Introduction to Inference

Comprehensive Lecture Notes & Formulations

Indian Statistical Institute

Delhi Center

Bachelor of Statistical Data Science (BSDS)

Prepared by: Utkarsh Bharadwaj

Spring 2026

Preface

This document serves as a comprehensive set of notes compiled during the Spring 2026 semester for the **Introduction to Inference** course under the Bachelor of Statistical Data Science (BSDS) program at the **Indian Statistical Institute, Delhi Center**.

The objective of this document is to provide a rigorous, structured, and highly readable foundation for statistical inference. The content spans from the foundational principles of statistical models, populations, and samples, through the intricacies of point and interval estimation, culminating in the robust framework of hypothesis testing. Special emphasis has been placed on the clarity of mathematical derivations, the practical application of Uniformly Minimum Variance Unbiased Estimators (UMVUE), and the fundamental theorems governing large sample theory and the exponential family.

I hope these notes serve as a valuable companion in your study of introductory statistical theory and inference techniques.

— Prepared by Utkarsh Bharadwaj

Contents

1 Statistical Models, Population, and Sample

1.1 Population

Definition

A **Population** is a large, non-random collection of measurements. The population can be represented mathematically by a **Cumulative Distribution Function (CDF)**, denoted as $F(x)$.

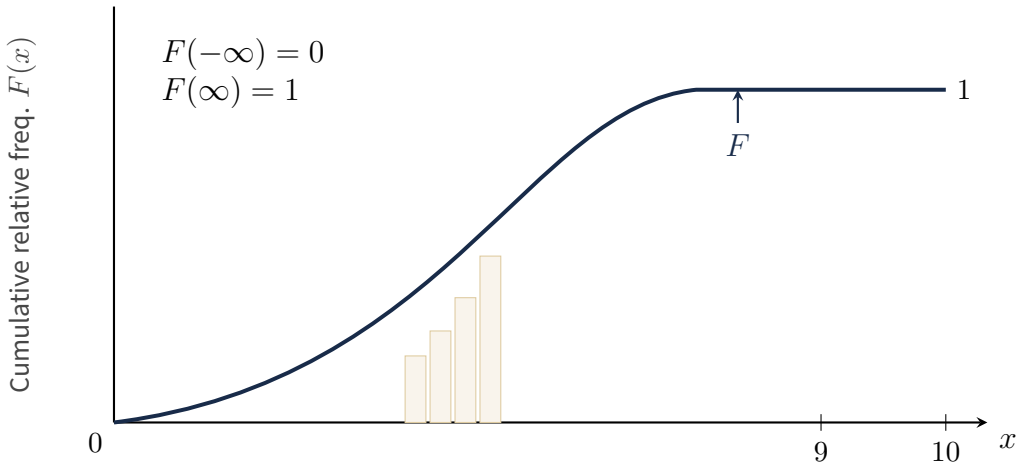

Figure 1: Representation of Population via CDF

Key Property

If the CDF is completely known, then the population is also known completely.

1.2 Sample

Definition

A **Sample** is a number of representative candidates of the population $\mathbb{P}$ to be investigated, denoted as:

$$X_1, X_2, \ldots, X_n$$

- These are random variables drawn from the population.

- Before observing X_1, we do not know the value of X_1. However, every value in the range of $\mathbb{P}$ is a possible value for X_1, but not every value is equally probable.
- The probability is given by:

$$P(X_1 \in (a, b]) = F(b) - F(a)$$

- Since X_i are identically distributed for all i:

$$P(X_i \in (a, b]) = F(b) - F(a) \quad \text{for } i = 2, \ldots, n$$

- Assuming independence, we write:

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} F$$

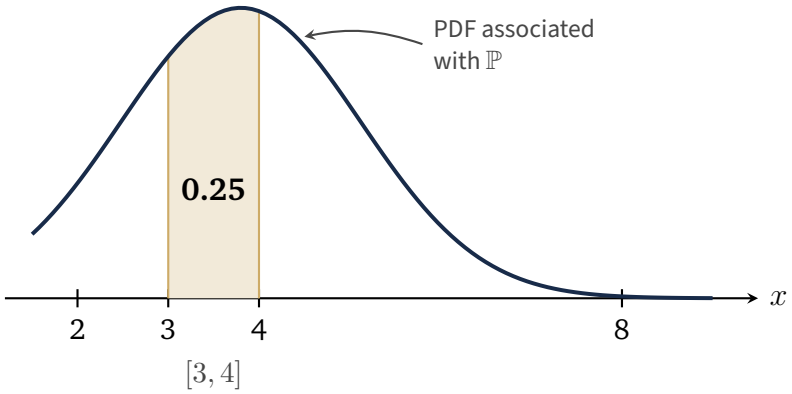

Figure 2: Probability Density Function (PDF) of a Sample

1.3 Model

In practice, the population is **unknown**, and consequently, F is also unknown. However, based on domain knowledge, one can approximate or "guess" the shape of F.

Example: Income Modeling

Consider the distribution of **Income per hour**.

- A significant portion of the population might have 0 income (unemployed), represented by a point mass at 0.
- The rest follow a continuous distribution (e.g., Gamma).

This suggests a **Mixture Model**:

$$\text{Density} = p \cdot \delta_{\{0\}}(x) + (1 - p) \cdot f(x \mid \text{Gamma}(\alpha, \beta))$$

Important Points

- The art of deriving a shape of the population is called 'modeling'.
- The assumed form $F(p, \alpha, \beta)$ with a finite number of unknown constants in it is called a parametric model, and the unknown constants (here p, α and β) are called parameters.

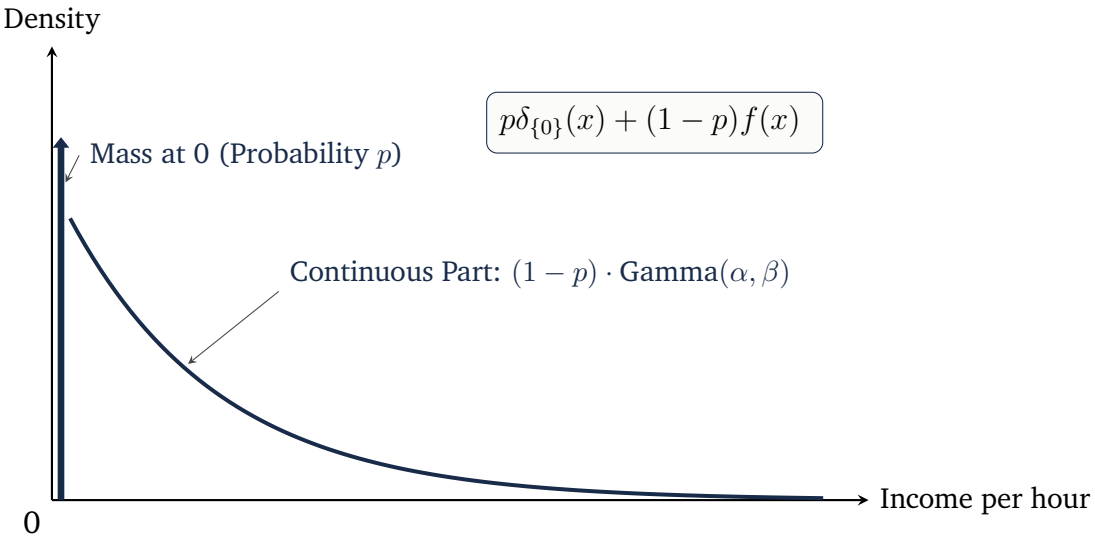

Figure 3: Mixture Model: Zero-inflated Gamma Distribution

2 Statistic and Inference

We observe a sample $X_1, \ldots, X_n \sim G(\underset{\sim}{\theta})$. Our goal is to infer the parameter vector $\underset{\sim}{\theta}$ based on the observed data $X_1, X_2, \ldots, X_n$, where $\underset{\sim}{\theta}$ represents an unknown constant or parameter of the distribution.

EXAMPLE | SAMPLE PROPORTION ESTIMATOR

Consider a binary case (Bernoulli trials). An estimator $T_n(\underset{\sim}{X})$ can be defined as the proportion of successes (1s):

$$T_n(\underset{\sim}{X}) = \frac{\text{Number of 1s in } X_1, \ldots, X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Here, $T_n(\underset{\sim}{X})$ serves as an estimator for the true probability.

Definition: Statistic

Any function of the samples $\{X_1, X_2, \ldots, X_n\}$ that does not depend on unknown parameters is called a **Statistic**. It is a function of $X_1, \cdots, X_n$ denoted by $T_n = T_n(\underline{X})$. **Common Examples:**

1. **Sample Mean:** $T_1(\underset{\sim}{X}) = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
2. **Sample Range:** $T_2(\underset{\sim}{X}) = \max_i X_i - \min_i X_i$

For example, considering the mixture distribution earlier, we had:

$$F_{\underline{\theta}} = p\delta_{\{0\}}(\cdot) + (1-p)f\left(\cdot \mid \text{Gamma}(\alpha, \beta)\right)$$

where the parameter vector is:

$$\underline{\theta} = \begin{pmatrix} p \\ \alpha \\ \beta \end{pmatrix}$$

Important Counter-Example

The following function is **NOT** a statistic:

$$T_3(\underset{\sim}{X}, \theta) = \sum_{i=1}^{n} (X_i - \theta)$$

Reason: It depends on the unknown parameter θ. A statistic must be calculable solely from the data.

2.1 Mathematical Formulation

Formally, a statistic is a mapping $T : \chi \to \mathbb{R}$, where χ is the sample space of $\{X_1, \ldots, X_n\}$. For the Bernoulli example mentioned above, the sample space χ consists of all possible binary vectors of length n:

$$\chi = \left\{ \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \right\}$$

Here, the cardinality of the space is $|\chi| = 2^n$.

2.2 Properties of a Statistic and Inference

- $T(\underset{\sim}{X})$ is random and hence has a distribution. Ex:

$$T(\underset{\sim}{X}) = \begin{cases} 0 & \text{w.p } P(T(\underset{\sim}{X}) = 0) = P(X_1 = 0, X_2 = 0, \cdots, X_n = 0) \end{cases}$$

- The task of parametric inference is to estimate the parameters, or, a function of parameters, from the sample.
- Usually, one uses the functions of sample to infer the population characterstic.
- As the samples are random, a statistic is also a random variable. The distribution of a statistic is called a **Sampling Distribution**.

3 Sampling Distributions

Definition

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F_\theta$ and we have an estimator say, $T_n = h(X_1, \cdots, X_n)$. The distribution of T_n is called the **sampling distribution**.

EXAMPLE | SAMPLING DISTRIBUTION OF THE MEAN

Consider $X_i \sim N(\mu, \sigma^2)$. We can consider a statistic $T_n = \frac{X_1+\cdots+X_n}{n}$. T_n has a distribution which is called the **sampling distribution** which also depends on θ. Say $n = 6$ and we have two cases: $N(0,1)$ and $N(2,5)$. So, $T_n = \overline{X}$ will follow:

- $N\left(0, \frac{1}{6}\right)$
- $N\left(2, \frac{5}{6}\right)$ respectively.

Conclusion:

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

There happen to be three primary techniques to find the sampling distribution:

1. **Distribution function technique:** For eg., distribution of $X_{(n)}$ and $X_{(1)}$.
2. **MGF Technique:** For eg., additive properties.
3. **Transformation technique:** This applies only when X is continuous.

3.1 Distribution Function (CDF) Technique

Let F_T be the CDF of T. So we have:

$$\begin{aligned} F_T(t) &= P(T \leq t) \\ &= P(g(X) \leq t) \end{aligned}$$

Note: Sometimes, it is easy to calculate the inverse of $g(X)$.

EXAMPLE | TRANSFORMATION OF CAUCHY VARIABLE

Let $X \sim \text{Cauchy}(0,1)$. The PDF is given by:

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2} \quad x \in \mathbb{R}$$

Consider the transformation $T = \dfrac{1}{1+X^2}$.

Step-by-Step Derivation

We calculate the CDF of T:

$$\begin{aligned}
F_T(t) = P(T \le t) &= P\left(\frac{1}{1+X^2} \le t\right) \\
&= P\left(1 + X^2 \ge \frac{1}{t}\right) \\
&= P\left(X^2 \ge \left(\frac{1}{t} - 1\right)\right) \\
&= 1 - P\left(X^2 < \left(\frac{1}{t} - 1\right)\right) \\
&= 1 - P\left(-\sqrt{\frac{1}{t} - 1} < X < \sqrt{\frac{1}{t} - 1}\right) \\
&= 1 - \left[F_X\left(\sqrt{\frac{1}{t} - 1}\right) - F_X\left(-\sqrt{\frac{1}{t} - 1}\right)\right] \\
&= 1 - \int_{-\sqrt{\frac{1}{t}-1}}^{\sqrt{\frac{1}{t}-1}} f_X(x)\, dx \\
&= 1 - \frac{1}{\pi}\int_{-\sqrt{\frac{1}{t}-1}}^{\sqrt{\frac{1}{t}-1}} \frac{1}{1+x^2}\, dx \\
&= 1 - \frac{2}{\pi}\arctan\left(\sqrt{\frac{1}{t} - 1}\right)
\end{aligned}$$

Thus, the CDF is:

$$F_T(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 - \frac{2}{\pi}\arctan\left(\sqrt{\frac{1}{t} - 1}\right) & \text{if } 0 \le t \le 1 \\ 1 & \text{if } t > 1 \end{cases}$$

Finding the PDF: We differentiate $F_T(t)$ with respect to t.

$$\begin{aligned}
f_T(t) &= \frac{d}{dt} F_T(t) \\
&= \frac{\Gamma(1)}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})} t^{\frac{1}{2}-1}(1-t)^{\frac{1}{2}-1} \\
&= \frac{1}{\pi}\frac{1}{\sqrt{t}}\frac{1}{\sqrt{1-t}}
\end{aligned}$$

Hence, we arrive at the conclusion that:

$$T \sim \text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)$$

EXAMPLE | PROBABILITY INTEGRAL TRANSFORM

Let X be a continuous random variable and $T = F_X(x)$. **Problem:** Prove that $T \sim \mathcal{U}(0, 1)$.

Proof: Probability Integral Transform

Let $F_T(t)$ be the CDF of the random variable $T = F_X(X)$. **1. Range:** Since F_X is a probability function, the values of T must lie in $[0, 1]$.

- If $t < 0$, $F_T(t) = 0$.
- If $t > 1$, $F_T(t) = 1$.

2. For $0 < t < 1$**:**

$$\begin{aligned} F_T(t) &= P(T \leq t) \\ &= P(F_X(X) \leq t) \end{aligned}$$

Since X is a continuous random variable, F_X is a strictly increasing function (over the support of X). We can apply the inverse function F_X^{-1} to both sides:

$$\begin{aligned} F_T(t) &= P\left(X \leq F_X^{-1}(t)\right) \\ &= F_X\left(F_X^{-1}(t)\right) \quad \text{(By definition of CDF)} \\ &= t \end{aligned}$$

Conclusion: We found that $F_T(t) = t$ for $0 < t < 1$. This is the specific Cumulative Distribution Function of a Standard Uniform variable.

$$\therefore T \sim \mathcal{U}(0, 1)$$

This technique can be generalized to the case where $T = T_n(\underline{X})$.

EXAMPLE | 5 (MINIMUM OF EXPONENTIALS)

$X_i \sim \exp(\lambda) \quad i = 1, \cdots, n$. Consider $T_n = \min\{X_1, \cdots, X_n\}$
Therefore, we have

$$\begin{aligned} F_{T_n}(t) &= P(T_n \leq t) = 1 - P(T_n > t) \\ &= 1 - P(X_1 > t, X_2 > t, \cdots, X_n) \\ &= 1 - P(X_1 > t) \cdots P(X_n > t) \end{aligned}$$

$$= 1 - e^{-\lambda t} \cdots e^{-\lambda t}$$

$$= 1 - e^{-n\lambda t}$$

Therefore, we have

$$f_{T_n}(t) = n\lambda e^{-n\lambda t}$$

Hence, $T_n \sim \exp(n\lambda)$

3.2 MGF Technique

We can use the uniqueness property to arrive at the required density function.

