1 Basic Concepts of Probability

- Sample space S = set of **all possible outcomes** = **sure event**, Sample space with no elements = $\varnothing$ = **null event**
- **Mutually exclusive/disjoint** if $A \cap B = \varnothing$
- **Contained:** $A \subset B \equiv B \supset A$, **all** of the elements in A are also in B.
- If $A \subset B$ and $B \supset A$, then $A = B$

1.1 Basic Properties

- $A \cap A' = \varnothing$
- $A \cap \varnothing = \varnothing$
- $A \cup A' = S$
- $(A \cap B)' = A' \cup B'$
- $(A \cup B)' = A' \cap B'$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup B = A \cup (B \cap A')$
- $A = (A \cap B) \cup (A \cap B')$
- $(A \cap B) \cup C \neq A \cap (B \cup C)$

1.2 De Morgan's Law

- $(\bigcup_{r=1}^{n} A_r)' = \bigcap_{r=1}^{n} (A_r)'$
- $(\bigcap_{r=1}^{n} A_r)' = \bigcup_{r=1}^{n} (A_r)'$

1.3 Counting Methods

1.3.1 Multiplication & Addition Principle

- **Multiplication Principal (OP1 $\wedge$ 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 2 operations can be performed together in $n_1 n_2$ ways.
- **Addition Principal (OP1 $\vee$ 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 procedures is $n_1 + n_2$ ways

1.3.2 Permutation

- An arrangement of r objects from a set of n objects, $r \le n$, order taken into consideration.
- n distinct objects taken r at a time $= {}_nP_r = \frac{n!}{(n-r)!}$
- In a circle: $(n-1)!$
- Not all are distinct: $\sum_{r=1}^{k} n_k = n$, ${}_nP_{n_1,n_2,\dots,n_k} = \frac{n!}{n_1!n_2!\dots n_k!}$

1.3.3 Combination

- # of ways of selecting r from n objects w/o regards to order
- $\binom{n}{r} = {}_nC_r = \frac{n!}{r!(n-r)!}$, ${}_nC_r \times r! = {}_nP_r$
- $\binom{n}{r}$ = binom coeff of the term $a^r b^{n-r}$ in binom expansion of $(a+b)^n$:
  - $\binom{n}{r} = \binom{n}{n-r}$ for $r = 0, 1, \dots, n$
  - $\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$ for $1 \le r \le n$
  - $\binom{n}{r} = 0$ for $r < 0$ pr $r > n$

1.4 Relative frequency (f_A)

$f_A = \frac{n_A}{n}$, is the relative frequency of A in n repetitions of experiment E, n_A = no of times that event A occurred among the n repetitions.

1.5 Axioms of Probability

- $0 \le Pr(A) \le 1$
- $\Pr(S) = 1$
- If $A_1, A_2, \dots$ are mutually exclusive (disjoint), i.e. $A_i \cap A_j = \varnothing$ when $i \neq j$, then $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$
  If events A and B are mutually exclusive, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

1.6 Properties of Probability

- $\Pr(\varnothing) = 0$
- If $A_1, A_2, \dots, A_n$ are mutually exclusive, then $\Pr(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \Pr(A_i)$
- $\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B')$
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
- $\Pr(A \cup B \cup C) = \Pr(A) + \Pr(B) + \Pr(C) - \Pr(A \cap B) - \Pr(B \cap C) - \Pr(A \cap C) + \Pr(A \cap B \cap C)$
- **The Inclusion-Exclusion Principle**
  $\Pr(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \Pr(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr(A_i \cap A_j) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} \Pr(A_i \cap A_j \cap A_k) - \dots$
- If $A \subset B$, then $\Pr(A) \le \Pr(B)$

1.7 Conditional Probability, $P(A \mid B)$

- $\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$, if $\Pr(A) \neq 0$
- For fixed A, $\Pr(B \mid A)$ satisfies the postulates of probability.
- False positive: $\Pr(+ \mid \text{condition})$
- Points to take note of:
  1. $\Pr(A|B) \neq \Pr(B|A)$
  2. $\Pr(B|A') \neq 1 - \Pr(B|A)$
  3. $\Pr(A'|B) = 1 - \Pr(A|B)$

1.7.1 Multiplication rule

- $\Pr(A \cap B) = \Pr(A)\Pr(B \mid A) = \Pr(B)\Pr(A \mid B)$, provided $\Pr(A) > 0, \Pr(B) > 0$
- $\Pr(A \cap B \cap C) = \Pr(A)\Pr(B \mid A)\Pr(C \mid A \cap B)$
- $\Pr(A_1 \cap \dots \cap A_n) = \Pr(A_1)\Pr(A_2 \mid A_1)\Pr(A_3 \mid A_1 \cap A_2)\dots\Pr(A_n \mid A_1 \cap \dots \cap A_{n-1})$

1.7.2 The Law of Total Probability

- Let $A_1, A_2, \dots, A_n$ be a partition of sample space S (mutually exclusive & exhaustive events s.t. $A_i \cap A_j = \varnothing$ for $i \neq j$ and $\cup_{i=1}^{n} A_i = S$).
- Then $\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap A_i) = \sum_{i=1}^{n} \Pr(A_i)\Pr(B \mid A_i)$
- e.g. $P(B) = P(A)P(B|A) + P(A')P(B|A')$

1.7.3 Bayes' Theorem

- Let $A_1, A_2, \dots, A_n$ be a partition of S
- $\Pr(A_k \mid B) = \frac{\Pr(A_k)\Pr(B|A_k)}{\sum_{i=1}^{n} \Pr(A_i)\Pr(B|A_i)} = \frac{\Pr(A_k)\Pr(B|A_k)}{\Pr(B)}$, $k \in [1, n]$

1.8 Independent Events

- **Definition:** iff $\Pr(A \cap B) = \Pr(A)\Pr(B)$

1.8.1 Properties

- Suppose $\Pr(A) > 0, \Pr(B) > 0$, A and B are independent:
  - $\Pr(B \mid A) = \Pr(B)$ and $\Pr(A \mid B) = \Pr(A)$
  - A and B cannot be mutually exclusive if they are independent (and vice versa)
- The sample space S and $\varnothing$ are independent of any event
- If $A \subset B$, then A and B are dependent unless $B = S$

Warning: Independent events can't be shown using Venn Diagram, so calc!!!

