributions
    - Discrete uniform distribution
    - Bernoulli and Binomial distributions
    - Negative binomial distribution
    - Poisson distribution
    - Poisson approximation to Binomial distribution
- Continuous distributions
    - Continuous uniform distribution
    - Exponential distribution
    - Normal distribution
    - Normal approximation to Binomial distribution

CHAPTER 5: SAMPLING & SAMPLING DISTRIBUTIONS

- Population and sample
- Random sampling
- Sampling distribution of sample mean
- CLT and applications
- Sampling distribution of difference of two sample means
- Chi-square distribution
- Sampling distribution of $\frac{(n-1)S^2}{\sigma^2}$
- t-distribution
- F-distribution

CHAPTER 6: ESTIMATION BASED ON NORMAL DISTRIBUTION

- Point estimation
- Parameter and statistic
- Unbiased estimator
- Interval estimation
- Confidence interval for the mean
- Sample size
- Confidence intervals for the difference between two means
- Confidence interval for variances and ratio of variances

CHAPTER 7: HYPOTHESIS TESTING BASED ON NORMAL DISTRIBUTION

- Hypothesis testing based on Normal distribution
- Type I and II Error
- Level of significance
- Hypotheses testing concerning mean
- Critical value approach and p-value approach
- Hypotheses testing concerning difference between two means
- Hypothesis testing concerning variances

CHAPTER 1: BASIC CONCEPTS OF PROBABILITY

Union and Intersection Events

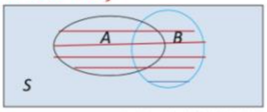
1.2 Operations with Events

1.2.1. Union and Intersection Events

Let S denote a sample space, A and B are any two events of S.

- **Union:** The **Union** of two events A and B, denoted by $A \cup B$, is the event containing all the elements that belong to A or B or to both. That is,

$$A \cup B = \{x : x \in A \textbf{ or } x \in B\}$$



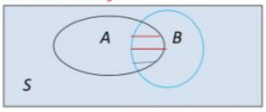
Operations with Events (Continued)

Union and Intersection Events (Continued)

Let S denote a sample space, A and B are any two events of S.

- **Intersection:** The **intersection** of two events A and B, denoted by $A \cap B$ or simply AB, is the event containing all elements that are common to A and B. That is

$$A \cap B = \{x : x \in A \textbf{ and } x \in B\}$$

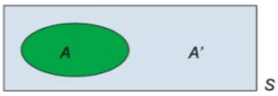


Complement Event

1.2.2 Complement Event

- **Complement:** The **complement** of event A with respect to S, denoted by A' or A^C, is the set of all elements of S that are not in A. That is

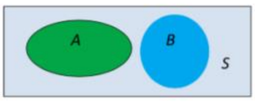
$$A' = \{x : x \in S \text{ and } \boldsymbol{x \notin A}\}$$



Mutually Exclusive Events

1.2.3 Mutually Exclusive Events

Two events A and B are said to be **mutually exclusive** or **mutually disjoint** if $A \cap B = \emptyset$, that is, if A and B have no elements in common.



Union and Intersection of n Events

1.2.4 Union of n events

Union:

The **Union** of n events $A_1, A_2, \cdots, A_n$, denoted by

$$A_1 \cup A_2 \cup \cdots \cup A_n,$$

is the event containing all the elements that belong to one or more of the events A_1, or A_2, or $\cdots$, or A_n. That is,

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n = \{x: x \in A_1 \text{ or} \cdots \text{or } x \in A_n\}$$

1.2.5 Intersection of n events

Intersection:

The **intersection** of n events $A_1, A_2, \cdots, A_n$, denoted by

$$A_1 \cap A_2 \cap \cdots \cap A_n,$$

is the event containing all the elements that are common to all the events A_1, and A_2, and $\cdots$, and A_n. That is,

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n = \{x: x \in A_1 \text{ and} \cdots \text{and } x \in A_n\}$$

Basic Properties of Operations of Events

1.2.6 Some Basic Properties of Operations of Events

1. $A \cap A' = \emptyset$.
2. $A \cap \emptyset = \emptyset$.
3. $A \cup A' = S$.
4. $(A')' = A$
5. $(A \cap B)' = A' \cup B'$

Some Basic Properties of Operations of Events (Continued)

6. $(A \cup B)' = A' \cap B'$
7. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
8. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
9. $A \cup B = A \cup (B \cap A')$
10. $A = (A \cap B) \cup (A \cap B')$

Note: A U B = (A’ INTERSECT B’)’

Note: A’ U B’ = (A INTERSECT B)’

De Morgan's Law

1.2.7 De Morgan's Law

For any n events $A_1, A_2, \cdots, A_n$,

1.
$$(A_1 \cup A_2 \cup \cdots \cup A_n)' = A_1' \cap A_2' \cap \cdots \cap A_n'$$

2.
$$(A_1 \cap A_2 \cap \cdots \cap A_n)' = A_1' \cup A_2' \cup \cdots \cup A_n'$$

Contains

1.2.8 Contained ($\subset$)

- If **all** of the elements in event A are also in event B, then event A is contained in event B, denoted by
$$A \subset B.$$
(or equivalently, $B \supset A$).

- If $A \subset B$ and $B \subset A$, then $A = B$.
(i.e. Event A is equivalent with event B).

Counting methods: Multiplication Principle

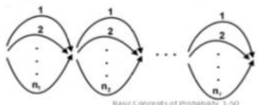
1.3 Counting Methods

1.3.1 Multiplication Principle

- If an operation can be performed in n_1 ways, and
- if for **each of these ways** a second operation can be performed in n_2 ways, then
- the two operations can be performed together in
$$n_1 n_2$$
ways.

Generalized Multiplication Rule

- If an operation can be performed in n_1 ways, and
- if for each of these ways, a second operation can be performed in n_2 ways, and
- for each of the first two ways, a third operation can be performed in n_3 ways, and so forth, then
- the sequence of k operations can be performed in
$$n_1 n_2 \cdots n_k$$
ways.



Addition Principle

1.3.2 Addition Principle

- Suppose that a procedure, designated by 1 can be performed in n_1 ways.
- Assume that a procedure, designated by 2 can be performed in n_2 ways.
- Suppose furthermore that it is **NOT possible** that both procedures 1 and 2 are **performed together**.
- Then the number of ways in which we can perform 1 **or** 2 is

$$n_1 + n_2$$

ways.

Addition Principle (Continued)

It may be generalized as follows.

- If there are k procedures and the i^{th} procedure may be performed in n_i ways, $i = 1, 2, \cdots, k$,
- then the number of ways in which we may perform procedure 1 **or** procedure 2 **or** $\cdots$ **or** procedure k is given by

$$n_1 + n_2 + \cdots + n_k,$$

assuming that **no** two procedures may be performed together.

Permutation

1.3.3 Permutation

A **permutation** is an arrangement of r objects from a set of n objects, where $r \le n$.

(Note that the **order** is taken into consideration in permutation.)

Permutation (Continued)

- In general n distinct objects can be arranged in

$$n(n-1)(n-2)\cdots \times 2 \times 1 = n!$$

ways which is read as ***n factorial*** ways.

Note: $1! = 1$ and $0! = 1$.

Permutation of n distinct objects taken r at a time

1.3.3.1 Permutations of n distinct objects taken r at a time

Number of **permutations** of n distinct objects taken r at a time is denoted by

$${}_nP_r = n(n-1)(n-2)\cdots(n-(r-1)) = n!/(n-r)!$$

We may consider by putting n distinct objects in r compartments:

r compartments

n	$n-1$	$n-2$	....	$n-(r-1)$
ways	ways	ways		ways

By the multiplication principle, there are

$$n(n-1)(n-2)\cdots(n-(r-1)) \text{ ways.}$$

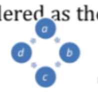
Permutations of n distinct objects arranged in a circle

1.3.3.2 Permutations of n distinct objects arranged in a circle

- The number of permutations of n distinct objects arranged in **a circle** is $(n-1)!$.
- By considering 1 person in a fixed position and arranging the other $n-1$ persons, therefore there are $(n-1)!$ ways.
- Consider the following 4 different ways of arrangement in a line

  $abcd \quad bcda \quad cdab \quad dabc$

  And these arrangements are considered as the same arrangement in a circle



Permutations when not all n objects are distinct

1.3.3.3 Permutations when not all n objects are distinct

- Suppose we have n objects such that there are n_1 of one kind, n_2 of second kind, $\cdots$, n_k of a k^{th} kind, where

$$n_1 + n_2 + \cdots + n_k = n\,.$$

- Then the number of distinct permutations of these n objects taken all together is given by

$${}_nP_{n_1,n_2,\cdots,n_k} = \frac{n!}{n_1!\,n_2!\cdots n_k!}$$

Combination

1.3.4 Combination

- In many problems we are interested in the number of ways of selecting r objects from n objects **without regard to the order**.
- These selections are called **combinations.**
- A combination creates a partition with 2 groups, one group containing the r objects selected and the other group containing the $n-r$ objects that are left.
- The number of such combinations is denoted by

$$\binom{n}{r} \text{ or } {}_nC_r \text{ or } C_r^n$$

Combination (Continued)

- The number of **combinations** of n **distinct** objects taken r at a time is

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

Binomial Coefficient

Binomial Coefficient

- The quantity $\binom{n}{r}$ is called a **binomial coefficient** because it is the coefficient of the term $a^r b^{(n-r)}$ in the binomial expansion of $(a+b)^n$.
- It can be verified that the following hold:
  1. $\binom{n}{r} = \binom{n}{n-r}$ for $r = 0, 1, \cdots, n$
  2. $\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$ for $1 \leq r \leq n$
  3. $\binom{n}{r} = 0$ for $r < 0$ or $r > n$

Axioms of Probability/Laws of Probability

Axioms of Probability (Continued)

Axiom 1: $0 \leq \Pr(A) \leq 1$.

Axiom 2: $\Pr(S) = 1$.

Axiom 3: If $A_1, A_2, \cdots$ are **mutually exclusive** (disjoint) events (that is, $A_i \cap A_j = \emptyset$ when $i \neq j$), then

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$$

In particular, if A and B are two mutually exclusive events then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

Basic Properties of Probability

1.5 Basic Properties of Probability

1.5.1 Some Basic properties of probability

1. $\Pr(\emptyset) = 0$.

Basic Properties of Probability (Continued)

2. If $A_1, A_2, \cdots, A_n$ are **mutually exclusive events**, then

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i)$$

Basic Properties of Probability (Continued)

3. For any event A,

$$\Pr(A') = 1 - \Pr(A).$$

Basic Properties of Probability (Continued)

4. For any two events A and B,

$$\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B').$$

Basic Properties of Probability (Continued)

5. For any two events A and B,

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B).$$

Basic Properties of Probability (Continued)

6. For any three events A, B, C,
$$\Pr(A \cup B \cup C) = \Pr(A) + \Pr(B) + \Pr(C) - \Pr(A \cap B) - \Pr(A \cap C) - \Pr(B \cap C) + \Pr(A \cap B \cap C).$$

Basic Properties of Probability (Continued)

The above property can be extended to n events

$$\Pr(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \Pr(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr(A_i \cap A_j) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} \Pr(A_i \cap A_j \cap A_k) - \cdots \cdots + (-1)^{n+1} \Pr(A_1 \cap A_2 \cap \cdots \cap A_n)$$

It can be proved by mathematical induction.

The above identity is also known as **"The Inclusion-Exclusion Principle"**.

Basic Properties of Probability (Continued)

7. If $A \subset B$, then $\Pr(A) \le \Pr(B)$.

Sample Spaces having Finite Outcomes

1.5.2 Sample Spaces Having Finite Outcomes

Consider the sample space S which contains a finite number of k outcomes. That is,

$$S = \{a_1, a_2, \cdots, a_k\}$$

Let $\Pr(a_i) = p_i$ be the probability of $\{a_i\}$ and

(1) $0 \le p_i \le 1$, for $i = 1, 2, \cdots, k$.

(2) $p_1 + p_2 + \cdots + p_k = 1$.

Sample Spaces Having Finite Outcomes (Continued)

Let an event A consists of r outcomes, $1 \le r \le k$, say

$$A = \{a_{j_1}, a_{j_2}, \cdots, a_{j_r}\}$$

where $j_1, j_2, \cdots, j_r$ represent any r indices from $1, 2, \cdots, k$.

Then

$$\Pr(A) = p_{j_1} + p_{j_2} + \cdots + p_{j_r},$$

where $\Pr(a_{j_l}) = p_{j_l}$, $l = 1, \cdots, r$.

That is, the probability of an event A equals the sum of the probabilities of the various individual outcomes making up the event A.