3.3 Transformation Technique

This can be applied only to continuous random variables. Only, one to one transformation. Consider $y = h(X)$ where h is a one-one and onto function. Now consider $h^{-1}(y)$, we have:

$$h^{-1}(y) = \{x : h(x) = y\}$$

Jacobian and PDF Formula

Now we need to find the Jacobian (J):

$$J = \left|\frac{dx}{dy}\right| = \left|\frac{dh^{-1}(y)}{dy}\right|$$

$$= \left|\frac{dh(x)}{dx}\right|^{-1}$$

Now consider the pdf of y, $f_Y(y)$:

$$f_Y(y) = f_X(h^{-1}(y)).\, J \qquad ; y \in$$

EXAMPLE | 1: SIN TRANSFORMATION

Consider $X \sim Uni(-\frac{\pi}{2}, \frac{\pi}{2})$. Hence we have

$$f_X(x) = \begin{cases} \frac{1}{\pi} & \text{if } -\frac{\pi}{2} \le x \le \frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = \sin X$. Thus, we have $y = h(x)$ where $y \in [-1, 1]$. We will now compute the Jacobian:

$$J = \left|\frac{dx}{dy}\right| = \left|\frac{d}{dx}\sin x\right|^{-1}$$

$$= \frac{1}{\sqrt{1-y^2}}$$

Thus we now have

$$f_Y(y) = \frac{1}{\pi}.\frac{1}{\sqrt{1-y^2}} \qquad y \in (-1,1)$$

EXAMPLE | 2: BETA DISTRIBUTION DERIVATION

Consider $X \sim Uni(-\frac{\pi}{2}, \frac{\pi}{2})$. Let $Z = \frac{1+Y}{2} = \frac{1+\sin X}{2}$. We will first compute the Jacobian (J):

$$J = \left| \frac{d\left(\frac{1+y}{2}\right)}{dy} \right|^{-1} = 2$$

Thus we now have,

$$f_Z(z) = \frac{1}{\pi\sqrt{(1-(2z-1))(1+2z-1)}}.2$$
$$= \frac{1}{\pi(1-z)^{1/2}z^{1/2}}$$

Thus we have

$$Z \sim \beta\left(\frac{1}{2}, \frac{1}{2}\right)$$

Note on CDF Approach: We can also further compute the cdf of Y as

$$F_Y(y) = \int_{-\infty}^{y} f_Y(y)dy$$
$$= \int_{-\infty}^{y} f_X(h^{-1}(y))\frac{dh^{-1}(y)}{dy}dy$$
$$= \int_{-\infty}^{h^{-1}(y)} f_X(z)dz$$

where we used the transformation $h^{-1}(y) = z$. If h is decreasing, $J = -\frac{dx}{dy}$

$$F_Y(y_0) = \int_{-\infty}^{y_0} f_Y(y)dy = \int_{-\infty}^{y_0} f_X(h^{-1}(y))\left(-\frac{dh^{-1}(y)}{dy}\right)dy$$
$$= \int_{h^{-1}(y_0)}^{\infty} f_X(z)dz$$

3.4 Common Sampling Distributions

The χ^2 Distribution

Let $X_i \overset{IID}{\sim} N(0,1)$. Consider the sum of squares:

$$Y_n = \sum_{i=1}^{n} X_i^2$$

So Y_n follows a chi-square distribution with n degrees of freedom:

$$Y_n \sim \chi^2_{(n)}$$

Properties: Gamma Relation: $\chi^2_{(n)} = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ **PDF & Moments:**

$$\boxed{f_{\chi^2_{(n)}}(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}$$

$$E(Y_n) = n \qquad Var(Y_n) = 2n$$

The F-Distribution

F-Distribution Construction: Consider $X \sim \chi^2_{(n)}$ and $Y \sim \chi^2_{(m)}$. So we have:

$$Z = \frac{X/n}{Y/m} \sim F_{n,m}$$

The following is the pdf for F-Distribution

$$\boxed{f_{F_{n,m}}(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \left(\frac{n}{m}\right)^{n/2} x^{\frac{n}{2}-1} \left(1 + \frac{n}{m}x\right)^{-\frac{n+m}{2}}}$$

Reciprocal Property: $\frac{1}{Z} \sim F_{m,n}$

The t-Distribution

t-Distribution Construction: Consider $X \sim N(0,1)$ and $Y \sim \chi^2_{(m)}$. So we have:

$$T = \frac{X}{\sqrt{Y/m}} \sim t_{(m)}$$

The following is the pdf for t-Distribution

$$f_{t_{(m)}}(t) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\,\Gamma\left(\frac{m}{2}\right)}\left(1+\frac{t^2}{m}\right)^{-\frac{m+1}{2}}$$

Relation to F: $T^2 \sim F_{1,m}$

3.5 Expectation of Functions

Say $Y = h(x)$. We can find $E(Y)$ without needing to find the distribution of Y:

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

EXAMPLE | 3: EXPECTATION OF SAMPLE MEAN

$X_i \overset{i.i.d.}{\sim} N(\theta, 1)$. Find $E(\overline{X})$.

Solution:

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \theta$$

Hence, we have

$$Y \sim N\left(\theta, \frac{1}{n}\right)$$

4 Estimation Framework

4.1 Basic Concepts of Estimation

Population and Parameters

The population is characterised by the cumulative distribution function $F_{\underset{\sim}{\theta}}$, where $\underset{\sim}{\theta}$ is the parameter vector. We are typically interested in estimating some specific property of this distribution, denoted as $g(\theta)$.

NOTE

- Estimate a function of θ say $g(\theta)$. In order to estimate $g(\theta)$, we put forward a function of $X_1, \ldots, X_n$ say $T(X_1, \ldots, X_n) = T_n$ which is called an *estimator of* $g(\theta)$.
- For a *realization* of the sample, say, $x_1, \ldots, x_n$, the values of T_n, $T(x_1, \ldots, x_n) = t_n$ is called an *estimate of* $g(\theta)$.
- The estimator $T_n = T(X_1, \ldots, X_n)$ is a random variable and it's distribution is sampling distribution which depends on θ.

The Estimator

To estimate $g(\theta)$, we consider a function of the random sample $X_1, \ldots, X_n$:

$$T_n = T(X_1, \ldots, X_n)$$

Sampling Distributions: We may derive the sampling distribution of the estimator itself:

$$T_n \sim f_{T,\theta}$$

We may also derive the distribution of the squared error (or loss):

$$\mathcal{Z} = (T_n - \theta)^2$$

4.2 Mean Squared Error (MSE)

Mean Squared Error (MSE)

Suppose $T_n = T(X_1, \ldots, X_n)$ be an estimator of θ. Then the MSE of T_n is $E\left[(T_n - \theta)^2\right]$
The Mean Squared Error of T_n in estimating θ is defined as:

$$E(\mathcal{Z}_{T_n,\theta}) = E\left([T_n - \theta]^2\right)$$

Comparison of Estimators: We prefer estimator T_n over S_n if:

$$E\left([T_n - \theta]^2\right) < E\left([S_n - \theta]^2\right)$$

Let $E\left[(\overline{X} - \theta)^2\right] = \frac{\theta}{n}$ and $E\left[(\overline{X}_{Me} - \theta)^2\right] = \frac{\theta}{n+1}$, then we'll consider the second estimator as it is giving a lower value of the mean squared error

EXAMPLE | 1 (MSE OF BINOMIAL)

Consider $X_1, \ldots, X_n \overset{iid}{\sim} Ber(\theta)$ and $T_n = \sum X_i$. We know that $T_n \sim Bin(n, \theta)$
Hence, we have

T_n	0	1	...	n
$P(T_n = t)$	$(1-\theta)^n$	$n\theta(1-\theta)^{n-1}$	...	θ^n

Also consider $\psi_T(\theta) = (T_n - \theta)$

$\psi_T(\theta)$	$-\theta$	$-\theta + 1$	...	$n - \theta$
$P(\psi_T(\theta) = \psi)$	$(1-\theta)^n$	$n\theta(1-\theta)^{n-1}$	...	θ^n

Also consider $\psi_T(\theta)^2 = (T_n - \theta)^2$

$\psi_T^2(\theta)$	θ^2	$(1-\theta)^2$	...	$(n-\theta)^2$
$P(\psi_T(\theta) = \psi)$	$(1-\theta)^n$	$n\theta(1-\theta)^{n-1}$	...	θ^n

We have MSE of $T_n = E[\psi_T(\theta)^2] = E[T_n - \theta]^2$

EXAMPLE | 3 (MSE OF NORMAL MEAN)

Let $X_i \overset{i.i.d.}{\sim} N(\theta, 1)$. We know that $\overline{X} = T \sim N\left(\theta, \frac{1}{n}\right)$.
We can also say that:

$$T - \theta \sim N\left(0, \frac{1}{n}\right) \qquad \text{and} \qquad \sqrt{n}(T - \theta) \sim N(0, 1)$$

Hence, using the fact that $Z^2 \sim \chi^2_{(1)}$, we can conclude:

$$n(T-\theta)^2 \sim \chi^2_{(1)} \implies (T-\theta)^2 \sim \frac{\chi^2_{(1)}}{n}$$

Therefore, the Mean Squared Error is:

$$MSE = E\left[(T-\theta)^2\right] = \frac{1}{n}E\left[\chi^2_{(1)}\right] = \frac{1}{n}\cdot 1 = \frac{1}{n}$$

4.3 Comparing Estimators (Variance Estimation)

EXAMPLE | VARIANCE ESTIMATION IN NORMAL DISTRIBUTION

Let $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$. The parameter vector is $\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$. We are interested in estimating $g(\theta) = \sigma^2$. Consider the sample variance estimator:

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

Result: If $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then:

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \sigma^2 \chi^2_{(n-1)}$$

Expectation: Let $Y_{n-1} \sim \chi^2_{(n-1)}$. We can rewrite S_n^2 in terms of this variable:

$$E(S_n^2) = E\left(\frac{\sigma^2}{n}Y_{n-1}\right) = \frac{\sigma^2}{n}E(Y_{n-1})$$

Since $E(\chi^2_{(k)}) = k$, we have:

$$E(S_n^2) = \frac{\sigma^2}{n}(n-1) = \left(\frac{n-1}{n}\right)\sigma^2$$

Thus, we have

$$E(T_n^2) = E\left(\frac{n}{n-1}S_n^2\right) = \sigma^2$$

Note: $T_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is the **unbiased sample variance**.

MSE of S_n^2

Consider the following expansion:

$$\begin{aligned} E\left(\left[S_n^2-\sigma^2\right]^2\right) &= E\left(\left[S_n^2-E(S_n^2)+E(S_n^2)-\sigma^2\right]^2\right) \\ &= E\left(\left[S_n^2-E(S_n^2)\right]^2-\frac{2}{n}\sigma^2\left[S_n^2-E(S_n^2)\right]+\frac{1}{n^2}\sigma^4\right) \\ &= Var(S_n^2)-0+\frac{1}{n^2}\sigma^4 \\ &= Var\left(\frac{\sigma^2}{n}Y_n\right)+\frac{1}{n^2}\sigma^4 \\ &= \frac{\sigma^4}{n^2}(Var(Y_n)+1) \\ &= \frac{2n-1}{n^2}\sigma^4 \end{aligned}$$

MSE of T_n^2

Consider the following:

$$\begin{aligned} E\left(\left[T_n^2-\sigma^2\right]^2\right) = Var(T_n^2) &= Var\left(\frac{\sigma^2}{n-1}Y_n\right) \\ &= \frac{2\sigma^4}{n-1} \end{aligned}$$

EXAMPLE | COMPARISON OF DIFFERENT MSE

We can now compare whether S_n^2 or T_n^2 is a better estimator by comparing their MSE. Hence, we have:

$$MSE_{\sigma^2}(S_n^2)=\frac{2n-1}{n^2}\sigma^4<\frac{2}{n-1}\sigma^4=MSE_{\sigma^2}(T_n^2)$$

Hence, S_n^2 is a better estimator than T_n^2.

5 Unbiased Estimators and UMVUE

5.1 Unbiasedness and Bias-Variance Decomposition

NOTE

- An estimator with smaller MSE is preferred
- In a pool of estimators of $g(\theta)$, the one with the smallest MSE is the *best estimator*
- However, the class of all estimators of $g(\theta)$ is usually very large, and it is NOT possible to find the one with the lowest MSE in this class. Hence, we focus on a sub-class os the class of all estimators i.e., the class of *unbiased estimators*

Bias-Variance Decomposition & Unbiasedness

We often have a class of estimators for $g(\theta)$. The best estimator is that one which has the lowest Mean Squared Error. Consider the following decomposition:

$$\begin{aligned} E\left[(T_n-\theta)^2\right] &= E\left[(T_n - E(T_n) + E(T_n) - \theta)^2\right] \\ &= Var(T_n) + \underbrace{[E(T_n)-\theta]^2}_{\text{Bias}^2} \end{aligned}$$

Due to the large size of the class, we often restrict our class to the class of **Unbiased Estimators**. An estimator $T_n = T(X_1, \ldots, X_n)$ is called unbiased for $g(\theta)$ if

$$E_\theta(T_n) = g(\theta) \quad \forall\, \theta \in \Theta$$

Since we are considering only the class of unbiased estimators, we have to consider $bias = 0$ as $E(T_n) = \theta$. Thus, now in our restricted class, the best estimator is that with **minimum variance**. If T_n is an unbiased estimator of $g(\theta)$, then

$$MSE_{T,n}(g(\theta)) = Var_\theta(T_n)$$

EXAMPLE | 2 (BIAS CHECK)

Let $X_1, \ldots, X_n \sim N(\theta, 1)$. Consider the following two statistics:

$$T_{1n} = \frac{X_1 + \ldots + X_n}{n} \qquad T_{2n} = \sum_{j=1}^{m}(X_{2j} - X_{2j-1})$$

where $n = 2m$. Hence, we have:

$$E_\theta(T_{1n}) = \theta \qquad E_\theta(T_{2n}) = 0$$

Note: T_{2n} is unbiased only when $\theta = 0$, not otherwise

5.2 Uniformly Minimum Variance Unbiased Estimator (UMVUE)

The best unbiased estimator of $g(\theta)$ must be the one with minimum variance.

Definition: UMVUE

An estimator T_n is called the *Uniformly Minimum Variance Unbiased Estimator (UMVUE)* of $g(\theta)$ if

1. T_n is **unbiased for** $g(\theta)$
2. If S_n be another unbiased estimator of $g(\theta)$, then

$$Var_\theta(T_n) \leq Var_\theta(S_n) \quad \forall\, \theta$$

EXAMPLE | 3 (COMPARING VARIANCES)

Consider $X_1, \ldots, X_n \sim Ber(\theta)$

$$T_{1n} = \overline{X} \qquad T_{2n} = \frac{X_2 + X_4 + \ldots + X_n}{m}$$

Both of them are unbiased, hence bias associated with both of them is 0. Now we calculate the variance for both of them. Hence, we have:

$$\begin{aligned} Var_\theta(T_{1n}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \\ &= \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n} \end{aligned}$$

Similar calculations can be performed on T_{2n} to get the following result

$$Var_\theta(T_{2n}) = \frac{2\theta(1-\theta)}{n}$$

Thus, clearly $Var_\theta(T_{1n}) < Var_\theta(T_{2n})$. Hence, T_{1n} is the best estimator

5.3 Relative Efficiency

Definition: Relative Efficiency

Let T_{1n} and T_{2n} be two unbiased estimator of $g(\theta)$, then the *relative efficiency* of T_{2n} compared to T_{1n} is:

$$RE_\theta(T_{1n}, T_{2n}) = \frac{Var_\theta(T_{2n})}{Var_\theta(T_{1n})} \quad ; \text{ for any } \theta$$

If $RE_\theta(T_{1n}, T_{2n}) > 1$, then T_{1n} is better unbiased estimator and vice-versa

6 Fundamental Properties and Lehmann-Scheffé

6.1 Completeness and Sufficiency

How to obtain the UMVUE?

We will focus on *Complete Sufficient statistics (CSS)* only. Suppose T_n is a CSS, if we find a function h of T^* such that

$$E(h(T_n^*)) = g(\theta)$$

then $h(T_n^*)$ must be the $UMVUE$

Completeness

For any function h and statistic T^*, it is said to be *Complete* when

$$E_\theta\left[h(T^*)\right] = 0 \implies h(T^*) = 0$$

Sufficiency

Sufficiency of T^* for θ requires that the *conditional distribution of* $\underline{X}$ *given* T^* *is free of* $\underline{\theta}$.

NOTE

Any one-to-one function of a complete-sufficient statistic (CSS) is also a CSS.

6.2 The Lehmann-Scheffé Theorem

Lehmann-Scheffé Theorem

The **Lehmann-Scheffé Theorem** states that if we have a *Complete Sufficient Statistic (CSS)*, if any function that we can find of it is unbiased for the desired parameter, then that function of the CSS is a *Uniformly Minimum Variance Unbiased Estimator or (UMVUE)*. If T^* is a CSS and there exists a function of T^*, say

$h(T^*)$ such that $E(h(T^*)) = g(\theta)$, then $h(T^*)$ is the *UMVUE* of $g(\theta)$.

6.3 Applications of Lehmann-Scheffé

EXAMPLE | 4 (UMVUE FOR BERNOULLI)

Consider $X_i \overset{iid}{\sim} Ber(\theta)$. Say we already know that $\sum_{i=1}^n X_i$ is CSS. We are interested in finding the UMVUE. **Solution:** Consider $E\left(\sum_{i=1}^n X_i\right)$. We have the following:

$$E\left(\sum_{i=1}^n X_i\right) = n\theta$$

$$E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \theta$$

Because $\frac{1}{n}\sum X_i = \overline{X}$ is a function of the CSS and is an unbiased estimator for θ, by the Lehmann-Scheffé theorem, $\overline{X}$ is the unique UMVUE for θ.

EXAMPLE | UNIFORM DISTRIBUTION

Given $X_i \overset{iid}{\sim} U(0, \theta)$ and also given that $X_{(n)} = T^*$ is a CSS for θ. Consider the following:

$$E_\theta(T^*) = \int_0^\theta t\frac{n}{\theta^n}t^{n-1}dt$$

$$= \frac{n}{n+1}\theta$$

Therefore, we have:

$$E_\theta\left(\frac{n+1}{n}T^*\right) = \theta$$

Thus, $\left(\frac{n+1}{n}T^*\right)$ is *UMVUE* for θ.

7 The Exponential Family

7.1 Definition and Form

Exponential Family

We have $n,\ iid$ random variables such that $X_1, \ldots, X_n \overset{iid}{\sim} F_\theta$. F_θ belongs to the exponential family if the joint pdf/pmf of $X_1, \ldots, X_n$ can be expressed as follows:

$$f_{X_1,\ldots,x_n}(x_1, \ldots, x_n) = \exp\left\{ A(\underline{\theta}) + H(\underline{X}) + \sum_{j=1}^{k} c_j(\underline{\theta}) T_j(\underline{X}) \right\}$$

where $A(\underline{\theta})$ and $c_j(\underline{\theta}) \forall\ j$ are functions of $\underline{\theta}$ only, and $H(\underline{X})$ and $T_j(\underline{X}) \forall\ j$ are functions of $\underline{X}$ only.

7.2 Complete Sufficient Statistics (CSS) Connection

Connection of CSS and Exponential Family

If $\Theta \in \mathbb{R}^d$ contains an open set in it, then $\{T_1(\underline{X}), \ldots, T_k(\underline{X})\}$ is jointly CSS.

