ST2334 Notes

Chapter 1: Basic Concepts of Probability

Basic Concepts and Definitions

Observation

Any recording of information, be it numerical or categorical

Statistical Experiment

Any procedure that generates a set of observations

Sample Space

The set of all possible outcomes of a statistical experiment, represented by the symbol S.

> 💡 $S = \{1, 2, 3, 4, 5, 6\}$ for a die toss.

> 💡 $S = \{\text{even}, \text{odd}\}$ for a number

> 💡 $S = \{(H, H), (H, T), (T, H), (T, T)\}$ for two coin flips. Note that both (H, T) and (T, H) are shown so that each outcome is equally likely, but this does not really matter.

Sample Points

Every outcome in a sample space

> 💡 (H, H) is a sample point for the above example

Events

A subset of a sample space

Simple Event

Consists of exactly one outcome or sample point

Compound Event

Consists of more than one outcome or sample point

> 💡 The event may even be expressed as $A = \{t : 0 \leq t < 5\}$.

Sure Event

The sample space itself

Null Event

Event with no outcomes or sample points, i.e. $\emptyset$.

Operations of Events

Complement Events

A' is the set of all elements in S that are not in A

Mutually Exclusive Events

A and B are mutually exclusive or disjoint if $A \cap B = \emptyset$

Union of Events

$A \cup B$ is the event containing all sample points in A or B or both

$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$

Intersection of Events

$A \cap B$ is the event containing all elements common to A and B

$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$

Basic Properties

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

De Morgan's Law

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

Contained

$A \subset B$ if all elements in event A are in event B

If $A \subset B$ and $B \subset A$ then $A = B$

In this module, we assume contained means it's a **proper subset**

Counting Methods

Multiplication Principle (OP1 ∧ OP2)

If an operation can be performed in n_1 ways, and 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

For k such operations, we have $n_1 n_2 \cdots n_k$ ways

Addition Principle (OP1 ∨ OP2)

If a first procedure can be performed in n_1 ways, and a second procedure in n_2 ways, and that it is not possible to perform both together, then the number ways we can perform either the first or second proceudres is $n_1 + n_2$ ways

For k such procedures, we have $n_1 + n_2 + \cdots + n_k$ ways

> 💡 **Application of the above concepts:**
> How many even three-digit numbers can we form from $0, 1, 2, 5, 6, 9$? Each digit can only be used once.
>
> **Case A:** 0 is used for the ones.
> Number of ways = $5 \times 4 = 20$ for the hundreds and tens places.
>
> **Case B:** 0 is not used for the ones
> Number of ways = $4 \times 4 \times 2 = 32$ as we cannot put 0 in the hundreds.
>
> Total ways = 20 + 32 = 52

Permutation

An arrangement of r objects from a set of n objects, where $r \leq n$

Number of permutations of n distinct objects taken r at a time = ${}_nP_r = \frac{n!}{(n-r)!}$

Permutations around a Circle

Number of ways = $(n-1)!$

Permutations when not all objects are distinct

If we have n_k elements of a k-th kind, where $n_1 + n_2 + \cdots + n_k = n$, then the number of distinct permutations is ${}_nP_{n_1, n_2, \cdots, n_k} = \frac{n!}{n_1!n_2!\cdots n_k!}$

Combination

Number of ways to select r objects from n objects without regard to the order

Number of combinations of n distinct objects taken r at a time = $\binom{n}{r} = {}_nC_r = \frac{n!}{r!(n-r)!}$

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1} \text{ for } 1 \leq r \leq n$$

> 💡 If 2 balls are randomly drawn from an urn containing 6 white and 5 black balls, what is the probability that 1 is white and 1 is black?
>
> We have $\frac{\binom{6}{1} \times \binom{5}{1}}{\binom{11}{2}} = \frac{30}{55}$

Approaches to Probability

Classical

Assume each outcome has equal probability, hence n outcomes = $1/n$ probability

Also known as axiomatic approach

Relative Frequency

$f_A = \frac{n_A}{n}$ is the relative frequency of A in the n repetitions of E

Not the same as probability, but we can assert that $\Pr(A) = \lim_{n \to \infty} f_A$

Subjective

For outcomes that cannot really be calculated, i.e. unrepeatable experiments

Axioms of Probability

1. $0 \leq \Pr(A) \leq 1$
2. $\Pr(S) = 1$
3. If $A_1, A_2, \cdots$ are **mutually exclusive** (disjoint) events, i.e. $A_i \cap A_j = \emptyset$ when $i \neq j$, then

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

Basic Properties of Probability

1. $\Pr(\emptyset) = 0.$
2. If $A_1, A_2, \cdots, A_n$ are **mutually exclusive** events, then $\Pr(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \Pr(A_i)$
3. For any event A, $\Pr(A') = 1 - \Pr(A)$
4. For any two events A and B, $\Pr(A) = \Pr(A \cap B) + \Pr(A \cup B')$
5. For any two events A and B, $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
6. For any three events, A, B, C, $\Pr(A \cup B \cup C) = \Pr(A) + \Pr(B) + \Pr(C) - \Pr(A \cap B) - \Pr(A \cap C) - \Pr(B \cap C) + \Pr(A \cap B \cap C)$ (See the Inclusion-Exclusion Principle)
7. If $A \subset B$, then $\Pr(A) \leq \Pr(B)$.

Inclusion-Exclusion Principle

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

Birthday Problem

$$p_n = \Pr(A) = 1 - q_n$$

Once you have 23 people, the probability of having two people sharing the same birthday exceeds 1/2.

n	q_n	p_n
1	1	0
2	0.99726	0.00274
3	0.99180	0.00820
10	0.88305	0.11695
15	0.74710	0.25290
20	0.58856	0.41144
21	0.55631	0.44369
22	0.52430	0.47570
23	0.49270	0.50730
30	0.29368	0.70632
40	0.10877	0.89123
50	0.029626	0.979374
100	$3.0725(10)^{-7}$	1
253	$6.9854(10)^{-53}$	1

Inverse Birthday Problem

How large does a group of randomly selected people have to be such that the probability that someone is sharing his or her birthday **with me** is larger than 0.5?