Conditional Probability

1.6 Conditional probability

1.6.1 Introduction

- We sometimes encounter in calculating probabilities of events when some **partial information** concerning the result of the experiment is available; in such a situation the desired probabilities are **conditional ones**.
- Let A and B be two events associated with an experiment E. We denote

$$\Pr(A|B)$$

the **conditional probability** of the event A, given that event B has occurred.

An Illustrative Example (Continued)

- Hence the conditional probability of A given B can be obtained by

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

- If we divide both the numerator and denominator by $\#(S)$, then

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

- Similarly,

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

1.6.2 Conditional Probability

Definition

- The **conditional probability of B given A**, is defined as

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}, \qquad \text{if } \Pr(A) \neq 0.$$

- Intuitively $\Pr(B|A)$ means that probability of the occurrence of the event B, under the assumption that event A has already occurred.

Remarks

Note: For fixed A, $\Pr(B|A)$ satisfies the various postulates of probability.

- That is, we have
  1. $0 \leq \Pr(B|A) \leq 1$.
  2. $\Pr(S|A) = 1$.
  3. If $B_1, B_2, \cdots$ are **mutually exclusive** (disjoint) events (that is, $B_i \cap B_j = \emptyset$ when $i \neq j$), then

     $$\Pr(\cup_{i=1}^{\infty} B_i\,|A) = \sum_{i=1}^{\infty} \Pr(B_i|A).$$

     In particular, if B_1 and B_2 are disjoint events, then

     $$\Pr(B_1 \cup B_2|A) = \Pr(B_1|A) + \Pr(B_2|A).$$

Multiplicative Rule of Probability

1.6.3 Multiplication Rule of Probability

$$\Pr(A \cap B) = \Pr(A)\Pr(B|A) \quad \text{or}$$
$$\Pr(A \cap B) = \Pr(B)\Pr(A|B),$$

providing $\Pr(A) > 0, \Pr(B) > 0$.

- This rule enables us to calculate the probability that two events will both occur.
- The probability that both events occur is the product of the probability of one event occurs and the conditional probability that the other event occurs given that the first event has occurred.

Multiplication Rule of Probability (Continued)

- It can be extended to more than 2 events:
  $$\Pr(A \cap B \cap C) = \Pr(A)\Pr(B|A)\Pr(C|A \cap B),$$
  providing that $\Pr(A \cap B) > 0$.

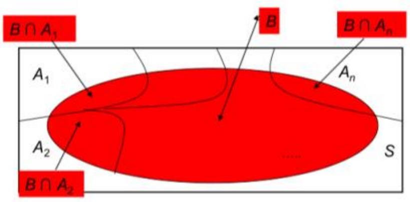
- In general
  $$\Pr(A_1 \cap \cdots \cap A_n) = \Pr(A_1)\Pr(A_2 \mid A_1)$$
  $$\Pr(A_3 \mid A_1 \cap A_2) \cdots \Pr(A_n \mid A_1 \cap \cdots \cap A_{n-1}),$$
  providing that $\Pr(A_1 \cap \cdots \cap A_{n-1}) > 0$.

The Law of Total Probability

1.6.4 The Law of Total Probability

- Let $A_1, A_2, \cdots, A_n$ be a **partition** of the sample space S.
- That is $A_1, A_2, \cdots, A_n$ are **mutually exclusive and exhaustive events** such that $A_i \cap A_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^{n} A_i = S$.
- Then for any event B
  $$\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap A_i) = \sum_{i=1}^{n} \Pr(A_i)\Pr(B|A_i)$$

The Law of Total Probability (Continued)



Bayes' Theorem

1.6.5 Bayes' Theorem

Bayes' Theorem

- Let $A_1, A_2, \cdots, A_n$ be **a partition** of the sample space S. Then
$$\Pr(A_k \mid B) = \frac{\Pr(A_k)\Pr(B \mid A_k)}{\sum_{i=1}^{n}\Pr(A_i)\Pr(B \mid A_i)}$$
for $k = 1, \cdots, n$.

Note : The denominator is just $\Pr(B)$. That is,
$$\Pr(A_k \mid B) = \frac{\Pr(A_k)\Pr(B \mid A_k)}{\Pr(B)}$$

Independent Events

1.7.2 Independent Events

Definition:

- Two events A and B are said to be ***independent*** if and only if
$$\Pr(A \cap B) = \Pr(A)\Pr(B).$$

- Two events A and B that are not independent are said to be **dependent**.

Independent Events (Continued)

- Hence the events A and B are independent if the occurrence of one event does not in any way influence (or associate with) the occurrence of the other event.
- Therefore
$$A \text{ is independent of } B$$
$$\Leftrightarrow B \text{ is independent of } A$$
$$\Leftrightarrow \Pr(A \cap B) = \Pr(A)\Pr(B).$$

1.7.3 Properties of Independent Events

1. Suppose $\Pr(A) > 0, \Pr(B) > 0$.
   If A and B are independent, then
   $$\Pr(B|A) = \Pr(B) \text{ and } \Pr(A|B) = \Pr(A).$$
   The above equalities are sometimes used as the definition of two independent events.

   That is, the conditional probability of event B given event A has occurred is the same as the unconditional probability of event B.

Properties of Independent Events (continued)

2. Suppose $\Pr(A) > 0, \Pr(B) > 0$.

   If A and B are **independent** events, then events A and B **cannot be mutually exclusive**.

   Proof: Since A and B are independent events, $\Pr(A) > 0$ and $\Pr(B) > 0$, therefore
   $$\Pr(A \cap B) = \Pr(A)\Pr(B) > 0.$$
   Since $\Pr(A \cap B) \neq 0$, therefore $A \cap B \neq \emptyset$.

   Hence A and B are not mutually exclusive events.

Properties of Independent Events (Continued)

3. Suppose $\Pr(A) > 0, \Pr(B) > 0$.

   If A and B are **mutually exclusive**, then A and B **cannot be independent**.

   Proof: A and B are **mutually exclusive** events implies
   $$\Pr(A \cap B) = 0.$$
   On the other hand, $\Pr(A) > 0$ and $\Pr(B) > 0$ implies
   $$\Pr(A)\Pr(B) > 0$$
   Therefore $\Pr(A \cap B) \neq \Pr(A)\Pr(B)$ and hence events A and B are not independent.

Properties of Independent Events (Continued)

4. The sample space S as well as the empty set $\emptyset$ are independent of any event.

   Proof: For any event A, $A \cap S = A$, and $A \cap \emptyset = \emptyset$.
   $$\Pr(A \cap S) = \Pr(A) = \Pr(A)\Pr(S)$$
   since $\Pr(S) = 1$.
   $$\Pr(A \cap \emptyset) = \Pr(\emptyset) = \Pr(\emptyset)\Pr(A)$$
   since $\Pr(\emptyset) = 0$.

Properties of Independent Events (Continued)

5. If $A \subset B$, then A and B are dependent unless $B = S$.

   Proof: $A \subset B$ implies that $A \cap B = A$.

   $\Pr(A \cap B) = \Pr(A) \neq \Pr(A)\Pr(B)$ unless $\Pr(B) = 1$ which implies that $B = S$.

Some remarks about independent events

- The properties of independence, unlike the mutually exclusive property, **cannot** be shown on a Venn diagram.

  This means you can't trust your intuition.

  In general, the only way to check for independence is by checking with the definition for independence, namely $\Pr(A \cap B) = \Pr(A)\Pr(B)$.
- Mutually exclusive events are dependent events since $\Pr(A)\Pr(B) \neq \Pr(A \cap B) = 0$.

Theorem

- If A and B are **independent**, then so are A and B', A' and B, A' and B'.

N Independent Events (Pairwise Independent Events)

1.7.4 n Independent Events

Pairwise Independent Events

Definition :

- A set of events $A_1, A_2, \cdots, A_n$ are said to be **pairwise independent** if and only if

$$\Pr(A_i \cap A_j) = \Pr(A_i)\Pr(A_j)$$

for $i \neq j$ and $i, j = 1, \cdots, n$.

n Mutually Independent Events

- The events $A_1, A_2, \cdots, A_n$ are called **mutually independent** (or simply **independent**)

  if and only if for **any subset** $\{A_{i_1}, A_{i_2}, \cdots, A_{i_k}\}$ of $A_1, A_2, \cdots, A_n$,

$$\Pr(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \Pr(A_{i_1})\Pr(A_{i_2})\cdots\Pr(A_{i_k})$$

Remarks

1. When one says that events $A_1, A_2, \cdots, A_n$ are mutually independent, it means that
   - firstly, for any pair of events A_j, A_k where $j \neq k$, the multiplication rule holds, and
   - secondly, for any three events A_i, A_j and A_k for distinct i, j, k, the multiplication rule holds, and so on.
   - Of course, the following multiplication rule also holds: $\Pr(A_1 \cap A_2 \cap \cdots \cap A_n) = \Pr(A_1)\Pr(A_2)\cdots\Pr(A_n)$
   - There are in total $2^n - n - 1$ different cases.

Remarks (Continued)

2. Mutually independence implies pairwise independence, but pairwise independence does **not** imply mutually independence.

3. Suppose $A_1, A_2, \cdots, A_n$ are mutually independent events. Let
$$B_i = A_i \text{ or } A_i', \ \ i = 1, 2, \cdots, n.$$
Then $B_1, B_2, \cdots, B_n$ are also mutually independent events.

CHAPTER 2: CONCEPTS OF RANDOM VARIABLE

Random Variable

2.1.1 Random Variable

Definition 2.1

- Let S be a sample space associated with the experiment, E.
- A *function* X, which assigns a number to every element $s \in S$, is called a **random variable**.

Random Variable (Continued)

Notes:

1. X is a real-valued function.
2. The range space of X is the set of real numbers
$$\boldsymbol{R_X} = \{x \mid x = X(s), s \in S\}.$$
Each possible value x of X represents an event that is a subset of the sample space S.
3. If S has elements that are themselves real numbers, we take $X(s) = s$. In this case $\boldsymbol{R_X} = S$.

Equivalent Events

2.1.2 Equivalent Events

Definition 2.2

- Let E be an experiment and S its sample space.
- Let X be a random variable defined on S and $\boldsymbol{R_X}$ be its range space.
  That is, $X : S \to \mathbb{R}$
- Let B be an event with respect to $\boldsymbol{R_X}$;
  That is $B \subset \boldsymbol{R_X}$.

Equivalent Events (Continued)

Definition 2.2 (Continued)

- Suppose that A is defined as
$$A = \{s \in S \mid X(s) \in B\}.$$
In words: A consists of all sample points, s, in S for which $X(s) \in B$.
- In this case we say that A and B are **equivalent events** and
$$\Pr(B) = \Pr(A).$$

Discrete Random Variable

2.2.1 Discrete Random Variable
Definition 2.3

- Let X be a random variable.
- If the number of possible values of X (i.e., $\boldsymbol{R_X}$, the range space) is **finite or countable infinite**, we call X a **discrete** random variable.
- That is, the possible values of X may be listed as $x_1, x_2, x_3, \cdots$.

Probability Function

2.2.2 Probability Function

- For a discrete random variable, each value of X has a certain probability $f(x)$.
- Such a function $f(x)$ is called the **probability function**, p.f. (or **probability mass function**, p.m.f.).
- The collection of pairs $(x_i, f(x_i))$ is called the **probability distribution** of X.

Probability Function (Continued)

The probability of $X = x_i$ denoted by $f(x_i)$ (i.e. $f(x_i) = \Pr(X = x_i)$, must satisfy the following two conditions.

(1) $f(x_i) \geq 0$ for all x_i.

(2) $\sum_{i=1}^{\infty} f(x_i) = 1$.

Continuous Random Variable

2.3.1 Continuous Random Variable
Definition 2.4

- Suppose that $\boldsymbol{R_X}$, the range space of a random variable, X, is an **interval or a collection of intervals**.
- Then we say that X is a **continuous random variable.**

Probability Density Function

2.3.2 Probability Density Function

Definition 2.5

- Let X be a continuous random variable.
- The **probability density function (p.d.f.)** $f(x)$, is a function, $f(x)$, satisfying the following conditions:
  1. $f(x) \geq 0$ for all $x \in \boldsymbol{R_X}$,
  2. $\int_{\boldsymbol{R_X}} f(x)\,dx = 1$ or $\int_{-\infty}^{\infty} f(x)\,dx = 1$

     since $f(x) = 0$ for x not in $\boldsymbol{R_X}$.

Probability Density Function (Continued)

Definition 2.5 (Continued)

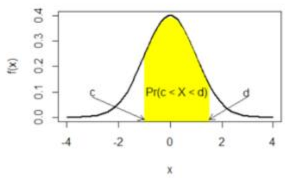
3. For any c and d such that $c < d$, (i.e. $(c, d) \subset \boldsymbol{R_X}$),

$$\Pr(c \leq X \leq d) = \int_c^d f(x)dx$$

Remarks

1. $\Pr(c \leq X \leq d) = \int_c^d f(x)dx$

   represents the area under the graph of the p.d.f. $f(x)$ between $x = c$ and $x = d$.