7.3 Common Distributions in the Exponential Family

EXAMPLE | BINOMIAL DISTRIBUTION

$X_i \overset{iid}{\sim} Bin(m, \theta)$. We can write it as:

$$\begin{aligned}
f_{X_1,\ldots,x_n}(x_1, \ldots, x_n) &= \prod_{i=1}^{n} f_{X_i}(x_i) \\
&= \prod_{i=1}^{n} \binom{m}{x_i} \theta^{x_i} (1-\theta)^{m-x_i} \\
&= \exp\left\{ \sum_{i=1}^{n} \left[\log \binom{m}{x_i} + x_i \log\theta + (m - x_i)\log(1-\theta) \right] \right\} \\
&= \exp\left\{ \sum_{i=1}^{n} \log \binom{m}{x_i} + \log\theta \sum_{i=1}^{n} x_i \right.
\end{aligned}$$

$$+ mn\log(1-\theta) - \log(1-\theta)\sum_{i=1}^{n} x_i \Bigg\}$$

$$= \exp\Bigg\{ \underbrace{\sum_{i=1}^{n} \log\binom{m}{x_i}}_{H(\underline{X})} + \underbrace{mn\log(1-\theta)}_{A(\theta)}$$

$$+ \underbrace{\left(\sum_{i=1}^{n} x_i\right)}_{T(\underline{X})} \underbrace{\log\left(\frac{\theta}{1-\theta}\right)}_{C(\theta)} \Bigg\}$$

Hence $T(\underline{X}) = \sum_{i=1}^{n} X_i$ is CSS for θ.

EXAMPLE | GAMMA DISTRIBUTION

$X_i \overset{iid}{\sim} \Gamma(\alpha, \beta)$. We can write it as:

$$f_{X_1,\ldots,x_n}(x_1,\ldots,x_n) = \exp\Bigg\{ \underbrace{n\alpha\log\beta - n\log\Gamma(\alpha)}_{A(\underline{\theta})}$$

$$+ \alpha\sum_{i=1}^{n}\log X_i - \beta\sum_{i=1}^{n} X_i - \sum_{i=1}^{n}\log X_i \Bigg\}$$

Thus, we have the following:

$$c_1(\underline{\theta}) = \alpha \qquad T_1(\underline{X}) = \sum_{i=1}^{n}\log X_i$$

$$c_2(\underline{\theta}) = -\beta \qquad T_2(\underline{X}) = \sum_{i=1}^{n} X_i$$

Thus, $T(\underline{X}) = (\sum_{i=1}^{n}\log X_i, \sum_{i=1}^{n} X_i)$ is jointly CSS for $\underline{\theta} = (\alpha, \beta)$.

EXAMPLE | NORMAL DISTRIBUTION

$X_i \overset{iid}{\sim} N(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma^2 > 0$. Thus we have:

$$f_{X_1,\ldots,x_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$= \exp\left\{ -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum x_i^2 + \frac{\mu}{\sigma^2}\sum x_i - \frac{\mu^2}{2\sigma^2}n \right\}$$

$$= \exp\left\{ \underbrace{-\frac{n}{2}\left(\log 2\pi\sigma^2 + \frac{\mu^2}{\sigma^2}\right)}_{A(\underline{\theta})} + \underbrace{\frac{\mu}{\sigma^2}\sum x_i - \frac{1}{2\sigma^2}\sum x_i^2}_{\sum c(\underline{\theta})T(\underline{X})} \right\}$$

Therefore, we have the following:

$$c_1(\underline{\theta}) = \frac{\mu}{\sigma^2} \qquad T_1(\underline{X}) = \sum_{i=1}^{n} X_i$$

$$c_2(\underline{\theta}) = -\frac{1}{2\sigma^2} \qquad T_2(\underline{X}) = \sum_{i=1}^{n} X_i^2$$

Hence we arrive at the conclusion that $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ are *jointly CSS* for $\underline{\theta} = (\mu, \sigma^2)$.
Consider the following for μ:

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

$$= n\mu$$

$$\implies E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = E(\overline{X}) = \mu$$

Therefore, $\overline{X}$ is the unique UMVUE for μ. Now, consider the following for σ^2: We need a function of our jointly CSS $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ that is unbiased for σ^2. Let's examine the sample variance, S^2:

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

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

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right)$$

Observe that S^2 is strictly a function of the CSS components $T_1 = \sum X_i$ and $T_2 = \sum X_i^2$. Since it is a well-known result that $E(S^2) = \sigma^2$, we can apply the Lehmann-Scheffé Theorem. Therefore, S^2 is the unique UMVUE for σ^2.

8 Methods of Point Estimation

8.1 Overview of Estimation Methods

> **NOTE**
>
> We have some sample points $X_1, \ldots, X_n \sim F_{\underline{\theta}}$
> We want to estimate some property of the population $g(\underline{\theta})$ based on $X_1 \ldots, X_n$
> We had a statistic as $T_n = T(X_1, \ldots, X_n)$. It has a distribution and it is a function of X_i
> We talked about the *Mean Squared Error (MSE)* $= \mathbb{E}\left([T_n - g(\theta)]^2\right)$. We then only consider the class of unbiased estimators of $g(\theta)$ such that $\mathbb{E}(T_n) = g(\theta)$

How to obtain an estimator for $g(\theta)$?

- Method of moments (MoM)
- Maximum Likelihood Estimation (MLE)

8.2 Method of Moments (MoM)

Method of Moments

Population Moment: $X \sim F_\theta$. We have the following:

$$r^{\text{th}} \text{ order raw moment} = E(X^r) = \mu'_r \quad r = 0, 1, \ldots$$
$$r^{\text{th}} \text{ order central moment} = E\left[(X - \mu'_1)^r\right] \quad r = 0, 1, \ldots$$

$$m'_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r \qquad \text{(sample raw moment)}$$

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - m'_1)^r \text{(sample central moment)}$$

The Method of Moments idea is to solve $m'_r = \mu'_r(\theta)$ in order to obtain an estimate of $\underline{\theta}$

EXAMPLE | GAMMA

$$m'_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \mu'_1(\theta) = \frac{\alpha}{\beta}$$

$$m_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu_2'(\theta) = \frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2}$$

Thus we have the following results

$$\hat{\alpha} = m_1'\hat{\beta} \qquad \hat{\beta} = \frac{m_1'}{m_2' - (m_1')^2}$$

NOTE

- If we want to estimate $\Psi(\theta) = \Psi\left(\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right)$ then the MoM estimator would be $\Psi(\hat{\theta}_{\text{MoM}})$. For example, if we want to estimate $\alpha + \beta$ then the MoM estimator will be $\hat{\alpha}_{MoM} + \hat{\beta}_{MoM}$
- Solve the equation $m_r' = \mu_r'(\underline{\theta})$ in order to obtain the estimate of $\underline{\theta}; r = 1, \ldots, k$ where $\underline{\theta} \in \mathbb{R}^k$

EXAMPLE | NORMAL DISTRIBUTION

$X_i \overset{iid}{\sim} N(\mu, \sigma^2)$. We need to solve the following two MoM equations

$$m_1' = \overline{X} = \mu$$

$$m_2' = \overline{X^2} = \sigma^2$$

EXAMPLE | UNIFORM DISTRIBUTION

$X_i \overset{iid}{\sim} U(0, \theta)$ Thus we have

$$\overline{X}_n = \frac{\theta}{2}$$

This implies

$$\hat{\theta}_{\text{MoM}} = 2\overline{X}_n$$

Recall that the MSE of $2\overline{X}_n$ is quite high as compared to the UMVUE which was $\frac{n+1}{n}X_{(n)}$

NOTE

- MoM is an adhoc method and not very usefull in modern days

8.3 Intuitive Introduction to MLE

Maximum Likelihood Estimator (MLE)

- Consider $X_i \sim Ber(\theta)$ where θ is the probability of appearance of Head. Say we know that $\theta \in \left\{\frac{1}{4}, \frac{3}{4}\right\}$. We look at the value of $X_i's$ and then estimate θ. Say we tossed the coin 10 times and obtained 6 heads. Thus we have

$$\mathbb{P}_{\theta=1/4}\left(\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{4}\right)^6 \left(\frac{3}{4}\right)^4$$

$$\mathbb{P}_{\theta=3/4}\left(\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right)^6$$

- X is the lifetime of a mobile. there are 3 types of mobiles
  - Good with average lifetime 6 Years
  - Medium with average lifetime 6 Years
  - Bad with average lifetime 6 Years

  Say $X \sim \exp(\theta)$. So $f_\theta(x) = \frac{1}{\theta}e^{-x/\theta}$ and $\mathbb{E}(X) = \theta$. Our parameter space is $\{1, 3, 6\}$. We observe $x = 4.5$. Thus, consider the following

$$f_1(4.5) = e^{-4.5} = 0.011$$

$$f_3(4.5) = \frac{1}{3}e^{-4.5/3} = 0.0743$$

$$f_6(4.5) = \frac{1}{6}e^{-4.5/6} = 0.078$$

9 Maximum Likelihood Estimation

9.1 The Likelihood Function and MLE Definition

EXAMPLE | EXPONENTIAL

$X \sim \exp(\theta)$ and $f_\theta(x) = \frac{1}{\theta}e^{-x/\theta}$. Given parameter space as $\theta \in \{2, 3, 6\}$. Observed value is 4.5. based on the calculations of $f_\theta(4.5)$ we estimate that the value of θ must be 6

Generalization 1: Suppose $n = 5$. Let the observations be $\{2, 3.2, 4.5, 4, 6\}$. We then consider the joint pdf

$$f_\theta(\underline{x}) = \frac{1}{\theta^5}\exp\left\{-\frac{\sum x_i}{\theta}\right\}$$
$$= \frac{1}{\theta^5}\exp\left\{-\frac{19.7}{\theta}\right\}$$

Compute $f_\theta(\underline{x})$ for $\theta = 2, 3, 6$ and $\hat{\theta}$ such that,

$$\hat{\theta} = \arg\max_{\theta=\{2,3,6\}} f_\theta(\underline{x})$$

After calculations, it turns out that $\hat{\theta} = 3$

Generalization 2: Let $\theta \in (0, \infty)$

$$\hat{\theta} = \arg\max_{\theta\in(0,\infty)} f_\theta(\underline{x})$$

This methos of estimation θ is called the *Maximum Likelihood estimation*

> **NOTE**
>
> The $\arg\max f_\theta(\underline{x})$ will not depend on θ

Likelihood Function

$X_1, \ldots, X_n \overset{iid}{\sim} F_{\underline{\theta}}$ having pdf/pmf as $f_{\underline{\theta}}$, then the likelihood function of $\underline{\theta}$, denoted by $L(\underline{\theta}; \underline{X})$ is given by

$$L(\underline{\theta}; \underline{X}) = \prod_{i=1}^{n} f_\theta(X_i)$$

Log-Likelihood Function: Sometimes it is convenient to work in log scale. Thus

we have

$$l(\theta; X) = \log L(\theta; X) = \sum_{i=1}^{n} \log f_\theta(\underline{X})$$

Maximum Likelihood Estimator

The MLE is the $\theta \in \Theta$, where Θ is the parameter space, say $\hat{\theta}_{ML}$ such that

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} L(\theta; \underline{X})$$

NOTE

1. As $\hat{\theta}_{ML}$ is $\arg\max$ of $L(\theta; \underline{X})$, so it is a function of $\underline{X}$ only, hence a *statistic*
2. As $\hat{\theta}_{ML}$ is $\arg\max$ over $\theta \in \Theta$, $\hat{\theta}_{ML}$ must lie in Θ. So, $\hat{\theta}_{ML}$ is a valid estimator of θ
3. If convenient, on can use the log-likelihood function of the likelihood because *it is a monotonically increasing function*

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} l(\underline{\theta}; \underline{X})$$

For θ_1, θ_2 if $L(\theta_1) > L(\theta_2) \implies l(\theta_1) > l(\theta_2)$

9.2 Deriving MLEs for Continuous Distributions

EXAMPLE | NORMAL DISTRIBUTION

$X_i \overset{iid}{\sim} N(\mu, \sigma^2)$. Consider the likelihood function

$$L(\theta, X) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2\right\}$$

$$l(\theta, X) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

Consider the following

$$\frac{\partial l(\theta, X)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = 0$$

$$\frac{\partial l(\theta, X)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

Thus, we have

$$\hat{\mu}_{ML} = \overline{X} \qquad \hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

The solution obtained here are solution to first order condition. After this, we calculate the Hessian Matrix which would be $\frac{\partial^2 l}{\partial\theta\partial\theta'}$

$$H(\mu,\sigma^2)=\begin{bmatrix}\frac{\partial^2 l}{\partial\mu^2} & \frac{\partial^2 l}{\partial\sigma^2\partial\mu}\\ \frac{\partial^2 l}{\partial\mu\partial\sigma^2} & \frac{\partial^2 l}{\partial(\sigma^2)^2}\end{bmatrix}$$

Check if $H(\hat{\mu},\hat{\sigma}^2)$ is negative definite
Thus, we have

$$H(\hat{\mu},\hat{\sigma}^2)=\begin{bmatrix}-\frac{n}{\hat{\sigma}^2} & 0\\ 0 & -\frac{n}{2\hat{\sigma}^4}\end{bmatrix}$$

NOTE

$$M=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$$

if $a<0$ and $\det(M)>0$, then M is negative definite

EXAMPLE | UNIFORM

$X_i \overset{iid}{\sim} U[0,\theta]$. Consider the likelihood function

$$L(\theta;\underline{X})=\begin{cases}\frac{1}{\theta^n} & \text{if } X_i\le\theta\forall i\\ 0 & \text{otherwise}\end{cases}$$

Thus we have

$$\hat{\theta}_{ML}=X_{(n)}$$

EXAMPLE | UNIFORM DISTRIBUTION

$X_i \overset{iid}{\sim} U(\alpha,\beta)$. Consider the joint density

$$f_{\underline{\theta}}(\underline{x})=\begin{cases}\frac{1}{(\beta-\alpha)^n} & \alpha\le x_i\le\beta \quad \forall i\\ 0 & \text{otherwise}\end{cases}$$

Thus we can also say that $\alpha\le x_{(1)}\ldots\le x_{(n)}\le\beta$
For any fixed β, the likelihood is maximized if we take $\alpha=x_{(1)}$ and for any fixed α, the likelihood is maximized if we take $\beta=x_{(n)}$
Thus, we have

$$\hat{\alpha}_{ML}=x_{(1)} \qquad \hat{\beta}_{ML}=x_{(n)}$$

such that

$$\arg\max_{(\alpha,\beta)} L(\underline{\theta}, \underline{X}) = \frac{1}{\left(x_{(n)} - x_{(1)}\right)^2}$$

EXAMPLE | DOUBLE EXPONENTIAL DISTRIBUTION

$X_i \overset{iid}{\sim}$ Double Exponential(μ, β). Consider the pdf

$$f_\theta(x) = \frac{1}{2\beta} \exp\left\{-\frac{|x-\mu|}{\beta}\right\} \qquad x, \mu \in \mathbb{R}, \beta > 0$$

Consider the log-likelihood function

$$l(\theta) = -n\log 2 - n\log\beta - \frac{1}{\beta}\sum_{i=1}^{n}|x_i - \mu|$$

This is differentiable for β for not for μ. **Estimating β:** So, first we fix $\mu \in \mathbb{R}$ and maximize $l(\theta; X)$ with respect to β. Thus, we have

$$\frac{\partial l}{\partial \beta} = 0 \implies -\frac{n}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n}|x_i - \mu| = 0$$

Hence

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \mu| \qquad \text{Solution to the FOC}$$

Consider

$$\frac{\partial^2 l}{\partial \beta^2} = -\frac{n}{\hat{\beta}} < 0$$

Thus, we attain a maxima
Estimating μ: Fix $\beta > 0$ and maximize $l(\theta, X)$ w.r.t μ. Observe that $l(\theta, X)$ is maximized w.r.t μ if $\sum|x_i - \mu|$ is minimized. $\sum|x_i - \mu|$ is minimized if μ is the median of $x_1, \ldots, x_n$. Thus, we have

$$\hat{\mu}_{ML} = \text{median}\{x_1, \ldots, x_n\} \qquad \hat{\beta}_{ML} = \frac{1}{n}\sum_{i=1}^{n}|X_i - \hat{\mu}_{ML}|$$

Suppose it is further known that $\beta \geq 1$. Thus, we have

$$\hat{\beta}_{ML} = \max\left\{1, \frac{1}{n}\sum_{i=1}^{n}|X_i - \hat{\mu}_{ML}|\right\}$$

9.3 Properties and Special Cases

NOTE

- MLE is not necessarily unbiased
- MLE may not be unique
- MLE may not exist

EXAMPLE | MLE - DISCRETE PARAMETER CASE

$X \sim Bin(\theta, p_0)$ where $\theta \in \{x, x+1, \ldots\}$. Thus we have

$$L(\theta, X) = \binom{\theta}{x} p_0^x (1-p_0)^{\theta - x}$$

Consider the following

$$\begin{aligned}\frac{L(\theta, X)}{L(\theta-1, X)} &= \frac{\binom{\theta}{x} p_0^x (1-p_0)^{\theta-x}}{\binom{\theta-1}{x} p_0^x (1-p_0)^{\theta-x-1}} \\ &= \frac{\theta}{\theta - x}(1-p_0) = \frac{\theta - \theta p_0}{\theta - x}\end{aligned}$$

Thus we have

$$\begin{aligned}L(\theta, X) > L(\theta-1, X) &\iff \theta < \frac{x}{p_0} \\ L(\theta, X) = L(\theta-1, X) &\iff \theta = \frac{x}{p_0} \\ L(\theta, X) < L(\theta-1, X) &\iff \theta > \frac{x}{p_0}\end{aligned}$$