We need n such that $1-(\frac{364}{365})^n \geq 0.5$.

Solving this, we get

$$n \geq \frac{\log(0.5)}{\log(\frac{364}{365})} = 252.7$$

Conditional Probability

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

Multiplicative Rule of Probability

$$\Pr(A \cap B) = \Pr(A)\Pr(B|A) = \Pr(B)\Pr(A|B)$$

$$\Pr(A \cap B \cap C) = \Pr(A)\Pr(B|A)\Pr(C|A \cap B)$$

Law of Total Probability

$$\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap A_i) = \sum_{i=1}^{n} \Pr(A_i)\Pr(B|A_i)$$

Assuming the events $A_1, \cdots, A_n$ are mutually exclusive and exhaustive events

Bayes' Theorem

Let $A_1, A_2, \cdots, A_n$ be a partition of the sample space S. Then

$$\Pr(A_k|B) = \frac{\Pr(A_k)\Pr(B|A_k)}{\sum_{i=1}^{n} \Pr(A_i)\Pr(B|A_i)}$$

Independent Events

Two events A and B are independent if and only if $\Pr(A \cap B) = \Pr(A)\Pr(B)$

Properties of Independent Events

1. $\Pr(B|A) = \Pr(B)$ and $\Pr(A|B) = \Pr(A)$
2. A and B cannot be mutually exclusive if they are independent, supposing $\Pr(A), \Pr(B) > 0$
3. A and B cannot be independent if they are mutually exclusive
4. The sample space S and the empty set $\emptyset$ are independent of any event
5. If $A \subset B$, then A and B are dependent unless $B = S$.

Theorem about Complementary of Independent Events

If A and B are independent, then so are A and B', A' and B, A' and B'

n Pairwise Independent Events

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 mutually independent if and only if for any subset $\{A_{i_1}, A_{i_2} \cdots, A_{i_k}\}$,

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

> 💡 Mutually independence basically means the multiplicative rule holds for any subset of the set of events.
>
> Mutually independence implies pairwise independence, but pairwise independence does not imply mutually independence!
>
> The complements of any number of the above events will also be mutually independent with the remaining events.

Chapter 2: Concepts of Random Variables

Random Variables

Random Variable

A real-valued function X which assigns a number to every element $s \in S$

Range space, $R_X = \{x | x = X(s), s \in S\}$

Equivalent Events

Let B be an event with respect to R_X, i.e. $B \subset R_X$

If $A = \{s \in S | X(s) \in B\}$, then A and B are equivalent events and $\Pr(A) = \Pr(B)$

Note that A contains **all** sample points that fit the criteria

Table Format

Rolling Two Die

x	2	3	4	5	6	7	8	9	10	11	12
Pr(X = x)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Discrete Probability Distributions

Discrete Random Variable

If the number of possible values of X is finite or countable infinite, we call X a discrete random variable

Probability Function

Each value of X has a certain probability $f(x)$, and this 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

It must satisfy the following two conditions

1. $f(x_i) \geq 0$ for all x_i
2. $\sum_{i=1}^{\infty} f(x_i) = 1$

Tossing Two Coins

x	0	1	2
f(x) = Pr(X = x)	1/4	1/2	1/4

If we plot these values on a graph, we get a probability histogram. The total area of all rectangles is 1.

Another View of Probability Function

We can also think of a probability function as specifying a mathematical model for a finite population.

Continuous Probability Distributions

Continuous Random Variable

If R_X, the range space of a random variable X, is an interval or a collection of intervals, then X is a continuous random variable

Probability Density Function

The probability density function (p.d.f.) $f(x)$ of a continuous random variable must satisfy the following conditions

1. $f(x) \geq 0$ for all $x \in R_X$.
    1. This also means that we may set $f(x) = 0$ for $x \notin R_X$, i.e. $\Pr(A) = 0$ does not imply $A = \emptyset$
2. $\int_{R_X} f(x)dx = 1$ or $\int_{\infty}^{\infty} f(x)dx = 1$ since $f(x) = 0$ for x not in R_X
3. For any $(c, d) \subset R_X, c < d$, $\Pr(c \leq X \leq d) = \int_c^d f(x)dx$
4. $\Pr(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$

For number 4, in the continuous case, the probability of X being equals to a fixed value is 0, and

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

We can use ≤ and < interchangeably for a probability density function.

Cumulative Distribution Function

Let X be a random variable (can be discrete or continuous).

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

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

CDF for Discrete Random Variables

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

This c.d.f. is a step function.

For any $a \leq b$, $\Pr(a \leq B \leq b) = \Pr(X \leq b) - \Pr(X < a) = F(b) - F(a^-)$

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

CDF for Continuous Random Variables

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

The reverse is also true, **if a derivative exists:**

$$f(x) = \frac{dF(x)}{dx}$$

For any $a \leq b$, $\Pr(a \leq X \leq b) = \Pr(a < X \leq b) = F(b) - F(a)$

Note that this c.d.f. is a non-decreasing function.

Expressing CDF

To express CDF, we can write in cases

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 0.3, & \text{if } 0 \leq x < 1, \\ 0.9, & \text{if } 1 \leq x < 2, \\ 1, & \text{if } 2 \leq x. \end{cases}$$

Expectation

Expected Value of Discrete Random Variable

The mean or expected value of X, denoted by $E(X)$ or μ_X is

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

Expected Value of Continuous Random Variable

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

Expectation of Functions of Random Variables

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

1. $E[g(X)] = \sum_x g(x) f_X(x)$ if X is discrete, providing the sum exists
2. $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx$ if X is continuous, providing the integral exists

Variance

The special function, $g(x) = (x - \mu_X)^2$, leads us to the definition of variance.