Remarks (Continued)

2. For any specified value of X, say x_0, we have

$$\Pr(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$$

Hence in the **continuous** case, the probability of X equals to a fixed value is 0 and

$\Pr(c \leq X \leq d) = \Pr(c \leq X < d) = \Pr(c < X \leq d) = \Pr(c < X < d)$.

Therefore in the continuous case, $\leq$ **and** $<$ can be used **interchangeably** in a probability statement.

*If x is continuous, Pr(X <= 5) is the same as Pr(X < 5)
If x is discrete, may not be the same

Remarks (Continued)

3. $\Pr(A) = 0$ does **not** necessary imply $A = \emptyset$.

4. If X assumes values only in some interval $[a, b]$, we may simply set $f(x) = 0$ for all X outside $[a, b]$.

Cumulative Distribution Function

2.4 Cumulative Distribution Function

Definition 2.6

- Let X be a random variable, discrete or continuous.

- We define $F(x)$ to be the **cumulative distribution function** of the random variable X (abbreviated as c.d.f.) where

$$F(x) = \Pr(X \leq x).$$

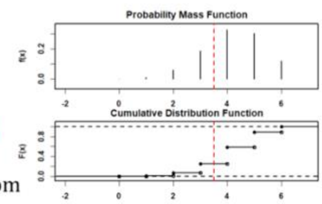
CDF for Discrete Random Variables

2.4.1 CDF for Discrete Random Variables

- If X is a **discrete random variable**, then

$$F(x) = \sum_{t \leq x} f(t)$$
$$= \sum_{t \leq x} \Pr(X = t)$$

- The c.d.f. of a discrete random variable is a step function.



CDF for Discrete Random Variables (Continued)

- For any two numbers a and b with $a \leq b$.

$$\Pr(a \leq X \leq b) = \Pr(X \leq b) - \Pr(X < a)$$
$$= F(b) - F(a^-)$$

where "a^-" represents the largest possible value of X value that is strictly less than a.

CDF for Discrete Random Variables (Continued)

- In particular, if the only possible values are **<u>integers</u>** and if a and b are integers, then

$$\Pr(a \leq X \leq b) = \Pr(X = a \text{ or } a+1 \text{ or } \cdots \text{ or } b)$$

Also $\Pr(a \leq X \leq b) = F(b) - F(a-1)$

- Taking $a = b$ yields

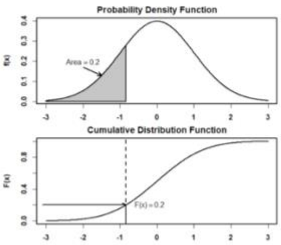
$$\Pr(X = a) = F(a) - F(a-1).$$

CDF for Continuous Random Variables

2.4.2 CDF for Continuous Random Variables

- If X is a **continuous random variable**, then

$$F(x) = \int_{-\infty}^{x} f(t)dt$$



CDF for Continuous Random Variables (Continued)

- For a **continuous random variable** X,

$$f(x) = \frac{d\,F(x)}{dx}$$

if the derivative exists.

- Also,

$$\Pr(a \leq X \leq b) = \Pr(a < X \leq b)$$
$$= F(b) - F(a).$$

CDF for Continuous Random Variables (Continued)

Remarks:

- $F(x)$ is a **non-decreasing** function.
  That is, if $x_1 < x_2$, then $F(x_1) \leq F(x_2)$.

- $0 \leq F(x) \leq 1$.

Mean or Expected Values of Discrete Random Variable

2.5.1 Expected Values

Definition 2.7a

- If X is a **discrete** random variable. taking on values $x_1, x_2, \cdots$ with probability function $f_X(x)$,
- then the **mean** or **expected value** (or **mathematical expectation**) of X, denoted by $\boldsymbol{E(X)}$ as well as by μ_X, is defined by

$$\mu_X = E(X) = \sum_i x_i\, f_X(x_i) = \sum_x x\, f_X(x)$$

NUS

Mean and Variance of a Random Variable (Continued)

2.5.1 Expected Values (Continued)

Note : The expected value is not necessarily a possible value of the random variable X.

Mean or Expected Values of Continuous Random Variable

NUS

Expected Values (Continued)

Definition 2.7b

- If X is a **continuous** random variable with probability density function $f_X(x)$, the mean of X is defined by

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

- Roughly speaking, the mathematical expectation is an "average" (or more precisely, a "weighted average").

NUS

Remarks

1. The expected value exists provided the sum or the integral in the above definitions exists.

2. In the discrete case, if $f_X(x) = 1/N$ for each of the N values of x, hence the mean,

$$E(X) = \sum_i x_i\, f(x_i) = \frac{1}{N}\sum_i x_i\,,$$

becomes the average of the N items.

Expectation of a Function of a Random Variable (note: μx is the mean)

2.5.2 Expectation of a Function of a RV

Definition 2.8

For any function $g(X)$ of a random variable X with p.f. (or p.d.f.) $f_X(x)$,

(a) $E[g(X)] = \sum_x g(x) f_X(x)$

if X is a **discrete** r.v. providing the sum exists; and

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

if X is a **continuous** r.v. providing the integral exists.

Some Special Cases

1. $g(x) = (x - \mu_X)^2$.

This leads to the definition of variance of a given random variable X.

Some Special Cases (Continued)

Definition 2.9

- Let X be a random variable with p.f. (or p.d.f.) $f(x)$, then the **variance** of X is defined as

$$\sigma_X^2 = V(X) = E[(X - \mu_X)^2]$$

$$= \begin{cases} \sum_x (x - \mu_X)^2 f_X(x), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Some Special Cases (Continued)

Remarks:

(a) $V(X) \geq 0$.

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

- The positive square root of the variance is called the **standard deviation** of X. That is

$$\sigma_X = \sqrt{V(X)}$$

Some Special Cases (Continued)

2. $g(x) = x^k$.

Then $E(X^k)$ is called the ***k*-th moment of *X***.

Properties of Expectation

2.5.3 Properties of Expectation

Property 1

$$E(aX + b) = a\,E(X) + b,$$

where a and b are constants.

Properties of Expectation (Continued)

Two special cases:

(a) Put $b = 0$, we have $E(aX) = a\,E(X)$.

(b) Put $a = 1$, we have $E(X + b) = E(X) + b$.

In general,

$$E[a_1 g_1(X) + a_2 g_2(X) + \cdots + a_k g_k(X)]$$
$$= a_1 E[g_1(X)] + a_2 E[g_2(X)] + \cdots + a_k E[g_k(X)]$$

where $a_1, a_2, \cdots, a_k$ are constants.

Properties of Expectation (Continued)

Property 2

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

Properties of Expectation (Continued)

Property 3

$$V(aX + b) = a^2 V(X),$$

where a and b are constants

Chebyshev's Inequality

2.6 Chebyshev's Inequality

- If we know the probability distribution of a random variable X, we may then compute $E(X)$ and $V(X)$.
- However, the converse is not true. From the knowledge of $E(X)$ and $V(X)$ we cannot reconstruct the probability distribution of X and
- hence, we cannot compute quantities such as
  $$\Pr(|X - E(X)| \le c\,),$$
  where c is a positive constant.
  [Note: $\Pr(|Y| \le a)$ is equivalent to $\Pr(-a \le Y \le a)$.

Chebyshev's Inequality (Continued)

- Nevertheless, the Russian mathematician Chebyshev gave a very useful upper (or lower) **bound** to such probability.
- This result is known as **Chebyshev's inequality**.

Chebyshev's Inequality (Continued)

- Let X be a random variable (discrete or continuous) with $E(X) = \mu$ and $V(X) = \sigma^2$.
- Then for any positive number k we have
$$\Pr(|X - \mu| \geq k\sigma) \leq 1/k^2.$$
- That is, the probability that the value of X lies at least k standard deviation from its mean is at most $1/k^2$.
- Alternatively,
$$\Pr(|X - \mu| < k\sigma) \geq 1 - 1/k^2.$$

Remarks

1. The quantity k in Chebyshev's Inequality can be any positive number.

2. This inequality is true for **all** distributions with finite mean and variance.

3. The theorem gives a **lower bound** on the probability that $|X - \mu| < k\sigma$. No guarantee that this lower bound is close to the exact probability.

Note: always check that the probability function is well defined i.e. all probabilities >= 0, all probabilities add up to 1.

CHAPTER 3: 2-D RANDOM VARIABLES AND CONDITIONAL PROBABILITY DISTRIBUTIONS

Two Dimensional Random Variables

Two Dimensional Random Variables (Continued)

Definition 3.3

1. (X, Y) is a two-dimensional **discrete** random variable if the possible values of $(X(s), Y(s))$ are finite or countable infinite.

   i.e. the possible values of $(X(s), Y(s))$ may be represented as
   $$(x_i, y_j), i = 1, 2, 3, \cdots; j = 1, 2, 3, \cdots$$

2. (X, Y) is a two-dimensional **continuous** random variable if the possible values of $(X(s), Y(s))$ can assume all values in some region of the Euclidean plane $\mathbb{R}^2$.

Joint Probability Function for Discrete Random Variables

3.2.1 Joint Probability Function for Discrete RVs

Definition 3.4

- Let (X, Y) be a 2-dimensional **discrete** random variable defined on the sample space of an experiment. With each possible value (x_i, y_j), we associate a number $f_{X,Y}(x_i, y_j)$ representing $\Pr(X = x_i, Y = y_j)$ and satisfying the following conditions:

1. $f_{X,Y}(x_i, y_j) \geq 0$ for all $(x_i, y_j) \in R_{X,Y}$.
2. $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \Pr(X = x_i, Y = y_j) = 1$ (3.1)

Joint Probability Function (Continued)

- The function $f_{X,Y}(x, y)$ defined for all pairs of values $(x_i, y_j) \in R_{X,Y}$ is called the joint probability function of (X, Y).
- Let A be any set consisting of pairs of (x, y) values. Then the probability $\Pr((X, Y) \in A)$ is defined by summing the joint probability function over pairs in A :

$$\Pr\big((X, Y) \in A\big) = \underbrace{\sum\sum}_{(x,y)\in A} f_{X,Y}(x, y)$$

Joint Probability Function (pdf) for Continuous Random Variables

3.2.2 Joint pdf for Continuous RVs

- Let (X, Y) be a 2-dimensional **continuous** random variable assuming all values in some region R of the Euclidean plane, $\mathbb{R}^{\mathbf{2}}$.
- $f_{X,Y}(x, y)$ is called a **joint probability density function** if it satisfies the following conditions:

Joint pdf for Continuous RVs (Continued)

1. $f_{X,Y}(x, y) \geq 0$ for all $(x, y) \in R_{X,Y}$.
2.

$$\iint_{(x,y)\in R_{X,Y}} f_{X,Y}(x, y)dx\, dy = 1$$

or

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)dx\, dy = 1.$$

Marginal Probability Distributions

3.3.1 Marginal probability distributions

Definition 3.6

- Let (X, Y) be a 2-dimensional discrete (or continuous) random variable with joint probability function (or joint probability density function) $f_{X,Y}(x, y)$.
- The **marginal probability distributions** of X and Y are respectively given by:

Marginal Distributions (Continued)

- For **discrete** case,

$$f_X(x) = \sum_y f_{X,Y}(x, y) \quad \text{and} \quad f_Y(y) = \sum_x f_{X,Y}(x, y)$$

- For **continuous** case,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$$

Conditional Distribution

Conditional Distribution (Continued)

Definition 3.7 (Continued)

- Then **the conditional distribution of Y given that $X = x$** is given by

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

for each x within the range of X.

Conditional Distribution (Continued)

Definition 3.7 (Continued)

- Similarly, the **conditional probability distribution of X given $Y = y$** is

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

for each y within the range of Y.

Remarks

1. The conditional p.f.'s (p.d.f.'s) satisfy all the requirements for a 1-dimensional p.f. (p.d.f.). Thus, we have
   (a) For a fixed y,
   $$f_{X|Y}(x|y) \geq 0$$
   and for a fixed x,
   $$f_{Y|X}(y|x) \geq 0.$$

Remarks (Continued)

1. (b)
   For discrete r.v.'s,
   $$\sum_x f_{X|Y}(x|y) = 1 \quad \text{and} \quad \sum_y f_{Y|X}(y|x) = 1.$$
   For continuous r.v.'s
   $$\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx = 1 \quad \text{and} \quad \int_{-\infty}^{\infty} f_{Y|X}(y|x)dy = 1.$$

Remarks (Continued)

2. For $f_X(x) > 0$,
$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x).$$

For $f_Y(y) > 0$,
$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y).$$

Independent Random Variables

3.4 Independent Random Variables

3.4.1 Definition of Independent RVs

Definition

- Random variables X and Y are **independent** if and only if

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y), \quad \text{for all } x, y.$$

Extension:

- Random variables $X_1, X_2, \cdots, X_n$ are independent if and only if

$$f_{X_1,X_2,\cdots,X_n}(x_1,x_2,\cdots,x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

for all $x_i, i = 1, \cdots, n$.