1.8.2 Theorem

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

1.8.3 n Independent Events

- **Pairwise Independent Events**:
  Events $A_1, A_2, \dots, A_n$ are pairwise independent iff $\Pr(A_i \cap A_j) = \Pr(A_i)\Pr(A_j)$ for $i \neq j$ and $i, j = 1, \dots, n$
- **Mutually Independent**:
  Events $A_1, A_2, \dots, A_n$ are (mutually) independent iff for any subset $\{A_{i_1}, A_{i_2}, \dots, A_{i_k}\}$ of $A_1, A_2, \dots, A_n$,
  $\Pr(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \Pr(A_{i_1})\Pr(A_{i_2})\dots\Pr(A_{i_k})$

1.8.4 Remarks

- $A_1, A_2, \dots, A_n$ are mutually independent $\Leftrightarrow$ for any pair of events A_j, A_k where $j \neq k$, the multiplication rule holds, for any 3 distinct events, the multiplication rule holds, and so on $\Pr(A_1 \cap A_2 \cap \dots \cap A_n) = \Pr(A_1)\Pr(A_2)\dots\Pr(A_n)$.
  In total there are $2^n - n - 1$ diff cases.
- Mutually independent $\Rightarrow$ pairwise independent **(not the converse)**
- Suppose $A_1, A_2, \dots, A_n$ are mutually independent events, let $B_i = A_i$ or A_i', $i \in [1, n]$. Then $B_1, B_2, \dots, B_n$ are also mutually independent events.

2 Concepts of Random Variables

2.1 Equivalent Events

2.1.1 Definition

- Let E be an experiment in sample space S. Let X be an R.V. defined on S, and R_X its range space, i.e. $X : S \to \mathbb{R}$
- Let B be an event w.r.t. R_X, i.e. $B \subset R_X$
- Suppose $A = \{s \in S \mid X(s) \in B\}$
  (A consists of all sample points s in S for which $X(s) \in B$)
- A and B are **equivalent events**, and $\Pr(B) = \Pr(A)$

2.1.2 Example

- Consider tossing a coin twice, $S = \{HH, HT, TH, TT\}$
- Let X be no of heads, then $R_X = \{0, 1, 2\}$
- $A_1 = \{HH\}$ equiv $B_1 = \{2\}$, $A_2 = \{HT, TH\}$ equiv $B_2 = \{1\}$, $A_3 = \{TT\}$ equiv $B_3 = \{0\}$, $A_4 = \{HH, HT, TH\}$ equiv $B_4 = \{2, 1\}$

2.2 Discrete Probability Distributions

2.2.1 Discrete R.V.

Let X be an RV. If R_X is **finite or countable infinite**, X is discrete RV

2.2.2 Probability Fn (p.f.) or Probability Mass Function (p.m.f.)

- For a discrete R.V., each value X has a certain probability $f(x)$. Such a function $f(x)$ is called the p.f (or **probability mass function**, p.m.f).
- The collection of pairs $(x_i, f(x_i))$ is probability distribution of X
- The probability of $X = x_i$ denoted by $f(x_i)$ must satisfy: $f(x_i) \ge 0 \, \forall x_i$ and $\sum_{i=1}^{\infty} f(x_i) = 1$

2.3 Continuous Probability Distributions

2.3.1 Continuous R.V.

Suppose that R_X is an **interval or a collection of intervals**, then X is a continuous R.V.

2.3.2 Probability Density Function (p.d.f.)

- Let X be a continuous R.V.
- p.d.f. $f(x)$ is a function satisfying:
  - $f(x) \ge 0 \, \forall x \in R_X$
  - $\int_{R_X} f(x)\,dx = 1$ or $\int_{-\infty}^{\infty} f(x)\,dx = 1$ as $f(x) = 0 \, \forall x \notin R_X$
  - $\forall c, d : c < d$ (i.e. $(c,d) \subset R_X$), $\Pr(c \le X \le d) = \int_c^d f(x)\,dx$
  - $\Pr(X = x_0) = \int_{x_0}^{x_0} f(x)\,dx = 0$

2.3.3 Remarks

- $\Pr(c \le X \le d) = \int_c^d f(x)\,dx$ represents area under the graph of the p.d.f. $f(x)$ between $x = c$ and $x = d$
- $\Pr(c \le X \le d) = \Pr(c \le X < d) = \Pr(c < X \le d) = \Pr(c < X < d)$
- $\Pr(A) = 0$ does not necessarily imply $A = \varnothing$
- $R_X \in [a, b] \Rightarrow f(x) = 0 \, \forall x \notin [a, b]$
- Note that **p.d.f can be more than 1!**

2.4 Cumulative Distribution Function (c.d.f.)

Let X be an R.V., disc or cont. $F(x)$ is a c.d.f of X where $F(x) = \Pr(X \le x)$

2.4.1 c.d.f. for Discrete R.V.