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

The positive square root of the variance is the standard deviation,

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

We also can calculate variance using

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

Moment

The special function, $g(x) = x^k$, leads us to the definition of moment. The k-th moment of X is $E(X^k)$

Properties of Expectation

1. $E(aX + b) = aE(X) + b$, where a and b are constants
2. $V(X) = E(X^2) - [E(X)]^2$
3. $V(aX + b) = a^2 V(X)$

Chebyshev's Inequality

We cannot reconstruct X from $E(X)$ and $V(X)$. However, we can derive some bounds.

Let X be a random variable, $E(X) = \mu$ and $V(X) = \sigma^2$. Then for **any positive number** k, we have

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Alternatively,

$$\Pr(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

> 💡 We are using the standard deviation σ, not the variance σ^2.

Chapter 3: Two-Dimensional Random Variables and Conditional Probability Distributions

2-Dimensional Random Variables

(X, Y) is a two-dimensional random variable, where X, Y are functions assigning a real number to each $s \in S$.

(X, Y) is also called a random vector.

Range Space

$R_{X,Y} = \{(x, y) | x = X(s), y = Y(s), s \in S\}$

The above definition can be extended to more than two random variables, i.e. n-dimensional random variable/vector.

Discrete and Continuous

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

(X, Y) is a 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 Functions for Discrete Random Variables

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. $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \Pr(X = x_i, Y = y_j) = 1$.

$f_{X,Y}$ is the **joint probability function** of (X, Y).

Expressing joint p.f.

x	y				Row Total
	0	1	2	3	
0	0	3/84	6/84	1/84	10/84
1	4/84	24/84	12/84	0	40/84
2	12/84	18/84	0	0	30/84
3	4/84	0	0	0	4/84
Column Total	20/84	45/84	18/84	1/84	1

Joint Probability Density Functions for Continuous Random Variables

$f_{X,Y}(x,y)$ is called a **joint probability density function** if it satisfies the following:

1. $f_{X,Y}(x,y) \geq 0$ for all $(x,y) \in R_{X,Y}$.
2. $\iint_{(x,y)\in R_{X,Y}} f_{X,Y}(x,y)dxdy = 1$ or $\int_{\infty}^{\infty}\int_{\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$.

Marginal Probability Distributions

For discrete random variables,

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

$f_Y(y) = \sum_x f_{X,Y}(x,y)$

For continuous random variables,

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

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

Basically we fix one of the two values, then either sum or integrate over the other. It gives the probabilities of various values of the variables in the subset without reference to the values of the other variables.

Conditional Probability Distributions

The conditional distribution of Y given that $X = x$ is given by $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)},\ \text{if}\ f_X(x) > 0$, for each x within the range of X. Flip the variables for X given $Y = y$.

The condition p.f. or p.d.f. is also a 1-dimensional p.f. or p.d.f.

1. $f_{Y|X}(y|x) \geq 0$ and $f_{X|Y}(x|y) \geq 0$.
2. The sum or integral of the p.f. or p.d.f. respectively is 1.
3. 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

Random variables X and Y are independent if and only if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x, y.

This can be extended to n random variables.

Product Space

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

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

Expectation

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

Covariance

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

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

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 independence.
3. $Cov(aX+b, cY+d) = acCov(X,Y)$.
4. $V(aX+bY) = a^2V(X) + b^2V(Y) + 2abCov(X,Y)$.

Correlation Coefficient

The correlation coefficient of X and Y, denoted by $Cor(X,Y)$ or $\rho_{X,Y}$ or ρ is defined by $\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$.

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.

Chapter 4: Special Probability Distributions

Discrete Distributions

Discrete Uniform Distribution

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 $f_X(x) = \frac{1}{k}$, $x = x_1, x_2, \cdots, x_k$, and 0 otherwise.

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

Bernoulli Distributions

Bernoulli experiments only have two possible outcomes, and we can code them as 1 and 0.

A random variable X is defined to have a Bernoulli distribution is the probability function of X is given by $f_X(x) = p^x(1-p)^{1-x}, x = 0, 1$, where $0 < p < 1$. $f_X(x) = 0$ for all other X values.

1. $(1-p)$ is often denoted by q.
2. $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p = q$.
3. Mean, $\mu = E(X) = p$
4. Variance, $\sigma^2 = V(X) = p(1-p) = pq$.

Parameter and Family of Distributions

If $f_X(x)$ depends on a quantity assigned to any one of some possible values, and each different value results in a different probability distribution, that quantity is called a parameter of the distribution.

p is the parameter in the Bernoulli distribution.

The collection of all probability distributions for different values of the parameter is called a family of probability distributions.

Binomial Distributions ~ $B(n, p)$

A random variable X is defined to have a binomial distribution with two parameters n and 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 $0 < p < 1$.

X is basically the number of successes that occur in n independent Bernoulli trials.

Bernoulli distribution is a special case of the binomial distribution, with $n = 1$.

1. Mean, $\mu = E(X) = np$.
2. Variance, $\sigma^2 = V(X) = np(1-p) = npq$.
3. Conditions for Binomial experiment: n repeated Bernoulli trials, probability of success is same, trials are independent.

Negative Binomial Distributions ~ $NB(k, p)$

Consider a binomial experiment, except that trials will be repeated until a fixed number of successes occur. We are interested in the probability of the k-th success occurring on the x-th trial, where x is the random variable.