Remark

- The product of 2 positive functions $f_X(x)$ and $f_Y(y)$ means a function which is positive on a **product space**.

- That is, if

$$f_X(x) > 0, \quad \text{for } x \in A_1 \quad \text{and}$$
$$f_Y(y) > 0, \quad \text{for } y \in A_2$$

then $f_X(x) f_Y(y) > 0$, for $(x, y) \in A_1 \times A_2$.

Expectation, covariance of X,Y

3.5 Expectation

Definition 3.5.1

- The expectation of $g(X,Y)$ is defined as

$$E[g(X,Y)] = \begin{cases} \displaystyle\sum_x \sum_y g(x,y) f_{X,Y}(x,y), & \text{for Discrete RV's,} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy, & \text{for Cont. RV's} \end{cases}$$

You need this for E(XY) to find Cov(X,Y)

A Special Case

- Let $g(X,Y) = (X - \mu_X)(Y - \mu_Y)$. This leads to the definition of covariance between two random variables.

Definition 3.5.2

- Let (X,Y) be a bivariate random vector with joint p.f. (or p.d.f.) $f_{X,Y}(x,y)$, then the **covariance** of (X,Y) is defined as

$$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

A Special Case (Continued)

- For **discrete** case

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f_{X,Y}(x,y)$$

- For **continuous** case

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f_{X,Y}(x,y)\,dx\,dy$$

Remarks

variance, covariance, expectation

1. $Cov(X,Y) = E(XY) - \mu_X \mu_Y$.
2. If X and Y are independent, then $Cov(X,Y) = 0$. However $Cov(X,Y) = 0$ does not imply X and Y are independent.
3. $Cov(aX+b, cY+d) = ac\,Cov(X,Y)$
4. $V(aX+bY) = a^2V(X) + b^2V(Y) + 2ab\,Cov(X,Y)$

Correlation Coefficient

Correlation coefficient

Definition 3.5.2

- The **correlation coefficient** of X and Y, denoted by $Cor(X,Y)$, $\rho_{X,Y}$ or ρ, is defined by

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

Remarks on Correlation coefficient

1. $-1 \le \rho_{X,Y} \le 1$.
2. $\rho_{X,Y}$ is a measure of the degree of **linear** relationship between X and Y.
3. If X and Y are independent, then $\rho_{X,Y} = 0$.
   On the other hand, $\rho_{X,Y} = 0$ does **not** imply independence.

Note: E(Y – X) = E(Y) – E(X)
Note: Cov(X, Y) = E(XY) – E(X)E(Y)
Note: Var(X) = E(X^2) – [E(X)]^2

CHAPTER 4: SPECIAL PROBABILITY DISTRIBUTIONS

Discrete Uniform Distribution (Discrete distribution)

4.1 Discrete Uniform Distribution

Definition 4.1

- If the random variable X assumes the values $x_1, x_2, \cdots, x_k$, with equal probability,
- then the random variable X is said to have a discrete uniform distribution and the probability function is given by

$$f_X(x) = \frac{1}{k}, \qquad x = x_1, x_2, \cdots, x_k,$$

and 0 otherwise.

Mean and Variance of Discrete Uniform Distribution

Theorem 4.1

The mean and variance of the **discrete uniform distribution** are

$$\mu = E(X) = \sum_{\text{all } x} x\, f_X(x) = \sum_{i=1}^{k} x_i \frac{1}{k} = \frac{1}{k}\sum_{i=1}^{k} x_i\,,$$

Mean and Variance of Discrete Uniform Distribution (Continued)

Theorem 4.1 (Continued)

$$\sigma^2 = V(X) = \sum_{\text{all } x} (x-\mu)^2 f_X(x) = \frac{1}{k}\sum_{i=1}^{k} (x_i - \mu)^2$$

or

$$\sigma^2 = E(X^2) - \mu^2 = \frac{1}{k}\left(\sum_{i=1}^{k} x_i^2\right) - \mu^2$$

Bernoulli Distribution (Discrete distribution) – Only two possible outcomes

4.2.1 Bernoulli Distribution

- A **Bernoulli experiment** is a random experiment with only two possible outcomes, say ‘success’ or ‘failure’ (e.g. head or tail, defective or non-defective, boy or girl, yes or no.).

4.2 Bernoulli and Binomial Distributions

Definition 4.2

- A random variable X is defined to have a Bernoulli distribution if the probability function of X is given by
  $f_X(x) = p^x(1-p)^{1-x}, \quad x = 0, 1;$
  where the parameter p satisfies $0 < p < 1$.
  $f_X(x) = 0$ for other X values.
- $(1-p)$ is often denoted by q.
- $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p = q$.

Bernoulli Distribution

Theorem 4.2

If X has a Bernoulli distribution, then the mean and variance of X are

$$\mu = E(X) = p,$$

and

$$\sigma^2 = V(X) = p(1-p) = pq.$$

Binomial Distribution (Discrete distribution) – Bernoulli but repeated n

4.2.2 Binomial Distributions

Definition 4.3

- A random variable X is defined to have a **binomial distribution** with two parameters n and p, (i.e. $X \sim B(n,p)$), if the probability function of X is given by
$$\Pr(X = x) = f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} p^x q^{n-x},$$
for $x = 0, 1, \cdots, n$, where p satisfies $0 < p < 1$, $q = 1 - p$, and n ranges over the positive integers.
- X is the **number of successes** that occur in n **independent Bernoulli trials**.

Binomial Distributions (Continued)

Note:

- When $n = 1$, the probability distribution of X becomes
$$f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1.$$
- Therefore Bernoulli distribution is a special case of the binomial distribution.

Mean and Variance of Binomial Distributions

Theorem 4.3

- If X has a binomial distribution with parameters n and p, (i.e. $X \sim B(n,p)$)
- then the mean and variance of X are
$$\mu = E(X) = np$$
and
$$\sigma^2 = V(X) = np(1-p) = npq.$$

Conditions for a Binomial Experiment

1. It consists of n repeated Bernoulli trials.
2. Only two possible outcomes: success and failure in **each** trial
3. Pr(success) $= p$ is the same constant in each trial.
4. Trials are **independent**.

- The random variable X is the **number of successes** among the ***n*** **trials** in a binomial experiment.
- Then $X \sim B(n,p)$.

Note: Pr(X = x) = nCx * p^x * (1-p)^(n-x)

Note: Pr(X >= 8) = 1 – Pr(X <= 7)

Negative Binomial Distribution (Binomial, but repeated until a fixed no. of success occur)
Note: The last trial must be a success

4.3 Negative Binomial Distribution

- Let us consider an experiment where the properties are the same as those listed for a binomial experiment, with the exception that the trials will be repeated until a **fixed** number of successes occur.
- We are interested in the probability of the k-th success occurs on the x-th trial where x is the random variable. (Notice that in Binomial distribution, we are interested in the probability of x successes in n trials)

Negative Binomial Distribution (Continued)

- Let X be a random variable represents the number of trials to produce the k successes in a sequence of independent Bernoulli trials
- The random variable X is said to follow a Negative Binomial distribution with parameters k and p (i.e. $NB(k,p)$.)
- The probability function of X is given by

$$\Pr(X = x) = f_X(x) = \binom{x-1}{k-1} p^k q^{x-k},$$

for $x = k, k+1, k+2, \cdots$

Negative Binomial Distribution (Continued)

- If $X \sim NB(k,p)$, then it can be shown that

$$E(X) = \frac{k}{p}$$

and

$$Var(X) = \frac{(1-p)k}{p^2}$$

Note: k is the number of successes to produce, p is the probability of success

Poisson Distribution (Discrete distribution) – no. of successes occurring during time interval

4.4 Poisson Distribution

- Experiments yielding numerical values of a random variable X, the number of successes occurring during a given time interval or in a specified region, are called **Poisson experiments**.
- The given time interval, t, may be of any length, such as a minute, a day, a week, a month, or even a year.

Poisson Experiment

A Poisson experiment is one that possesses the following properties:

1. The number of successes occurring in one time interval or specified region are **independent** of those occurring in any other disjoint time interval or region of space.
2. The probability of a single success occurring during a very short time interval or in a small region is **proportional** to the length of the time interval or the size of the region and does not depend on the number of successes occurring outside this time interval or region.

Poisson Experiment (Continued)

3. The probability of more than one success occurring in such a short time interval or falling in such a small region is negligible.

Poisson Distribution

Definition 4.4

- The number of successes X in a Poisson experiment is called a **Poisson** random variable.
- The probability distribution of the Poisson random variable X, is called the Poisson distribution and the probability function is given by

$$f_X(x) = \Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, 3, \cdots$$

where λ is the average number of successes occurring in the given time interval or specified region and $e \approx 2.1718281818\cdots$

Mean and Variance of Poisson RV

Theorem 4.4

If X has a **Poisson** distribution with parameter λ, then

$$E(X) = \lambda$$

and

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

Note: λ = average number of successes in given time period
Eg. Average number of robberies in a day is 4.
To find probability of six robberies occurring in two days, Pr(X=6), where X~P(2*4=8)

If want to find probability of X in n time periods, first find the probability of X in one time period, then a new variable Y will follow the binomial distribution with p = the probability

Poisson Approximation to Binomial Distribution

4.5 Poisson Approximation to the Binomial Distribution

Theorem 4.5

- Let X be a **Binomial** random variable with parameters n and p. That is
$$\Pr(X = x) = f_X(x) = {}_nC_x\, p^x q^{n-x}, \text{ where } q = 1 - p.$$
- Suppose that $n \to \infty$ and $p \to 0$ in such a way that $\lambda = np$ remains a constant as $n \to \infty$.
- Then X will have approximately a Poisson distribution with parameter np. That is
$$\lim_{\substack{p\to 0\\ n\to\infty}} \Pr(X = x) = \frac{e^{-np}(np)^x}{x!}$$

Continuous Uniform Random Variable (Continuous variable)

4.6 Continuous Uniform Distribution

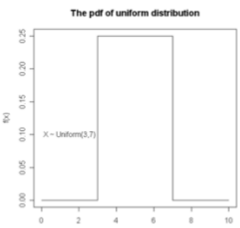
Definition 4.5

- A random variable is said to have a **uniform** distribution over the interval $[a, b], -\infty < a < b < \infty$, denoted by $U(a, b)$, if its probability density function is given by
$$f_X(x) = \frac{1}{b-a}, \qquad \text{for } a \le x \le b,$$
and 0 otherwise.

Continuous Uniform Distribution (Continued)

Definition 4.5 (Continued)

- This distribution is also referred to as rectangular distribution because of the rectangular shape of the p.d.f.



Mean and Variance of Cont Uniform RV

Theorem 4.6

If X is uniformly distributed over $[a, b]$, then
$$E(X) = \frac{a+b}{2}, \qquad \text{and} \quad V(X) = \frac{1}{12}(b-a)^2.$$

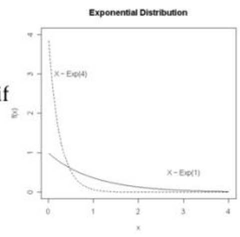
Eg. X~U(0, 2)
Then, fx(X) = ½ for 0 <= x <= 2
Pr(1 <= X <= 3/2) = Int(1/2 dx) from 1 to 3/2

Exponential Distribution (Continuous variable)

4.7 Exponential Distribution

Definition 4.6

- A continuous random variable X assuming all nonnegative values is said to have an exponential distribution with parameter $\alpha > 0$ if its probability density function is given by
  $f_X(x) = \alpha e^{-\alpha x}$, for $x > 0$.
  and 0 otherwise.
- Note : $\int_{-\infty}^{\infty} f(x)\, dx = 1$



Mean & Variance of Exponential RV

Theorem 4.7

- If X has an **Exponential** distribution with parameter $\alpha > 0$, then

$$E(X) = \frac{1}{\alpha}$$

and

$$V(X) = \frac{1}{\alpha^2}.$$

i.e. a = 1/E(X), i.e. X ~ Exp(1/E(X))

Remark

- The p.d.f. can be written in the form

$$f_X(x) = \frac{1}{\mu} e^{-x/\mu}, \qquad \text{for } x > 0.$$

  and 0 otherwise.
- Then

$$E(X) = \mu \quad \text{and} \quad V(X) = \mu^2.$$

No Memory Property of Exponential Distribution

No Memory Property of Exponential Distribution

Theorem 4.8

Suppose that X has an **exponential** distribution with parameter $\alpha > 0$.