- $F(x) = \sum_{t \le x} f(t) = \sum_{t \le x} \Pr(X = t)$
- c.d.f. of a discrete R.V. is a step function
- $\forall a, b$ s.t. $a \le b$, $\Pr(a \le X \le b) = \Pr(X \le b) - \Pr(X < a) = F(b) - F(a^-)$ where a^- is the largest possible value of X strictly less than a

2.4.2 c.d.f. for Continuous R.V.

- $F(x) = \int_{-\infty}^{x} f(t)\,dt$
- $f(x) = \frac{dF(x)}{dx}$ if the derivative exists
- $\Pr(a \le X \le b) = \Pr(a < X \le b) = F(b) - F(a)$
- $F(x)$ is a non-decreasing function: $x_1 < x_2 \Rightarrow F(x_1) \le F(x_2)$; and $0 \le F(x) \le 1$

2.5 Mean and Variance of an R.V.

2.5.1 Expected Value / Mean / Mathematical Expectation

- **Discrete:** $E(X) = \mu_X = \sum_i x_i f_X(x_i) = \sum_x x f_X(x)$
- If $f(x) = \frac{1}{N}$ for each of the N values of x, $E(X) = \frac{1}{N} \sum_i x_i$
- **Continuous:** $E(X) = \mu_X = \int_{-\infty}^{\infty} x f_X(x)\,dx$
- **Remark:** The expected value exists if the sum/integral exists

2.5.2 Expectation of a function of an R.V.

$\forall g(X)$ with p.f. $f_X(x)$

- **Discrete:** $E[g(X)] = \sum_x g(x) f_X(x)$
- **Continuous:** $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$
- Provided the sum/integral exists.

2.5.3 Variance ($\sigma_X^2 = V(X)$)

- $g(x) = (x - \mu_X)^2$, Let X be an R.V. with p.f. $f(x)$
- $\sigma_X^2 = V(X) = E[(X - \mu_X)^2]$
- $E[(X - \mu_X)^2] = \begin{cases} \sum_x (x - \mu_X)^2 f_X(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$
- $V(X) \ge 0$, $V(X) = E(X^2) - [E(X)]^2$
- **Standard deviation** $= \sigma_X = \sqrt{V(X)}$

2.5.4 K-th moment of X

- **Definition:** $E(X^k)$, use $g(x) = x^k$ in expectation of a fn

2.5.5 Properties of Expectation

- $E(aX + b) = aE(X) + b$
- $V(X) = E(X^2) - [E(X)]^2$
- $V(aX + b) = a^2 V(X)$

2.6 Chebyshev's Inequality

- Let X be an R.V. with $E(X) = \mu$, $V(X) = \sigma^2$
- $\forall k > 0$, $\Pr(|X - \mu| \ge k\sigma) \le \frac{1}{k^2}$ OR $\Pr(|X - \mu| < k\sigma) \ge 1 - \frac{1}{k^2}$
- Holds for **all** distributions with finite mean and variance
- Gives a **lower bound** but not exact probability.

3 2D RV & Conditional Probability Distributions

3.1 2D RV Definition (Random Vector)

- Let E be experiment and S sample space associated with E. Let X and Y be 2 functions each assigning a real number to each $s \in S$. (X, Y) is a 2D RV
- **Range Space:** $R_{X,Y} = \{(x, y) \mid x = X(s), y = Y(s), s \in S\}$
- The definition can be extended to n-dimensional RV (or n-dimensional random vector) for $X_1, X_2, \dots, X_n$.
- (X, Y) is a 2D discrete RV if the possible values of $(X(s), Y(s))$ are **finite or countable infinite**.
- (X, Y) is a 2D continuous RV if the possible values of $(X(s), Y(s))$ can assume **all values in some region** of the Euclidean plane $\mathbb{R}^2$

3.2 Joint Probability Density Function

3.2.1 For Discrete RV

Let (X, Y) be a 2D **discrete** RV. 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:

- $f_{X,Y}(x_i, y_j) \ge 0 \, \forall (x_i, y_j) \in R_{X,Y}$
- $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \Pr(X = x_i, Y = y_j) = 1$

The function $f_{X,Y}(x, y)$ defined $\forall (x_i, y_j) \in R_{X,Y}$ is called **joint probability function of** (X, Y).

Let A be any set consisting of pairs of (x, y) values, then:
$\Pr((X, Y) \in A) = \sum\sum_{(x,y) \in A} f_{X,Y}(x, y)$

3.2.2 For Continuous RV

Let (X, Y) be a 2D continuous RV assuming all values in some region R of the Euclidean plane $\mathbb{R}^2$.
$f_{X,Y}(x, y)$ is called joint pdf if it satisfies:

- $f_{X,Y}(x, y) \ge 0 \, \forall (x, y) \in R_{X,Y}$
- $\iint_{(x,y) \in R_{X,Y}} f_{X,Y}\,dy\,dx = 1$ or $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy\,dx = 1$

3.3 Marginal and Conditional Probability Distributions

3.3.1 Marginal Probability Distributions

Let (X, Y) be a 2D RV with joint pdf $f_{X,Y}(x, y)$. The **marginal probability distributions** of X and Y are:

- **Discrete:** $f_X(x) = \sum_y f_{X,Y}(x, y)$ and $f_Y(y) = \sum_x f_{X,Y}(x, y)$
- **Cont:** $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx$
- Basically fix one of the values, then sum/integrate over the other. Gives the probabilities of various values of the variables in the subset without reference to the values of the other variables