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

1. Mean, $E(X) = \frac{k}{p}$.
2. Variance, $\sigma^2 = \frac{(1-p)k}{p^2}$.

Geometric Distribution ~ $Geometric(p)$

This is a negative binomial distribution with $k = 1$, i.e. we stop after the first success.

Poisson Distribution ~ $P(\lambda)$

Poisson experiments yield the number of successes occurring during a given time interval or in a specified region.

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 size of the region and does not depend on the number of success occurring outside this time interval or region.
3. The probability of more than one success occurring in such a short time interval or falling in such a small region is negligible.

The probability distribution of a Poisson random variable X is called the Poisson distribution and the function is given by $f_X(x) = \Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ 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 $e \approx 2.171828...$

1. Mean, $E(X) = \lambda$
2. Variance, $\sigma^2 = V(X) = \lambda$.

Poisson Approximation to Binomial Distribution

Let X be a binomial random variable with parameters n and p. Thus $\Pr(X = x) = f_X(x) = \binom{n}{x} p^x q^{n-x}$. Suppose that $n \to \infty$ and $p \to 0$ such 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 Distributions

Continuous Uniform Distribution ~ $U(a, b)$

A random variable has a uniform distribution over the interval $[a, b], -\infty < a < b < \infty$, denoted by $U(a, b)$, with a probability density function given by $f_X(x) = \frac{1}{b-a}, a \le x \le b$, and 0 otherwise.

1. Mean, $E(X) = \frac{a+b}{2}$
2. Variance, $\sigma^2 = \frac{1}{12}(b-a)^2$

Exponential Distribution ~ $Exp(\alpha)$

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.

1. Mean, $E(X) = \frac{1}{\alpha}$
2. Variance, $\sigma^2 = \frac{1}{\alpha^2}$
3. $\int_{\infty}^{\infty} f(X) dx = 1$
4. $\Pr(X > t) = e^{-\alpha t}$
5. $\Pr(X \le t) = 1 - e^{-\alpha t}$

The p.d.f. can be rewritten in the form $f_X(x) = \frac{1}{\mu} e^{-x/\mu}$ for $x > 0$, then $E(X) = \mu$ and $\sigma^2 = \mu^2$.

No Memory Property

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

Normal Distribution ~ $N(\mu, \sigma^2)$

The random variable X assuming all real values, $-\infty < x < \infty$, has a normal distribution if its probability density function is given by $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}), -\infty < x < \infty$, where $-\infty < \mu < \infty$ and $\sigma > 0$.

1. Symmetrical about vertical line $x = \mu$
2. Maximum point is at $x = \mu$ and value is $\frac{1}{\sqrt{2\pi}\sigma}$
3. Total area under the curve is 1
4. Mean and variance is as given, μ and σ^2
5. As σ increases, the curve flattens, and as σ decreases, the curve sharpens
6. If X has distribution $N(\mu, \sigma^2)$, and if $Z = \frac{(X-\mu)}{\sigma}$, then Z has the $N(0, 1)$ distribution (standardized normal distribution), and $E(Z) = 0$ and $V(Z) = \sigma_Z^2 = 1$.
7. $x_1 < X < x_2 = (x_1 - \mu)/\sigma < Z < (x_2 - \mu)/\sigma$

Statistical Tables

Statistical tables give values $\Phi(z)$ for a given z, where $\Phi(z)$ is the cumulative distribution function of a standardized normal random variable Z. $1 - \Phi(z)$ is the upper cumulative probability for a given z.

1. $\Phi(z) = \Pr(Z \le z)$

2. $1 - \Phi(z) = \Pr(Z > z)$

Some statistical tables give the 100α percentage points, z_α, of a standardized normal distribution, where $\alpha = \Pr(Z \geq z_\alpha) = \int_{z_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) dx$

1. $\Pr(Z \geq z_\alpha) = \Pr(Z \leq -z_\alpha) = \alpha$

Normal Approximation to Binomial Distribution

When $n \to \infty$ and $p \to 1/2$, we can use normal distribution to approximate the binomial distribution. A good rule of thumb is to use normal approximation only when $np > 5$ and $n(1-p) > 5$.

If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$, then as $n \to \infty$, $Z = \frac{X-np}{\sqrt{npq}}$ is approximately ~$N(0,1)$. In other words, we want to consider Y~ $N(\mu, \sigma^2)$.

Continuity Correction

1. $\Pr(X = k) \approx \Pr(k - \frac{1}{2} < X < k + \frac{1}{2})$
2. $\Pr(a \leq X \leq b) \approx \Pr(a - \frac{1}{2} < X < b + \frac{1}{2})$
3. $\Pr(a < X \leq b) \approx \Pr(a + \frac{1}{2} < X < b + \frac{1}{2})$
4. $\Pr(a \leq X < b) \approx \Pr(a - \frac{1}{2} < X < b - \frac{1}{2})$
5. $\Pr(a < X < b) \approx \Pr(a + \frac{1}{2} < X < b - \frac{1}{2})$
6. $\Pr(X \leq c) = \Pr(0 \leq X \leq c) \approx \Pr(-\frac{1}{2} < X < c + \frac{1}{2})$
7. $\Pr(X > c) = \Pr(c < X \leq n) \approx \Pr(c + \frac{1}{2} < X < n + \frac{1}{2})$

Chapter 5: Sampling and Sampling Distributions

Population

The totality of all possible outcomes or observations of a survey or experiment is called a population.

Every outcome or observation can be recorded as a numerical or categorical value. Thus, each member of a population is a value of a random variable.

Finite Population

Consists of a finite number of elements, e.g. all citizens of Singapore.

Infinite Population

Consists of an infinitely (countable and uncountable) large number of elements, e.g. the results of all possible rolls of a pair of dice.

Sample

A sample is any subset of a population.

Random Sampling

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.