Then for any two positive numbers s and t, we have

$$\Pr(X > s + t \mid X > s) = \Pr(X > t).$$

Note:

$$\Pr(X > t) = \int_t^{\infty} \alpha e^{-\alpha x}\, dx = [-e^{-\alpha x}]_{x=t}^{\infty} = e^{-\alpha t}$$

for t > 0

No Memory Property of Exponential Distribution (Continued)

The above theorem states that the **exponential** distribution has '**no memory**' in the following sense:

- Let X denote the life length of a bulb.
- Given that the bulb has lasted s time units (i.e. $X > s$),
- then the probability that it will last for the next t units (i.e. $X > s + t$) is the same as the probability that it will last for the first t units as brand new.

Normal Distribution (Continuous variable)

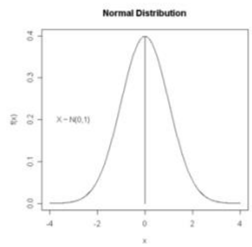
4.8 Normal Distribution

Definition 4.7

- The random variable X assuming all real values, $-\infty < x < \infty$, has a **normal** (or **Gaussian**) distribution if its probability density function is given by
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad -\infty < x < \infty,$$
where $-\infty < \mu < \infty$ and $\sigma > 0$.
- It is denoted by $N(\mu, \sigma^2)$.
- μ and σ are called parameters of the normal distribution.

Properties of the normal distribution

1. The graph of this distribution is of bell-shaped and called the normal curve and it is symmetrical about the vertical line $x = \mu$.

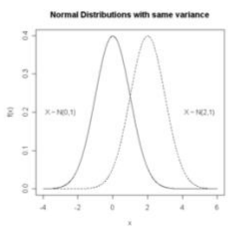


Properties of the normal distribution (Continued)

2. The maximum point occurs at $x = \mu$ and its value is
$$\frac{1}{\sqrt{2\pi}\sigma}$$
3. The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.
4. The total area under the curve and above the horizontal axis is equal to 1.
$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 1.$$

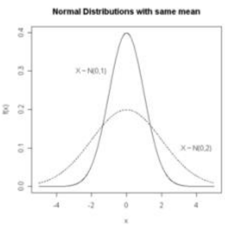
Properties of the normal distribution (Continued)

5. It can be shown that $E(X) = \mu$ and $V(X) = \sigma^2$.
6. Two normal curves are identical in shape if they have the same σ^2. But they are centered at different positions when their means are different.



Properties of the normal distribution (Continued)

7. As σ increases, the curve flattens; and as σ decreases, the curve sharpens.



Standardization of Normal Distribution

Properties of the normal distribution (Continued)

8. If X has distribution $N(\mu, \sigma^2)$, and if
$$Z = \frac{(X - \mu)}{\sigma}$$
then Z has the $N(0, 1)$ distribution.
That is, $E(Z) = 0$ and $V(Z) = 1$.
We say that Z has a standardized normal distribution.
That is, the p.d.f. of Z may be written as
$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).$$

NOTE:ALWAYS STANDARDIZE IT

Properties of the normal distribution (Continued)

- The importance of the standardized normal distribution is the fact that it is tabulated.
- Whenever X has distribution $N(\mu, \sigma^2)$, we can always simplify the process of evaluating the values of $\Pr(x_1 < X < x_2)$ by using the transformation $Z = (X - \mu)/\sigma$.
  Hence $x_1 < X < x_2$ is equivalent to
  $$(x_1 - \mu)/\sigma < Z < (x_2 - \mu)/\sigma.$$
- Let $z_1 = (x_1 - \mu)/\sigma$ and $z_2 = (x_2 - \mu)/\sigma$. Then
  $$\Pr(x_1 < X < x_2) = \Pr(z_1 < Z < z_2).$$

Linear Interpolation (to get z-value from probability in table)
https://www.johndcook.com/interpolator.html

Example 1 (Continued)

* <u>Linear interpolation</u>

- Let $\Pr(Z > a) = 0.12$.
- From the normal table, we have
  $\Pr(Z \geq 1.17) = 0.121$ and $\Pr(Z \geq 1.18) = 0.119$.
- Hence $\frac{a - 1.17}{1.18 - 1.17} = \frac{0.12 - 0.121}{0.119 - 0.121}$
  $$\Rightarrow \quad a = 1.17 + 0.01\left(\frac{-0.001}{-0.002}\right) = 1.175.$$

Normal Approximation to Binomial Distribution

4.9 Normal approximation to the binomial distribution

- When $n \to \infty$ and $p \to 0$, we may use Poisson distribution to approximate a binomial distribution as has been shown in Section 4.5.
- When $n \to \infty$ and $p \to 1/2$, we can also use normal distribution to approximate the binomial distribution. In fact, even when n is small and p is not extremely close to 0 or 1, the approximation is fairly good.
- A good rule of thumb is to use the normal approximation only when

$$np > 5 \quad \text{and} \quad n(1-p) > 5.$$

Theorem

Theorem

- If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$,
- then as $n \to \infty$,

$$Z = \frac{X - np}{\sqrt{npq}} \text{ is approximately } \sim N(0,1)$$

CHAPTER 5: SAMPLING & SAMPLING DISTRIBUTIONS

Simple Random Sample

5.2 Random Sampling

5.2.1 Simple Random Sample

- A set of n members taken from a given population is called a **sample** of size n.
- A **simple random sample** of n members is a sample that is chosen in such a way that every subset of n observations of the population has the same probability of being selected.

Sampling from Finite population Without replacement

5.2.2 Sampling from a finite population

1. Sampling without replacement

- Given a population, say, {A, B, C, D}, we have the following ${}_4C_2 = 6$ possible samples of size 2.
- They are

  {A, B}, {A, C}, {A, D}, {B, C}, {B, D}, {C, D}.
- Here the order of the letters is disregarded. It is the case of combinations.

Sampling from a finite population (Continued)

1. Sampling without replacement (Continued)

- In general, there are ${}_NC_n$ samples of size n that can be drawn from a finite population of size N **without replacement**.
- Each sample has equal chance of being selected. Hence each sample has a probability of $\frac{1}{{}_NC_n}$, of being selected.

Sampling from Finite population With replacement

Sampling from a finite population (Continued)

2. Sampling with replacement

- Using the same population, {A, B, C, D}, there are $4^2 =$ 16 samples of size 2 as follows:

  (A, A), (A, B), (A, C), (A, D), (B, A), (B, B), (B, C), (B, D),
  (C, A), (C, B), (C, C), (C, D), (D, A), (D, B), (D, C), (D, D).

Sampling from a finite population (Continued)

2. Sampling with replacement

- Here the order of the letters is taken into consideration. Hence, (A, B) and (B, A) are considered as two different samples. (Why?)
- In general, there are N^n samples of size n that can be drawn from a finite population of size N **with replacement**. Therefore, each sample has the same probability, $1/N^n$, of being selected.

Sampling from infinite population

Example 2

- We would also be sampling from an infinite population if we sample with replacement from a finite population, and our sample would be random if
  (1) in each draw all elements of the population have the **same** probability of being selected, and
  (2) successive draws are **independent**.

Sampling Distribution of Sample Mean

Sampling Distribution of Sample Mean

Theorem 5.1

- For random samples of size n taken from an **infinite population** or from a **finite population with replacement** having population mean μ and population standard deviation σ,
- the **sampling distribution of the sample mean** $\bar{X}$ has its mean and variance given by

$$\mu_{\bar{X}} = \mu_X \quad \text{and} \quad \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}.$$

Theorem 5.1 (Continued)

That is,

$$E(\bar{X}) = E(X) \quad \text{and} \quad V(\bar{X}) = \frac{V(X)}{n}.$$

Law of Large Numbers

Law of Large Number (LLN)

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population having any distribution with mean μ and **finite** population variance σ^2. Then for any $\epsilon \in \mathbb{R}$,

$$P(|\bar{X} - \mu| > \epsilon) \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

Central Limit Theorem (CLT)

5.4 Central Limit Theorem and its applications

Central Limit Theorem

- Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population having any distribution with mean μ and finite population variance σ^2.
- The sampling distribution of the sample mean $\bar{X}$ is **approximately normal** with mean μ and variance σ^2/n if ***n* is sufficiently large**.
- Hence

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ follows approximately } N(0, 1)$$

n > 30 is considered sufficiently large

Theorem 5.2

- If $X_i, i = 1, 2, \cdots, n$ are $N(\mu, \sigma^2)$, then $\bar{X}$ is $N(\mu, \frac{\sigma^2}{n})$ regardless of the sample size n.
- Similarly, if $X_i, i = 1, 2, \cdots, n$ are approximately $N(\mu, \sigma^2)$, then $\bar{X}$ is approximately $N(\mu, \frac{\sigma^2}{n})$ regardless of the sample size n.

Sampling Distribution of Difference of Two Sample Means

5.5 Sampling distribution of the difference of two sample means

Theorem 5.3

- If independent samples of sizes n_1 (≥ 30) and n_2 (≥ 30) are drawn from two populations,
- with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively,
- then the sampling distribution of the differences of the sample means, $\bar{X}_1$ and $\bar{X}_2$, is approximately normally distributed with mean and standard deviation given by

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2 \quad \text{and} \quad \sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Remarks

1. Note that if both n_1 and n_2 are greater than or equal to 30, the normal approximation for the distribution of $\bar{X}_1 - \bar{X}_2$ is very good regardless of the shapes of the two population distributions.
2. $$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \text{ approx} \sim N(0, 1).$$

Chi-square Distribution

5.6 Chi-square distribution

Definition 5.3

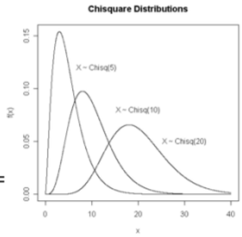
- If Y is a random variable with probability density function

$$f_Y(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}, \qquad \text{for } y > 0,$$

and 0 otherwise,
- then Y is defined to have a **chi-square distribution with n degrees of freedom**, denoted by $\chi^2(n)$, where n is a positive integer, and $\Gamma(\cdot)$ is the gamma function.

Chi-square distribution (Continued)

- The gamma function, $\Gamma(\cdot)$, is defined by
  $\Gamma(n) = \int_0^{\infty} x^{n-1}\, e^{-x}\, dx = (n-1)!$
  for $n = 1, 2, 3, \cdots$
- The p.d.f. of χ^2-distribution with $n = 5, 10$ and 25 are shown in the graph on the right side.



Some properties of Chi-square distributions

1. If $Y \sim \chi^2(n)$, then $E(Y) = n$ and $V(Y) = 2n$.
2. For large n, $\chi^2(n)$ approx $\sim N(n, 2n)$.
3. If $Y_1, Y_2, \cdots, Y_k$ are independent chi-square random variables with $n_1, n_2, \cdots, n_k$ degrees of freedom respectively, then $Y_1 + Y_2 + \cdots + Y_k$ has a chi-square distribution with $n_1 + n_2 + \cdots + n_k$ degrees of freedom. That is,
   $$\sum_{i=1}^{k} Y_i \sim \chi^2\left(\sum_{i=1}^{k} n_i\right)$$

Theorem 5.5

1. If $X \sim N(0,1)$, then $X^2 \sim \chi^2(1)$.
2. Let $X \sim N(\mu, \sigma^2)$, then $[(X-\mu)/\sigma]^2 \sim \chi^2(1)$.
3. Let $X_1, X_2, \cdots, X_n$ be a random sample from a normal population with mean μ, and variance σ^2. Define
   $$Y = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$$
   Then $Y \sim \chi^2(n)$

Excel Functions for χ^2-distribution

- The values on previous slide can be obtained using by built-in functions in Microsoft Excel:
- "=CHISQ.INV(α; n)" gives c such that $\Pr(W \le c) = \alpha$, where $W \sim \chi^2(n)$
- "=CHISQ.DIST(c; n; true)" gives α where $\Pr(W \le c) = \alpha$.
- For example,
  We have "=CHISQ.INV(0.95; 10)" gives 18.30703 and "=CHISQ.DIST(18.30703; 10; true)" gives 0.95.