3.3.2 Conditional Distribution

Let (X, Y) be a 2D RV with joint pdf $f_{X,Y}(x, y)$, let $f_X(x)$ and $f_Y(y)$ be the marginal probability functions of X and Y respectively.
Then the **conditional distribution of** Y **given that** $X = x$:

$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$, if $f_X(x) > 0$ for each $x \in$ range of X

Similarly, the **conditional distribution of** X **given** $Y = y$:

$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$, if $f_Y(y) > 0$ for each $y \in$ range of Y

Remarks:

- The conditional p.d.f (or p.f) satisfy all the requirements for a 1D p.d.f:
  - For a fixed y, $f_{X|Y}(x \mid y) \ge 0$, for a fixed x, $f_{Y|X}(y \mid X) \ge 0$
  - For discrete RV: $\sum_x f_{X|Y}(x \mid y) = 1$ and $\sum_y f_{Y|X}(y \mid x) = 1$
  - For cont RV: $\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)\,dx = 1$ and $\int_{-\infty}^{\infty} f_{Y|X}(y \mid x)\,dy = 1$
- For $f_X(x) > 0$, $f_{X,Y}(x, y) = f_{Y|X}(y \mid x) f_X(x)$. For $f_Y(y) > 0$, $f_{X,Y}(x, y) = f_{X|Y}(x \mid y) f_Y(y)$

3.4 Independent RV

RV X and Y are independent iff $f_{X,Y}(x, y) = f_X(x) f_Y(y) \, \forall x, y$
This definition can be extended to RV $X_1, X_2, \dots, X_n$

- The product of 2 positive functions $f_X(x)$ and $f_Y(y)$ means a function which is positive on a **product space**.
- i.e. if $f_X(x) > 0$ for $x \in A_1$ and $f_Y(y) > 0$ for $x \in A_2$, then $f_X(x) f_Y(y) > 0$ for $(x, y) \in A_1 \times A_2$

3.5 Expectation

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

3.5.1 Covariance ($\sigma_{x,y}$)

Let $g(X, Y) = (X - \mu_X)(Y - \mu_Y)$.
Let (X, Y) be a bivariate RV with joint pdf $f_{X,Y}(x, y)$, then the **covariance** of X, Y is $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$

- **Discrete:** $Cov(X, Y) = \sum_x \sum_y (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y)$
- **Cont:** $Cov(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y)\,dx\,dy$

Remarks:

- $Cov(X, Y) = E(XY) - \mu_X \mu_Y = E(XY) - E(X)E(Y)$
- If X, Y are independent, then $Cov(X, Y) = 0$. But $Cov(X, Y) = 0 \not\Rightarrow X$ and Y are independent
- $Cov(aX + b, cY + d) = acCov(X, Y)$
- $V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2abCov(X, Y)$

3.5.2 Correlation Coefficient

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

- $-1 \le \rho_{X,Y} \le 1$
- $\rho_{X,Y}$ = measure of degree of **linear** relationship between X and Y
- If X, Y are independent, then $\rho_{X,Y} = 0$. But $\rho_{X,Y} = 0 \not\Rightarrow$ independence

4 Special Probability Distributions

4.1 Discrete Uniform Distribution

If RV X assumes the values $x_1, x_2, \dots, x_k$ with equal probability, then X has a discrete **uniform** distribution, and the probability function is $f_X(x) = \frac{1}{k}$, $x = x_1, x_2, \dots, x_k$, and 0 otherwise.

4.1.1 Mean and Variance of Discrete Uniform Distribution

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

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

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

4.2 Bernoulli and Binomial Distribution

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

4.2.1 Bernoulli Distribution

- A random experiment with only 2 possible outcomes.
- RV X has a Bernoulli distribution if the probability function of X is $f_X(x) = p^x (1-p)^{1-x}$, $x = 0, 1$ where $0 < p < 1$, 0 for other X values.

Remarks

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

4.2.2 Binomial Distributions $\sim B(n, p)$

- RV X has a **Binomial** distribution with 2 parameters n and p, if the probability function of X is $\Pr(X = x) = f_X(x) = \binom{n}{x} p^x q^{n-x}$ for $x = 0, 1, \dots, n$ where $0 < p < 1$
- X is the # of successes in n independent Bernoulli trials.
- Bernoulli distribution is a special case of Binomial distribution when $n = 1$
- Mean, $\mu = E(X) = np$
- Variance, $\sigma^2 = V(X) = npq$
- Conditions: (1) consists of n repeated Bernoulli trials, (2) Only 2 possible outcomes in each trial, (3) $\Pr(\text{success}) = p$ is constant in each trial, (4) trials are **independent**

4.2.3 Negative Binomial Distribution $\sim NB(k, p)$

- Like binomial, but trials will be repeated until a **fixed** # of successes occur (interested in the probability of the k-th success occurs on the x-th trial)
- Let X be a RV represents # of trials to produce k successes in a sequence of independent Bernoulli trials
- $\Pr(X = x) = f_X(x) = \binom{x-1}{k-1} p^k q^{x-k}$ for $x = k, k+1, k+2, \dots$
- Mean, $\mu = E(X) = \frac{k}{p}$
- Variance, $\sigma^2 = V(X) = \frac{(1-p)k}{p^2}$

- **Special case:** # of trials required to have the first success (i.e. $k=1$) is **Geometric** distribution ($X \sim NB(1,p) \equiv X \sim Geom(p)$)