Random Sampling in General

The below sampling examples can be generalized as such:

Let X be a random variable with certain probability distribution, $f_X(x)$. Let $X_1, X_2, \cdots, X_n$ be n independent random variables each having the same distribution as X. Then $(X_1, X_2, \cdots, X_n)$ is a random sample of size n from a population with distribution $f_X(x)$.

The joint p.f. or p.d.f. of $(X_1, X_2, \cdots, X_n)$ is given by $f_{X_1, X_2, \cdots, X_n}(x_1, x_2, \cdots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$.

Sampling from a Finite Population

Sampling Without Replacement

There are $\binom{N}{n}$ samples of size n that can be drawn from a finite population of size N without replacement.

Each sample has a probability of $\frac{1}{\binom{N}{n}}$ of being selected.

Sampling With Replacement

Order of selection matters here. Hence, there are N^n samples of size n that can be drawn from a finite population of size N with replacement.

Each sample has a probability of $\frac{1}{N^n}$ being selected.

Sampling from an Infinite Population

Unfortunately, the concept of a random sample from an infinite population is more difficult to explain.

Refer to Chapter 5 slides 16-20 for some very unclear examples.

Sampling Distribution of Sample Mean

The main purpose in selecting random samples is to elicit information about unknown population parameters. Values calculated from the sample is used to make some inference concerning the true value of the population.

Statistic

A function of a random sample $(X_1, X_2, \cdots, X_n)$ is called a statistic. For example, the mean is a statistic. Hence, a statistic is a random variable, and it is meaningful to consider the probability distribution of a statistic, which is also called a sampling distribution.

Sample Mean

For some random sample of size n represented by $X_1, X_2, \cdots, X_n$, the sample mean is defined by the statistic $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

If the values in the random sample are observed and they are $x_1, x_2, \cdots, x_n$, then the realization of the statistic $\overline{X}$ is given by $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Sampling Distribution

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

1. $\mu_{\overline{X}} = \mu_X$, i.e. $E(\overline{X}) = E(X)$.
2. $\sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n}$, i.e. $V(\overline{X}) = \frac{V(X)}{n}$.

Law of Large Number

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.

For any $\epsilon \in \mathbb{R}$, $\mathrm{P}(|\overline{X} - \mu| > \epsilon) \to 0$ as $n \to \infty$.

In other words, as the sample size increases, the probability that the sample mean differs from the population mean goes to zero.

Central Limit Theorem

The sampling distribution of the sample mean $\overline{X}$ is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ if n is sufficiently large.

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

Sampling distribution properties of $\overline{X}$:

1. Central Tendency: $\mu_{\overline{X}} = \mu$.
2. Variation: $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$.

Normal Sampling Distributions

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

Sampling Distribution of Difference of Two Sample Means

If independent samples of sizes $n_1 (\geq 30)$ and $n_2 (\geq 30)$ 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 $\overline{X}_1$ and $\overline{X}_2$ is approximately normally distributed with mean and standard deviation given by:

1. $\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$.
2. $\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.
3. $\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ approx ~ $N(0, 1)$.

Chi-square Distribution ~ $\chi^2(n)$

If Y is a random variable with probability density function $f_Y(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{(n/2)-1} e^{-y/2}$, 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.

The gamma function, $\Gamma(\cdot)$, is defined by $\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx = (n-1)!$ for $n = 1, 2, 3, \cdots$.

1. If Y~$\chi^2(n)$, then $E(Y) = n$ and $V(Y) = 2n$.
2. For large n, $\chi^2(n)$ approx ~ $N(n, 2n)$.
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, $\sum_{i=1}^{k} Y_i$ ~ $\chi^2(\sum_{i=1}^{k} n_i)$.

From Normal to Chi-square

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 $Y \sim \chi^2(n)$.

χ^2-distribution Statistical Table

Let c be a constant satisfying $\Pr(Y \geq c) = \int_c^\infty f_Y(y)dy = \alpha$, where $Y \sim \chi^2(n)$. We use the notation $\chi^2(n;\alpha)$ to denote this constant c. That is, $\Pr(Y \geq \chi^2(n;\alpha)) = \int_{\chi^2(n;\alpha)}^\infty f_Y(y)dy = \alpha$.

Similarly, $\chi^2(n; 1-\alpha)$ is the constant satisfying

$\Pr(Y \leq \chi^2(n;1-\alpha)) = \int_0^{\chi^2(n;1-\alpha)} f_Y(y)dy = \alpha$.

Thus, we have:

1. $\chi^2(10;0.9)$ means $\Pr(Y \geq \chi^2(10;0.9)) = 0.9$ or $\Pr(Y \leq \chi^2(10;0.9)) = 0.1$.
2. From the statistical table on χ^2-distribution, we have $\chi^2(10;0.9) = 4.865$.

Sampling Distribution of $(n-1)S^2/\sigma^2$

The statistic $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is the sample variance. However, it has little practical application. Instead, we shall consider the sampling distribution of the random variable $\frac{(n-1)S^2}{\sigma^2}$ when $X_i \sim N(\mu, \sigma^2)$ for all i.

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

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.

$\frac{Z}{\sqrt{U/n}} \sim t(n)$.

The p.d.f. is given by

$$f_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{t^2}{n})^{\frac{n+1}{2}}, -\infty < t < \infty$$

1. The graph of the t-distribution is symmetric about the vertical axis and resembles the graph of the standard normal distribution.
2. It can be shown that the p.d.f. of t-distribution with n d.f. (degrees of freedom) is approaching to the p.d.f. of standard normal distribution when $n \to \infty$. That is, $\lim_{n\to\infty} f_T(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$.
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. For example, $\Pr(T \geq t_{10;0.05}) = 0.05$ gives $t_{10;0.05} = 1.812$.
4. If $T \sim t(n)$, then $E(T) = 0$ and $V(T) = \frac{n}{n-2}$ for $n > 2$.

t-distribution from Random Sample from Normal Population

If the random sample was selected from a normal population, then $Z = \frac{(\overline{X}-\mu)}{\sigma/\sqrt{n}} \sim N(0,1)$ and $U = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. It can be shown that $\overline{X}$ and S^2 are independent, and so are Z and U.