- Case 1: When $\frac{x}{p_0}$ is an integer
  We don't have unique MLE in this case
- Case 2: When $\frac{x}{p_0}$ isn't an integer
  We then have a unique MLE which is $\hat{\theta}_{ML} = \left\lfloor \frac{x}{p_0} \right\rfloor$

Invariance of MLE

Suppose ψ is a function of θ, $\psi = h(\theta)$ and MLE of θ is $\hat{\theta}_{ML}$ then MLE of ψ is $\hat{\psi}_{ML} = h(\hat{\theta}_{ML})$

EXAMPLE | POISSON DISTRIBUTION

$X_i \overset{iid}{\sim} Pois(\theta)$. It is easy to verify that $\hat{\theta}_{ML} = \overline{X}$. Define

$$\psi = P_\theta(X = 1) = e^{-\theta}\theta$$

Thus, we have

$$\hat{\psi}_{ML} = \hat{\theta}e^{-\hat{\theta}}$$

10 Inequalities and Large Sample Theory

10.1 Fundamental Inequalities

Markov's Inequality

If X is a non-negative random variable, then

$$P(X > t) \leq \frac{E(X)}{t}$$

Proof. X is a non-negative continuous random variable. Consider the following

$$\begin{aligned} E(X) &= \int_0^\infty x f_X(x) dx \\ &= \int_0^t x f_X(x) dx + \int_t^\infty x f_X(x) dx \\ &\geq \int_t^\infty x f_X(x) dx \\ &\geq t \int_t^\infty f(x) dx = tP(X > t) \end{aligned}$$

Thus we have

$$P(X > t) \leq \frac{E(X)}{t}$$

□

Chebyshev's Inequality

If X is a random variable with expectation $E(X) = \mu$ and finite variance $Var(X) = \sigma^2$, then

$$P\left(|X - \mu| > t\right) \leq \frac{\sigma^2}{t^2}$$

Proof. Consider

$$\begin{aligned} P\left(|X - \mu| > t\right) &= P\left((x - \mu)^2 > t^2\right) \\ &\leq \frac{E[(X - \mu)^2]}{t^2} = \frac{\sigma^2}{t^2} \end{aligned}$$

□

We can also say that

$$P\left(|x-\mu|>t\right) \leq \frac{E[|X-\mu|^r]}{t^r} \quad t>0, r=1,2,\ldots$$

EXAMPLE | 1

$X_i \overset{iid}{\sim} F_\theta$. Say $E(X)=\mu$ and $Var(X)=1$
What is the minimum sample size which ensures

$$P\left(|\overline{X}_n-\mu|\leq 1\right) \geq 0.95$$

Solution

By Chebyshev's inequality, we have

$$P\left(|\overline{X}_n-\mu|>1\right) \leq \frac{1}{n} \qquad \left(\because Var(\overline{X}_n)=\frac{1}{n}\right)$$

This implies,

$$P\left(|\overline{X}_n-\mu|\leq 1\right) \geq 1-\frac{1}{n}$$

Thus we have, $1-\frac{1}{n}=0.95$. This implies $n=20$

10.2 Modes of Convergence

Notion of Convergence

Recall that Chebyshev's Inequality suggests that

$$P\left(|X-E(X)|>t\right) \leq \frac{Var(X)}{t^2}$$

Note: *We have the following definition of convergence of a sequence* $\{a_n\}$ *of real number*

For any $\epsilon>0, \exists n \geq N$ such that $|a_n-a|<\epsilon$

For any estimator (T_n) to be a good estimator of θ, it must *concentrate around* θ
Note: *Since* T_n *is a random variable and theta is a real number, we can't apply the conventional definiton of convergence*

Convergence in Mean

Given T_n as a sequence of random variables, it is considered a good estimator of $g(\theta)$ if $T_n \xrightarrow{L^r} g(\theta)$ as $n \to \infty$
We write this as $E\left[|T_n - T|^r\right] \to 0$ as $n \to \infty$

EXAMPLE | UNIFORM DISTRIBUTION

Consider $X_i \overset{iid}{\sim} U(0, \theta)$. Let $T_n = X_{(n)}$
We know that

$$f_{T_{(n)}}(t) = \frac{n}{\theta^n} t^{n-1} \qquad E_\theta[T_n] = \frac{n}{n+1}\theta$$

Thus we have

$$\begin{aligned} E\left[|T_n - \theta|\right] &= \theta - E(T_n) \\ &= \theta - \frac{n}{n+1}\theta \\ &= \frac{\theta}{n+1} \end{aligned}$$

Clearly, $E[|T_n - \theta|] \xrightarrow{L^1} 0$ when $n \to \infty$

NOTE

1. $E(T_n) \to E(T)$ as $n \to \infty$ *doesn't imply that* $T_n \xrightarrow{L^r} T$
2. $T_n \xrightarrow{L^r} T$as $n \to \infty \implies T_n \xrightarrow{L^s} T$as $n \to \infty$ for $s < r$ *but the converse is not true*

EXAMPLE | 1

$$X_n = \begin{cases} \sqrt{n} & \text{w.p. } \frac{1}{n} \\ 0 & \text{w.p. } \left(1 - \frac{1}{n}\right) \end{cases}$$

Convergence in L^1

Consider the following

$$\begin{aligned} E(|X_n - 0|) &= E(X_n) \\ &= \sqrt{n}.\frac{1}{n} + 0.\left(1 - \frac{1}{n}\right) \\ &= \frac{1}{\sqrt{n}} \end{aligned}$$

We can see that $\frac{1}{\sqrt{n}} \to 0$ as $n \to \infty$
Thus, we have

$$\boxed{X_n \xrightarrow{L^1} 0}$$

Convergence in L^2

Consider the following

$$\begin{aligned} E(|X_n - 0|^2) &= E(X_n^2) \\ &= n.\frac{1}{n} + 0.\left(1 - \frac{1}{n}\right) \\ &= 1 \end{aligned}$$

Clearly, $1 \not\to 0$ as $n \to \infty$
Hence, we can conculde that

$$\boxed{X_n \xrightarrow{L^2} 0}$$

Hence, we can verify that the converse of note 2 is not true

Convergence in Probability

$T_n \to T$ in probability as $n \to \infty$ is for any $\epsilon > 0$,

$$P\left(|T_n - T| > \epsilon\right) = p_n \to 0 \quad \text{as } n \to \infty$$

Thus we write,

$$T_n \xrightarrow{p} T \quad \text{as } n \to \infty$$

Similarly, $T_n \xrightarrow{p} g(\theta)$ if for any $\epsilon > 0$,

$$P\left(|T_n - g(\theta)| > \epsilon\right) \to 0 \quad \text{as } n \to \infty$$

We can also write that

$$T_n = g(\theta) \quad \text{w.p. } 1$$

EXAMPLE | 2

$X_i \overset{iid}{\sim} U(0, \theta)$
Does $T_n = X_{(n)} \xrightarrow{p} \theta$?

Solution:

For $0 < \epsilon < \theta$

$$\begin{aligned} P\left(|T_n - \theta| \leq \epsilon\right) &= P(\theta - \epsilon \leq T_n \leq \theta) \\ &= \int_{\theta-\epsilon}^{\theta} \frac{nt^{n-1}}{\theta^n} dt \\ &= 1 - \left(1 - \frac{\epsilon}{\theta}\right)^n \end{aligned}$$

Thus, we have

$$P(|T_n - \theta| > \epsilon) = \left(1 - \frac{\epsilon}{\theta}\right)^n$$

This we have

$$P(|T_n - \theta| > \epsilon) \to 0$$

For $\epsilon > \theta$, we have

$$\begin{aligned} P(|T_n - \theta| \leq \epsilon) &= P(\theta - \epsilon \leq T_n \leq \theta) \\ &= P(0 \leq T_n \leq \theta) \\ &= 1 \end{aligned}$$

Thus we have

$$P(|T_n - \theta| > \epsilon) = 0$$

NOTE

1. If $T_n \xrightarrow{L^r} T \implies T_n \xrightarrow{p} T$ as $n \to \infty$. Hence, we can say that L^r *convergence is stronger than probability convergence*

Proof. Fix $\epsilon > 0$, we have

$$P(|T_n - T|^r > \epsilon) \leq \frac{E|T_n - T|^r}{\epsilon^r} \to 0$$

□

The converse may not be true

EXAMPLE | 3

$$X_n = \begin{cases} n & \text{w.p. } \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}$$

Clearly, $X_n \xrightarrow{L^r} 0$ but $X_n \xrightarrow{p} 0$
$E|X_n| = 1 \not\to 0$ as $n \to \infty$
Fix, $\epsilon > 0$, we have

$$P\left(|X_n| > \epsilon\right) = p_n$$

We can verify that $p_n = \frac{1}{n} \; \forall \; n \geq \epsilon$

Consistency

An estimator T_n of $g(\theta)$ is said to be consistent if

$$T_n \xrightarrow{p} g(\theta) \quad \text{as } n \to \infty$$

Example: $X_{(n)}$ is consistent for $X_i \overset{iid}{\sim} U(0, \theta)$
Sufficient Condition for Consistency
If

$$E(T_n) \to g(\theta) \quad \text{and} \quad Var(T_n) \to 0 \qquad \text{as } n \to \infty$$

then we can say that $T_n \xrightarrow{p} g(\theta)$ and hence T_n is consistent.

Proof. Fix $\epsilon > 0$

$$P(|T_n - g(\theta)|^2 > \epsilon^2) \leq \frac{E\left[(T_n - g(\theta))^2\right]}{\epsilon^2} \qquad \text{(as per chebyshev's inequality)}$$

$$= \frac{\{E(T_n) - g(\theta)\}^2 + Var_\theta(T_n)}{\epsilon^2}$$

$$\to 0$$

Thus, we have $T_n \xrightarrow{p} g(\theta)$

□

EXAMPLE | 4

$X_i \overset{iid}{\sim} F_\theta$ with $E(X_1) = \mu_\theta$ and $Var(X_1) = \sigma^2_\theta$

Analysis of Sample Mean

We know as facts that for $T_n = \overline{X}_n$

$$E(T_n) = \mu_\theta \qquad Var(T_n) = \frac{\sigma^2_\theta}{n} \to 0 \text{ as } n \to \infty$$

Thus, we can conclude that $T_n = \overline{X}_n$ is consistent for θ

EXAMPLE | 5

$X_i \overset{iid}{\sim} N(\mu, \sigma^2)$.

Analysis of $\hat{\sigma}^2_{MLE}$

Consider the following

$$T_n = \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum(X_i - \overline{X}_n)^2$$

We have the following

$$E(T_n) = \frac{n-1}{n}\sigma^2 \to 0 \quad \text{as } n \to \infty$$

$$Var(T_n) = \frac{2(n-1)}{n^2}\sigma^4 \to 0 \quad \text{as } n \to \infty$$

Weak Law of Large Numbers

Theorem: If $X_i \overset{iid}{\sim} F_\theta$ with $E(X_i) = \mu_\theta$ then

$$\overline{X}_n \xrightarrow{p} \mu_\theta \quad \text{as } n \to \infty$$

i.e., $\overline{X}_n$ is consistent for μ_θ

Proof. Consider this as homework. Assume $E(|X_1 - \mu_\theta| < \infty)$ □

Theorem: Let $X_i \overset{ind}{\sim} F_\theta$ with $E(X_i) = \mu_i$ and $Var(X_i) = \sigma^2_i < \infty$. If

$n^{-2}\sum_{i=1}^{n}\sigma_i^2 \to 0$ as $n \to \infty$ then we have

$$\left(\overline{X}_n - \overline{\mu}_n\right) \xrightarrow{p} 0$$

where $\overline{\mu}_n = \frac{1}{n}\sum \mu_i$

Proof. Fix $\epsilon > 0$. Thus we have

$$\begin{aligned}(\overline{X}_n - \overline{\mu}_n)^2 &= \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i)\right]^2 \\ &= \frac{1}{n^2}\left[\sum_{i=1}^{n}(X_i-\mu_i)^2 + \sum_{i\neq j}(X_i-\mu_i)(X_j-\mu_j)\right]\end{aligned}$$

According to chebyshev's inequality, we have

$$\begin{aligned}P(|\overline{X}_n - \overline{\mu}_n| > \epsilon) &\leq \frac{E[(\overline{X}_n - \overline{\mu}_n)^2]}{\epsilon^2} \\ &= \frac{\sum \sigma_i^2 + \sum_{i\neq j} Cov(X_i, X_j)}{n^2\epsilon^2} \\ &= \frac{\sum_{i=1}^{n}\sigma_i^2}{n^2\epsilon^2} \to 0 \quad \text{as } n \to 0\end{aligned}$$

□

NOTE

We do not need independence of $X_i's$, we just need $Cov(X_i, X_j) = 0\ \forall i \neq j$

EXAMPLE | 6

$X_i \overset{ind}{\sim} N(\mu_i, \sqrt{i})$. Does $\left(\overline{X}_n - \overline{\mu}_n\right) \xrightarrow{p} 0$

Solution

We have $\overline{X}_n - \overline{\mu}_n = \frac{1}{n}\sum(X_i - \mu_i)$
Verify: If $n^{-2}\sum\sigma_i^2 = \frac{1}{n^2}\sum\sqrt{i}$
We can use any on of the following two results for further calculations

1. $\sqrt{n-1} \leq \int_{n-1}^{n}\sqrt{x}dx \leq \sqrt{n}$
2. $\sum_{i=1}^{n} \leq n\sqrt{n}$

Say we use the 2^{nd} result, thus we have

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 &= \frac{1}{n^2}\sum_{i=1}^{n}\sqrt{i}\\ &\leq \frac{1}{n^2}n\sqrt{n}\\ &= \frac{1}{\sqrt{n}} \to 0 \quad \text{as } n \to \infty\end{aligned}$$

Therefore, we have

$$\left(\overline{X}_n - \overline{\mu}_n\right) \xrightarrow{p} 0$$

10.3 Convergence in Distribution

Convergence in Distribution

Definition: Let T_n be a sequence of random variables with CDF F_n and T be another random variable with CDF F, thus we have

$$F_n(x) \to F(x) \quad \text{as } n \to \infty$$

$$P\left(T_n \leq x\right) \to P\left(T \leq x\right)$$

for all x in the set of continuity points of T

NOTE

Continuity points of F x_0 is a discontinuity point of F if

$$F(x_0^-) < F(x_0) = D(x_0^+) = \lim_{h \downarrow 0} F(x_0 - h)$$

The collection of all discontinuity point of F is $\mathcal{D}$
$\therefore$ the set of continuity points of F is $\mathbb{R} - \mathcal{D}$
Claim: There are almost countable no. of discontinuity points of F

$$\left\{x_0 : F(x_0) - F(x_o^-) > \frac{1}{n}\right\} = D_n$$

such that $D_1 \cup D_2 \cup \ldots = \mathcal{D}$
We can Also cay that $|D_n| < n$

$$\mathcal{D} = \left\{x_0 : F(x_0) - F(x_0^-) > 0\right\} = D_1 \cup D_2 \cup \ldots$$

Countable union of countable sets is countable

EXAMPLE | UNIFORM DISTRIBUTION

$X_i \overset{iid}{\sim} U(0,\theta); T_n = X_{(n)}\ ;\ T_n \xrightarrow{d} \theta$ as $n \to \theta$
Thus we have

$$T = \begin{cases} \theta & \text{w.p } 1 \end{cases}$$

Thus, the CDF of T is the following

$$F_T(t) = \begin{cases} 0 & \text{if } t < \theta \\ 1 & \text{if } t \geq \theta \end{cases}$$

Also, we have

$$F_{T_n}(t) = \begin{cases} 0 & \text{if } t < 0 \\ \left(\frac{t}{\theta}\right)^n & \text{if } t \in [0,\theta) \\ 1 & \text{if } t \geq \theta \end{cases}$$

Thus we have the following result

$$t < 0 \quad F_{T_n}(t) = 0 = F_T(t)$$

$$t \in [0,\theta) \quad F_{T_n}(t) = \left(\frac{t}{\theta}\right)^n \to 0 = F_T = t$$

$$t > \theta, \quad F_{T_n}(t) = 1 = F(t)$$

$\therefore T_n \xrightarrow{d} T$

NOTE

Remarks

1. $T_n \xrightarrow{p} T$ then $T_n \xrightarrow{d} T$ But the converse is *NOT* true

   ### EXAMPLE | BERNOULLI

   Consider $T_n \sim Ber(\theta)$ and $T \sim Ber(\theta)$. We can show that $T_n \xrightarrow{d} T$ but $T_n \not\xrightarrow{p} T$

2. Continuous Mapping Theorem
   (a) If $T_n \xrightarrow{p} T$ then $g(T_n) \xrightarrow{p} g(T)$ where g is a continuous function
   (b) If $T_n \xrightarrow{d} T$ then $g(T_n) \xrightarrow{d} g(T)$ provided g is continuous
3. $T_n \xrightarrow{p} c \iff T_n \xrightarrow{d} c$
4. Slutsky's Lemma

   $$X_n \xrightarrow{d} X \quad \text{and} \quad Y_n \xrightarrow{p/d} c$$

   then we have

(a)

$$X_n + Y_n \xrightarrow{d} X + c$$

(b)

$$X_n Y_n \xrightarrow{d} cX$$

(c) provided $c \neq 0$,

$$\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$$

11 Central Limit Theorem

11.1 The Core Theorem

Central Limit Theorem

Let $X_1, \ldots, X_n$ be IID random variables from some distribution with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2 < \infty$, then

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} Z, \quad Z \sim N(0,1)$$

EXAMPLE | NORMAL DISTRIBUTION

$X_i \overset{IID}{\sim} N(0, \sigma^2)$. Consider $T_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$
Let $Y_i = X_i^2$. Thus we have

$$E(Y_1) = \mu_Y = \sigma^2 \qquad Var(Y_1) = \sigma_Y^2 = 2\sigma^4$$

Thus we can see

$$\frac{\sqrt{n}(T_n - \sigma^2)}{\sqrt{2\sigma^4}} \xrightarrow{d} N(0,1)$$

11.2 Asymptotic Distributions

Asymptotic Distribution of T_n

By asymptotic distribution of T_n, we mean two sequence of real numbers $\{a_n\}$ and $\{b_n\}$ and a non-degenerate probability distribution F such that

$$\frac{T_n - a_n}{b_n} \xrightarrow{d} Z \quad (Z \sim F)$$

Note on Degenerate Random Variables:

- A non-degenerate distribution is one that doesn't take any value with probability 1.
- Thus, we can say that Z is degenerate if $\exists$ a number a_0 such that $P(Z = a_0) = 1$.
- We can also say that for a degenerate random variable, only 1 number, a_0, is in the support of it.