4.3 Poisson Distribution $\sim P(\lambda)$

- R.V. X, # of successes occurring during a given time interval/in a specified region
- $\Pr(X=x) = f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0,1,2,3,...$ where λ = average no of successes occurring in the given time interval/specified region
- Mean, $\mu = E(X) = \lambda$
- Variance, $\sigma^2 = V(X) = \lambda$
- **Properties:**
  1. # of successes in one time interval/specified region are **independent** of those in any other disjoint time interval/region of space
  2. The probability of a single success during a short time interval/in a small region is **proportional** to length of time interval/size of region, and does not depend on no of successes outside this time interval/region
  3. The prob of more than one success in such a short time interval/falling in such a small region is **negligible**

4.4 Poisson Approximation to the Binomial Distribution

- Let $X \sim B(n,p)$, suppose that $n \to \infty$ and $p \to 0$ such that $\lambda = np$ remains a constant as $n \to \infty$, then $X \approx$ Poisson distribution with parameter np
- $\lim_{\substack{p\to 0 \\ n\to\infty}} \Pr(X=x) = \frac{e^{-np}(np)^x}{x!}$
- If $p \to 1$, can still use Poisson distribution to approximate binomial probabilities by swapping success & failure s.t. $p \to 0$

4.5 Continuous Uniform Distribution $\sim U(a,b)$

- RV has **uniform** distribution over interval $[a,b]$, $-\infty < a < b < \infty$, denoted by $U(a,b)$ if its p.d.f is $f_X(x) = \frac{1}{b-a}$ for $a \le x \le b$ and 0 otherwise.
- Mean, $\mu = E(X) = \frac{a+b}{2}$
- Variance, $\sigma^2 = V(X) = \frac{1}{12}(b-a)^2$

4.6 Exponential Distribution $\sim Exp(\alpha)$

- Continuous RV X assuming all non-negative values has an exponential distribution with parameter $\alpha > 0$ if its p.d.f is $f_X(x) = \alpha e^{-\alpha x}$ for $x > 0$ and 0 otherwise.
- Mean, $\mu = E(X) = \frac{1}{\alpha}$
- Variance, $\sigma^2 = V(X) = \frac{1}{\alpha^2}$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$
- p.d.f can be written in the form $f_X(x) = \frac{1}{\mu}e^{-x/\mu}$ for $x > 0$ and 0 otherwise. Then $E(X) = \mu$, $V(X) = \mu^2$
- $\Pr(X > t) = e^{-\alpha t}$, $\Pr(X \le t) = 1 - e^{-\alpha t}$

4.6.1 No Memory Property of Exponential Distribution

Suppose $X \sim Exp(\alpha)$ where $\alpha > 0$, then for any 2 positive numbers s and t, $\Pr(X > s+t \mid X > s) = \Pr(X > t)$

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

RV X assuming all real values, $-\infty < x < \infty$, has a normal distribution if its p.d.f is $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(x-\mu)^2}{2\sigma^2})$ where $-\infty < x < \infty$, $-\infty < \mu < \infty$ and $\sigma > 0$

4.7.1 Properties

- Graph of the distribution is bell-shaped and symmetrical about the vertical line $x = \mu$ ($\Pr(Z \ge z_\alpha) = \Pr(Z \le -z_\alpha) = \alpha$)
- Max point occurs at $x = \mu$, its value is $\frac{1}{\sqrt{2\pi}\sigma}$
- $E(X) = \mu$, $V(X) = \sigma^2$
- Total area under the curve and above the horizontal axis is 1.
- The normal curve approaches the horizontal axis asymptotically in either direction away from mean.
- 2 normal curves are identical in shape with same σ^2, but centered around their means.
- As σ increases, the curve flattens; as σ decreases, the curve sharpens.
- If $X \sim N(\mu,\sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$, then $Z \sim N(0,1)$ (standardized normal distribution), and $E(Z) = 0$ and $V(Z) = 1$

4.8 Normal Approximation to the Binomial Distribution

When $np > 5$ and $nq > 5$ ($n \to \infty$, $p \to \frac{1}{2}$)

If $X \sim B(np, npq)$, then as $n \to \infty$, $Z = \frac{X-np}{\sqrt{npq}}$ is approx. $\sim N(0,1)$

4.8.1 Continuity Correction (Apply when doing normal approximation)

- $\Pr(X=k) \approx \Pr(k - \frac{1}{2} < X < k + \frac{1}{2})$
- $\Pr(a \le X \le b) \approx \Pr(a - \frac{1}{2} < X < b + \frac{1}{2})$, $\Pr(a < X \le b) \approx \Pr(a + \frac{1}{2} < X < b + \frac{1}{2})$
  $\Pr(a \le X < b) \approx \Pr(a - \frac{1}{2} < X < b - \frac{1}{2})$, $\Pr(a < X < b) \approx \Pr(a + \frac{1}{2} < X < b - \frac{1}{2})$
- $\Pr(X \le c) = \Pr(0 \le X \le c) \approx \Pr(-\frac{1}{2} < X < c + \frac{1}{2})$
- $\Pr(X > c) = \Pr(c < X \le n) \approx \Pr(c + \frac{1}{2} < X < n + \frac{1}{2})$

5 Sampling and Sampling Distributions

5.1 Population and Sample

- Totality of all possible outcomes is called **population**
- A **sample** is any subset of a population
- 2 kinds of population: (1) **Finite** population, consisting of a finite number of elements, (2) **Infinite** population, consisting of an infinitely (countable and uncountable) large number of elements
- A set of n observations from a given population is called a **sample** of size n
- Each observation in population can be considered as a value of a RV with p.d.f $f_X(x)$

5.2 Random Sampling

5.2.1 Simple Random Sampling (SRS)

A SRS of n members is a sample chosen in such a way that **every subset** of n observations of the population has the **same probability of being selected**

5.2.2 Sampling from a Finite Population

- **Sampling without replacement**: There are $\binom{N}{n}$ samples of size n to be drawn from a population of size N **without replacement**. Each sample has the same probability of $1/\binom{N}{n}$ of being selected.
- **Sampling with replacement**: There are N^n samples of size n drawn from a population of size N **with replacement**. Each sample has the same probability $\frac{1}{N^n}$ of being selected.

