ESTIMATION:

7.4 POINT ESTIMATION

Definition 7.4.1 Let $X_1, ..., X_n$ be a random sample from a distribution with mass/density function f_X that depends on a (possibly vector) parameter θ. Then $f_{X_1}(x_1) = f_X(x_1; \theta)$, so that

$$f_{X_1,...,X_k}(x_1, ..., x_k) = \prod_{i=1}^{k} f_X(x_i; \theta).$$

A statistic $T = t(X_1, ..., X_n)$ that is used to represent or estimate a function $\tau(\theta)$ of θ based on an observed sample of the random variables $x_1, ..., x_n$ is an **estimator**, and $t = t(x_1, ..., x_n)$ is an **estimate**, $\hat{\tau}(\theta)$, of $\tau(\theta)$. The estimator $T = t(X_1, \ldots, X_n)$ is said to be **unbiased** if $E(T) = \theta$, otherwise T is **biased**.

7.4.1 ESTIMATION TECHNIQUES I: METHOD OF MOMENTS

Suppose that $X_1, ..., X_n$ is a random sample from a probability distribution with mass/density function f_X that depends on a vector parameter θ of dimension k, and suppose that a sample $x_1, ..., x_n$ has been observed. Let the rth moment of f_X be denoted μ_r, and let the rth sample moment, denoted m_r be defined for $r = 1, 2, \ldots$ by

$$m_r = \frac{1}{n} \sum_{i=1}^{n} x_i^r.$$

Then m_r is an **estimate** of μ_r, and

$$M_j = \frac{1}{n} \sum_{i=1}^{n} X_i^r$$

is an **estimator** of μ_r.

PROCEDURE : The method of moments technique of estimation involves matching the theoretical moments $\mu_r \equiv \mu_r(\theta)$ to the sample moments $m_r, r = 1, 2, \ldots, l$, for suitable l, and solving for θ. In most situations taking $l = k$, the dimension of θ, suffices: we obtain k equations in the k elements of vector θ which may be solved simultaneously to find the parameter estimates.

We may, however, need $l > k$. Intuitively, and recalling the **Weak Law of Large Numbers**, it is reasonable to suppose that there is a close relationship between the theoretical properties of a probability distribution and estimates derived from a large sample. For example, we know that, for large n, the sample mean converges in probability to the theoretical expectation.

7.4.2 ESTIMATION TECHNIQUES II: MAXIMUM LIKELIHOOD

Definition 7.4.2 Let random variables $X_1, ..., X_n$ have joint mass or density function, denoted $f_{X_1,...,X_k}$, that depends on a vector parameter $\theta = (\theta_1, ..., \theta_k)$. Then the joint/mass density function considered as a function of θ for the (fixed) observed values $x_1, ..., x_n$ of the variables is the **likelihood function**, $L(\theta)$:

$$L(\theta) = f_{X_1,...,X_n}(x_1, ..., x_n; \theta).$$

If $X_1, ..., X_n$ represents a random sample from joint/mass density function f_X

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i; \theta).$$

Definition 7.4.3 Let $L(\theta)$ be the likelihood function derived from the joint/mass density function of random variables $X_1, ..., X_n$, where $\theta \in \Theta \subseteq \mathbb{R}^k$, say, and Θ is termed the **parameter space**. Then for a fixed set of observed values $x_1, ..., x_n$ of the variables, the estimate of θ termed the **maximum likelihood estimate** (MLE) of θ, $\widehat{\theta}$, is defined by

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} L(\theta).$$

That is, the maximum likelihood estimate is the value of θ for which $L(\theta)$ is maximized in the parameter space Θ.
DISCUSSION : The method of estimation involves finding the value of θ for which $L(\theta)$ is maximized. This is generally done by setting the first partial derivatives of $L(\theta)$ with respect to θ_j equal to zero, for $j = 1, ..., k$, and solving the resulting k simultaneous equations. But we must be alert to cases where the likelihood function $L(\theta)$ is not differentiable, or where the maximum occurs on the boundary of Θ! Typically, it is easier to obtain the MLE by maximising the (natural) logarithm of $L(\theta)$: we maximise $l(\theta) = \log L(\theta)$, the **log-likelihood**.

THE FOUR STEP ESTIMATION PROCEDURE: Suppose a sample $x_1, ..., x_n$ has been obtained from a probability model specified by mass or density function $f_X(x; \theta)$ depending on parameter(s) θ lying in parameter space Θ. The maximum likelihood estimate is produced as follows;

1. Write down the likelihood function, $L(\theta)$.
2. Take the natural log of the likelihood, collect terms involving θ.
3. Find the value of $\theta \in \Theta$, $\widehat{\theta}$, for which $\log L(\theta)$ is maximized, for example by differentiation. Note that, if parameter space Θ is a bounded interval, then the maximum likelihood estimate may lie on the boundary of Θ. If the parameter is a k vector, the maximization involves evaluation of partial derivatives.

4. Check that the estimate $\widehat{\theta}$ obtained in STEP 3 truly corresponds to a maximum in the (log) likelihood function by inspecting the second derivative of $\log L(\theta)$ with respect to θ. In the single parameter case, if the second derivative of the log -likelihood is negative at $\theta = \widehat{\theta}$, then $\widehat{\theta}$ is confirmed as the MLE of θ (other techniques may be used to verify that the likelihood is maximized at $\hat{\theta}$).