NOTE

Support of Random Variable X A number $a \in \mathbb{R}$ is in the support of X if $\forall h > 0$,

$$P\left(X \in (a-h, a+h)\right) > 0$$

EXAMPLE | UNIFORM DISTRIBUTION

Given $X_n \overset{IID}{\sim} U(0,\theta)$. Let $T_n = X_{(n)}$.
We know that $T_n \xrightarrow{p} \theta$ and hence we have $T_n - \theta \xrightarrow{p} 0$. Consider the following:

$$n(\theta - T_n) = W_{n,\theta}$$

Let F_n be the CDF of $W_{n,\theta}$. Thus we have:

$$\begin{aligned} F_n(w) &= P(W_{n,\theta} \leq w) \\ &= P(n(\theta - T_n) \leq w) \\ &= P\left(T_n \geq (\theta - w/n)\right) \\ &= \begin{cases} 1 & \text{if } w > n\theta \\ 1 - \left(1 - \frac{w}{n\theta}\right)^n & \text{if } w \in (0, n\theta) \end{cases} \end{aligned}$$

Thus we have:

$$\begin{aligned} \lim_{n\to\infty} F_n(w) &= 1 - e^{-\frac{w}{\theta}} \quad \text{for any } w > 0 \\ &= F(w) \end{aligned}$$

If $W \sim \exp(1/\theta)$ with $f_W(w) = \frac{1}{\theta} e^{-\frac{w}{\theta}}; w \geq 0$, then the CDF of W is F. Thus we have:

$$n(\theta - X_{(n)}) \xrightarrow{d} W$$

where $W \sim \exp(1/\theta)$. When comparing to the asymptotic limit, we have:

$$a_n = \theta \qquad b_n = -\frac{1}{n} \qquad (n = 1, 2, \ldots)$$

11.3 Applications of CLT

EXAMPLE | USE OF CLT IN SAMPLE SIZE DETERMINATION

Let $X_i \overset{IID}{\sim} \exp(1/\theta)$. Find the minimum sample size that ensures $\overline{X}_n \in [\theta(1-0.1), \theta(1+0.1)]$ with probability at least 0.99.

Solution

We require the following:

$$P(|\overline{X}_n - \theta| > 0.1 \times \theta) \leq 0.01$$

Now, by Chebyshev's inequality, we have the following:

$$P(|\overline{X}_n - \theta| > 0.1 \times \theta) \leq \frac{Var(\overline{X}_n)}{0.01 \times \theta^2} = \frac{1}{0.01n}$$

Thus, we have the following needed n:

$$\boxed{\frac{1}{0.01n} \leq 0.01 \implies n \geq 10,000}$$

Now, by CLT, we have:

$$\frac{\sqrt{n}(\overline{X}_n - \theta)}{\theta} \xrightarrow{d} N(0,1)$$

Thus we have the following:

$$\begin{aligned} P\left(|\overline{X}_n - \theta| \leq 0.1 \times \theta\right) &= P\left(\sqrt{n}\left|\frac{\overline{X}_n - \theta}{\theta}\right| \leq \sqrt{n} \times 0.1\right) \\ &\approx P(|Z| \leq \sqrt{n} \times 0.1) \\ &= P(-\sqrt{n} \times 0.1 \leq Z \leq \sqrt{n} \times 0.1) \\ &= \Phi(\sqrt{n} \times 0.1) - \Phi(-\sqrt{n} \times 0.1) \\ &= 2\Phi(\sqrt{n} \times 0.1) - 1 \\ &\geq 0.99 \end{aligned}$$

where $Z \sim N(0,1)$. Thus we need minimum sample size n such that:

$$\Phi(\sqrt{n} \times 0.1) \geq \frac{1.99}{2} = 0.995$$

We know that $\Phi(2.78) \approx 0.995$. Thus we have:

$$\sqrt{n} \times 0.1 \approx 2.78 \implies \boxed{n \geq 773}$$

EXAMPLE | ASYMPTOTIC DISTRIBUTION EXAMPLE

In the previous example, what is the asymptotic distribution of $\frac{1}{n}\sum_{i=1}^{n} X_i^2$?

Solution

We have the following prior information:

$$E(X_1^2) = 2\theta^2 \qquad Var(X_1^2) = 20\theta^4$$

Now consider the following:

$$\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\theta^2\right)}{2\sqrt{5}\theta^2} \xrightarrow{d} N(0,1)$$

We thus have that $\frac{1}{n}\sum_{i=1}^{n} X_i^2 = T_n$ is asymptotically normal with $2\theta^2$ as the asymptotic mean of T_n and $\frac{20\theta^4}{n}$ as the asymptotic variance.

12 Testing of Hypothesis

12.1 Basic Definitions

Statistical Hypothesis: It is a statement about the population. In a parametric case, it becomes a statement about the parameter or a function of that parameter $g(\theta)$.

Null Hypothesis (H_0) and Alternative Hypothesis (H_1)

Null Hypothesis (H_0)**:** It is a statement about the parameter or the population which is being tested. We are interested in accepting or rejecting H_0.
Alternative Hypothesis (H_1)**:** Any statement which contradicts H_0 is an Alternative Hypothesis. We will be facing situations like:

$$H_0 : \text{Statement A} \quad \text{vs} \quad H_1 : \text{Statement B}$$

Based on our sample $X_1, \ldots, X_n$, we will be accepting or rejecting H_0.

One-Sided and Two-Sided Hypothesis

Say we have the following Alternative hypotheses:

$$H_{11} : \theta < \theta_0$$

$$H_{12} : \theta > \theta_0$$

$$H_{13} : \theta \neq \theta_0$$

Here, H_{11} and H_{12} are **one-sided hypotheses** and H_{13} is a **two-sided hypothesis**.

Simple or Composite Hypothesis

Under a hypothesis, H, if the population is completely specified then H is a **simple hypothesis**. Otherwise, it is a **composite hypothesis**.
Consider the following two cases for $X_i \overset{IID}{\sim} Ber(\theta)$:

1. $H_0 : \theta = 1/2$. This gives us the defined population distribution. It is always $Ber(1/2)$. Thus, it is a simple hypothesis.
2. $H_1 : \theta \neq 1/2$. This can result in multiple different population distributions. Hence, this is not a simple hypothesis.

Example (Gamma Distribution): Consider $H_0 : \alpha = 0.5$.

1. $X_i \overset{IID}{\sim} \Gamma(\alpha, 1)$. Here, H_0 is a simple hypothesis.

2. $X_i \overset{IID}{\sim} \Gamma(\alpha, \beta)$. Here, H_0 is a composite hypothesis (because β remains unspecified).

12.2 Test Functions and Regions

EXAMPLE | TOSSING OF A COIN (CONSTRUCTING A TEST FUNCTION)

$X_1, \ldots, X_n \overset{IID}{\sim} Bern(\theta)$. Say we have the following two statements about θ:

$$H_0 \text{ (Null Hypothesis) } : \theta = \frac{1}{2}$$

$$H_A \text{ (Alternative Hypothesis) } : \theta > \frac{1}{2}$$

Say we have the following observation for n tosses:

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

When comparing between the given H_0 and H_A, we will have to accept H_0. Now say we have the following observation:

$$\underbrace{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{pmatrix}}_{\text{6 heads}}$$

In this case, we will reject the null hypothesis and accept the alternative hypothesis H_A. Thus, more generally, we have the following:

$$\phi(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \underset{\sim}{x} \text{ such that } \sum x_i \geq 6 \\ 0 & \text{if } \underset{\sim}{x} \text{ such that } \sum x_i \leq 5 \end{cases}$$

where ϕ is the probability of rejecting H_0. ϕ is called a **Test Function**, $\phi : \chi \to [0, 1]$. Now, say we have a very huge penalty for rejecting H_0; in that scenario we might reconsider the conditions on our test function. We may construct the following new test function:

$$\phi(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \underset{\sim}{x} \text{ such that } \sum x_i \geq 8 \\ 0 & \text{if } \underset{\sim}{x} \text{ such that } \sum x_i \leq 7 \end{cases}$$

EXAMPLE | TESTING OF HYPOTHESIS (EXPONENTIAL)

$X_1, \ldots, X_n \sim \exp(\lambda)$. Let $T_n = \overline{X}_n$. Consider we have the following:

$$H_0 : \lambda = 1 \quad \text{vs} \quad H_{11} : \lambda > 1$$
$$H_{12} : \lambda < 1$$
$$H_{13} : \lambda \neq 1$$

Thus we have the following test function:

$$\phi(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \overline{X}_n < c \\ 0 & \text{if } \overline{X}_n \geq c \end{cases}$$

Critical and Acceptance Region

Critical Region (C) is a subset of the sample space such that:

$$C = \left\{ \underset{\sim}{x} \in \chi : \phi(\underset{\sim}{x}) = 1 \right\}$$

i.e., H_0 is rejected if $\underset{\sim}{x}$ appears. **Acceptance Region (A)** is the complement of Critical Region:

$$A = \left\{ \underset{\sim}{x} \in \chi : \phi(\underset{\sim}{x}) = 0 \right\}$$

i.e., H_0 is accepted if $\underset{\sim}{x}$ appears. **Example:** Consider the strict coin toss test from the previous example:

$$C = \left\{ \underset{\sim}{x} \in \chi : \sum x_i \geq 8 \right\} \quad \text{and} \quad A = \left\{ \underset{\sim}{x} \in \chi : \sum x_i \leq 7 \right\}$$

12.3 Types of Errors and Level of Significance

Types of Errors

When executing a test, there are two fundamental errors we can make based on the truth of our hypothesis:

Action \ Truth	H_0 **is true**	H_0 **is false**
Accept H_0	✓	Type II error
Reject H_0	Type I error	✓

EXAMPLE | CALCULATING ERROR PROBABILITIES

$X_1, \ldots, X_n \overset{IID}{\sim} N(\mu, 1)$. Consider the following testing problem:

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu = 2$$

Consider the following test function:

$$\phi^*(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \overline{X}_n \geq c \\ 0 & \text{otherwise} \end{cases}$$

Thus, we have the following:

$$\begin{aligned} P\left(\text{Type I error}\right) &= P\left(\text{Reject } H_0 | H_0 \text{ is true}\right) \\ &= P_{\mu=0}(\overline{X}_n \geq c) \\ &= P\left(\overline{X}_n \geq c | \overline{X}_n \sim N\left(0, \frac{1}{n}\right)\right) \\ &= 1 - \Phi(c\sqrt{n}) \end{aligned}$$

Similarly, we have:

$$\begin{aligned} P(\text{Type II error}) &= P\left(\text{Accepting } H_0 | H_1 \text{ is true}\right) \\ &= P_{\mu=2}(\overline{X}_n < c) \\ &= P\left(\overline{X}_n < c | \overline{X}_n \sim N\left(2, \frac{1}{n}\right)\right) \\ &= \Phi\left(\sqrt{n}(c-2)\right) \end{aligned}$$

Note:

1. One should take a larger c to reduce the probability of Type I error.
2. One should take a smaller c to reduce the probability of Type II error.

Level of Significance (α)

The **Level of Significance** is the upper-permissible limit of $P(\text{Type I error})$, denoted by α. We usually take $\alpha = 0.05$ or $\alpha = 0.01$. Thus we have:

$$P(\text{Type I error}) \leq 0.05$$

Using the Normal distribution example from above, this implies that:

$$\begin{aligned} 1 - \Phi(c\sqrt{n}) &\leq 0.05 \\ \Phi(c\sqrt{n}) &\geq 0.95 \qquad \ldots (1) \end{aligned}$$

$$c\sqrt{n} \geq 1.645$$

$$c \geq \frac{1.645}{\sqrt{n}}$$

We define the following: τ_α: upper α point in $N(0,1)$ distribution. Thus we have:

$$c \geq \frac{\tau_{0.05}}{\sqrt{n}}$$

Equality holds in (1) if $c = \frac{\tau_{0.05}}{\sqrt{n}} = c_0$. Thus, for all $c \in [c_0, \infty)$, condition (1) is satisfied.

1. Among the tests which satisfy the level condition (i.e, P (Type I error) $\leq \alpha$), we should consider that test which has the smallest possible probability of Type II error.
2. Thus, we should take $c = c_0$ as smaller values of c reduce the probability of Type II error, and c_0 is the smallest possible permissible value of c.

12.4 Randomized Tests

EXAMPLE | EXACT LEVEL AND RANDOMIZED TESTS

$X_1, \ldots, X_{10} \sim Bern(p)$. Consider the following testing problem:

$$H_0 : p = \frac{1}{2} \quad \text{vs} \quad H_1 : p = \frac{3}{4}$$

Let $T_n = \sum_{i=1}^{n} X_i$. We know that $T_n \sim Bin(n,p)$. Thus $T_{10} \sim Bin(10,p)$. Consider the following test function:

$$\underset{\sim}{\phi(x)} = \begin{cases} 1 & \text{if } T_{10} \geq c \\ 0 & \text{otherwise} \end{cases}$$

Let $\alpha = 0.05$. We require:

$$P \text{ (Type I error)} \leq 0.05$$

We have the Rejection region as $\{c : T_{10} \geq c\}$.

Analysis for $c = 10$

If $c = 10$:

$$P \text{ (Type I error)} = P \text{ (Reject } H_0 | H_0 \text{ is true)}$$

$$= P_{p=1/2}(\text{Reject } H_0)$$

$$= \sum_{t=0}^{10} P_{p=1/2} \text{ (Reject } H_0 | T_{10} = t) \, P_{p=1/2}(T_{10} = t)$$

$$= \sum_{t=0}^{10} P\left(\text{Reject } H_0 \bigcap T_{10} = t\right)$$

$$= 0 \cdot P_{p=1/2}(T_{10} = 0) + \ldots + 0 \cdot P_{p=1/2}(T_{10} = 9) + 1 \cdot P_{p=1/2}(T_{10} = 10)$$

$$= P_{p=1/2}(T_{10} = 10)$$

$$= \binom{10}{10}\left(\frac{1}{2}\right)^{10} = \frac{1}{1024} < 0.05$$

Also, consider the following for Type II error:

$$P(\text{Type II error}) = P\left(\text{Accept } H_0 | H_1 \text{ is true}\right)$$

$$= P_{p=3/4}(\text{Accept } H_0)$$

$$= 1 - P_{p=3/4}(\text{Reject } H_0)$$

$$= 1 - \sum_{t=0}^{10} P_{p=3/4}\left(\text{Reject } H_0 | T_{10} = t\right) P_{p=3/4}(T_{10} = t)$$

$$= 1 - \sum_{t=0}^{10} P\left(\text{Reject } H_0 \bigcap T_{10} = t\right)$$

$$= 1 - \big(0 \cdot P_{p=3/4}(T_{10} = 0) + \ldots + 0 \cdot P_{p=3/4}(T_{10} = 9) + 1 \cdot P_{p=3/4}(T_{10} = 10)\big)$$

$$= 1 - P_{p=3/4}(T_{10} = 10)$$

$$= 1 - \binom{10}{10}\left(\frac{3}{4}\right)^{10} = 1 - \left(\frac{3}{4}\right)^{10} = 0.9436$$

Analysis for $c = 9$ and $c = 8$

Now say, $c = 9$. We'll arrive at the following result:

$$P(\text{Type I error}) = 0.0106 \ll 0.05 \qquad P(\text{Type II error}) = 0.7558$$

Now say, $c = 8$. We'll arrive at the following result:

$$P(\text{Type I error}) = 0.0547 > 0.05 \qquad P(\text{Type II error}) = 0.475$$

Thus, we have the following new randomized test function:

$$\phi^*(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } T_{10} \geq 9 \\ \beta & \text{if } T_{10} = 8 \\ 0 & \text{if } T_{10} \in \{0, \ldots, 7\} \end{cases}$$

Derivation of β for ϕ^*

Consider the following:

$$\begin{aligned} P(\text{Type I error for } \phi^*) &= P_{p=1/2}(\text{Reject } H_0) \\ &= \sum_{t=0}^{10} P_{p=1/2}\left(\text{Reject } H_0 | T_{10} = t\right) P_{p=1/2}(T_{10} = t) \\ &= 0 \cdot P_{p=1/2}(T_{10} = 0) + \ldots + 0 \cdot P_{p=1/2}(T_{10} = 7) \\ &\quad + \beta \cdot P_{p=1/2}(T_{10} = 8) + 1 \cdot P_{p=1/2}(T_{10} = 9) \\ &\quad + 1 \cdot P_{p=1/2}(T_{10} = 10) \\ &= \beta \cdot P_{p=1/2}(T_{10} = 8) + P_{p=1/2}(T_{10} = 9) + P_{p=1/2}(T_{10} = 10) \\ &= \beta \binom{10}{8} \left(\frac{1}{2}\right)^{10} + \binom{10}{9} \left(\frac{1}{2}\right)^{10} + \binom{10}{10} \left(\frac{1}{2}\right)^{10} \\ &= \frac{45\beta + 10 + 1}{1024} \end{aligned}$$

From the level condition, we have:

$$\begin{aligned} & P(\text{Type I error}) = 0.05 \\ \implies & \frac{45\beta + 11}{1024} = 0.05 \\ \implies & 45\beta + 11 = 2^{10} \times 0.05 \\ \implies & \beta = 0.89 \end{aligned}$$

Thus, to perform ϕ^*:

1. If the realized value of T_{10} is 9 or 10, we reject H_0.
2. If the realized value of T_{10} is 8, we take a $U(0,1)$ draw.
   - If the $U(0,1)$ draw is $\leq \beta$, we reject H_0.
   - If the $U(0,1)$ draw is $> \beta$, we accept H_0.
3. If the realized value of T_{10} is ≤ 7, we accept H_0.