Microsoft Excel Worksheet

Theorem 5.5

- If S^2 is the variance of a random sample of size n taken from a **normal** population having the variance σ^2,
- then the random variable
  $$\frac{(n-1)S^2}{\sigma^2}$$
- has a chi-square distribution with $n-1$ degrees of freedom. That is,
  $$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

T-distribution

5.8 The t-distribution

Definition 5.4

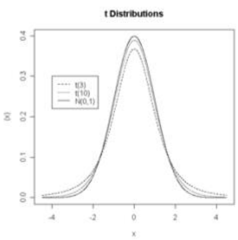
- Suppose $Z \sim N(0,1)$ and $U \sim \chi^2(n)$.
- If Z and U are independent, and let
$$T = \frac{Z}{\sqrt{U/n}}$$
- then the random variable T follows the t-distribution with n degrees of freedom. That is,
$$\frac{Z}{\sqrt{U/n}} \sim t(n)$$

The p.d.f of a t-distribution

- If T follows a t-distribution with n degrees of freedom,
- then its p.d.f. is given by
$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \qquad -\infty < t < \infty.$$
- The gamma function, $\Gamma(\cdot)$, is defined by
$$\Gamma(n) = \int_0^\infty x^{n-1}\, e^{-x}\, dx = (n-1)! \quad \text{for } n = 1, 2, 3, \cdots$$

Properties of a t-distribution

1. The graph of the t-distribution is symmetric about the vertical axis and resembles the graph of the standard normal distribution.



Properties of a t-distribution (Continued)

2. It can be shown that the p.d.f. of t-distribution with n d.f. is approaching to the p.d.f. of standard normal distribution when $n \to \infty$.

   That is
$$\lim_{n\to\infty} f_T(t) = \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2},$$
   as $n \to \infty$.

Properties of a t-distribution (Continued)

3. The values of
$$\Pr(T \geq t) = \int_t^\infty f_T(x)dx,$$
   for selected values of n and t are given in a statistical table.

Properties of a *t*-distribution (Continued)

3. (Continued)

   For example,

   - $\Pr\left(T \geq t_{10;0.05}\right) = 0.05$ gives $t_{10;0.05} = 1.812$.
   - $\Pr\left(T \geq t_{10;0.01}\right) = 0.01$ gives $t_{10;0.01} = 2.602$.

4. If $T \sim t(n)$, then
   $E(T) = 0$ and $V(T) = n/(n-2)$ for $n > 2$.

Adobe Acrobat Document

Excel Functions for *t*-distribution

The values of $\Pr(T \leq t)$ can be obtained by the functions from Microsoft Excel:

"=T.DIST(t; n;TRUE)" gives c.d.f. $\Pr(T \leq t)$. FALSE gives p.d.f $f_T(t)$

"=T.INV(p; n)" gives *p-th* quantile of a $t(n)$ distribution

For examples

"=T.DIST(2; 10; TRUE)" gives 0.963306

"=T.INV(0.95; 10)" gives 1.812461

Microsoft Excel Worksheet

Remark

- If the random sample was selected from a normal population, then

$$Z = \frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} \sim N(0, 1)$$

and

$$U = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

- It can be shown that $\bar{X}$ and S^2 are independent, and so are Z and U.

Remark (Continued)

- Therefore,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}}$$

$$= \frac{Z}{\sqrt{U/(n-1)}} \sim t_{n-1},$$

- That is, T has a t-distribution with $n - 1$ d.f.

F-Distribution

The *F*-distribution

Definition 5.5

- Let U and V be independent random variables having $\chi^2(n_1)$ and $\chi^2(n_2)$, respectively,
- then the distribution of the random variable,
$$F = \frac{U/n_1}{V/n_2},$$
is called a F distribution with (n_1, n_2) degrees of freedom.

The *F*-distribution (Continued)

Definition 5.5

- The p.d.f. F is given by
$$f_F(x) = \frac{n_1^{n_1/2} n_2^{n_2/2} \Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \frac{x^{n_1/2-1}}{(n_1 x + n_2)^{(n_1+n_2)/2}},$$
for $x > 0$ and 0 otherwise.
- It can be shown that $E(X) = n_2/(n_2 - 2)$, with $n_2 > 2$ and
$V(X) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2-2)^2(n_2-4)}$, with $n_2 > 4$

Example 1

- Suppose that random samples of sizes n_1 and n_2 are selected from two normal populations with variances σ_1^2 and σ_2^2 respectively.
- From Section 5.7, we know that
$$U = \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1)$$
and
$$V = \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1)$$
are independent random variables.

Example 1 (continued)

- Therefore we have
$$F = \frac{U/(n_1-1)}{V/(n_2-1)} = \frac{\frac{(n_1-1)S_1^2/\sigma_1^2}{(n_1-1)}}{\frac{(n_2-1)S_2^2/\sigma_2^2}{(n_2-1)}}$$
$$= \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

Theorem 5.7

- If $F \sim F(n, m)$, then $1/F \sim F(m, n)$.
- This theorem follows immediately from the definition of F-distribution.
- Values of the F-distribution can be found in the statistical tables.
- The table gives the values of $F(n_1, n_2; \alpha)$ such that $\Pr(F > F(n_1, n_2; \alpha)) = \alpha$.

Adobe Acrobat Document

ST2334 Probability and Statistics CYM Sampling and Sampling Distributions 5-84

*** F(a,b) = 1/F(b,a) (use this if cannot find probability in the f table)

Theorem 5.7 (Continued)

For example

- $F(5, 4; 0.05) = 6.26$ means $\Pr(F > 6.26) = 0.05$, where $F \sim F(5, 4)$.
- $F(4, 5; 0.025) = 7.39$ means $\Pr(F > 7.39) = 0.025$, where $F \sim F(4, 5)$.

<u>Excel commands for F-distribution</u>

Adobe Acrobat Document

"=F.DIST(f; n_1; n_2; true)" gives $\Pr(F \leq f)$

e.g. "=F.DIST(6.26;5;4; true)" gives 0.95

"=F.INV(p; n_1;n_2)" gives c satisfies $\Pr(F \leq c) = p$

e.g. "=F.INV(0.95;4;5)" gives 5.192

ST2334 Probability and Statistics CYM Sampling and Sampling Distributions 5-85

Theorem 5.8

$$F(n_1, n_2; 1 - \alpha) = 1/\, F(n_2, n_1; \alpha).$$

For example,

- $F(10, 5; 0.95) = 1/F(5, 10; 0.05) = 1/3.33 = 0.30$
  which means $\Pr(F > 0.30) = 0.95$,
  where $F \sim F(10, 5)$.

ST2334 Probability and Statistics CYM Sampling and Sampling Distributions 5-86

CHAPTER 6: ESTIMATION BASED ON NORMAL DISTRIBUTION

Point Estimation

6.1.2 Estimation

The estimation can be made in two ways: **Point estimation** and **Interval estimation**

- **Point estimation** is to let the value of some statistic, say
$$\widehat{\Theta} = \widehat{\Theta}(X_1, X_2, \cdots, X_n),$$
to estimate the unknown parameter θ;
such a statistic
$$\widehat{\Theta}(X_1, X_2, \cdots, X_n),$$
is called **a point estimator**.

ST2334 Probability and Statistics CYM Estimation based on Normal Distribution 6-5

Statistic (Continued)

- Let
$$W = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$
Then W is not a statistic if μ is not known.

- However W is a statistic if μ is known.

I.e. statistic must not have unknown parameters

Eg. sample mean is μ = 5, then point estimate of population mean μ is 5.

Interval estimation

Interval Estimation

- Interval estimation is to define two statistics, say,
$$\widehat{\Theta}_L \text{ and } \widehat{\Theta}_U, \qquad \text{where } \widehat{\Theta}_L < \widehat{\Theta}_U$$
so that $(\widehat{\Theta}_L, \widehat{\Theta}_U)$ constitutes a random interval for which the probability of containing the unknown parameter θ can be determined.

Interval Estimation (Continued)

- We shall seek a random interval
$$(\widehat{\Theta}_L, \widehat{\Theta}_U)$$
containing θ with a given probability $1 - \alpha$.
- That is
$$\Pr(\widehat{\Theta}_L < \theta < \widehat{\Theta}_U) = 1 - \alpha.$$

Interval Estimation (Continued)

- Then the interval $\hat{\theta}_L < \theta < \hat{\theta}_U$, computed from the selected sample is called a $(1-\alpha)100\%$ confidence interval for θ,
- and the fraction $(1-\alpha)$ is called **confidence coefficient** or **degree of confidence**,
- and the end points $\hat{\theta}_L$ and $\hat{\theta}_U$ are called lower and upper confidence limits respectively.

Unbiased estimator

6.1.3 Unbiased Estimator

Definition 6.1 (Unbiased estimator)

- A statistic $\widehat{\Theta}$ is said to be an **unbiased estimator** of the parameter θ if

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

Unbiased Estimator (Continued)

Example 1

$\bar{X}$ is an unbiased estimator of μ. That is, $E(\bar{X}) = \mu$.

Example 2

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an **unbiased** estimator of σ^2.

That is,

$$E(S^2) = \sigma^2$$

Confidence Interval for Mean – Known variance

6.3 Confidence Intervals for the Mean

6.3.1 Known Variance Case

- Confidence interval for mean with
  (i) known variance and
  (ii) the population is normal
  or n is sufficiently large (say $n \geq 30$)

C.I. for Mean with Known Variance (Continued)

- If $\bar{X}$ is the mean of a random sample of size n from a population with known variance σ^2,
- a $(1-\alpha)100\%$ confidence interval for μ is given by

$$\left(\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right)$$

Sample Size for Estimating μ

- Most of the time, $\bar{X}$ will not be exactly equal to μ and the point estimate is in error.
- The size of this error will be $|\bar{X} - \mu|$.
- We know that

$$\Pr\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

or

$$\Pr\left(|\bar{X} - \mu| < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Sample Size for Estimating μ (Continued)

- Let e denote the margin of error.
- We want the error $|\bar{X} - \mu|$ does not exceed the margin of error, e, with a probability larger than $1 - \alpha$.
- That is,

$$\Pr(|\bar{X} - \mu| \leq e) \geq 1 - \alpha$$

Sample Size for Estimating μ (Continued)

- Since $\Pr\left(|\bar{X}-\mu| < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$, therefore

$$e \geq z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

- Hence for a given margin of error e, the sample size is given by

$$n \geq \left(z_{\alpha/2}\,\frac{\sigma}{e}\right)^2.$$

Confidence Interval for Mean – Unknown variance

6.3.2 Unknown Variance Case

Confidence interval for mean with

(i) unknown population variance and

(ii) the population is normal or very closed to a normal distribution

(iii) the sample size is small

Unknown Variance Case (Continued)

- Let

$$T = \frac{(\bar{X}-\mu)}{S/\sqrt{n}},$$

where S^2 is the sample variance.

- We know that $T \sim t_{n-1}$.

C.I. for Mean with Unknown Variance (Continued)

- If $\bar{X}$ and S are the sample mean and standard deviation of a random sample of size $n < 30$ from an approximate normal population with unknown variance σ^2,
- a $(1-\alpha)100\%$ confidence interval for μ is given by

$$\bar{X} - t_{n-1;\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < \bar{X} + t_{n-1;\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$$

C.I. for Mean with Unknown Variance (Continued)

- For large n (say $n > 30$),
- the t-distribution is approximately the same as the $N(0, 1)$ distribution. Hence,
  - when σ^2 is unknown,
  - population is normal and
  - $n > 30$,
- a $(1-\alpha)100\%$ confidence interval for μ is given by

$$\bar{X} - z_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$$

Note: 90% confidence interval is Z0.05 = 1.645 (z-value)
Note: 95% confidence interval is Z0.025 = 1.96 (z-value)
Note: 96% confidence interval is Z0.02 = 2.05 (z-value)
Note: 99% confidence interval is Z0.005 = 2.5728 (z-value)
Means, Pr(Z > 2.05) = 0.02, where Z ~ N(0, 1)
Note: n=7, 95% confidence interval is t6, 0.025 = 2.447 (t-value)
Note: n=21, 90% confidence interval is t20, 0.05 = 1.7247 (t-value)
Note: n= 10, 98% confidence interval is t9, 0.01 = 2.821 (t-value)

Confidence Interval for Difference between Two Means – Known variances

6.4 Confidence Intervals for the Difference between Two Means

- If we have two populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 respectively,
- Then
$$\bar{X}_1 - \bar{X}_2$$
is the point estimator of $\mu_1 - \mu_2$.