5.2.3 Sampling from an Infinite Population

Random if (1) In each draw all elements of the population have the **same probability of being selected**, (2) Successive draws are **independent**

5.2.4 Theorem

Let X be an RV with p.d.f $f_X(x)$, $X_1, X_2, ..., X_n$ be n independent RV each having the same distribution as X.
Then $(X_1, X_2, ..., X_n)$ is called a **random sample** of size N from a population with distribution $f_X(x)$.
The joint p.d.f is $f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n) = f_{X_1}(x_1)x_{X_2}(x_2)...f_{X_n}(x_n)$

5.3 Sampling distribution of sample mean ($\overline{X}$)

5.3.1 Statistic and Sampling Distribution

Sampling distribution = probability distribution of a statistic

5.3.2 Sample Mean

$X_1, X_2, ..., X_n$ is a random sample of size $n \Rightarrow$ sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$

5.3.3 Theorem

For random samples of size n taken from **infinite** population or from **finite population with replacement** having population mean μ and population standard deviation σ, $\overline{X}$ has its mean and standard deviation: $\mu_{\overline{X}} = \mu_X$ and $\sigma^2_{\overline{X}} = \frac{\sigma^2_X}{n}$, i.e. $E(\overline{X}) = E(X)$ and $V(\overline{X}) = \frac{V(X)}{n}$

5.3.4 Law of Large Number (LLN)

Let $X_1, X_2, ..., X_n$ 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 $\epsilon \in \mathbb{R}$, $\Pr(|\overline{X} - \mu| > \epsilon) \to 0$ as $n \to \infty$ (basically saying that as $n \to \infty$, $\overline{X}$ will be very close to μ)

5.4 Central Limit Theorem

Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population having any distribution with mean μ, finite population variance σ^2. $\overline{X}$ is **approximately normal** with mean μ and variance $\frac{\sigma^2}{n}$ if n is sufficiently large ($n > 30$). Hence, $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ follows approx. $N(0,1)$

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

5.4.1 Theorem

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

5.5 Sampling Distribution of the Difference of 2 Sample Means

5.5.1 Theorem

If independent samples of sizes n_1 and n_2 (each ≥ 30) are drawn from 2 populations, with means μ_1, μ_2, variances σ_1^2, σ_2^2, then the sampling distribution of the differences of means $\overline{X}_1$ and $\overline{X}_2$ is approx. normally distributed with

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. $\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

5.6 Chi-square distribution $\sim \chi^2(n)$

If Y is RV with p.d.f $f_Y(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}$ for $y > 0$ and 0 otherwise, then Y has a **chi-square distribution with** n **degrees of freedom**, denoted $\chi^2(n)$, where $n \in \mathbb{Z}^+$ and $\Gamma(\cdot)$ is the gamma function

5.6.1 Gamma function

$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x}\,dx = (n-1)!$ for $n \in \mathbb{Z}^+$

5.6.2 Properties of Chi-square Distribution

- If $Y \sim \chi^2(n)$, then $E(Y) = n$ and $V(Y) = 2n$
- For large n, $\chi^2(n)$ approx. $\sim N(n, 2n)$
- If $Y_1, Y_2, ..., Y_k$ are **independent** chi-square RV with $n_1, n_2, ..., n_k$ degrees of freedom, then $Y_1 + Y_2 + ... + Y_k$ has a chi-square distribution with $n_1 + n_2 + ... + n_k$ degrees of freedom: $\sum_{i=1}^k Y_i \sim \chi^2(\sum_{i=1}^k n_i)$
- $\Pr(Y \ge \chi^2(n;\alpha)) = \alpha$ where $Y \sim \chi^2(n)$. $\Pr(Y \le \chi^2(n; 1-\alpha)) = \alpha$

5.6.3 From Normal to Chi-square

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

5.6.4 The sampling distribution of $\frac{(n-1)S^2}{\sigma^2}$

Let $X_1, X_2, ..., X_n$ be a random sample from a population, then $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ is the sample variance

Theorem: If S^2 is the variance of a random sample of size n taken from a **normal** population, then $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

5.7 The t-distribution

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

5.7.1 The p.d.f of a t-distribution

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

5.7.2 Properties

- Graph of t-distribution is symmetric about the vertical axis, resembles standard normal distribution
- p.d.f of t-distribution is approaching p.d.f of std normal distribution when $n \to \infty$, $\lim_{n\to\infty} f_T(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$
- $E(T) = 0$ and $V(T) = \frac{n}{n-2}$ for $n > 2$
- **Remark:** if the random sample was selected from a normal population, then $Z \sim N(0,1)$ and $U \sim \chi^2(n-1)$, then $T = \frac{Z}{\sqrt{U/(n-1)}} \sim t_{n-1}$

5.8 The F-distribution $\sim F(n_1, n_2)$

Let $U \sim \chi^2(n_1)$ and $V \sim \chi^2(n_2)$, then $F = \frac{U/n_1}{V/n_2}$ is called **F-distribution** with (n_1, n_2) degrees of freedom.