12.5 Randomized Tests and Power

Non-Randomized vs. Randomized Tests

Non-Randomized Test: A test function ϕ, which takes only two discrete values, viz., $\{0, 1\}$.
Randomized Test: A test which is NOT non-randomized (i.e., it can take fractional values representing probabilities).

EXAMPLE | EXECUTING A RANDOMIZED TEST

Let $X_1, \ldots, X_{10} \overset{IID}{\sim} Bern(p)$. Consider testing the following:

$$H_0 : p = 0.5 \qquad \text{vs} \qquad H_1 : p = 0.75$$

Suppose we derived the following randomized test function based on an exact level α:

$$\phi(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{10} x_i \in \{9, 10\} \\ 0.89 & \text{if } \sum_{i=1}^{10} x_i = 8 \\ 0 & \text{otherwise} \end{cases}$$

Execution Rule for the Boundary Case: If our realized sample yields exactly $\sum_{i=1}^{10} x_i = 8$, we must randomize our decision:

- Draw a random variable $U \sim \mathcal{U}(0, 1)$.
- If $U \leq 0.89$, we **reject** H_0.
- If $U > 0.89$, we **accept** H_0.

Note: Among all tests satisfying $P(\text{Type I error}) \leq \alpha$, we consider the test ϕ^* as the "best" if its $P(\text{Type II error})$ is minimized.

Size and Power of a Test

Let $\phi : \chi \to [0, 1]$ be a test function. We define two fundamental metrics:
1. Size of ϕ:

$$\text{Size} = P(\text{Type I error of } \phi) = P(\text{Rejecting } H_0 \mid H_0 \text{ is true})$$

2. Power of ϕ:

$$\begin{aligned} \text{Power} &= 1 - P(\text{Type II error of } \phi) \\ &= 1 - P(\text{Accepting } H_0 \mid H_1 \text{ is true}) \\ &= P(\text{Rejecting } H_0 \mid H_1 \text{ is true}) \end{aligned}$$

Core Principle: A good test must have a *small size* and a *large power*.

Power Function

Let ϕ be a test-function. The **power function** of ϕ, evaluated at a specific parameter value, is simply the expected value of the test function:

$$\text{Power Function} = E\left[\phi(\underset{\sim}{X})\right]$$

12.6 The Neyman-Pearson Lemma

Most Powerful (MP) Level-α Test

Consider the problem of testing a simple null hypothesis against a simple alternative hypothesis:

$$H_0 : X \sim f_0 \qquad \text{vs} \qquad H_1 : X \sim f_1$$

based on a random sample $X_1, \ldots, X_n$.
A level-α test ϕ must satisfy:

$$\text{Size of } \phi \leq \alpha \qquad \ldots (1)$$

The **Most Powerful (MP) level-α test**, denoted as ϕ^*, satisfies condition (1) and mathematically guarantees that:

$$\text{Power of } \phi^* \geq \text{Power of } \phi \quad \text{for any } \phi \text{ satisfying } (1)$$

Neyman-Pearson's Lemma

The Neyman-Pearson (NP) lemma provides a definitive method for obtaining the Most Powerful test *for simple vs. simple hypotheses*.
Theorem: Consider the problem of testing $H_0 : X \sim f_0$ vs $H_1 : X \sim f_1$, based on a random sample $X_1, \ldots, X_n$. Then, a test ϕ^* is the MP level-α test if it satisfies both:

$$\phi^*(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } f_1(\underset{\sim}{X}) > k f_0(\underset{\sim}{X}) \\ 0 & \text{if } f_1(\underset{\sim}{X}) < k f_0(\underset{\sim}{X}) \end{cases} \qquad \ldots (1)$$

and

$$E_{H_0}\left[\phi^*(\underset{\sim}{X})\right] = \alpha \qquad \ldots (2)$$

Furthermore, any MP level-α test *must* satisfy conditions (1) and (2).

Proof Intuition for Non-Randomized Tests: Suppose ϕ^* is non-randomized, yielding a critical region $\mathbb{C}$ and an acceptance region $\overline{\mathbb{C}}$:

$$\phi^*(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \underset{\sim}{X} \in \mathbb{C} \\ 0 & \text{if } \underset{\sim}{X} \in \overline{\mathbb{C}} \end{cases}$$

Thus, the expected value under the null perfectly matches the size:

$$\begin{aligned} E_{H_0}\left[\phi^*(\underset{\sim}{X})\right] &= 1 \cdot P_{H_0}(\underset{\sim}{X} \in \mathbb{C}) + 0 \cdot P_{H_0}(\underset{\sim}{X} \in \overline{\mathbb{C}}) \\ &= P_{H_0}(\underset{\sim}{X} \in \mathbb{C}) \\ &= \text{Size of } \phi^* \end{aligned}$$

EXAMPLE | 1: APPLYING NP LEMMA (DISCRETE CASE)

Consider testing $H_0 : X \sim f_0$ vs $H_1 : X \sim f_1$ with the following distributions:

x	**1**	**2**	**3**	**4**	**5**	**6**
$f_0(x)$	0.01	0.01	0.01	0.01	0.01	0.95
$f_1(x)$	0.05	0.04	0.03	0.02	0.01	0.85
$\lambda(x) = \frac{f_1(x)}{f_0(x)}$	5	4	3	2	1	$\frac{85}{95}$

As per the NP lemma, the MP test prioritizes rejecting H_0 for the largest values of the likelihood ratio $\lambda(x)$.

Case A: Level $\alpha = 0.01$ To achieve an exact size of 0.01, we simply select the point with the highest $\lambda(x)$, which is $x = 1$. Since $f_0(1) = 0.01$, our test function is:

$$\phi^*(x) = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{if } x \in \{2, \ldots, 6\} \end{cases}$$

Verification: $E_{H_0}[\phi^*(X)] = 1 \cdot P_{H_0}(X = 1) = f_0(1) = 0.01 = \alpha$. Since it satisfies the NP conditions, it is the MP level-0.01 test.

Case B: Level $\alpha = 0.05$ We move down the sorted $\lambda(x)$ values until our cumulative size reaches 0.05. We select $x \in \{1, 2, 3, 4, 5\}$.

$$\phi^{**}(x) = \begin{cases} 1 & \text{if } x \in \{1, 2, 3, 4, 5\} \\ 0 & \text{if } x = 6 \end{cases}$$

Verification: $E_{H_0}[\phi^{**}(X)] = \sum_{x=1}^{5} f_0(x) = 5 \times 0.01 = 0.05 = \alpha$.

EXAMPLE | 2: APPLYING NP LEMMA (CONTINUOUS CASE)

Consider the hypothesis test: $H_0 : X \sim N(0,1)$ vs $H_1 : X \sim DE(0,1)$ (Double Exponential).
The densities are:

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad \text{and} \qquad f_1(x) = \frac{1}{2} e^{-|x|}$$

Deriving the Critical Region

We form the likelihood ratio $\lambda(x) > k$:

$$\lambda(x) = \frac{f_1(x)}{f_0(x)} = \sqrt{\frac{\pi}{2}} e^{-|x| + \frac{x^2}{2}} > k$$

Taking the natural logarithm and simplifying:

$$e^{-|x| + \frac{x^2}{2}} > k\sqrt{\frac{2}{\pi}}$$

$$-2|x| + x^2 > 2\log\left(k\sqrt{\frac{2}{\pi}}\right)$$

Completing the square on the left side by adding 1:

$$|x|^2 - 2|x| + 1 > 1 + 2\log\left(k\sqrt{\frac{2}{\pi}}\right)$$

$$(|x| - 1)^2 > k_0 \qquad \text{(where } k_0 \text{ is our new constant)}$$

Thus, the MP test function takes the form:

$$\phi^*(x) = \begin{cases} 1 & \text{if } (|x| - 1)^2 > k_0 \\ 0 & \text{if } (|x| - 1)^2 < k_0 \end{cases}$$

Evaluating the Expected Value (Size)

To find α, we evaluate the expectation under H_0:

$$E_{H_0}[\phi^*(X)] = P_{H_0}\left((|X|-1)^2 > k_0\right)$$
$$= P_{H_0}\left(||X|-1| > \sqrt{k_0}\right)$$

This absolute value inequality splits into two distinct bounds:

$$|X| - 1 > \sqrt{k_0} \implies |X| > 1 + \sqrt{k_0}$$
$$|X| - 1 < -\sqrt{k_0} \implies |X| < 1 - \sqrt{k_0}$$

Solving these bounds for X:

$$P(|X| > 1 + \sqrt{k_0}) = P(X > 1 + \sqrt{k_0}) + P(X < -(1 + \sqrt{k_0}))$$
$$P(|X| < 1 - \sqrt{k_0}) = P\left(-(1 - \sqrt{k_0}) < X < 1 - \sqrt{k_0}\right)$$

Expressing this purely in terms of the standard normal CDF, Φ:

$$\alpha = \left[1 - \Phi(1 + \sqrt{k_0})\right] + \Phi(-(1 + \sqrt{k_0}))$$
$$+ \left[\Phi(1 - \sqrt{k_0}) - \Phi(-(1 - \sqrt{k_0}))\right]$$

Using the symmetry property $\Phi(-z) = 1 - \Phi(z)$, this simplifies to:

$$\alpha = 2 - 2\Phi(1 + \sqrt{k_0}) + 2\Phi(1 - \sqrt{k_0}) - 1$$
$$\alpha = 1 - 2\Phi(1 + \sqrt{k_0}) + 2\Phi(1 - \sqrt{k_0})$$

12.7 The Generalized NP Lemma (Randomized Boundaries)

Generalized NP Lemma with Randomization

In discrete distributions, it is often impossible to find a critical region that exactly equals our desired level of significance α. To achieve exactly α, the Most Powerful (MP) test ϕ^* incorporates a randomization probability γ at the boundary.
A test is the MP level-α test if it satisfies:

1. **The Test Function:**

$$\phi^*(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \lambda(\underset{\sim}{X}) = \frac{f_1(\underset{\sim}{X})}{f_0(\underset{\sim}{X})} > k \\ \gamma & \text{if } \lambda(\underset{\sim}{X}) = k \\ 0 & \text{if } \lambda(\underset{\sim}{X}) < k \end{cases}$$

2. **The Size Condition:**

$$\text{Size } = E_{H_0}\left[\phi^*(\underset{\sim}{X})\right] = \alpha$$

Any test that satisfies both conditions is guaranteed to be a Most Powerful Level-α test.

EXAMPLE | 1: NORMAL DISTRIBUTION MEAN SHIFT (CONTINUOUS)

Let $X_i \overset{IID}{\sim} N(\mu, 1)$. We are testing a left-sided shift:

$$H_0 : \mu = 0 \qquad \text{vs} \qquad H_1 : \mu = -2$$

Deriving the Critical Region

We evaluate the likelihood ratio $\lambda(\underset{\sim}{X}) > k$:

$$\lambda(\underset{\sim}{X}) = \frac{f_1(\underset{\sim}{X})}{f_0(\underset{\sim}{X})} = \exp\left\{-2\sum_{i=1}^{n} X_i - 2n\right\} > k$$

Taking the natural log and simplifying to isolate the sample mean $\overline{X}_n$:

$$-2n[\overline{X}_n + 1] > \log k$$

$$\overline{X}_n + 1 < -\frac{1}{2n}\log k \qquad \text{(Inequality flips due to division by negative)}$$

$$\overline{X}_n < \underbrace{-1 - \frac{1}{2n}\log k}_{k_0}$$

Thus, the MP test function simplifies to:

$$\phi^*(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \overline{X}_n \leq k_0 \\ 0 & \text{if } \overline{X}_n > k_0 \end{cases}$$

To find k_0, we use the size condition under H_0 (where $\overline{X}_n \sim N(0, 1/n)$):

$$\alpha = E_{H_0}\left[\phi^*(\underset{\sim}{X})\right] = 1 \cdot P_{H_0}\left(\overline{X}_n \leq k_0\right)$$

$$\alpha = \Phi(\sqrt{n}k_0)$$

EXAMPLE | 2: POISSON DISTRIBUTION (DISCRETE)

Let $X_i \overset{IID}{\sim} Pois(\lambda)$. Consider testing:

$$H_0 : \lambda = \lambda_0 \qquad \text{vs} \qquad H_1 : \lambda = \lambda_1 \quad (\text{where } \lambda_1 > \lambda_0)$$

Deriving the Test Structure

Forming the likelihood ratio:

$$\frac{f_1(\underset{\sim}{X})}{f_0(\underset{\sim}{X})} = \frac{e^{-n\lambda_1}\lambda_1^{\sum x_i}/\prod x_i!}{e^{-n\lambda_0}\lambda_0^{\sum x_i}/\prod x_i!}$$

$$= e^{n(\lambda_0-\lambda_1)}\left(\frac{\lambda_1}{\lambda_0}\right)^{\sum_{i=1}^n x_i} > k$$

Taking the logarithm and rearranging:

$$\left(\sum_{i=1}^n x_i\right)(\log\lambda_1 - \log\lambda_0) > \log k + n(\lambda_1 - \lambda_0)$$

$$\sum_{i=1}^n x_i > \frac{\log k + n(\lambda_1-\lambda_0)}{\log\lambda_1 - \log\lambda_0} = k_0$$

Let $T_n = \sum X_i$. Under the null hypothesis, we know $T_n \overset{H_0}{\sim} Pois(n\lambda_0)$. Our randomized test function is:

$$\phi^*(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } T_n > k_0 \\ \gamma & \text{if } T_n = k_0 \\ 0 & \text{if } T_n < k_0 \end{cases}$$

Calculating γ (Numerical Example)

Suppose $n = 5, \lambda_0 = 1, \lambda_1 = 2$, and we want $\alpha = 0.05$. Under H_0, $T_5 \sim Pois(5)$.
We need to find an integer k_0 such that $P_{H_0}(T_5 > k_0) < 0.05$:

- If $k_0 = 8$, $P(T_5 > 8) = 0.0681$ (Too large, > 0.05).
- If $k_0 = 9$, $P(T_5 > 9) = 0.0318$ (Valid, < 0.05).

We set $k_0 = 9$. Now we solve for the boundary probability γ:

$$\alpha = P_{H_0}(T_5 > 9) + \gamma \cdot P_{H_0}(T_5 = 9)$$

$$0.05 = 0.0318 + \gamma \cdot P_{H_0}(T_5 = 9)$$

$$\gamma = \frac{0.05 - 0.0318}{P_{H_0}(T_5 = 9)} = 0.501$$

Final Rule: If $\sum X_i > 9$, reject H_0. If $\sum X_i = 9$, reject H_0 with probability 0.501. If $\sum X_i < 9$, accept H_0.

12.8 Composite Hypotheses & UMP Tests

Simple vs. Composite Hypothesis

Let $X_i \overset{IID}{\sim} F_\theta$. When the alternative hypothesis covers a range of values, it is a **composite hypothesis**.

$$H_0 : \theta = \theta_0 \qquad \text{vs} \qquad H_1 : \begin{cases} \theta > \theta_0 & \text{(Right-tailed)} \\ \theta < \theta_0 & \text{(Left-tailed)} \\ \theta \neq \theta_0 & \text{(Two-tailed)} \end{cases}$$

We denote the entire alternative parameter space as Θ_1.
Because $\theta \in \Theta_1$ is not a single point, the **Power of the test** is no longer a single number. It is a *function* evaluated for each specific $\theta \in \Theta_1$:

$$\text{Power Function: } \beta_\phi(\theta) = 1 - P_\theta(\text{Type II Error})$$

Uniformly Most Powerful (UMP) Test

Consider testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta \in \Theta_1$.
A test ϕ is called a **level-α test** if its size does not exceed α:

$$E_{H_0}\left[\phi(\underset{\sim}{X})\right] \leq \alpha \qquad \ldots (1)$$

A test ϕ^* is called the **Uniformly Most Powerful (UMP) level-α test** if it is the absolute best test regardless of which specific value in the alternative space is the true parameter. It must satisfy:

1. ϕ^* is a level-α test (satisfies condition 1).
2. $\beta_{\phi^*}(\theta) \geq \beta_\phi(\theta)$ for *all* $\theta \in \Theta_1$, compared to any other level-α test ϕ. (i.e., It maximizes power uniformly across the entire alternative space).

NOTE

Caveats on UMP Tests

1. Not all Simple vs. Simple MP tests can be naturally extended into UMP tests.
2. A UMP test *may not always exist*.
3. **Classic Example of Non-Existence:** For a Normal distribution $N(\mu, 1)$ testing $H_0 : \mu = \mu_0$ vs a two-sided alternative $H_1 : \mu \neq \mu_0$, no single test is uniformly most powerful for both $\mu < \mu_0$ and $\mu > \mu_0$ simultaneously.

12.9 Power Functions and UMP Tests

The Power Function (β_ϕ)

Let $X_1, \ldots, X_n \overset{IID}{\sim} F_\theta$ and consider the hypothesis test:

$$H_0 : \theta \in \Theta_0 \qquad \text{vs} \qquad H_1 : \theta \in \Theta_1$$

For a given test ϕ, the **Power Function** $\beta_\phi(\theta)$ evaluates the expected value of the test function across the entire parameter space. It represents the probability of rejecting H_0 given the true parameter θ:

$$\beta_\phi(\theta) = E_\theta\left[\phi(\underset{\sim}{X})\right] = P_\theta(\text{Rejecting } H_0)$$

Interpretation based on the True Parameter:

- If $\theta_0 \in \Theta_0$ (Null is true): $\beta_\phi(\theta_0) = P_{\theta_0}(\text{Type I Error}) \leq \alpha$.
- If $\theta_1 \in \Theta_1$ (Alternative is true): $\beta_\phi(\theta_1)$ is the actual **Power** of the test at θ_1.

