CSL003P1M : Probability and Statistics Lecture 39 (Estimation-II (Interval Estimates-I))

Sumit Kumar Pandey

December 13, 2021

Introduction

- Suppose that $X_1, \ldots, X_n$ is a sample from a normal population having unknown mean μ and variance σ^2.
- It has been shown that $\bar{X} = \sum_{i=1}^{n} X_i / n$ is the maximum likelihood estimator for μ.
- However, we don't expect that the sample mean $\bar{X}$ will exactly equal μ, but rather that it will "be close".
- Hence, rather than a point estimate, it is sometimes more valuable to be able to specify an interval for which we have a certain degree of confidence that μ lies within.
- To obtain such an interval estimator, we make use of the probability distribution of the point estimator.

Since the point estimator $\bar{X}$ is normal with mean μ and variance σ^2/n, it follows that

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

has a standard normal distribution. Therefore,

$$P\left\{-1.96 < \sqrt{n}\frac{(\bar{X}-\mu)}{\sigma} < 1.96\right\} = 0.95$$

or,

$$P\left\{\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right\} = 0.95$$

Interval Estimates

"With 95 percent confidence" we assert that the true mean lies within $1.96\sigma/\sqrt{n}$ of the observed sample mean. The interval

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

is called a 95 percent confidence interval estimator of μ.

$Exercise$-1

Suppose that when a signal having value μ is transmitted from location A the value received at location B is normally distributed with mean μ and variance 4. That is, if μ is sent, then the value received is $\mu + N$ where N, representing noise, is normal with mean 0 and variance 4. To reduce error, suppose the same value is sent 9 times. If the successive vlaues received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, construct a 95 percent confidence interval for μ.

Solution: Since

$$\bar{x} = \frac{81}{9} = 9.$$

It follows, under the assumption that the values received are independent, that a 95 percent confidence interval for μ is

$$\left(9 - 1.96\frac{\sigma}{3}, 9 + 1.96\frac{\sigma}{3}\right) = (7.69, 10.31)$$

The interval

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

is called **a two-sided confidence interval**. Sometimes, however, we are interested in determining a value so that we can assert with, say, 95 percent confidence, that μ is at least as large as that value. Note that if Z is a standard normal random variable then

$$P\{Z < 1.645\} = 0.95$$

As a result,

$$P\left\{\sqrt{n}\frac{(\bar{X} - \mu)}{\sigma} < 1.645\right\} = 0.95$$

or

$$P\left\{\bar{X} - 1.645\frac{\sigma}{\sqrt{n}} < \mu\right\} = 0.95$$

Thus a 95 percent **one-sided upper confidence interval** for μ is

$$\left(\bar{x} - 1.645\frac{\sigma}{\sqrt{n}}, \infty\right)$$

where $\bar{x}$ is the observed value of the sample mean.

A **one-sided lower confidence interval** is obtained similarly; when the observed value of the sample mean is $\bar{x}$, then the 95 percent one-sided lower confidence interval for μ is

$$\left(-\infty, \bar{x} + 1.645\frac{\sigma}{\sqrt{n}}\right)$$

Suppose that when a signal having value μ is transmitted from location A the value received at location B is normally distributed with mean μ and variance 4. That is, if μ is sent, then the value received is $\mu + N$ where N, representing noise, is normal with mean 0 and variance 4. To reduce error, suppose the same value is sent 9 times. If the successive vlaues received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, construct a 95 percent upper and lower confidence interval for μ.

Solution: Since

$$\bar{x} = \frac{81}{9} = 9.$$

Since

$$1.645\frac{\sigma}{\sqrt{n}} = \frac{3.29}{3} = 1.097$$

The 95 percent upper confidence interval is

$$(9 - 1.097, \infty) = (7.903, \infty)$$

and the 95 percent lower confidence interval is

$$(-\infty, 9 + 1.097) = (-\infty, 10.097)$$

Confidence Interval for a Normal Mean when the Variance is Unknown

Suppose now that $X_1, \ldots, X_n$ is a sample from a normal distribution with unknown mean μ and unknown variance σ^2, and that we wish to construct a $100(1-\alpha)$ percent confidence interval for μ.

- Since σ is unknown, we can no longer base our interval on the fact that $\sqrt{n}(\bar{X}-\mu)/\sigma$ is a standard normal random variable.
- However, by letting $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ denote the sample variance, then it follows that

$$\sqrt{n}\frac{(\bar{X}-\mu)}{S}$$

is a t-random variable with $n-1$ degrees of freedom.