5.8.1 The p.d.f of a F-distribution

$f_F(x) = \frac{n_1^{n_1/2} n_2^{n_2/2} \Gamma(\frac{n_1+n_2}{2}) x^{n_1/2-1}}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})(n_1 x + n_2)^{(n_1+n_2)/2}}$ 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 $n_2 > 4$

5.8.2 Theorem

$F \sim F(n,m) \to \frac{1}{F} \sim F(m,n)$ and $F(n_1, n_2; 1-\alpha) = \frac{1}{F(n_2, n_1; \alpha)}$

6 Estimation based on Normal Distribution

Some characteristics of elements in a population can be represented by an RV X with p.d.f $f_X(x;\theta)$ where the form is assumed known, values of random sample can be observed except unknown parameters θ

6.1 Point Estimation of Mean and Variance

- **Point estimator** is to let the value of some statistic $\hat{\theta} = \hat{\theta}(X_1, X_2, ..., X_n)$ to estimate unknown parameters θ
- **A statistic** is a function of the random sample which does not depend on any unknown parameters. (e.g. sum/max of observations)
- An **estimator** is the statistic used to obtain a point estimate. ($\overline{X}$ is an estimator of μ. The value of $\overline{X}$, $\overline{x}$ is an estimate of μ)

6.1.1 Unbiased Estimator

- A statistic $\hat{\Theta}$ is an **unbiased estimator** of the parameters θ if $E(\hat{\Theta}) = \theta$
- $\overline{X}$ is an unbiased estimator of μ
- $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ is unbiased estimator of σ^2, i.e. $E(S^2) = \sigma^2$

6.2 Interval Estimation

- Form: $\hat{\theta}_L < \theta < \hat{\theta}_U$ where $\hat{\theta}_L$ and $\hat{\theta}_U$ depend on (1) value of the stat $\hat{\Theta}$ for a particular sample, (2) the sampling distribution of $\hat{\Theta}$
- $\hat{\theta}_L$ and $\hat{\theta}_U$ = lower and upper confidence limit, $\hat{\theta}$ = point estimate
- Seek a random interval s.t. $\Pr(\hat{\Theta}_L < \theta < \hat{\Theta}_U) = 1 - \alpha$
- The interval computed from the selected sample is $(1-\alpha)100\%$ confidence interval for α
- $(1-\alpha)$ is **confidence coefficient** or **degree of confidence**

6.3 Confidence Intervals for the Mean

6.3.1 Known Variance Case

- With (i) known variance, (ii) the population is normal, or $n \ge 30$
- $\Pr(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = \Pr(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$
- If $\overline{X}$ is sample mean from a popn with variance σ^2, a $(1-\alpha)100\%$ confidence interval for μ is as expression inside $\Pr$ above (middle)

Sample size for Estimating μ

For a given margin of error e, **sample size** is $n \ge (z_{\alpha/2}\frac{\sigma}{e})^2$

6.3.2 Unknown Variance Case

- With (i) unknown population variance, (ii) the population is normal or very close to normal, (iii) sample size is small ($n < 30$)
- Let $T = \frac{(\overline{X}-\mu)}{S/\sqrt{n}}$ where S^2 is sample variance, then $T \sim t_{n-1}$.
- $\Pr(-t_{n-1;\alpha/2} < T < t_{n-1;\alpha/2}) = \Pr(\overline{X} - t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{n-1;\alpha/2}\frac{S}{\sqrt{n}})$ $= 1 - \alpha$
- $\overline{X}$ and S are sample mean and standard deviation, a $(1-\alpha)100\%$ confidence interval for μ is as expressed inside $\Pr$ above (middle)
- For large $n > 30$, the t-distribution approx. $N(0,1)$. Hence the confidence interval is given by $\overline{X} - z_{\alpha/2}(\frac{S}{\sqrt{n}}) < \mu < \overline{X} + z_{\alpha/2}(\frac{S}{\sqrt{n}})$

6.4 Confidence Intervals for the Difference between 2 Means

2 populations with means μ_1, μ_2, variances σ_1^2, σ_2^2 then $\overline{X}_1 - \overline{X}_2$ is the point estimator of $\mu_1 - \mu_2$

6.4.1 Known variance

- When $\sigma_1^2 \ne \sigma_2^2$ and (2 populations are normal or n_1, n_2 both ≥ 30)
- $(\overline{X}_1 - \overline{X}_2) \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$
- $\Pr\left(-z_{\alpha/2} < \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_{\alpha/2}\right) = 1 - \alpha$
- $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is
  $(\overline{X}_1 - \overline{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{X}_1 - \overline{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

6.4.2 Large sample Confidence Interval for Unknown Variances

σ_1^2, σ_2^2 are unknown, n_1, n_2 both ≥ 30, replace σ_1^2, σ_2^2 by their estimates S_1^2, S_2^2 in **6.4.1**

6.4.3 Unknown but Equal Variances

- $\sigma_1^2 = \sigma_2^2$, 2 populations are normal, n_1, n_2 both ≤ 30
- Pooled sample variance $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \sim \chi^2_{n_1+n_2-2}$
- $T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$
- $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is
  $(\overline{X}_1 - \overline{X}_2) - t_{n_1+n_2-2;\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2$
  $< (\overline{X}_1 - \overline{X}_2) + t_{n_1+n_2-2;\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