EXAMPLE | POWER FUNCTION OF A NORMAL MEAN TEST

Let $X_1, \ldots, X_n \overset{IID}{\sim} N(\mu, 1)$. Consider the one-sided testing problem:

$$H_0 : \mu = 0 \qquad \text{vs} \qquad H_1 : \mu > 0$$

Suppose we use the following right-tailed level-α test function (for $\alpha = 0.05$):

$$\phi^*(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \overline{X}_n \geq 1.645/\sqrt{n} \\ 0 & \text{if } \overline{X}_n < 1.645/\sqrt{n} \end{cases}$$

Deriving the Power Function

We evaluate the expectation under a general parameter μ:

$$\begin{aligned} \beta_{\phi^*}(\mu) = E_\mu\left[\phi^*(\underset{\sim}{X})\right] &= P_\mu\left(\overline{X}_n \geq \frac{1.645}{\sqrt{n}}\right) \\ &= P_\mu\left(\frac{\overline{X}_n - \mu}{1/\sqrt{n}} \geq 1.645 - \mu\sqrt{n}\right) \end{aligned}$$

Since $\frac{\overline{X}_n - \mu}{1/\sqrt{n}} \sim N(0,1)$ under true parameter μ:

$$\begin{aligned} \beta_{\phi^*}(\mu) &= 1 - \Phi(1.645 - \mu\sqrt{n}) \\ &= \Phi(\mu\sqrt{n} - 1.645) \qquad (\text{Using } 1 - \Phi(z) = \Phi(-z)) \end{aligned}$$

EXAMPLE | UMP TEST FOR THE EXPONENTIAL DISTRIBUTION

Let $X_1, \ldots, X_n \overset{IID}{\sim} \exp(\lambda)$ (where $E[X] = 1/\lambda$). We want to find a Uniformly Most Powerful (UMP) test for:

$$H_0 : \lambda = \lambda_0 \qquad \text{vs} \qquad H_1 : \lambda > \lambda_0$$

Deriving the Most Powerful (MP) Test

Fix a specific alternative $\lambda_1 > \lambda_0$. By the NP Lemma, the MP test ϕ^* rejects H_0 when $\lambda(\underset{\sim}{X}) > k$:

$$\frac{f_1(\underset{\sim}{X})}{f_0(\underset{\sim}{X})} = \left(\frac{\lambda_1}{\lambda_0}\right)^n \exp\left\{-(\lambda_1 - \lambda_0)\sum_{i=1}^n X_i\right\} > k$$

$$-(\lambda_1 - \lambda_0)\sum_{i=1}^n X_i > \log k - n\log\left(\frac{\lambda_1}{\lambda_0}\right)$$

Because $\lambda_1 > \lambda_0$, the term $-(\lambda_1 - \lambda_0)$ is strictly negative. Dividing both sides by it **flips the inequality**:

$$\sum_{i=1}^n X_i < \frac{n\log(\lambda_1/\lambda_0) - \log k}{\lambda_1 - \lambda_0} = k_0$$

Thus, our test function is:

$$\phi^*(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \sum X_i \le k_0 \\ 0 & \text{if } \sum X_i > k_0 \end{cases}$$

Evaluating the Size and Constant k_0: Under H_0, the sum of IID exponentials is Gamma distributed: $T_n = \sum X_i \sim \Gamma(n, \lambda_0)$.

$$\alpha = E_{\lambda_0}\left[\phi^*(\underset{\sim}{X})\right] = P_{\lambda_0}\left(T_n \le k_0\right)$$

Therefore, k_0 must be the α-quantile of the Gamma distribution: $k_0 = \Gamma_{\alpha;n,\lambda_0}$.
Why this is UMP: Notice that the final test condition ($T_n \le \Gamma_{\alpha;n,\lambda_0}$) does **NOT** depend on the specific choice of λ_1, so long as $\lambda_1 > \lambda_0$. Since this single test is the Most Powerful test for *every* possible $\lambda_1 \in H_1$, it is the **Uniformly Most Powerful (UMP)** test for the composite alternative.

The Power Function of ϕ^*

For any $\lambda > \lambda_0$, $T_n \sim \Gamma(n, \lambda)$.

$$\beta_{\phi^*}(\lambda) = P_\lambda \left(T_n \leq \Gamma_{\alpha;n,\lambda_0}\right)$$

$$= P_\lambda \left(\frac{\lambda}{\lambda_0} T_n \leq \frac{\lambda}{\lambda_0} \Gamma_{\alpha;n,\lambda_0}\right)$$

Since multiplying T_n by its rate parameter λ transforms it to a standard Gamma, this isolates the exact probabilistic bound.

Composite vs. Composite Hypotheses

We can extend the previous result to a fully composite null hypothesis. Consider the updated testing scenario:

$$H_0^* : \lambda \leq \lambda_0 \qquad \text{vs} \qquad H_1 : \lambda > \lambda_0$$

For all the "newly added points" in H_0^* (i.e., any $\lambda_0' < \lambda_0$), the probability of rejecting the null drops because the distribution shifts further into the acceptance region:

$$P(\text{Type I error at } \lambda_0') = \beta_{\phi^*}(\lambda_0') < \alpha$$

Because the size of the test remains strictly $\leq \alpha$ across the entire expanded null space, the UMP test derived for the simple null $H_0 : \lambda = \lambda_0$ is automatically the valid **UMP level-α test** for the composite null H_0^*.

The Non-Existence of a UMP Test (Two-Sided Alternative)

A global UMP test may not always exist. Consider testing a normal mean with a known variance $N(\mu, 1)$:

$$H_0 : \mu = 0 \qquad \text{vs} \qquad H_1 : \mu \neq 0$$

This two-sided composite alternative is essentially a combination of two one-sided alternatives: $H_{11} : \mu > 0$ and $H_{12} : \mu < 0$.

Conflicting Optimal Tests:

- The UMP test for H_{11} ($\mu > 0$) is ϕ_{11}, which rejects if $\overline{X}_n \geq 1.645/\sqrt{n}$.
- The UMP test for H_{12} ($\mu < 0$) is ϕ_{12}, which rejects if $\overline{X}_n \leq -1.645/\sqrt{n}$.

To test the two-sided alternative H_1, we intuitively combine these into a two-tailed

test ϕ^{**}:

$$\phi^{**}(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \overline{X}_n \geq 1.645/\sqrt{n} \textbf{ or } \overline{X}_n < -1.645/\sqrt{n} \\ 0 & \text{otherwise} \end{cases}$$

The Size Dilemma: If we construct ϕ^{**} using these exact critical boundaries, the combined size under H_0 doubles:

$$E_{H_0}\left[\phi^{**}(\underset{\sim}{X})\right] = 2\left[1 - \Phi(1.645)\right] = 2\alpha$$

To fix the size back to exactly α (e.g., to 0.05), we must push the critical boundaries further out (to 1.96). However, doing this **shrinks** the rejection region on both sides.
Because the critical region is strictly smaller on the right, this new two-tailed test will have strictly **less power** for $\mu > 0$ compared to the dedicated one-sided test ϕ_{11}. Similarly, it will have strictly **less power** for $\mu < 0$ compared to ϕ_{12}. Thus, no single test can maximize power uniformly across both positive and negative alternatives simultaneously.

12.10 Unbiasedness and UMPU Tests

EXAMPLE | POWER FUNCTION OF A VARIANCE TEST

Consider $X_1, \ldots, X_n \overset{IID}{\sim} N(0, \sigma^2)$. We want to test the hypothesis:

$$H_0 : \sigma \leq 1 \qquad \text{vs} \qquad H_1 : \sigma > 1$$

Consider the following test function:

$$\phi(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } S_n^2 > 2 \\ 0 & \text{otherwise} \end{cases}$$

Evaluating the Power Function

We can evaluate the power function of this test, $\beta_\phi(\sigma)$:

$$\beta_\phi(\sigma) = E_\sigma[\phi(\underset{\sim}{X})] = P_\sigma(S_n^2 > 2)$$

$$= P_\sigma\left(\frac{nS_n^2}{\sigma^2} > \frac{2n}{\sigma^2}\right)$$

Since $\frac{nS_n^2}{\sigma^2} \sim \chi^2_{(n)}$, we can rewrite the probability in terms of the Chi-Square CDF:

$$\beta_\phi(\sigma) = 1 - P\left(\chi^2_{(n)} \leq \frac{2n}{\sigma^2}\right) = 1 - F_{\chi^2_{(n)}}\left(\frac{2n}{\sigma^2}\right)$$

Interpretation:

- For any $\sigma^* \leq 1$ (under H_0), $\beta_\phi(\sigma^*) = P_{\sigma^*}(\text{Rejecting } H_0) = P(\text{Type I error at } \sigma^*)$.
- For any $\sigma^{**} > 1$ (under H_1), $\beta_\phi(\sigma^{**}) = P_{\sigma^{**}}(\text{Rejecting } H_0) = \text{Power at } \sigma^{**}$.

EXAMPLE | NORMAL MEAN TESTING AND COMBINED TESTS

Consider the problem of testing:

$$H_0 : \mu = 0 \qquad \text{vs} \qquad H_1 : \mu \neq 0$$

given $X_1, \ldots, X_n \overset{IID}{\sim} N(\mu, 1)$.
The UMP tests for H_0 vs $H_{11}(\mu > 0)$ and $H_{12}(\mu < 0)$ respectively are:

$$\phi_A(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \overline{X}_n \geq 1.67/\sqrt{n} \\ 0 & \text{otherwise} \end{cases} \qquad \text{vs} \qquad \phi_B(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \overline{X}_n \leq -1.67/\sqrt{n} \\ 0 & \text{otherwise} \end{cases}$$

Consider the following power functions:

$$\begin{aligned} \beta_{\phi_A}(\mu) = E_\mu\left(\phi_A(\underset{\sim}{X})\right) &= P_\mu\left(\overline{X}_n \geq 1.67/\sqrt{n}\right) \\ &= 1 - \Phi(1.67 - \mu\sqrt{n}) \\ &= \Phi(\mu\sqrt{n} - 1.67) \end{aligned}$$

Similarly, we have the following for ϕ_B:

$$\beta_{\phi_B}(\mu) = 1 - \Phi(\mu\sqrt{n} + 1.67)$$

Thus, we can construct a combined two-sided test:

$$\phi_C(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \overline{X}_n \geq 1.67/\sqrt{n} \text{ or } \overline{X}_n \leq -1.67/\sqrt{n} \\ 0 & \text{otherwise} \end{cases}$$

We have the following power for $\phi_C(\mu)$:

$$\beta_{\phi_C}(\mu) = \Phi(\mu\sqrt{n} - 1.67) + 1 - \Phi(\mu\sqrt{n} + 1.67)$$

We can clearly see that $\beta_{\phi_C}(0) = 2\alpha$. Thus, ϕ_C is not a level-α test; it is a level-2α test. We now try to modify ϕ_C such that it becomes a level-α test.
Consider the following updated test function:

$$\phi_C^*(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \overline{X}_n \geq 2.68/\sqrt{n} \text{ or } \overline{X}_n \leq -2.68/\sqrt{n} \\ 0 & \text{otherwise} \end{cases}$$

Now, $\beta_{\phi_C^*}(0) = \alpha$. However, because the critical regions were shrunken, there exists **NO** UMP test for H_0 vs H_1 overall.

EXAMPLE | THE NEED FOR UNBIASEDNESS

Let's look at a flaw in using a one-sided test for a two-sided alternative. Suppose we use ϕ_A (which is UMP for $H_{11} : \mu > 0$). What if the true mean is actually $\mu_1 = -5$ (which falls under H_{12})?
The power of ϕ_A at $\mu_1 = -5$ is:

$$\beta_{\phi_A}(-5) = \Phi(-5\sqrt{n} - 1.67) \approx 0$$

However, the probability of a Type I error under H_0 ($\mu = 0$) is α. Thus, we face a scenario where:

$$\text{Power at } \mu_1 = 0 < \alpha = P(\text{Type I error})$$

- $P(\text{Rejecting } H_0 \mid H_1 \text{ is true}) < P(\text{Rejecting } H_0 \mid H_0 \text{ is true})$.
- **This is bad!!** A wrong decision is more probable than a correct one in certain regions of the alternative space.
- Ideally, we want the power at any θ under $H_1 \geq$ Probability of Type I error at θ under H_0.

Unbiased Test

A test is called *Unbiased* if the power of the test is larger than or equal to the size of the test for $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$.
For any $\theta_0 \in \Theta_0$, there is a $P(\text{Type I error}) = P(\text{Rejecting } H_0 \mid H_0 \text{ is true})$. Thus, we have:

$$\sup_{\theta \in \Theta_0} P(\text{Type I error under } \theta) = \text{Size} = \sup_{\theta \in \Theta_0} \beta_\phi(\theta)$$

Similarly, for any $\theta_1 \in \Theta_1$, there is a $P(\text{Type II error under } \theta_1)$. Also, there exists a power under θ_1 such that $\text{Power} = 1 - P(\text{Type II error under } \theta_1)$. Thus, we have:

$$\text{smallest possible power } = \inf_{\theta \in \Theta_1} (1 - P(\text{Type II error under } \theta_1)) = \inf_{\theta \in \Theta_1} \beta_\phi(\theta)$$

Thus, we have the following **Unbiased criterion**:

$$\boxed{\sup_{\theta \in \Theta_0} \beta_\phi(\theta) \leq \inf_{\theta \in \Theta_1} \beta_\phi(\theta)}$$

Uniformly Most Powerful Unbiased (UMPU) Test

To avoid the pitfalls of biased tests, we should search for the "best" unbiased level-α test. A test ϕ is called a **Uniformly Most Powerful Unbiased (UMPU)** test if:

i. It is an unbiased test.
ii. It is uniformly most powerful among the class of all unbiased level-α tests.

12.11 The p-value Framework

EXAMPLE | MOTIVATING EXAMPLE: RIGHT-TAILED TEST

Consider the problem of testing a null hypothesis (H_0) against an alternative hypothesis (H_1):

$$H_0 : \mu = 0 \qquad \text{vs} \qquad H_1 : \mu > 0$$

based on 100 samples where $X_1, \ldots, X_{100} \overset{IID}{\sim} N(\mu, 1)$. We have the following Uniformly Most Powerful (UMP) test:

$$\phi^*(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \overline{X}_n \geq 1.67/\sqrt{100} \iff \sum_{i=1}^{100} X_i \geq 16.7 \\ 0 & \text{otherwise} \end{cases}$$

Analyzing Realizations

Consider two sets of observed realizations:

$$\underset{\sim\text{obs}}{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_{100} \end{pmatrix} \qquad \underset{\sim\text{obs}}{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_{100} \end{pmatrix}$$

Suppose their sums are:

$$\sum_{i=1}^{100} x_{i,\text{obs}} = 17.25 \qquad \sum_{i=1}^{100} y_{i,\text{obs}} = 25.7$$

Say we also have a third realization, $\underset{\sim\text{obs}}{z}$, where:

$$\sum_{i=1}^{100} z_{i,\text{obs}} = -25$$

Core Concept: Understanding the p-value

How do we obtain a method to measure the extremity of a realization?

- **Identify the Basics:** For a given testing problem (H_0 and H_1), we first need to establish:
  1. An appropriate test statistic (e.g., $\overline{X}_n$ or $\sum X_i$).
  2. The direction of extremity (e.g., whether larger or smaller values indicate a more extreme observation).
- **The Problem with Simple Likelihood:** We might try to calculate how unlikely it is to observe the realization under H_0 using the joint distribution/likelihood of $\underset{\sim}{X}$ evaluated at our observation:

$$P(\underset{\sim}{X} = \underset{\sim\text{obs}}{x}) \qquad \text{or} \qquad f_{H_0}(\underset{\sim\text{obs}}{x})$$

However, if we just calculate $f_{H_0}(\underset{\sim\text{obs}}{x})$, we miss the *direction* of extremity. For example, both $f_{H_0}(\underset{\sim\text{obs}}{y})$ and $f_{H_0}(\underset{\sim\text{obs}}{z})$ will be very small under H_0. But in our right-tailed test, only $\underset{\sim\text{obs}}{y}$ indicates an extreme observation that challenges H_0, while $\underset{\sim\text{obs}}{z}$ does not.

- **The Solution (p-value):** To accommodate the direction of the alternative hypothesis, we calculate the probability of seeing our observation *or something even more extreme*, assuming the null hypothesis is true.

$$p\text{-value} = p(\underset{\sim\text{obs}}{y}) = P_{H_0}(\text{observing } \underset{\sim\text{obs}}{y} \text{ or a more extreme observation})$$
$$= P_{H_0}\left(\sum_{i=1}^{100} X_i \geq \sum_{i=1}^{100} y_{i,\text{obs}}\right)$$

EXAMPLE | ALTERNATIVE SCENARIO: LEFT-TAILED TEST

Suppose the testing problem is reversed:

$$H_0 : \mu = 0 \qquad \text{vs} \qquad H_1 : \mu < 0$$

(Based on the same 100 samples $X_1, \ldots, X_{100} \overset{IID}{\sim} N(\mu, 1)$)
Because the direction of extremity is now in the negative direction, the p-value calculation adapts accordingly:

$$p\text{-value} = p(\underset{\sim\text{obs}}{x}) = P_{H_0}\left(\sum_{i=1}^{100} X_i \leq \sum_{i=1}^{100} x_{i,\text{obs}}\right)$$

KEY TAKEAWAY

We only require the **test statistic** and the **direction of rejection** to calculate the p-value.

