ASSIGNMENT TOPIC : Fisher's Lemma and Study of Chi-Square (χ^2) Distribution

Course Code : STAT-2201
Course Tittle : Engineering Statistics

Submitted By :
MD. SABBIR HOSSAIN
Roll : 220603
Session : 2021-22
Dept. of Information and Communication Engineering

Submitted To :
Dr. Md. Sarwar Hosain
Associate Professor

Dept. of Information and Communication Engineering

Submission Date : 23-04-2025

Department of
INFORMATION AND COMMUNICATION ENGINEERING
PABNA UNIVERSITY OF SCIENCE AND TECHNOLOGY

Fisher’s Lemma and Study of Chi-Square (χ^2) Distribution

Introduction

Statistics as a discipline involves analyzing data, making inferences, and testing hypotheses about population parameters based on sample data. Among the many theorems and distributions foundational to statistical theory, Fisher’s Lemma and the Chi-Square (χ^2) distribution are critically important in both theoretical and applied contexts.

This assignment discusses in detail Fisher’s Lemma and the Chi-Square distribution. Fisher’s Lemma plays an essential role in understanding the behavior of estimators in normally distributed data. The χ^2 distribution is a special distribution widely used in hypothesis testing, especially in the context of variance and categorical data.

Fisher’s Lemma

Historical Background

Fisher’s Lemma is named after Sir Ronald A. Fisher, one of the most influential statisticians of the 20th century. He laid the groundwork for modern statistical methods, especially in estimation theory, experimental design, and hypothesis testing.

Theoretical Statement

Let X_1,X_2,...,X_n be independent and identically distributed random variables drawn from a normal distribution:

$$X_i \sim \text{N}(\mu,\sigma2), \qquad i = 1,2,\ldots\ldots,n.$$

Then Fisher's Lemma tells us:

- The sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is independent of the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- The scaled sample variance $\frac{(n-1)S^2}{\sigma^2} \sim X_{n-1}^2$.

This is crucial because it means that in a normal population, the sample mean and sample variance provide independent information about the population parameters μ and σ2.

Mathematical Proof Sketch

Let us define the total sum of squares:

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \mu)^2.$$

Using algebraic manipulation and properties of expectations, this can be rewritten as:

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2.$$

This identity is the essence of Fisher's Lemma—it decomposes the total variation into the variation due to sampling error (within-sample variability) and the error due to estimation of the mean (between-sample variability). It also leads directly to the independence of $\bar{X}$ and S^2.

Application of Fisher's Lemma

Example

Let's take a small data set:

$$X = \{5,7,9,6,8\}, \quad n = 5$$

Calculate the sample mean:

$$\bar{X} = \frac{5 + 7 + 9 + 6 + 8}{5} = 7$$

Calculate the sum of squares:

$$\sum(X_i - \bar{X})^2 = (5-7)^2 + (7-7)^2 + (9-7)^2 + (6-7)^2 + (8-7)^2$$
= 4 + 0 + 4+ 1+ 1 =10

Sample variance:

$$S^2 = \frac{10}{4} = 2.5$$

Then:

$$\frac{(n-1)S^2}{\sigma^2} = \frac{10}{\sigma^2} \sim X_4^2$$

This shows how Fisher's Lemma connects the sample variance with the chi-square distribution.

Chi-Square (χ^2) Distribution

Definition

A random variable is said to follow a chi-square distribution with k degrees of freedom if it is the sum of the squares of k independent standard normal variables:

$$X_k^2 = \sum_{i=1}^{k} Z_i^2 \quad , \qquad where\ Z_i \sim N(0,1).$$

This distribution is denoted as X_k^2 .

Properties

- **Non-Negativity:** Since it is a sum of squared terms, $\chi^2 \geq 0$.
- **Shape:** Positively skewed, more so for smaller k.
- **Mean:** $E\,[X_k^2] = k$.
- **Variance:** $Var[X_k^2] = 2\,k$
- **Additivity:** If χ_{k1}^2 and χ_{k2}^2 are independent, then χ_{k1}^2 + χ_{k2}^2 ~ χ_{k1+k2}^2

PDF (Probability Density Function)

The PDF of the chi-square distribution is:

$$f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} \ , \ x > 0$$

This function becomes more bell-shaped as k increases.

Applications of χ^2Distribution

Goodness-of-Fit Test

Used to test whether a sample data matches a population with a specific distribution.

Example: A die is rolled 60 times. The frequencies of outcomes are compared to expected frequencies using:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where O_i is observed frequency and E_i is expected frequency.

Test of Independence

Used in contingency tables to test whether two categorical variables are independent.

Example: In a 2×2 table showing gender and voting preference, chi-square can be used to test independence.

Confidence Interval for Variance

If $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2), then$

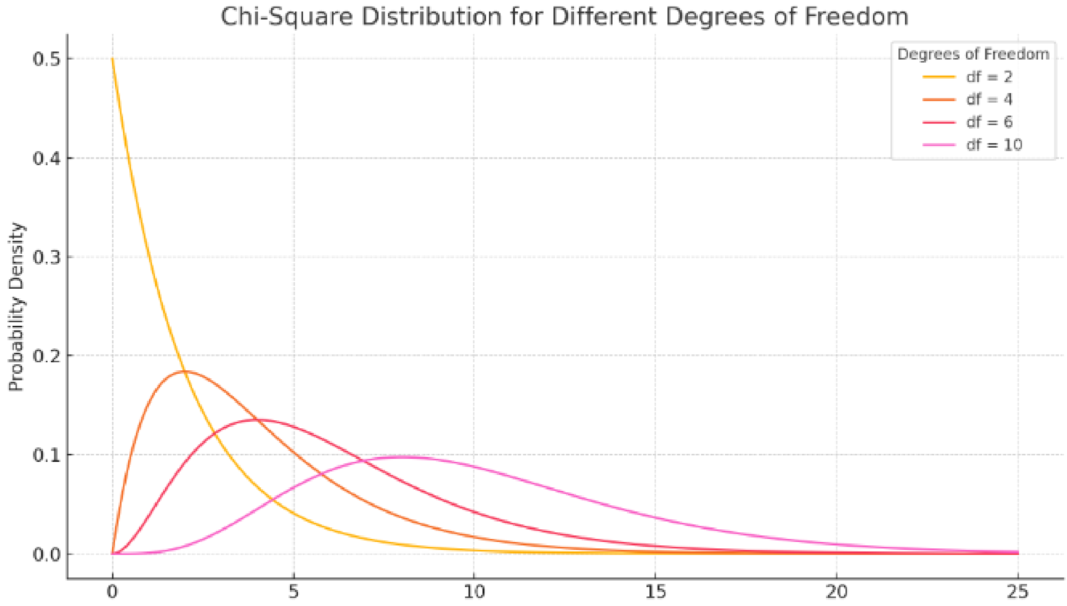
$$\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}$$

This provides a confidence interval for the population variance.

Graphical Representation

The following key features are seen in the χ^2-curve:

- Skewed to the right for small degrees of freedom.
- Becomes symmetric as degrees of freedom increase.
- Mode at $k - 2 \; for \; k \geq 2$.



Conclusion

Fisher’s Lemma and the Chi-Square distribution are pillars of inferential statistics. Fisher’s Lemma gives deep insight into the independence of estimators in normal models, laying the groundwork for t and F distributions. The χ^2-distribution serves as a foundational tool in hypothesis testing and estimation, especially in categorical data analysis and variance testing.

Understanding these concepts thoroughly is essential for any student or practitioner of statistics, as they recur frequently in both theoretical derivations and real-world applications