CSL003P1M : Probability and Statistics
Lecture 37 (Distribution of Sampling Statistics)

Sumit Kumar Pandey

December 07, 2021

The Chi-Square Distribution

The Chi-Square Distribution

If $Z_1, Z_2, \ldots, Z_n$ are **independent standard normal** random variables, then X, defined by

$$X = Z_1^2 + Z_2^2 + \ldots + Z_n^2$$

is said to have a chi-square distribution with n degrees of freedom.

Additive Property of Chi-Square Distribution

If X_1 and X_2 are independent chi-square random variables with n_1 and n_2 degrees of freedom, respectively, then $X_1 + X_2$ is chi-square with $n_1 + n_2$ degrees of freedom.

If X is a chi-square random variable with n degrees of freedom, then for any $\alpha \in (0,1)$, the quantity $\chi^2_{\alpha,n}$ is defined to be such that

$$P\{X \geq \chi^2_{\alpha,n}\} = \alpha$$

Exercise-1

Suppose that we are attempting to locate a target in three-dimensional space, and that the three coordinate errors (in meters) of the point chosen are independent normal random variables with mean 0 and standard deviation 2. Find the probability that the distance between the point chosen and the target exceeds 3 meters.

Solution: If D is the distance, then

$$D^2 = X_1^2 + X_2^2 + X_3^2$$

where X_i is the error in the ith coordinate. Since $Z_i = X_i/2$, $i = 1, 2, 3$, are all standard normal random variables, it follows that

$$\begin{aligned} P\{D^2 > 9\} &= P\{Z_1^2 + Z_2^2 + Z_3^2 > 9/4\} \\ &= P\{\chi_3^2 > 9/4\} \\ &= 0.5222 \end{aligned}$$

Introduction

The science of statistics deals with drawing conclusions from observed data. For instance,

- A typical situation in a technological study arises when one is confronted with a large collection, or population, of items that have measurable values associated with them.
- By suitably sampling from this collection, and then analyzing the sampled items, one hopes to be able to draw some conclusions about the collection as a whole.

Assumptions

To use sample data to make inferences about an entire population, it is necessary to make some assumptions about the relationship between the two.

- One such assumption, which is often quite reasonable, is that there is an underlying (population) probability distribution such that the measurable values of the items in the population can be thought of as being independent random variables having this distribution.
- If the sample data are then chosen in a random fashion, then it is reasonable to suppose that they too are independent values from the distribution.

Definitions

Sample or Random Sample

If $X_1, \ldots, X_n$ are independent random variables having a common distribution D, then we say that they constitute a sample (sometimes called a random sample) from the distribution D.

Parametric and Nonparametric Inference

In most applications, the population distribution D will not be completely specified and one will attempt to use the data to make inference about D.

- Sometimes it will be supposed that D is specified up to some unknown parameters (for instance, one might suppose that D was a normal distribution function having an unknown mean and variance, or that it is a Poisson distribution function whose mean is not given).
- And at other times it might be assumed that almost nothing is known about D (except maybe for assuming that it is a continuous, or a discrete distribution).

Parametric and Nonparametric Inference

Parametric Inference Problems

Problems in which the form of the underlying distribution is specified up to set of unknown parameters are called parametric inference.

Nonparametric Inference Problems

Those problems in which nothing is assumed about the form of D are called nonparametric inference problems.

The Sample Mean

The quantities μ and σ^2 are called the **population mean** and **population variance**, respectively.

Sample Mean

Let $X_1, X_2, \ldots, X_n$ be a sample of values from this population. The sample mean is defined by

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

Since the values of the sample mean $\bar{X}$ is determined by the values of the random variables in the sample, it follows that $\bar{X}$ is also a random variable.

Expectation and Variance of the Sample Mean

$$
\begin{aligned}
E[\bar{X}] &= E\left[\frac{X_1 + \cdots + X_n}{n}\right] \\
&= \frac{1}{n}(E[X_1] + \cdots + E[X_n]) \\
&= \mu
\end{aligned}
$$

and

$$
\begin{aligned}
Var(\bar{X}) &= Var\left(\frac{X_1 + \cdots + X_n}{n}\right) \\
&= \frac{1}{n^2}[Var(X_1) + \cdots + Var(X_n)] \quad \text{(by independence)} \\
&= \frac{n\sigma^2}{n^2} \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

The Sample Variance

Sample Variance

Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Let $\bar{X}$ be the sample mean. The statistic S^2, defined by

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

is called the sample variance. $S = \sqrt{S^2}$ is called the sample standard deviation.

Show that

$$E[S^2] = \sigma^2.$$

Joint Distribution of $\bar{X}$ and S^2

Theorem

If $X_1, \ldots, X_n$ is a sample from a normal population having mean μ and variance σ^2, then $\bar{X}$ and S^2 are independent random variables, with $\bar{X}$ being normal with mean μ and variance σ^2/n and $(n-1)S^2/\sigma^2$ being chi-square with $n-1$ degrees of freedom.

Proof: Since the sum of independent normal random variables is normally distributed, it follws that $\bar{X}$ is normal with mean

$$E[\bar{X}] = \sum_{i=1}^{n} \frac{E[X_i]}{n} = \mu$$

and variance

$$Var(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \sigma^2/n$$

Therefore, $\bar{X}$ is a normal random variable with mean equal to the population mean but variance reduced by a factor of $1/n$. It follows from this that

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

is a standard normal random variable.

Now we look at the distribution of S^2

Joint Distribution of $\bar{X}$ *and* S^2

For numbers $x_1, \ldots, x_n$, let $y_i = x_i - \mu$, $i = 1, \ldots, n$. Then as $\bar{y} = \bar{x} - \mu$, it follows from the identity

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

that

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \mu)^2 - n(\bar{x} - \mu)^2$$

Now, if $X_1, \cdots, X_n$ is a sample from a normal population having mean μ and variance σ^2, then we obtain from the preceding identity that

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

or, equivalently

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}+\left[\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right]^2 = \frac{(n-1)S^2}{\sigma^2}+\left[\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}\right]^2$$

- Because $(X_i-\mu)/\sigma,\ i=1,\ldots,n$ are independent standard normals, it follws that left side of the above equation is a chi-square random variable with n degrees of freedom.
- Also, $(\bar{X}-\mu)/\sigma/\sqrt{n}$ is a standard normal random variable and so its square is chi-square random variable with 1 degree of freedom.

Joint Distribution of $\bar{X}$ and S^2

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} + \left[\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right]^2$$

$$= \frac{(n-1)S^2}{\sigma^2} + \left[\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}\right]^2$$

- Since the sum of two independent chi-square random variable is a chi-square with a degree of freedom equal to the sum of the two degrees of freedom, therefore $(n-1)S^2/\sigma^2$ is independent of $\bar{X}$ and moreover a chi-square random variable with $n-1$ degrees of freedom.

Thank You