SECTION 4 : Point and Interval Estimation

In the previous sections we have discussed how to derive a **point estimator** of a parameter – i.e. a single number, $\hat{\theta}$, which we associate with the *true* (but unknown) value of some parameter, θ. We derived $\hat{\theta}$ by applying e.g. the principle of maximum likelihood or the principle of least squares. In order to assess the likely *range* of true values of of θ, we can derive the dispersion, $\sigma_{\hat{\theta}}$, of the estimator $\hat{\theta}$ (equal to the square root of the variance of the estimator) – or more generally the *covariance* in the multivariate case, where we are simultaneously estimating several parameters and our estimates may not be independent. Thus, we can adopt

$$\hat{\theta} \quad \pm \quad \sigma_{\hat{\theta}}$$

as a suitable measure of the range of likely true values of θ.

We can approach the problem in a complementary fashion, by deriving an **interval estimate** for θ.

4.1 : Defining Confidence Intervals

We illustrate the construction of confidence intervals for a specific example. Consider an iid random sample of size n from the normal distribution $N(\mu, \sigma^2)$, and suppose that the dispersion, σ is known a priori. If we define the variable

$$z \quad = \quad \frac{\sqrt{n}\,(\hat{\mu} - \mu)}{\sigma}$$

then it follows that $z \sim N(0,1)$. We therefore know that

$$\text{Prob}\,[-1.96 \leq z \leq 1.96] \quad = \quad 0.95$$

After some algebra one may easily show that this probability statement is precisely equivalent to

$$\text{Prob}\left[\hat{\mu} - 1.96\frac{\sigma}{\sqrt{n}} \;\leq\; \mu \;\leq\; \hat{\mu} + 1.96\frac{\sigma}{\sqrt{n}}\right] \quad = \quad 0.95$$

We call the interval

$$\left[\hat{\mu} - 1.96\frac{\sigma}{\sqrt{n}}\;,\; \hat{\mu} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

a **95% confidence interval** for the true mean, μ. We refer to the two ends of the confidence interval as **95% confidence limits** for μ.

Confidence intervals involving the standard normal pdf, $N(0,1)$, are particularly common. Thus

$$\hat{\mu} \pm 2.58\frac{\sigma}{\sqrt{n}} \qquad \hat{\mu} \pm 2.64\frac{\sigma}{\sqrt{n}} \qquad \hat{\mu} \pm \frac{\sigma}{\sqrt{n}}$$

are 99%, 90% and 68% confidence intervals for μ respectively.

4.2 : Interpreting Confidence Intervals

We must be careful in interpreting the meaning of probability statements concerning confidence intervals.

In the above example of sampling from $N(\mu, \sigma^2)$, for a *given* sample of actual values from the pdf $\hat{\mu}$ is a unique number, so the probability that the true mean value, μ, lies within the chosen confidence interval is either zero or unity.

The meaning of a confidence interval requires one to think in terms of repeating the process of random sampling from the pdf a large number of times – each time obtaining a different value of of the sample mean, and hence different confidence limits for μ, for the *same* fixed (but unknown) true mean, μ. The probability statement

$$\text{Prob}\left[\hat{\mu} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \hat{\mu} + 1.96\frac{\sigma}{\sqrt{n}}\right] \quad = \quad 0.95$$

means that we would expect μ to lie within the confidence limits in 95% of the large number of different samples. We are, thus, **95% confident** that μ lies within in the interval that we obtain with our actual, observed, value of $\hat{\mu}$.

4.3 : Shortest Confidence Intervals

Finally, note that confidence intervals of a given percentage level are *not* unique. One can prove, however, that the **shortest** confidence interval for the mean of a normal pdf corresponds to the case where the sample mean is taken to lie precisely in the centre of the confidence interval.

Thus, the shortest $100(1-\alpha)\%$ confidence interval for μ in the above example, is

$$\hat{\mu} \quad \pm \quad z_\alpha \frac{\sigma}{\sqrt{n}}$$

where z_α satisfies

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_\alpha} e^{-t^2/2} dt \quad = \quad \frac{\alpha}{2}$$