MA181 INTRODUCTION TO STATISTICAL MODELLING
CONFIDENCE INTERVALS

Suppose the random variable X follows a normal distribution with mean μ and variance σ^2, and it is desired to estimate μ from a random sample of observations. The best point estimator to use is, in many respects, the sample mean $\overline{X}$. However, a value $\overline{x}$ given by this estimator carries with it no measure of its precision; a value of 13.6 from a sample of 10 values looks just like a value of 13.6 from a sample of 100, yet the latter is more precise, or reliable, than the former. One way to remedy this is to construct an interval around $\overline{x}$, the length of which reflects its reliability or, in other words, our confidence about its containing μ.
The starting point for the construction of this interval is the distribution of $\overline{X}$, which is $N(\mu, \sigma^2/n)$, where n is the sample size. On standardising this result, we obtain

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1). \tag{1}$$

Variance known If the variance σ^2 is known, the distribution of Z given at (1), i.e. $N(0, 1)$, can be used to construct the required interval. For a given probability α, let c be the point such that $\frac{\alpha}{2} = P(Z < -c) = P(Z > c)$. Then, from (1), we have

$$1 - \alpha = P\left(-c \le \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le c\right) = P(\overline{X} - c\sigma/\sqrt{n} \le \mu \le \overline{X} + c\sigma/\sqrt{n}).$$

This is not a probability statement about the population mean μ, which remains constant throughout, but about the interval $[\overline{X} - c\sigma/\sqrt{n}, \overline{X} + c\sigma/\sqrt{n}]$, which in repeated sampling contains μ with probability $1 - \alpha$. For a single sample, with observed mean $\overline{x}$, the interval $[\overline{x} - c\sigma/\sqrt{n}, \overline{x} + c\sigma/\sqrt{n}]$ is called a $100(1 - \alpha)\%$ *confidence interval* for μ. Our level of confidence springs from the fact that this interval is either one of the $100(1 - \alpha)\%$ of intervals that would contain μ if we continued drawing samples of size n, or one of the $100\alpha\%$ of intervals that would fail to contain it. The end-points of the interval, $\overline{x} - c\sigma/\sqrt{n}$ and $\overline{x} + c\sigma/\sqrt{n}$, are known as the *confidence limits* to μ, and the fraction $100(1 - \alpha)\%$ as the *confidence coefficient*.

As is common with confidence interval construction, a value of $\alpha = 0.05$ is chosen for general use, leading to a 95% interval, while a value of

$\alpha = 0.01$, which leads to a longer 99% interval, is chosen for increased confidence that the interval contains μ. Occasionally $\alpha = 0.1$ is used.

Note that a $100(1-\alpha)\%$ confidence interval for μ contains those values that would not be rejected by a two-sided test with significance level α.

Examples (i) The height of an adult male is normally distributed with standard deviation 2.56 inches, and a random sample of 120 men yields a mean height of 68.21 inches. For a 95% confidence interval for μ, the mean height of men in the population from which the sample is drawn, we need to set $c = 1.9600$. Then the confidence limits are found to be

$$\overline{x} \mp c\sigma/\sqrt{n} = 68.21 \mp 1.9600(2.56)/\sqrt{120} = 68.21 \mp 0/46.$$

Hence the 95% confidence interval for μ is $[67.75, 68, 67]$.

For a confidence coefficient of 99%, $c = 2.5758$, which leads to the interval $[67.61, 68.81]$.

Not surprisingly, the desire to be more confident about the interval's containing μ results in an increase in its length.

(ii) The length of a confidence interval is equal to $2c\sigma/\sqrt{n}$, which does not depend on $\overline{x}$. Consequently, it is possible to determine the sample size required to obtain an interval with a maximum specified length. To take an example, if $\sigma = 3$ and we desire a 90% confidence interval for μ with length at most 1.3, then $c = 1.6449$ and we need to satisfy the inequality

$$2c\sigma/\sqrt{n} = 2(1.6449)(3)/\sqrt{n} \leq 1.3,$$

which leads to $n \geq 57.6$. Since n must in practice be an integer, we need to take a sample of size at least 58.

Variance unknown More often than not, the variance σ^2 of X is not known. This in no way nullifies the truth of the statement made in (1), but the random variable Z can no longer be used to construct an interval estimate of μ since it depends on σ. The obvious solution to this problem is to replace σ by an estimate based on the sample. The estimate usually adopted is the sample standard deviation defined by

$$s = \left[\sum(x_i - \overline{x})^2/(n-1)\right]^{\frac{1}{2}}.$$

Replacing σ by s in (1) leads to the new statistic

$$T = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}. \tag{2}$$

In repeated sampling, T no longer follows a normal distribution, in view of its dependence not only on $\overline{X}$ but also on s, which itself varies in value from one sample to another. The distribution of T was derived in 1908 by W.S.Gosset writing under the pseudonym Student. Consequently T is said to follow the Student t distribution (or t distribution for short). There is, however, not a single t distribution but a family of them, indexed by the number of degrees of freedom, which is the number of linearly independent components of s. The symbol $t_{(\nu)}$ is used to denote the t distribution with ν degrees of freedom, and the distribution of T required in (2) is the t distribution with $(n-1)$ degrees of freedom, where n in the sample size.

The probability density function of $t_{(\nu)}$ is similar in shape to that of the standard normal distribution in that it is symmetric about zero and bell-shaped. It is, however, somewhat "fatter" in the tails, and, since inferences about μ depend on values from the tails of a distribution, it is important that the correct distribution is used, at least for small sample sizes. In fact, as $nu \to \infty$, the distribution of $t_{(\nu)}$ converges to that of $N(0,1)$.

A $100(1-\alpha)\%$ confidence interval for μ can be derived in a similar manner to before by finding, this time from the t distribution with $(n-1)$ degrees of freedom, the point c such that $\frac{\alpha}{2} = P(T < -c) = P(T > c)$. Then, from (2), we have

$$1 - \alpha = P\left(-c \leq \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \leq c\right) = P(\overline{X} - cs/\sqrt{n} \leq \mu \leq \overline{X} + cs/\sqrt{n}).$$

For an observed sample, the interval $[\overline{x} - c\sigma/\sqrt{n} \leq \mu \leq \overline{x} + c\sigma/\sqrt{n}]$ is a $100(1-\alpha)\%$ confidence interval for μ. Note that its length, $2cs/\sqrt{n}$, is not a constant as before, but varies from sample to sample in view of its dependence on s.

Example In 1928, the LNER ran the locomotive Lemberg with an experimental boiler pressure of 220lb in five trial runs and measured the coal consumption in lb per draw-bar horse-power hour. The results were as follows:

$$3.27,\ 3.17,\ 3.24,\ 2.92,\ 2.99.$$

Denoting the ith observation by x_i, we have,

$$n = 5, \sum x_i = 15.59 \text{ and } \sum x_i^2 = 48.7059,$$

so that $\overline{x} = 15.59/5 = 3.118$ and $s = (48.7059 - 15 - 59^2/5)/4 = 0.02407$. For a 95% confidence interval for μ, the mean coal consumption of the locomotive, we find, from the table of the $t_{(4)}$ distribution, that $c = 2.776$. Hence the confidence limits are

$$\overline{x} \mp cs/\sqrt{n} = 3.118 \mp 2.776\sqrt{(0.0240/5)} = 3.118 \mp 0/193$$

and the 95% confidence interval for μ is $[2.925, 3.311]$.