The t-Distribution

The t-distribution

If Z and χ_n^2 are independent random variables, with Z having a standard normal distribution and χ_n^2 having a chi-square distribution with n degrees of freedom, then the random variable T_n defined by

$$T_n = \frac{Z}{\sqrt{\chi_n^2/n}}$$

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

The t-Distribution

- Like the standard normal density, the t-density is symmetric about zero.
- In addition, as n becomes larger, it becomes more and more like a standard normal density.

t-Distribution and Standard Normal Distribution

As n becomes larger, it becomes more and more like a standard normal density.

To understand why, recall that χ_n^2 can be expressed as the sum of the squares of n standard normals, and so on

$$\frac{\chi_n^2}{n} = \frac{Z_1^2 + \cdots + Z_n^2}{n}$$

where $Z_1, \ldots, Z_n$ are independent standard normal random variables.

It now follows from the weak law of large numbers that, for large n, χ_n^2/n will, with probability close to 1, be approximately equal to $E[Z_i^2] = Var(Z_i) = 1$.

Hence for n large, $T_n = Z/\sqrt{\chi_n^2/n}$ will have approximately the same distribution as Z.

Mean and Variance of t-distribution

The mean and variance of T_n can be shown to equal

$$E[T_n] = 0, \quad n > 1$$
$$Var(T_n) = \frac{n}{n-2}, \quad n > 2$$

Thus, the variance of T_n decreses to 1 – the variance of a standard normal random variable – as n increases to ∞.

For α, $0 < \alpha < 1$, let $t_{\alpha,n}$ be such that

$$P\{T_n \geq t_{\alpha,n}\} = \alpha$$

It follows from the symmetry about zero of the t-density function that $-T_n$ has the same distribution as T_n, and so

$$\begin{aligned} \alpha &= P\{-T_n \geq t_{\alpha,n}\} \\ &= P\{T_n \leq -t_{\alpha,n}\} \\ &= 1 - P\{T_n > -t_{\alpha,n}\} \end{aligned}$$

Therefore,

$$P\{T_n \geq -t_{\alpha,n}\} = 1 - \alpha$$

leading to the conclusion that

$$-t_{\alpha,n} = t_{1-\alpha,n}$$

Confidence Interval for a Normal Mean when the Variance is Unknown

Suppose now that $X_1, \ldots, X_n$ is a sample from a normal distribution with unknown mean μ and unknown variance σ^2, and that we wish to construct a $100(1-\alpha)$ percent confidence interval for μ.

- Since σ is unknown, we can no longer base our interval on the fact that $\sqrt{n}(\bar{X}-\mu)/\sigma$ is a standard normal random variable.
- However, by letting $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ denote the sample variance, then it follows that

$$\sqrt{n}\frac{(\bar{X}-\mu)}{S}$$

is a t-random variable with $n-1$ degrees of freedom.

Confidence Interval for a Normal Mean when the Variance is Unknown

Hence, from the symmetry of the t-density function, we have that for any $\alpha \in (0, 1/2)$,

$$P\left\{-t_{\alpha/2,n-1} < \sqrt{n}\frac{(\bar{X}-\mu)}{S} < t_{\alpha/2,n-1}\right\} = 1-\alpha$$

or, equivalently,

$$P\left\{\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right\} = 1-\alpha$$

Thus, if it is observed that $\bar{X} = \bar{x}$ and $S = s$, then we can say that "with $100(1-\alpha)$ percent confidence"

$$\mu \in \left(\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$$

Exercise-3

Consider the previous exercise but let us now suppose that when the value μ is transmitted at location A then the value received at location B is normal with mean μ and variance σ^2 but with σ^2 being unknown. If 9 successive values are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, compute a 95 percent confidence interval for μ.

Solution: A simple calculation yields that

$$\bar{x} = 9$$

and

$$s^2 = \frac{\sum x_i^2 - 9(\bar{x})^2}{8} = 9.5$$

or

$$s = 3.082$$

Exercise-3

Here, $100(1-\alpha) = 95$, so $\alpha = 0.05$. Hence, as $t_{0.025,8} = 2.306$, a 95 percent confidence interval for μ is

$$\left[9 - 2.306\frac{(3.082)}{3}, 9 + 2.306\frac{(3.082)}{3}\right] = (6.63, 11.37)$$

Thank You