Definition and Application of p-value

The p-value is the probability of observing data at least as extreme as the realized value, assuming H_0 is true. Given a realization $\underset{\sim\text{obs}}{x}$, it is defined as:

$$p = p(\underset{\sim\text{obs}}{x}) = \inf\left\{\alpha : \text{given } \underset{\sim\text{obs}}{x}, H_0 \text{ is rejected against } H_1 \text{ at level } \alpha\right\}$$

How to test using p-value: For a right-tailed test:

$$p(\underset{\sim\text{obs}}{x}) = P_{H_0}(\overline{X}_n \geq \overline{x}_{n,\text{obs}}) = \rho(\underset{\sim\text{obs}}{x})$$

If the p-value of $\underset{\sim\text{obs}}{x}$ satisfies $p(\underset{\sim\text{obs}}{x}) < \alpha$, then we **reject** H_0.

EXAMPLE | CALCULATING p-VALUE (BINOMIAL)

Suppose we have the following sample $X_1, \ldots, X_{10} \overset{IID}{\sim} Ber(p)$:

$$H_0 : p = \frac{1}{4} \qquad \text{vs} \qquad H_1 : p > \frac{1}{4}$$

Consider the realization $\{0,1,0,0,1,1,1,0,0,1\} = \underset{\sim}{x}_{\text{obs}}$ and the following test function:

$$\phi^*(\underset{\sim}{X}) = \begin{cases} 1 & \text{if } \sum X_i > k_0 \\ \gamma & \text{if } \sum X_i = k_0 \\ 0 & \text{otherwise} \end{cases}$$

Consider the following calculation:

$$\begin{aligned} p(\underset{\sim}{x}_{\text{obs}}) &= P_{p=1/4}\left(\sum_{i=1}^{10} X_i \geq 5\right) \\ &= P\left(T_{10} \geq 5 \mid T_{10} \sim Bin(10, 0.25)\right) \\ &= 0.0781 > 0.05 \end{aligned}$$

Since $0.0781 > 0.05$, we **accept** H_0 given $\underset{\sim}{x}_{\text{obs}}$.

EXAMPLE | p-VALUE FOR TWO-SIDED ALTERNATIVES

Let $X_1, \ldots, X_n \overset{IID}{\sim} N(\mu, 1)$. Consider the following UMPU test function:

$$H_0 : \mu = 0 \qquad \text{vs} \qquad H_1 : \mu \neq 0$$

Recall:

$$\phi(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \overline{x}_n \geq c_0 \text{ or } \overline{x}_n \leq -c_0 \\ 0 & \text{otherwise} \end{cases}$$

Consider an observation $\underset{\sim}{x}_{\text{obs}}$. We evaluate both tails:

$$P_{H_0}\left(\overline{X}_n \geq \overline{x}_{\text{obs}}\right) = p_1 \qquad P_{H_0}\left(\overline{X}_n \leq \overline{x}_{\text{obs}}\right) = p_2$$

The two-sided p-value is given by:

$$p(\underset{\sim}{x}_{\text{obs}}) = 2\min\{p_1, p_2\}$$

We reject H_0 if $p(\underset{\sim}{x}_{\text{obs}}) < \alpha$.

13 Interval Estimation

13.1 Fundamental Concepts

EXAMPLE | INTRODUCTORY EXAMPLE

Let $X_1, \ldots, X_n \overset{IID}{\sim} N(\mu, 1)$. We know that $\overline{X}_n$ is the UMVUE of μ, and that $\overline{X}_n \sim N\left(\mu, \frac{1}{n}\right)$.
Thus, we can say that:

$$T_n = \sqrt{n}(\overline{X}_n - \mu) \sim N(0, 1)$$

Evaluating probabilities based on the standard normal distribution:

$$\begin{aligned} & P_\mu\left(-1.67 \leq T_n \leq 1.67\right) && = 0.9 \quad \ldots(1) \\ \implies & P_\mu\left(-1.67 \leq \sqrt{n}(\overline{X}_n - \mu) \leq 1.67\right) && = 0.9 \\ \implies & P_\mu\left(\overline{X}_n - \frac{1}{\sqrt{n}}1.67 \leq \mu \leq \overline{X}_n + \frac{1}{\sqrt{n}}1.67\right) && = 0.9 \end{aligned}$$

Thus, we can say that the interval

$$\left[\overline{X}_n - \frac{1}{\sqrt{n}}1.67, \overline{X}_n + \frac{1}{\sqrt{n}}1.67\right]$$

contains μ with probability 0.9. The width of the interval is $\frac{2 \times 1.67}{\sqrt{n}}$.

Interval Estimate

Let $X_1, \ldots, X_n \overset{IID}{\sim} F_\theta$. An **interval estimate** of $g(\theta)$ is a pair of statistics $\left[L(\underset{\sim}{X}), U(\underset{\sim}{X})\right]$ such that $L(\underset{\sim}{X}) \leq U(\underset{\sim}{X})$ with probability 1.
Note: Sometimes, $L(\underset{\sim}{X}) = -\infty$ or $U(\underset{\sim}{X}) = +\infty$. In such cases, we get a one-sided interval.

Confidence Coefficient (CC) and Confidence Interval

The **Confidence Coefficient (CC)** associated with an interval estimate $\left[L(\underset{\sim}{X}), U(\underset{\sim}{X})\right]$ for the parameter $g(\theta)$ is a number $\beta = (1-\alpha) \in [0,1]$ such that:

$$P_\theta\left[L(\underset{\sim}{X}) \leq g(\theta) \leq U(\underset{\sim}{X})\right] \geq \beta = (1-\alpha) \quad \forall \theta \in \Theta$$

or

$$\inf_{\theta \in \Theta} P_\theta\left[L(\underset{\sim}{X}) \leq g(\theta) \leq U(\underset{\sim}{X})\right] = \beta = (1-\alpha)$$

An interval estimate of $g(\theta)$ and its associated confidence coefficient together is called a **Confidence Interval**. We typically write that $\left[L(\underset{\sim}{X}), U(\underset{\sim}{X})\right]$ is a $100(1-\alpha)\%$ confidence interval for $g(\theta)$.

Interpretation

If we repeat the experiment N times, then in approximately $N \times \beta$ number of cases, the obtained interval will successfully capture the true parameter $g(\theta)$.

HOW TO OBTAIN A CONFIDENCE INTERVAL

The two primary methods for finding an interval estimate are:
1. Method of Pivots
2. Test Inversion

13.2 The Method of Pivots

Pivots

Definition: A random variable $T_n(\theta) = g(\underset{\sim}{X}, \theta)$, which is a function of $\underset{\sim}{X}$ and θ, and whose distribution is **free of** θ, is called a *Pivot*.

EXAMPLE | EXAMPLES OF PIVOTS

1. Let $X_1, \ldots, X_n \overset{IID}{\sim} N(\mu, \sigma_0^2)$ where σ_0 is known. We have the following pivot $T_n(\mu)$:

$$T_n(\mu) = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma_0} \sim N(0,1)$$

2. Let $X_1, \ldots, X_n \overset{IID}{\sim} N(\mu, \sigma^2)$.

$$T_n(\sigma^2) = \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

3. Let $X_1, \ldots, X_n \overset{IID}{\sim} \exp(\lambda)$. For $n = 1$:

$$F_X(x) = U_\lambda(x) = P(X \leq x) = \lambda \int_0^x e^{-\lambda y} dy = 1 - e^{-\lambda x}$$

We know that $U_\lambda(X) \sim U(0, 1)$.
For $n > 1$, consider an appropriate statistic, for example $T_n = \sum X_i$. The CDF of T_n, $U_{T_n}(\lambda)$, is also a pivot as $U_{T_n}(\lambda) \sim U(0, 1)$.

Steps for the Method of Pivots

Let $X_1, \ldots, X_n \overset{IID}{\sim} F_\theta$. We need to find a $100(1-\alpha)\%$ confidence interval for $g(\theta)$.

1. **Find an appropriate pivot:**
   (a) It should be a function of a 'good estimate' of $g(\theta)$ and $g(\theta)$ itself.
   (b) Say the pivot is $T_n(\theta)$.
   (c) $T_n(\theta)$ should be solvable w.r.t $g(\theta)$.
   (d) $T_n(\theta)$ should be a monotonic function w.r.t $g(\theta)$.
2. **Find a pair of constants** (a, b) **such that:**

$$P_\theta(a \leq T_n(\theta) \leq b) = (1 - \alpha)$$

3. **Solve** $T_n(\theta) = a$ **and** $T_n(\theta) = b$**:** Let the solutions be

$$g(\theta) = S_a \qquad \text{and} \qquad g(\theta) = S_b$$

respectively. *(Note: S_a and S_b are statistics).*
4. **Construct the Interval:** Evaluate the condition $\{a \leq T_n(\theta) \leq b\}$.
   - Suppose $T_n(\theta)$ is a **strictly increasing** function of $g(\theta)$. Then:

$$\{a \leq T_n(\theta) \leq b\} \Leftrightarrow \{S_a \leq g(\theta) \leq S_b\}$$

Thus:

$$P_\theta(a \leq T_n(\theta) \leq b) = (1 - \alpha) = P_\theta\left(\underbrace{S_a}_{L(\underset{\sim}{X})} \leq g(\theta) \leq \underbrace{S_b}_{U(\underset{\sim}{X})}\right)$$

   - Suppose $T_n(\theta)$ is a **strictly decreasing** function of $g(\theta)$. Then:

$$\{a \leq T_n(\theta) \leq b\} \Leftrightarrow \{S_b \leq g(\theta) \leq S_a\}$$

Thus:

$$P_\theta(a \leq T_n(\theta) \leq b) = (1 - \alpha) = P_\theta\left(\underbrace{S_b}_{L(\underset{\sim}{X})} \leq g(\theta) \leq \underbrace{S_a}_{U(\underset{\sim}{X})}\right)$$

EXAMPLE | APPLYING THE METHOD OF PIVOTS

1. Normal Mean with Known Variance: $X_1, \ldots, X_n \overset{IID}{\sim} N(\mu, \sigma_0^2)$

- **Step 1:** $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma_0} = T_n(\mu) \sim N(0,1)$
- **Step 2:** $(a, b) : P_\theta(a \leq T_n(\mu) \leq b) = (1-\alpha)$

$$a = -\tau_{\alpha/2} = \tau_{1-\alpha/2} : \text{lower-}\alpha/2 \text{ point}$$

$$b = \tau_{\alpha/2} : \text{upper-}\alpha/2 \text{ point}$$

- **Step 3:** Solve the following equations:

$$S_a = \overline{X}_n - \frac{\tau_{1-\alpha/2}}{\sqrt{n}}\sigma_0 \qquad S_b = \overline{X}_n - \frac{\tau_{\alpha/2}}{\sqrt{n}}\sigma_0$$

- **Step 4:** As $T_n(\mu)$ is a monotonically decreasing function of μ, we have:

$$P_\mu\left(\overline{X}_n - \frac{\tau_{\alpha/2}}{\sqrt{n}}\sigma_0 \leq \mu \leq \overline{X}_n - \frac{\tau_{1-\alpha/2}}{\sqrt{n}}\sigma_0\right) = 1-\alpha$$

(Note: Since standard normal is symmetric, $-\tau_{1-\alpha/2} = \tau_{\alpha/2}$, recovering the classic CI format).

> **CHOOSING A "GOOD" PIVOT**
>
> We could have also considered the pivot $T_n^*(\mu) = \dfrac{X_1 - \mu}{\sigma_0} \sim N(0,1)$. We would have ended up with the interval estimate:
>
> $$P_\mu\left(X_1 - \tau_{\alpha/2}\sigma_0 \leq \mu \leq X_1 - \tau_{1-\alpha/2}\sigma_0\right) = 1-\alpha$$
>
> However, the interval width in this case is much larger than when using $\overline{X}_n$. This demonstrates why choosing a 'good pivot' based on a sufficient/efficient estimator is vital.

2. Uniform Distribution: $X_1, \ldots, X_n \overset{IID}{\sim} U(0, \theta)$ for estimating $g(\theta) = \theta$

- **Step 1 (Pivot):** Should involve $X_{(n)}$ and θ. The CDF of $X_{(n)}$ follows a $U(0,1)$ distribution. Thus:

$$T_n(\theta) = \left(\frac{X_{(n)}}{\theta}\right)^n \sim U(0,1)$$

- **Step 2 (Find a, b):** Consider $a = \alpha/2$ and $b = 1 - \alpha/2$.
- **Step 3 (Solve):** $T_n(\theta) = \alpha/2$ and $T_n(\theta) = 1 - \alpha/2$. Thus:

$$S_a = \left(\frac{2}{\alpha}\right)^{1/n}\left(X_{(n)}\right) \qquad S_b = \frac{X_{(n)}}{(1-\alpha/2)^{1/n}}$$

- **Step 4 (Construct):** $T_n(\theta)$ is a monotonically decreasing function of θ. Thus, we have:

$$P_\theta\left(S_b \leq \theta \leq S_a\right) = (1-\alpha)$$

13.3 Location, Scale, and Location-Scale Families

Location Family

Let $X_1, \ldots, X_n \sim F_\theta$ such that $X_i = W_i + \theta$, where $W_i \overset{IID}{\sim} F$ and F is free of θ.
Any function of $W_1, \ldots, W_n \iff X_1 - \theta, \ldots, X_n - \theta$ is free of θ and thus can serve as a pivot.

EXAMPLE | EXAMPLES OF LOCATION FAMILY

1. $X_i \overset{IID}{\sim} N(\mu, 1)$. Here, $W_i = X_i - \mu \sim N(0, 1)$.
2. $X_i \overset{IID}{\sim} U(\theta - 1/2, \theta + 1/2)$. Here, we have $W_i = X_i - \theta \sim U(-1/2, 1/2)$.

Scale Family

Let $X_1, \ldots, X_n \overset{IID}{\sim} F_\theta$ such that:

$$X_i = \frac{W_i}{\theta} \qquad (\text{or } X_i = \theta W_i)$$

where $W_i \overset{IID}{\sim} F$, which is free of θ.
Any function of $W_1, \ldots, W_n \iff \theta X_1, \ldots, \theta X_n$ (or $X_1/\theta, \ldots$) is a valid pivot.

EXAMPLE | EXAMPLES OF SCALE FAMILY

1. $X \sim \exp(\lambda) \implies W = \lambda X \sim \exp(1)$.
2. $X \sim N(0, \sigma^2) \implies W = \frac{X}{\sigma} \sim N(0, 1)$.

Location-Scale Family

Let $X_i \overset{IID}{\sim} F_\theta$ such that:

$$X_i = \frac{W_i - \theta_1}{\theta_2}$$

where θ_1 is the location parameter and θ_2 is the scale parameter.
$W_i \overset{IID}{\sim} F$ is free from any $\underset{\sim}{X}$. Any function of:

$$W_1, \ldots, W_n \iff \frac{X_1 - \theta_1}{\theta_2}, \ldots, \frac{X_n - \theta_1}{\theta_2}$$

is a valid pivot.

EXAMPLE | EXAMPLES OF INFERENCE IN FAMILIES

1. Location-Scale Family (Normal): $X_1, \ldots, X_n \overset{IID}{\sim} N(\mu, \sigma^2)$. We know:

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim N(0,1) \qquad \text{and} \qquad \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

By dividing these independent variables, we eliminate the scale parameter σ:

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \times \sqrt{\frac{\sigma^2(n-1)}{\sum_{i=1}^n (X_i - \overline{X}_n)^2}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sqrt{\frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n-1}}}$$

$$= \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n^*} \sim t_{(n-1)}$$

Using the quantiles of the t-distribution ($\tau_{\alpha/2;n-1}$ and $-\tau_{\alpha/2;n-1}$), we eventually get the following $100(1-\alpha)\%$ confidence interval estimate for μ when σ is unknown:

$$\left[\overline{X}_n - \tau_{\alpha/2}\frac{S_n^*}{\sqrt{n}}, \overline{X}_n + \tau_{\alpha/2}\frac{S_n^*}{\sqrt{n}}\right]$$

2. Scale Family (Exponential): $X_1, \ldots, X_n \overset{IID}{\sim} \exp(\lambda)$. We should choose an appropriate function of $\lambda X_1, \ldots, \lambda X_n$.

- **Step 1:** $\lambda \sum_{i=1}^n X_i = T_n(\lambda) \sim \Gamma(n,1)$
- **Step 2:** $(a,b) \equiv \left(\Gamma_{1-\frac{\alpha}{2};n,1}, \Gamma_{\frac{\alpha}{2};n,1}\right)$. Thus:

$$P_\lambda\left(\Gamma_{1-\frac{\alpha}{2};n,1} \leq T_n(\lambda) \leq \Gamma_{\frac{\alpha}{2};n,1}\right) = (1-\alpha)$$

- **Step 3:**

$$T_n(\lambda) = a \iff \lambda = \frac{a}{\sum X_i} = \frac{na}{\overline{X}_n} = S_a$$

$$T_n(\lambda) = b \iff \lambda = \frac{b}{\sum X_i} = \frac{nb}{\overline{X}_n} = S_b$$

- **Step 4:** $T_n(\lambda)$ is an increasing function of λ. Thus, we have the following result:

$$P_\lambda\left(\frac{n\Gamma_{1-\frac{\alpha}{2};n,1}}{\sum X_i} \leq \lambda \leq \frac{n\Gamma_{\frac{\alpha}{2};n,1}}{\sum X_i}\right) = (1-\alpha)$$

3. Location Family (Homework): $X_1, \ldots, X_n \overset{IID}{\sim} U(\theta, \theta+1)$. Consider the transformation:

$$W_i = X_i - \theta \overset{IID}{\sim} U(0,1)$$

We can construct a pivot using the known relation $-\log U(0,1) \sim \exp(1)$:

$$\sum_{i=1}^n (-\log W_i) \sim \Gamma(n,1)$$