7.5 INTERVAL ESTIMATION

The techniques of the previous section provide just a single point estimate of the value of an unknown parameter. Instead, we can attempt to provide a set or interval of values which expresses our uncertainty over the unknown value.

Definition 7.5.1 Let $X = (X_1, \ldots, X_n)$ be a random sample from a distribution depending on an unknown scalar parameter θ. Let $T_1 = l_1(X_1, \ldots, X_n)$ and $T_2 = l_2(X_1, \ldots, X_n)$ be two statistics satisfying $T_1 \leq T_2$ for which $P(T_1 < \theta < T_2) = 1 - \alpha$, where P denotes probability when X has distribution specified by parameter value θ, whatever the true value of θ, and where $1 - \alpha$ does not depend on θ. Then the random interval (T_1, T_2) is called a $1 - \alpha$ **confidence interval** for θ, and T_1 and T_2 are called lower and upper confidence limits, respectively.

Given data, $x = (x_1, \ldots, x_n)$, the realised value of X, we calculate the interval (t_1, t_2), where $t_1 = l_1(x_1, \ldots, x_n)$ and $t_2 = l_2(x_1, \ldots, x_n)$. **NOTE:** θ here is fixed (it has some true, fixed but unknown value in Θ), and it is T_1, T_2 that are random. The calculated interval (t_1, t_2) either does, or does not, contain the true value of θ. Under repeated sampling of X, the random interval (T_1, T_2) contains the true value a proportion $1 - \alpha$ of the time. Typically, we will take $\alpha = 0.05$ or $\alpha = 0.01$, corresponding to a 95% or 99% confidence interval.

If θ is a vector, then we use a **confidence set**, such as a sphere or ellipse, instead of an interval. So, $C(X)$ is a $1 - \alpha$ confidence set for θ if $P(\theta \in C(X)) = 1 - \alpha$, for all possible θ.

The key problem is to develop methods for constructing a confidence interval. There are two general procedures.

7.5.1 PIVOTAL QUANTITY

Definition 7.5.2 Let $X = (X_1, \ldots, X_n)$ be a random sample from a distribution specified by (scalar) parameter θ. Let $Q = q(X; \theta)$ be a function of X and θ. If Q has a distribution that does not depend on θ, then Q is defined to be a **pivotal quantity**.

If $Q = q(X; \theta)$ is a pivotal quantity and has a probability density function, then for any fixed $0 < \alpha < 1$, there will exist q_1 and q_2 depending on α such that $P(q_1 < Q < q_2) = 1 - \alpha$. Then, if for each possible sample value $x = (x_1, \ldots, x_n)$, $q_1 < q(x_1, \ldots, x_n; \theta) < q_2$ if and only if $l_1(x_1, \ldots, x_n) < \theta < l_2(X_1, \ldots, x_n)$, for functions l_1 and l_2, not depending on θ, then (T_1, T_2) is a $1 - \alpha$ confidence interval for θ, where $T_i = l_i(X_1, \ldots, X_n), i = 1, 2$.

7.5.2 INVERTING A TEST STATISTIC

A second method utilises a correspondence between hypothesis testing and interval estimation.

Definition 7.5.3 For each possible value θ_0, let $A(\theta_0)$ be the acceptance region of a test of $H_0 : \theta = \theta_0$, of significance level α, so that $A(\theta_0)$ is the set of data values x such that H_0 is accepted in a test of significance level α. For each x, define a set $C(x)$ by

$$C(x) = \{\theta_0 : x \in A(\theta_0)\}.$$

Then the random set $C(X)$ is a $1 - \alpha$ confidence set.

Example. Let $X = (X_1, \ldots, X_n)$, with the X_i IID $N(\mu, \sigma^2)$, with σ^2 known. Both the above procedures yield a $1 - \alpha$ confidence interval for μ of the form $(\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{\alpha/2}\sigma/\sqrt{n})$, where $\Phi(z_{\alpha/2}) = 1 - \alpha/2$.

We know that

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1),$$

and so is a pivotal quantity. We have

$$P(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha.$$

Rearranging gives

$$P(\bar{X} - z_{\alpha/2}\sigma/\sqrt{n} < \mu < \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}) = 1 - \alpha,$$

which defines a $1 - \alpha$ confidence interval.

If we test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ a reasonable test has rejection region of the form $\{x : |\bar{x} - \mu_0| > z_{\alpha/2}\sigma/\sqrt{n}\}$. So, H_0 is *accepted* for sample points with $|\bar{x} - \mu_0| < z_{\alpha/2}\sigma/\sqrt{n}$, or,

$$\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} < \mu_0 < \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}.$$

The test is constructed to have size (significance level) α, so $P(H_0 \text{ is accepted} | \mu = \mu_0) = 1 - \alpha$. So we can write

$$P(\bar{X} - z_{\alpha/2}\sigma/\sqrt{n} < \mu_0 < \bar{X} + z_{\alpha/2}\sigma/\sqrt{n} \mid \mu = \mu_0) = 1 - \alpha.$$

This is true *for every* μ_0, so

$$P(\bar{X} - z_{\alpha/2}\sigma/\sqrt{n} < \mu < \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}) = 1 - \alpha$$

is true for every μ, so the confidence interval obtained by inverting the test is the same as that derived by the pivotal quantity method.