6.4.1 Known Variances

- **σ_1^2 and σ_2^2 are known and not equal, and the two populations are normal,**
- **or when σ_1^2 and σ_2^2 are known and not equal, but n_1, n_2 are sufficiently large ($n_1 \geq 30, n_2 \geq 30$)**
- According to Section 5.5, we have
$$(\bar{X}_1 - \bar{X}_2) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right).$$

ST2334 Probability and Statistics CYM Estimation based on Normal Distribution 6-44

Known Variances (Continued)

which leads to the following $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

ST2334 Probability and Statistics CYM Estimation based on Normal Distribution 6-46

Confidence Interval for Difference between Two Means – Unknown variances

6.4.2 Large Sample C.I. for Unknown Variances

- **σ_1^2 and σ_2^2 are unknown**
- **n_1, n_2 are sufficiently large ($n_1 \geq 30, n_2 \geq 30$)**
- we may replace by σ_1^2 and σ_2^2 by their estimates, S_1^2 and S_2^2,

Large Sample C.I. for Unknown Variances (Continued)

- A $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is given by:

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}.$$

ST2334 Probability and Statistics CYM Estimation based on Normal Distribution 6-52

Note: question says “standard deviation” instead of “population standard deviation”

Confidence Interval for Difference between Two Means – Unknown but equal variances

6.4.3 Unknown but Equal Variances

- **σ_1^2 and σ_2^2 are unknown but equal and**
- **the two populations are normal**
- **Small sample sizes ($n_1 \leq 30$ and $n_2 \leq 30$)**
- Let $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then

$$(\bar{X}_1 - \bar{X}_2) \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right).$$

Unknown but Equal Variances (Continued)

- σ^2 can be estimated by the pooled sample variance

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

with S_1^2 and S_2^2 being the sample variances of the first and second samples respectively.

Unknown but Equal Variances (Continued)

- Therefore a $(1 - \alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{X}_1 - \bar{X}_2) - t_{n_1+n_2-2;\alpha/2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2$$

$$< (\bar{X}_1 - \bar{X}_2) + t_{n_1+n_2-2;\alpha/2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Unknown but Equal Variances for Large Samples

- Note that for large samples such that $n_1 \geq 30$ and $n_2 \geq 30$, we can replace $t_{n_1+n_2-2;\alpha/2}$ by $z_{\alpha/2}$ in the above formula.
- Therefore a $(1 - \alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2$$

$$< (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Confidence Interval for Difference between Two Means for Paired data

C.I. for the difference between two means for paired data (Continued)

- These differences are the values of a random sample $d_1, d_2, \cdots, d_n$ from a population that we shall assume to be normal with mean μ_D and unknown variance σ_D^2.
- In fact $\mu_D = \mu_1 - \mu_2$ and the point estimate of μ_D is given by

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)$$

- The point estimate of σ_D^2 is given by

$$s_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2.$$

6.4.4.1 Small Sample and Approximate Normal Population

- A $(1-\alpha)100\%$ confidence interval for μ_D can be established by writing
$$\Pr\left(-t_{n-1;\alpha/2} < T < t_{n-1;\alpha/2}\right) = 1-\alpha,$$
where $T = \frac{\bar{d}-\mu_D}{s_d/\sqrt{n}} \sim t_{n-1}$ distribution.
- Therefore a $(1-\alpha)100\%$ confidence interval for $\mu_D = \mu_1 - \mu_2$ is given by
$$\bar{d} - t_{n-1;\alpha/2}\left(\frac{s_D}{\sqrt{n}}\right) < \mu_D < \bar{d} + t_{n-1;\alpha/2}\left(\frac{s_D}{\sqrt{n}}\right).$$

For large sample (n > 30)

- For **sufficiently large** sample, we may replace $t_{n-1;\alpha/2}$ by $z_{\alpha/2}$ and
- a $(1-\alpha)100\%$ confidence interval for $\mu_D = \mu_1 - \mu_2$ is given by
$$\bar{d} - z_{\alpha/2}\left(\frac{s_D}{\sqrt{n}}\right) < \mu_D < \bar{d} + z_{\alpha/2}\left(\frac{s_D}{\sqrt{n}}\right).$$

Confidence Interval for Variances and Ratio of Variances – Known Pop. Mean

6.5 C.I. for Variances and Ratio of Variances

6.5.1 Confidence intervals for a variance (of a normal population)

- Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a (approximately) $N(\mu, \sigma^2)$ distribution.
- Then the sample variance
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$
is a **point estimate** of σ^2.

Case 1 μ is known

- When μ is known, we have
$$\frac{X_i - \mu}{\sigma} \sim N(0,1) \quad \text{for all } i$$
or
$$\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(1) \quad \text{for all } i$$
and hence
$$\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n).$$

Case 1 μ is known (Continued)

- Therefore, a $(1-\alpha)100\%$ confidence interval for σ^2 of $N(\mu, \sigma^2)$ population with μ known is
$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;\,\alpha/2}} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;1-\alpha/2}}$$

Confidence Interval for Variances and Ratio of Variances – Unknown Pop. Mean

Case 2 μ is unknown

- When μ is unknown, we have

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1).$$

Case 2 μ is unknown (Continued)

- Therefore, a $(1-\alpha)100\%$ confidence interval for σ^2 of $N(\mu, \sigma^2)$ population with μ unknown is

$$\frac{(n-1)S^2}{\chi^2_{n-1;\,\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}$$

where S^2 is the sample variance.

This is true for both small and large n

Confidence Interval for Ratio of Two Variances (Normal population) – Unknown means

6.5.2 C.I. for the ratio of two variances (of normal population) with unknown means

- Let $X_1, X_2, \cdots, X_{n_1}$ be a random sample of size n_1 from a (or approximately) $N(\mu_1, \sigma_1^2)$ population and
- $Y_1, Y_2, \cdots, Y_{n_2}$ be a random sample of size n_2 from a (or approximately) $N(\mu_2, \sigma_2^2)$ population,
- where μ_1 and μ_2 are unknown.

C.I. for the ratio of two variances (of normal population) with unknown means (Continued)

- Then

$$\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1) \text{ and } \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1)$$

where $S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$ and $S_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(Y_j - \bar{Y})^2$.

C.I. for the ratio of two variances (of normal population) with unknown means (Continued)

- Hence, a $100(1-\alpha)\%$ confidence interval for the ratio σ_1^2/σ_2^2 when μ_1 and μ_2 are unknown

$$\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;\,\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}F_{n_2-1,n_1-1;\,\alpha/2}$$

since $F_{n_1-1,n_2-1;1-\alpha/2} = \frac{1}{F_{n_2-1,n_1-1;\,\alpha/2}}$. (See p6.95)

C.I. for the ratio of two variances (of normal population) with unknown means (Continued)

$$\Pr\left(\boldsymbol{W} > F_{n_1-1,n_2-1;1-\alpha/2}\right) = 1 - \alpha/2$$

$$\Rightarrow \Pr\left(\frac{1}{\boldsymbol{W}} < \frac{1}{F_{n_1-1,n_2-1;1-\alpha/2}}\right) = 1 - \alpha/2$$

where $W \sim F_{n_1-1,n_2-1}$ and $\frac{1}{W} \sim F_{n_2-1,n_1-1}$.

C.I. for the ratio of two variances (of normal population) with unknown means (Continued)

But

$$\Pr\left(\frac{1}{W} < F_{n_2-1,n_1-1;\alpha/2}\right) = 1 - \alpha/2$$

Hence

$$\frac{1}{F_{n_1-1,n_2-1;1-\alpha/2}} = F_{n_2-1,n_1-1;\alpha/2}.$$

CHAPTER 7: HYPOTHESIS TESTING BASED ON NORMAL DISTRIBUTION

Statistical Hypothesis

Null and Alternative Hypotheses (Continued)

7.1.1 Statistical Hypothesis (Continued)

- It is important to understand that the rejection of a hypothesis is to conclude that it is false, while the acceptance of a hypothesis merely implies that we have insufficient evidence to believe otherwise.
- Because of this terminology, the statistician or experimenter will often choose to state the hypothesis in a form that hopefully will be rejected.

Null and Alternative Hypotheses (Continued)

Null hypothesis:

- Hypothesis that we formulate with the hope of rejecting, denoted by H_0.
- A null hypothesis concerning a population parameter will always be stated to specify an exact value of the parameter.

Alternative hypothesis:

- The rejection of H_0 leads to the acceptance of an alternative hypothesis, denoted by H_1.
- It allows for the possibility of several values.

Types of Error (Type I Error and Type II Error)

7.1.2 Types of Error

- Two types of errors in the hypothesis testing:

	State of Nature	
Decision	H_0 is true	H_0 is false
Reject H_0	**Type I error** Pr(Reject H_0 given that H_0 is true) = α	Correct decision Pr(Reject H_0 given that H_0 is false) = $1 - \beta$
Do not reject H_0	Correct decision Pr(Do not reject H_0 given that H_0 is true) = $1 - \alpha$	**Type II error** Pr(Do not reject H_0 given that H_0 is false) = β

Types I and II Error

Type I error

- Rejection of H_0 when H_0 is true is called a type I error.
- It is considered as a serious type of error

Type II Error

- Not rejecting H_0 when H_0 is false is called a type II error.

Types I and II Error (Continued)

- α = level of significance
  = Pr(type I error)
  = Pr(reject H_0 when H_0 is true)
  = Pr(reject H_0 | H_0).

Types I and II Error (Continued)

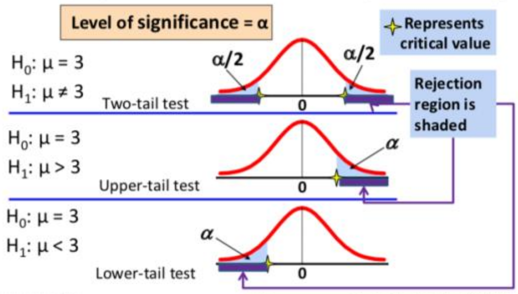
- β = Pr(type II error)
  = Pr(do not reject H_0 when H_0 is false)
  = Pr(do not reject $H_0 \mid H_1$).

- $1 - \beta$ = Power of a test = Pr(reject $H_0 \mid H_1$)

Note: Probability of committing a type II error (β) is impossible to compute, unless we have a specific alternative hypothesis, e.g. H0: p = ¼, H1: p = ½ (instead of p != ½)

Acceptance and Rejection Regions

Level of Significance and the Rejection Region



ST2334 Probability and Statistics CYM Hypothesis Testing based on Normal Distribution 7-16

Note: Hypothesis testing may not just be about mean, can be probability (See L7, P17)

Hypothesis Testing on Mean – Known variance

7.2 Hypotheses Testing Concerning Mean

7.2.1 Hypo. Testing on Mean with Known Variance

Consider the problem of testing the hypothesis concerning the mean, μ, of a population with

1. **Variance, σ^2, known and**
2. **Underlying distribution is normal or n is sufficiently large (say $n > 30$)**

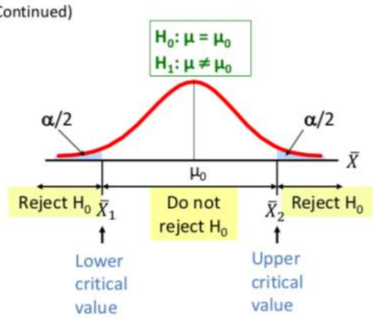
7.2.1.1 Two-sided Test

- Test $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$.
- When the population is normal or the sample size is large (then by the Central Limit Theorem), we can expect that
$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$
- Hence under $H_0: \mu = \mu_0$, we have
$$\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right).$$

ST2334 Probability and Statistics CYM Hypothesis Testing based on Normal Distribution 7-23

Two-sided Test (Continued)

- There are two cutoff values (critical values), defining the regions of rejection



Critical value approach

Finding critical values

The critical region can be given in terms of z values by means of the transformation

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Note: μ_0 is the value of μ under H_0.

Hence $\bar{x}_1 = \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ and $\bar{x}_2 = \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Hypothesis testing process

- From the population we select a random sample of size n and compute the sample mean.
- If $\bar{X}$ falls in the acceptance region $\bar{x}_1 < \bar{X} < \bar{x}_2$, we conclude that $\mu = \mu_0$; otherwise we reject H_0 and accept the $H_1: \mu \neq \mu_0$.
- Since $Z = (\bar{X} - \mu_0)/(\sigma/\sqrt{n})$, therefore $\bar{x}_1 < \bar{X} < \bar{x}_2$ is equivalent to $-z_{\alpha/2} < Z < z_{\alpha/2}$.
- The critical region is usually stated in terms of Z rather then $\bar{X}$.

Equivalent to finding confidence interval

Example 1 (Continued)

- For $\bar{x} = 34.5, \sigma = 1.5$ and $n = 49$, the 95% confidence interval is:

$$34.5 - (1.96)\frac{1.5}{\sqrt{49}} < \mu < 34.5 + (1.96)\frac{1.5}{\sqrt{49}}$$

$$34.08 \leq \mu \leq 34.92$$

- Since this interval does not contain the hypothesized mean, μ_0 (= 35), we reject the null hypothesis at $\alpha = 0.05$.

p-value Approach

p-Value Approach to Testing

- *p*-value: Probability of obtaining a test statistic more extreme ($\leq$ or $\geq$) than the observed sample value given H_0 is true
  - Also called **observed level of significance**

Note: level of significance is probability of rejecting null hypothesis, given it is true
p-value is a probability

p-Value Approach to Testing (Continued)

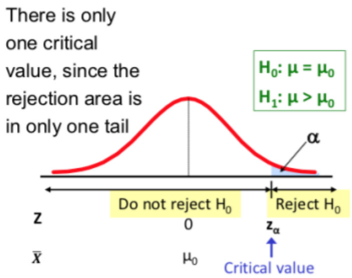
- Convert a sample statistic (e.g., $\bar{X}$) to a test statistic (e.g., Z statistic)
- Obtain the *p*-value
- Compare the *p*-value with $\boldsymbol{\alpha}$
  - If *p*-value $< \boldsymbol{\alpha}$, reject H_0
  - If *p*-value $\geq \boldsymbol{\alpha}$, do not reject H_0

One sided Test

7.2.1.2 One sided test

(a) Test H_0: $\mu = \mu_0$ against H_1: $\mu > \mu_0$.