6.4.4 Unknown but Equal Variances for Large Samples

For n_1, n_2 both ≥ 30, replace $t_{n_1+n_2-2;\alpha/2}$ by $z_{\alpha/2}$ in **6.4.3**

6.4.5 C.I. for the Difference between 2 Means for Paired (Dependent) Data

- E.g. same individual before and after (related observations)
- Point estimate of $\mu_D = \mu_1 - \mu_2$ is given by $\overline{d} = \frac{1}{n}\sum_{i=1}^n d_i = \frac{1}{n}\sum_{i=1}^n (x_i - y_i)$
- Point estimate of σ_D^2 is given by $s_D^2 = \frac{1}{n-1}\sum_{i=1}^n (d_i - \overline{d})^2$
- **Small sample, approximate normal population**:
  $T = \frac{\overline{d} - \mu_D}{s_d/\sqrt{n}} \sim t_{n-1}$
  $(1-\alpha)100\%$ CI for $\mu_D = \overline{d} - t_{n-1;\alpha/2}(\frac{S_D}{\sqrt{n}}) < \mu_D < \overline{d} + t_{n-1;\alpha/2}(\frac{S_D}{\sqrt{n}})$
- **For large sample** ($n > 30$), CI $= \overline{d} - z_{\alpha/2}(\frac{S_D}{\sqrt{n}}) < \mu_D < \overline{d} + z_{\alpha/2}(\frac{S_D}{\sqrt{n}})$

6.5 C.I. for Variances and Ratio of Variances

6.5.1 C.I. for a Variance of a Normal Population

Sample var $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{1}{n-1}(\sum_{i=1}^n X_i^2 - n\overline{X}^2)$ is pt est of σ^2

CI when μ **is known:** $\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n;\alpha/2}} < \sigma^2 < \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{n;1-\alpha/2}}$

CI when μ **is unknown:** $\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}$

6.5.2 C.I. for the Ratio of 2 Variances of Norm Population with Unknown Means

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

7 Hypotheses Testing based on Normal Distribution

7.1 Null and Alternative Hypotheses

- Null Hypothesis, H_0: Hypothesis formulated with the hope of rejecting, which leads to acceptance of the alternative hypothesis, 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.
- We often choose to state the hypothesis in a form that hopefully will be rejected, i.e. usually H_0 will be the status quo.

7.1.1 Types of errors

- **Type I** (serious): $\Pr(\text{Reject } H_0 \mid H_0 \text{ is true}) = \alpha$ = level of significance
- **Type II**: $\Pr(\text{Do not reject } H_0 \mid H_0 \text{ is false}) = \beta$, Power of a test $= 1-\beta$

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

7.1.2 Acceptance and Rejection Regions

Rejection (critical) region and acceptance region are separated by **critical value**

7.2 Hypotheses Testing Concerning Mean

7.2.1 Hypo Testing on Mean with Known Variance

Variance σ^2 is known and underlying distribution is normal or $n > 30$

Two-sided test:

- Test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. Under H_0, we have $\overline{X} \sim N(\mu_0, \frac{\sigma^2}{n})$
- $\overline{x}_1 < \overline{X} < \overline{x}_2$ or $-z_{\alpha/2} < Z < z_{\alpha/2}$ defines acceptance region.
- The two tails, $\overline{X} < \overline{x}_1$ and $\overline{X} > \overline{x}_2$ constitute the critical or rejection region.
- $\overline{x}_1 = \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ and $\overline{x}_2 = \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
- If $\overline{X}$ falls in acceptance region, conclude $\mu = \mu_0$. Else reject H_0 & accept H_1
- Basically, if the $(1-\alpha)100\%$ confidence interval covers μ_0, null hypothesis is accepted, else it's rejected.

One-sided test: Test $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$ or $H_1 : \mu < \mu_0$. The rest is the same.

7.2.2 p-value Approach to Testing (observed level of significance)

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

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 $\alpha/2$ (or α). If p-value < $\alpha/2$ (or α), reject H_0.

7.2.3 Hypo Testing on Mean with Unknown Variance

Variance unknown and underlying distribution is normal

Two-sided test:

Let $T = \frac{\overline{X}-\mu_0}{S/\sqrt{n}}$ where S^2 is sample variance.Reject H_0 if $> t_{n-1;\alpha/2}$ or $< -t_{n-1;\alpha/2}$

One-sided test:

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

7.3 Hypo Testing Concerning Difference Between 2 Means

7.3.1 Known Variances

Known variances, normal distribution, or n_1, n_2 both ≥ 30, use section 6.4.1. Generally, since variance is known, we will use Z distribution.

7.3.2 Large Sample Testing with Unknown Variances

Unknown variances, both n_1, n_2 both ≥ 30, use section 6.4.2

7.3.3 Unknown but Equal Variances

$\sigma_1^2 = \sigma_2^2$, populations are normal, n_1, n_2 both ≤ 30, use section 6.4.3

7.3.4 Paired Data

Use section 6.4.5

7.4 Hypo Testing Concerning Variance

7.4.1 One Variance Case

- Assume normal distribution where σ^2 is unknown.
- $H_0 : \sigma^2 = \sigma_0^2$, use test statistic: $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$
- Reject H_0 if within critical region:

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 $\chi^2 > \chi^2_{n-1;\alpha/2}$

7.4.2 Hypo Testing Concerning Ratio of Variances

- Assume normal distribution, unknown mean.
- $H_0 : \sigma_1^2 = \sigma_2^2$, use test statistic: $F = \frac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1)$
- Reject H_0 if within 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 $F > F_{n_1-1,n_2-1;\alpha/2}$