Therefore,

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

$\sim t_{n-1}$. That is, T has a t-distribution with $n-1$ d.f.

F-distribution ~ $F(n_1, n_2)$

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 p.d.f. F is given by

$$f_F(x) = \frac{n_1^{n_1/2} n_2^{n_2/2}\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}\frac{x^{(n_1/2)-1}}{(n_1x+n_2)^{(n_1+n_2)/2}}$$

for $x > 0$ and 0 otherwise.

1. $E(X) = n_2/(n_2 - 2)$, with $n_2 > 2$.
2. $V(X) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$, with $n_2 > 4$.
3. If $F \sim F(n, m)$, then $1/F \sim F(m, n)$.

F-distribution Statistical Table

Values of the F-distribution can be found in the statistical table, which gives the values of $F(n_1, n_2; \alpha)$ such that $\Pr(F > F(n_1, n_2; \alpha)) = \alpha$.

For example, $F(5, 4; 0.05) = 6.26$ means $\Pr(F > 6.26) = 0.05$, where $F \sim F(5, 4)$.

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

Chapter 6: Estimation based on Normal Distribution

Parameter

Assume that some characteristics of the elements in a population can be presented by a random variable X whose p.d.f. or p.f. is $f_X(x; \theta)$, where the form of the p.d.f. or p.f. is assumed known except that it contains an unknown parameter θ.

Further assume that the values $x_1, x_2, \cdots, x_n$ of a random sample $X_1, X_2, \cdots, X_n$ from $f_X(x; \theta)$ can be observed.

On the basis of the observed sample values $x_1, x_2, \cdots, x_n$, it is desired to **estimate** the value of the unknown parameter θ.

Statistic

A statistic is a function of the random sample which does not depend on any unknown parameters. For example, $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ or $X_{(n)} = \max(X_1, X_2, \cdots, X_n)$ are some examples of a statistic.

Let $W = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$, then W is a statistic if and only if μ is known.

Point 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 is called a point estimator.

Point Estimate of Mean

Suppose μ is the population mean. The statistic that one uses to obtain a point estimate is called an estimator.

For example, $\overline{X}$ is an estimator of μ. The value of $\overline{X}$, denoted by $\overline{x}$, is an estimate of μ.

Unbiased Estimator

A statistic $\widehat{\Theta}$ is said to be an unbiased estimator of the parameter θ if $E(\widehat{\Theta}) = \theta$.

Interval Estimation

We define two statistics, $\widehat{\Theta}_L$ and $\widehat{\Theta}_U$, 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.

For example, suppose σ^2 is known. Let $\widehat{\Theta}_L = \overline{X} - 2\frac{\sigma}{\sqrt{n}}$ and $\widehat{\Theta}_U = \overline{X} + 2\frac{\sigma}{\sqrt{n}}$. Then $(\overline{X} - 2\frac{\sigma}{\sqrt{n}}, \overline{X} + 2\frac{\sigma}{\sqrt{n}})$ is an interval estimator for μ.

Confidence Interval from Interval Estimation

An interval estimate of a population parameter θ is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_U$, where $\hat{\theta}_L$ and $\hat{\theta}_U$ depend on

1. The value of the statistic $\widehat{\Theta}$ for a particular sample, and
2. The sampling distribution of $\widehat{\Theta}$.

$\hat{\theta}_L$ is also known as the lower confidence limit, $\hat{\theta}$ the point estimate, and $\hat{\theta}_U$ the upper confidence limit.

Not all intervals will contain the parameter θ, since it depends on the sample. We thus seek a random interval $(\widehat{\Theta}_L, \widehat{\Theta}_U)$ containing θ with a given probability $1 - \alpha$. That is, $\Pr(\widehat{\Theta}_L < \theta < \widehat{\Theta}_U) = 1 - \alpha$.

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 the confidence coefficient or degree of confidence.

Interpretation

This means that if samples of the same size n are taken, then in the long run, $(1-\alpha)100\%$ of the intervals will contain the unknown parameter θ, and hence with a confidence of $(1-\alpha)100\%$, we can say that the interval covers θ.

Confidence Interval for the Mean

Known Variance Case

We can compute the confidence interval for mean with

1. Known variance and
2. The population is normal OR n is sufficiently large (≥ 30)

When the population is normal or by the CLT, we can expect that $\overline{X}$ ~ $N(\mu, \frac{\sigma^2}{n})$.

Thus $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ ~ $N(0,1)$. Hence

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

$\Pr(\overline{X} - z_{\alpha/2}(\frac{\sigma}{\sqrt{n}}) < \mu < \overline{X} + z_{\alpha/2}(\frac{\sigma}{\sqrt{n}})) = 1-\alpha$.

$z_{0.025} = 1.96$

Confidence Interval for Mean with Known Variance

If $\overline{X}$ 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

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

Sample Size for Estimating μ

The size of the error with the point estimate is $|\overline{X}-\mu|$, i.e. $\Pr(|\overline{X}-\mu| < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$.

Let e denote a margin of error that we do not want the error to exceed with a probability larger than $1-\alpha$. Thus, we have $e \geq z_{\alpha/2}(\frac{\sigma}{\sqrt{n}})$.

Hence, for a given margin of error e, the sample size is given by $n \geq (z_{\alpha/2}\frac{\sigma}{e})^2$.

Unknown Variance Case

This case applies for when

1. Unknown population variance and
2. The population is normal or very close to a normal distribution
3. The sample size is small

Let $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$, where S^2 is the sample variance. We know that T ~ t_{n-1}.