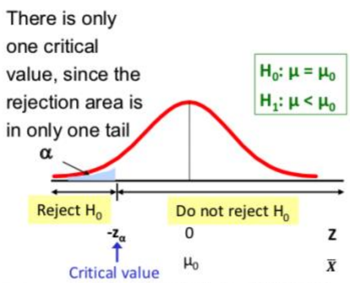
- Let $Z = \frac{(\bar{X}-\mu_0)}{\sigma/\sqrt{n}}$.
- Then H_0 is rejected if the observed values of Z, say z, is greater than z_α.



7.2.1.2 One sided test

(b) Test H_0: $\mu = \mu_0$ against H_1: $\mu < \mu_0$.

- Let $Z = \frac{(\bar{X}-\mu_0)}{\sigma/\sqrt{n}}$.
- Then H_0 is rejected if the observed values of Z, say z, is less than $-z_\alpha$.



Hypothesis Testing on Mean with Unknown Variance

7.2.2 Hypothesis Testing on Mean with Variance Unknown

Consider the problem of testing the hypothesis concerning the mean, μ, of a population with

1. **Variance unknown and**
2. **Underlying distribution is normal**

   Refer to Section 6.3.2

Test for mean with unknown variance

(1) Two sided test

- Test $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$.
- Let

$$T = \frac{(\bar{X} - \mu_0)}{S/\sqrt{n}}$$

where S^2 is the sample variance.

- Then H_0 is rejected if the observed value of T, say t, $> t_{n-1;\alpha/2}$ or $< -t_{n-1;\alpha/2}$.

Test for mean with unknown variance (Continued)

(2) One sided test

- Test $H_0: \mu = \mu_0$ against $H_1: \mu > \mu_0$.
- Then H_0 is rejected if $t > t_{n-1;\alpha}$.
- Test $H_0: \mu = \mu_0$ against $H_1: \mu < \mu_0$.
- Then H_0 is rejected if $t < -t_{n-1;\alpha}$.

See L7 P55 for example of using t-table and t-value

Hypothesis Testing concerning Difference between Two Means – Known variances

7.3 Hypotheses Testing Concerning Difference Between Two Means

7.3.1 Known Variances

1. **Variances σ_1^2 and σ_2^2 are known and**
2. **Underlying distribution is normal or both n_1 and n_2 are sufficiently large**

   (say $n_1 \geq 30, n_2 \geq 30$)

   Refer to Section 6.4.1

Solution to Example 1

Step 1

- Let μ_1 and μ_2 be the mean strength of cold-rolled steel and two-side galvanized steel respectively.
- Test $H_0: \mu_1 - \mu_2 = 0$ against $H_1: \mu_1 - \mu_2 \neq 0$.

Step 2

- Set $\alpha = 0.01$.

Solution to Example 1 (Continued)

Step 3

- Since σ_1^2 and σ_2^2 are known, therefore the test statistic

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

is used.

- $\alpha = 0.01$ implies $z_{\alpha/2} = z_{0.005} = 2.5728$.

Solution to Example 1 (Continued)

Step 3 (Continued)

- Critical region: $z < -2.5758$ or $z > 2.5758$, where

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

Solution to Example 1 (Continued)

Step 4

- **Computations:** $\bar{x}_1 = 29.8, \bar{x}_2 = 34.7, \sigma_1^2 = 16, \sigma_2^2 = 25,$ $n_1 = 20$ and $n_2 = 25$, so

$$z = \frac{[(29.8 - 34.7) - 0]}{\sqrt{16/20 + 25/25}} = -3.652.$$

- **p-value** $= 2 \times \min\{\Pr(Z > -3.652), \Pr(Z < -3.652)\} = 2(0.00013) = 0.00026.$

Hypothesis Testing concerning Difference of Two Means - Large Sample Testing with Unknown Variance

7.3.2 Large Sample Testing with Unknown Variances

1. **Variances σ_1^2 and σ_2^2 are unknown and**
2. **both n_1 and n_2 are sufficiently large (say $n_1 \geq 30, n_2 \geq 30$)**

   Refer to Section 6.4.2

Solution to Example 2 (Continued)

Step 3

- Since σ_1^2 and σ_2^2 are unknown and the sample sizes are large, therefore the test statistic

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

is used.

- $\alpha = 0.01$ implies $z_\alpha = z_{0.01} = 2.3263$.

Solution to Example 2 (Continued)

Step 3 (Continued)

- **Critical region**: $z > 2.3263$, where

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

Solution to Example 2 (Continued)

Step 4

- **Computations**: $\bar{x}_1 = 0.497, \bar{x}_2 = 0.359, s_1^2 = 0.187^2, s_2^2 = 0.158^2, n_1 = n_2 = 35$, so
$$z = \frac{[(0.497 - 0.359) - 0]}{\sqrt{0.187^2/35 + 0.158^2/35}} = 3.335.$$
- p-value $= \Pr(Z > 3.335) = 0.00043$.

Hypothesis Testing concerning Difference of Two Means – Small sample size, unknown but equal variances

7.3.3 Unknown but Equal Variances

1. **σ_1^2 and σ_2^2 are unknown but equal and**
2. **the populations are normal**
3. **Small sample sizes (say $n_1 \leq 30, n_2 \leq 30$)**

   Refer to Section 6.4.3

Solution to Example 3

Step 1

- Let μ_1 and μ_2 be the average grades students taking this course by the classroom and programmed presentations, respectively.
- Test $H_0: \mu_1 - \mu_2 = 0$ against $H_1: \mu_1 - \mu_2 \neq 0$.

Step 2

- Set $\alpha = 0.1$.

Solution to Example 3 (Continued)

Step 3

- $n_1 = 12$ and $n_2 = 10$ implies $t_{n_1+n_2-2;\alpha} = t_{20;0.05} = 1.725$.
- **Critical region :** $t < -1.725$ or $t > 1.725$, where
$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$
  with
$$S_p^2 = \frac{1}{n_1 + n_2 - 2}[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2].$$

Solution to Example 3 (Continued)

Step 4 (Continued)

- Hence
$$t = \frac{[(85 - 81) - 0]}{\sqrt{20.05(1/12 + 1/10)}} = 2.086$$
- p-value $= 2 \times \min\{\Pr(T_{20} > 2.086), \Pr(T_{20} < 2.806)\}$
  $= 2(0.025) = 0.05$.

Solution to Example 2 (Continued)

Step 5

- **Conclusion**: Since the observed t-value = 2.086 which falls inside the critical region, hence $H_0: \mu_1 = \mu_2$ is rejected at the 10% level of significance

  and conclude that the two methods of learning are not equal.
- Since p-value = 0.05 is less than 0.10, therefore we reject H_0 at the 10% level of significance and conclude that the two methods of learning are not equal.

Hypothesis Testing concerning Paired Data

Solution to Example 4

Step 1

- Let μ_d be the average difference in percentage between methods 1 and 2.
- Test $H_0: \mu_d = 0$ against $H_1: \mu_d < 0$. (why?)

Step 2

- Set $\alpha = 0.05$.

Solution to Example 4 (Continued)

Step 3

- $n = 12$ implies $t_{n-1;\alpha} = t_{11;0.05} = 1.796$.
- Critical region $t < -1.796$, where

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}, \qquad \text{with } d_i = X_{1i} - X_{2i}$$

Solution to Example 4 (Continued)

Step 4

- **Computations:** From the data, we have $\sum_i d_i = -0.2$ and $\sum_i d_i^2 = 0.0112$. Hence $\bar{d} = -0.0167$ and $s_d^2 = [0.0112 - 12(-0.0167)^2]/11 = 0.00072$.
- Therefore

$$t = [(-0.0167) - 0]/\sqrt{0.00072/12} = -2.156.$$

- p-value $= \Pr(\boldsymbol{T_{11}} < -2.156) = 0.027$.

  [or p-value is between 0.05 and 0.025 since 2.156 is between $t_{11;0.05} = 1.796$ and $t_{11;0.025} = 2.201$ if statistical table is used.]

Solution to Example 4 (Continued)

Step 5

- Since the observed t-value = −2.156 falls in the critical region, hence $H_0: \mu_d = 0$ is rejected at the 5% level of significance and conclude that sufficient evidence exists to permit us to conclude that Method 2 yields a higher average percentage than does Method 1.
- Since p-value = 0.027 is less than 0.05, therefore we reject H_0 and conclude that sufficient evidence exists to permit us to conclude that Method 2 yields a higher average percentage than does Method 1.

Hypothesis Testing concerning Variance – Assume underlying distribution is normal
(Eg. "Is the standard deviation > 0.9?")

7.4 Hypotheses Testing Concerning Variance

7.4.1 One Variance case

Assumption: Underlying distribution is normal

- Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a (approximate) $N(\mu, \sigma^2)$ distribution, where σ^2 is unknown.
- We wish to test null hypothesis
$$H_0: \sigma^2 = \sigma_0^2.$$
- We know that
$$\chi^2 = \frac{(n_1 - 1)S^2}{\sigma_0^2} \sim \chi^2(n-1).$$

Hypothesis Testing for σ^2 (Continuous)

Hence

H_0	Test Statistic
$\sigma^2 = \sigma_0^2$	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$

Hypothesis Testing for σ^2 (Continuous)

- $H_0: \sigma^2 = \sigma_0^2$ is rejected if the observed χ^2-value

H_1	Critical Region
$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{n-1;\alpha}$
$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{n-1;1-\alpha}$
$\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi^2_{n-1;1-\alpha/2}$ or $\chi^2 > \chi^2_{n-1;\alpha/2}$

where $\Pr(W > \chi^2_{n-1;\alpha}) = \alpha$ with $W \sim \chi^2(n-1)$

Hypothesis Testing concerning Ratio of Variance – Assume underlying distribution is normal
(Eg. "test the variances are equal")
See L7 P103 for example on how to use F-statistic

7.4.2 H.T. Concerning Ratio of Variances

Assumption:
1. **Underlying distributions is normal**
2. **Means are unknown**

H.T. Concerning Ratio of Variances (Continued)

Examples
- When we are comparing the precision of one measuring device with that of another,
- the variability in grading practices of one teacher with that of another, and
- the consistency of one production process with that of another,
- we are testing about the difference between two population variances (or standard deviations).

H.T. Concerning Ratio of Variances (Continued)

- We know that when two independent samples of sizes n_1 and n_2 are randomly selected from two normal populations then

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1).$$

- Under H_0: $\sigma_1^2 = \sigma_2^2$,

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1).$$

H.T. Concerning Ratio of Variances (Continued)

- Hence

H_0	Test Statistic
$\sigma_1^2 = \sigma_2^2$	$F = \frac{S_1^2}{S_2^2}$

H.T. Concerning Ratio of Variances (Continued)

- H_0: $\sigma_1^2 = \sigma_2^2$ is rejected if the observed F-value falls in the critical region

H_1	Critical Region
$\sigma_1^2 > \sigma_2^2$	$F > F_{(n_1-1, n_2-1\,; \alpha)}$
$\sigma_1^2 > \sigma_2^2$	$F < F_{(n_1-1, n_2-1\,; 1-\alpha)}$
$\sigma_1^2 \neq \sigma_2^2$	$F < F_{(n_1-1, n_2-1\,; 1-\alpha/2)}$ or $F > F_{(n_1-1, n_2-1\,; \alpha/2)}$

where $\Pr\left(W > F_{v_1, v_2; \alpha}\right) = \alpha$ with $W \sim F(v_1, v_2)$

ADDITIONAL

Variance

Population Variance	Sample Variance
$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$	$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$
$\sigma^2 =$ population variance $x_i =$ value of i^{th} element $\mu =$ population mean $N =$ population size	$s^2 =$ sample variance $x_i =$ value of i^{th} element $\bar{x} =$ sample mean $n =$ sample size

Formula >

$$\mathrm{SE} = \frac{\sigma}{\sqrt{n}}$$

SE = standard error of the sample

σ = sample standard deviation

n = number of samples

standard error