Hence, $\Pr(-t_{n-1;\alpha/2} < T < t_{n-1;\alpha/2}) = 1-\alpha$, or

$\Pr(-t_{n-1;\alpha/2} < \frac{(\overline{X}-\mu)}{S/\sqrt{n}} < t_{n-1;\alpha/2}) = 1-\alpha$, or

$\Pr(-t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} < \overline{X}-\mu < t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}) = 1-\alpha$, or

$\Pr(\overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}) = 1-\alpha$.

Confidence Interval for Mean with Unknown Variance

If $\overline{X}$ 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

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

For large $n > 30$, the t-distribution is approximately the same as the $N(0,1)$ distribution, hence for large n, a $(1-\alpha)100\%$ confidence interval for μ is given by

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

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 $\overline{X}_1 - \overline{X}_2$ is the point estimator of $\mu_1 - \mu_2$.

Known Variances Case

This case applies when

- σ_1^2 and σ_2^2 are known and not equal
- The two populations are normal OR $n_1 \geq 30, n_2 \geq 30$.

We have

$(\overline{X}_1 - \overline{X}_2)$ ~ $N(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$

We can further assert that

$$\Pr(-z_{\alpha/2} < \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_{\alpha/2}) = 1 - \alpha$$

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

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

Unknown Variances Case

This case applies when

- σ_1^2 and σ_2^2 are unknown
- $n_1 \geq 30, n_2 \geq 30$

We replace σ_1^2 and σ_2^2 by their estimates, S_1^2 and S_2^2, giving us the following $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$

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

Unknown but Equal Variances (Small Samples) Case

This case applies when

- σ_1^2 and σ_2^2 are unknown but equal
- The two populations are normal
- Small sample sizes, $n_1 \leq 30, n_2 \leq 30$

We let $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then
$(\overline{X}_1 - \overline{X}_2) \sim N(\mu_1 - \mu_2, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$
and we obtain a standard normal variable

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

But this requires the actual population variance. We thus need to estimate σ^2 using the pooled sample variance:

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

Since the two populations are normal, then the two sample variances can also be used to obtain a Chi-squared distribution. We can actually combine the two sample variances to get
$\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$
Finally, we can substitute S_p^2 for σ^2, giving us the statistic:
$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} \sim t_{n_1+n_2-2}$

Note that we are using t-distribution because the variance is unknown.

We can assert that

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

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

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

Unknown but Equal Variances (Large Samples) Case

For large samples, we can replace $t_{n_1+n_2-2;\alpha/2}$ by $z_{\alpha/2}$. Thus, a $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is

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

Paired (Dependent) Data

When our two samples are dependent on each other, e.g. before and after, we need to work with the differences $d_i = x_i - y_i$ of paired observations.

We assume these differences $d_1, d_2, \cdots, d_n$ are normal with a mean μ_D and unknown variance σ_D^2.

$\mu_D = \mu_1 - \mu_2$

The point estimate of μ_D is

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

The point estimate of variance σ_D^2 is given by

$s_D^2 = \frac{1}{n-1}\sum_{i=1}^n (d_i - \overline{d})^2$

We can thus establish

$$\Pr(-t_{n-1;\alpha/2} < T < t_{n-1;\alpha-2}) = 1 - \alpha$$

where $T = \frac{\overline{d}-\mu_D}{s_d/\sqrt{n}}$ ~ t_{n-1} distribution.

We have a $(1-\alpha)100\%$ confidence interval for $\mu_D = \mu_1 - \mu_2$

$$\overline{d} - t_{n-1;\alpha/2}(\frac{S_D}{\sqrt{n}}) < \mu_D < \overline{d} + t_{n-1;\alpha/2}(\frac{S_D}{\sqrt{n}})$$

For sufficiently large sample $n > 30$, we can replace $t_{n-1;\alpha/2}$ by $z_{\alpha/2}$ and get

$$\overline{d} - z_{\alpha/2}(\frac{S_D}{\sqrt{n}}) < \mu_D < \overline{d} + z_{\alpha/2}(\frac{S_D}{\sqrt{n}})$$

Confidence Interval for Variances

The following applies for a (approximately) $N(\mu, \sigma^2)$ distribution.

The sample variance,

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

is a point estimate of σ^2

Known Mean Case

When μ is known, we have

$\frac{X_i-\mu}{\sigma}$ ~ $N(0,1)$ for all i

or $(\frac{X_i-\mu}{\sigma})^2$ ~ $\chi^2(1)$ for all i

and hence $\sum_{i=1}^n \frac{(X_i-\mu)^2}{\sigma^2}$ ~ $\chi^2(n)$.

We get a $(1-\alpha)100\%$ confidence interval for σ^2 of $N(\mu, \sigma^2)$ population with μ known:

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n;\alpha/2}} < \sigma^2 < \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n;1-\alpha/2}}$$

For standard derivation, we just need to square root both sides.

$$\sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n;\alpha/2}}} < \sigma < \sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n;1-\alpha/2}}}$$

Unknown Mean Case

When μ is unknown, we have

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

This is true for both small and large n. Hence, we have

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

where S^2 is the sample variance.

For standard derivation, we just need to square root both sides

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

Confidence Interval for Ratio of Variances

Let us have two random samples from two approximately normal populations with unknown means.

We thus have $\frac{(n_1-1)S_1^2}{\sigma_1^2}$ ~ $\chi^2(n_1-1)$ and $\frac{(n_2-1)S_2^2}{\sigma_2^2}$ ~ $\chi^2(n_2-1)$, where $S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\overline{X})^2$ and $S_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(Y_j-\overline{Y})^2$.

Hence,

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

~ $F(n_1-1, n_2-1)$.

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

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

To get a confidence interval for σ_1/σ_2, we just square root both sides.

Chapter 7: Hypotheses Testing Based on Normal Distribution

Null and Alternative Hypotheses

Null hypothesis, H_0, is the hypothesis we formulate with the hope of rejecting.

The rejection of H_0 leads to the acceptance of an alternative hypothesis, denoted by H_1.

When we reject a hypothesis, we conclude that it is false. But if we accept it, it merely means we have insufficient evidence to believe otherwise.

Types I and II Error

Type I error occurs when we reject H_0 given that H_0 is true. This is considered as a serious type of error.

Type II error occurs when we do not reject H_0 given that H_0 is false.

$\text{Pr}(\text{Type I}) = \text{Pr}(\text{reject } H_0|H_0) = \alpha$, where α is the level of significance, usually 5% or 1%

$\text{Pr}(\text{Type II}) = \beta$, such that $1-\beta = \text{Pr}(\text{reject } H_0|H_1) = \text{Power of a test}$

Level of Significance

The level of significant separates all possible values of the test statistic into two regions, the rejection region (or critical region) and the acceptance region.

The value that separates the rejection and acceptance regions is called the critical value.

Hypotheses Testing Concerning Mean

Known Variance (Two-sided) - Critical Value

This is for

1. Variance, σ^2, is known, AND
2. Underlying distribution is normal OR $n > 30$

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

We can expect that $\overline{X}$ ~ $N(\mu, \frac{\sigma^2}{n})$, hence $\overline{X}$ ~ $N(\mu_0, \frac{\sigma^2}{n})$.

By using a significance level of α, we can find two critical values $\overline{x}_1$ and $\overline{x}_2$ such that

1. $\overline{x}_1 < \overline{X} < \overline{x}_2$ defines the acceptance region
2. The two tails, $\overline{X} < \overline{x}_1$ and $\overline{X} > \overline{x}_2$ constitute the critical or rejection region.

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

If $\overline{X}$ falls in the acceptance region, we accept the null hypothesis, else reject. The critical region is often stated in terms of Z instead of $\overline{X}$, where

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

Basically, if the $(1-\alpha)100\%$ confidence interval covers μ_0, null hypothesis is accepted, else it's rejected.

Known Variance (Two-sided) - p-Value

The p-value is the probability of obtaining a test statistic more extreme (≤ or ≥) than the observed sample value **given H_0 is true**. It is also called the observed level of significance.

Here are the steps:

1. Convert a sample statistic e.g. $\overline{X}$ into a test statistic e.g. Z statistic

2. Obtain the p-value
3. Compare the p-value with $\alpha/2$. If p-value < $\alpha/2$, reject H_0, else ≥, do not reject.

Known Variance (One-sided) - Critical Value

Same as before but the alternative hypothesis is now either $H_1 : \mu > \mu_0$ or $H_1 : \mu < \mu_0$.

In both cases, let $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$. Then we need to check if the observed values of Z is greater than z_α or less than $-z_\alpha$ respectively.

Known Variance (One-sided) - p-Value

Same as the two-sided known variance approach, just that we will compare against the relevant side, and against α itself.

Unknown Variance (Two-sided) - Critical Value

We use this case for

1. Variance unknown, AND
2. Underlying distribution is normal

Let $T = \frac{\overline{X}-\mu_0}{S/\sqrt{n}}$, 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 $< -t_{n-1;\alpha/2}$.

Unknown Variance (One-sided) - Critical Value

We test the relevant side, $t > t_{n-1;\alpha}$ or $t < -t_{n-1;\alpha}$.

Hypotheses Testing Concerning Difference Between Two Means

Known Variances

1. Variances σ_1^2 and σ_2^2 are known and
2. Underlying distribution is normal or both $n_1 \geq 30, n_2 \geq 30$.

Refer to section before on difference between two means with known variables. Generally, since variance is known, we will be using the Z distribution.

Unknown Variances (Large Samples)

1. Variances σ_1^2 and σ_2^2 are unknown and
2. Both $n_1 \geq 30, n_2 \geq 30$.

Refer to section before.

Unknown but Equal Variances (Small Samples)

1. Variances σ_1^2 and σ_2^2 are unknown but equal and
2. The populations are normal and
3. Both are small samples $n_1 \leq 30, n_2 \leq 30$.

Refer to section before.

Paired Data

Just refer to section before.

Hypotheses Testing Concerning Variances

One Variance

The assumption is that the underlying distribution is normal.

We wish to test $H_0 : \sigma^2 = \sigma_0^2$. We know that $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ ~ $\chi^2(n-1)$, and we will use it as our test statistic.

1. For $\sigma^2 > \sigma_0^2$, we have critical region $\chi^2 > \chi^2_{n-1;\alpha}$
2. For $\sigma^2 < \sigma_0^2$, we have critical region $\chi^2 < \chi^2_{n-1;1-\alpha}$
3. For $\sigma^2 \neq \sigma_0^2$, we have $\chi^2 < \chi^2_{n-1;1-\alpha/2}$ or $\chi^2 > \chi^2_{n-1;\alpha/2}$

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

Ratio of Variances

1. Underlying distributions are normal
2. Means are unknown

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

Under $H_0 : \sigma_1^2 = \sigma_2^2$, $F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$, which is our test statistic.

1. For $\sigma^2 > \sigma_0^2$, we have critical region $F > F_{n_1-1,n_2-1;\alpha}$
2. For $\sigma^2 < \sigma_0^2$, we have critical region $F < F_{n_1-1,n_2-1;1-\alpha}$
3. For $\sigma^2 \neq \sigma_0^2$, we have $F < F_{n_1-1,n_2-1;1-\alpha/2}$ or $F > F_{n_1-1,n_2-1;\alpha/2}$

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