PROPERTIES OF ESTIMATION AND METHODS OF ESTIMATION (Chapter 9)

We've already looked at several properties of point estimators in Chapter 8, for example bias, variance and MSE. Let's now discuss some others.

Efficiency

Suppose that *A* and *B* are two unbiased estimators of a parameter θ .

If $VarA < VarB$, then we say that *A* is **more efficient** than *B*.

The **efficiency** of A relative to B is defined as $Eff\,(A,B) = \dfrac{VarB}{VarA}$.

We say that *A* is *more efficient* than *B* if $Eff(A,B) > 1$,
or equivalently, if $VarA < VarB$.

NB: If either *A* or *B* is biased, the efficiency of *A* relative to *B* is not defined.

Example 1 Two numbers *X* and *Y* are to be randomly and independently chosen from between 0 and *c*.

Consider $U = X + Y$ and $W = 1.5\max(X,Y)$ as two estimators of *c*.

Find the efficiency of *U* relative to *W*.

In Example 2 of Chapter 8, we showed that *U* and *W* are both unbiased for *c*.

We also showed that $VarU = c^2/6$ and $VarW = c^2/8$.

Therefore $Eff\,(U,W) = \dfrac{VarW}{VarU} = \dfrac{c^2/8}{c^2/6} = 0.75$.

(Thus *W* is more efficient than *U*. *W*'s variance is only 3/4 the variance of *U*.)

We will next talk about a property of estimators called *consistency*.
But before this, a new concept needs to be introduced.

Convergence in probability

Suppose that $X = X_n$ is a random variable and k is a constant such that, for any $\varepsilon > 0$:

$$P(|X-k| > \varepsilon) \to 0 \text{ as } n \to \infty .$$

Then we say that X **converges in probability** to k,
and write $X \xrightarrow{p} k$.

Example 2 Suppose that $X \sim Expo(1/n)$.

Show that $X \xrightarrow{p} 0$.

$$P(|X-0| > \varepsilon) = P(X > \varepsilon) = \int_{\varepsilon}^{\infty} ne^{-nx} dx = e^{-n\varepsilon} \to 0 .$$

Therefore $X \xrightarrow{p} 0$.

(This makes sense, since $EX = 1/n \to 0$ and $VarX = 1/n^2 \to 0$.)

Consistency

Consider an estimator A of θ based on a sample of size n, and suppose that

$$A \xrightarrow{p} \theta \text{ as } n \to \infty .$$

Then we say that A is a **consistent** estimator of θ .

Example 3 Consider a random sample of n numbers from between 0 and c,
and let $U = 2\bar{Y}$ (twice the sample mean).

Is U a consistent estimator of c?

Observe that: $\mu = EU = 2E\bar{Y} = 2(c/2) = c$ (U is unbiased for c)

$$\sigma^2 = VarU = 2^2 Var\bar{Y} = 4\frac{VarY_1}{n} = 4\frac{(c-0)^2/12}{n} = \frac{c^2}{3n}.$$

So $P(|U-c|>\varepsilon) = P(|U-\mu|>k\sigma)$ where $k = \varepsilon/\sigma$

$= P(|U-\mu| \geq k\sigma)$ since U is a continuous random variable

$\leq \frac{1}{k^2}$ by Chebyshev's theorem

$= \frac{1}{(\varepsilon/\sigma)^2} = \frac{\sigma^2}{\varepsilon^2} = \frac{c^2}{3n\varepsilon^2} \to 0$ as $n \to \infty$.

Thus $U \xrightarrow{p} c$.

So yes, U is a consistent estimator of c.

The logic used here can be generalised, as follows.

Theorem Suppose that A is an unbiased estimator of θ such that $VarA \to 0$ as $n \to \infty$.

Then A is also a consistent estimator of θ.

The proof of this theorem is very similar to the working in Example 3:

$$\begin{aligned}
P(|A-\theta|>\varepsilon) &= P(|A-\theta|>k\sigma) \text{ where } k=\varepsilon/\sigma \text{ and } \sigma^2 = VarA \\
&\leq P(|A-\theta| \geq k\sigma) \\
&\leq 1/k^2 \quad \text{by Chebyshev's theorem} \\
&= \sigma^2/\varepsilon^2 \\
&\to 0 \quad \text{as } n\to\infty \quad \text{since } \sigma^2 \to 0.
\end{aligned}$$

So $A \xrightarrow{p} \theta$, or in other words, A is a consistent estimator of θ.

(Also see Theorem 9.1 in the text.)

The most important implication of the above theorem is the following result, which is called the *law of large numbers* (or more precisely, the *weak law of large numbers*).

The (weak) law of large numbers

Consider a random sample $Y_1,\ldots,Y_n$ from a distribution with finite mean μ and finite variance σ^2. Then the sample mean $\bar{Y}$ is a consistent estimator of μ.

Proof: $E\bar{Y} = \mu$. Thus $\bar{Y}$ is an unbiased estimator of μ.

Also, $Var\bar{Y} = \sigma^2 / n \to 0$ as $n \to \infty$.

It follows by the above theorem that $\bar{Y}$ is a consistent estimator of μ.

Example 4 Consider a bent coin, and suppose that we are interested in *p*, the probability of a head coming up on a single toss.

We toss the coin *n* times and observe $\hat{p}$, the proportion of heads.

Is $\hat{p}$ a consistent estimator of *p*?

We may also write $\hat{p}$ as $\bar{Y} = (1/n)(Y_1 + \ldots + Y_n)$,
where $Y_1,\ldots,Y_n \sim iid\ Bern(p)$.

Here: $\mu = EY_i = p$.
$\sigma^2 = VarY_i = p(1-p) < \infty$.

So by the law of large numbers, $\hat{p}$ is a consistent estimator of *p*.

That is, $\hat{p} \xrightarrow{p} p$, or equivalently, $P(|\hat{p} - p| > \varepsilon) \to 0$ as $n \to \infty$.

For example, if the coin is fair then $P(|\hat{p} - 0.5| > 0.01) \to 0$.

This is the same as saying $P(|\hat{p} - 0.5| \le 0.01) \to 1$.

That is, the probability that the proportion of heads will lie between 0.49 and 0.51 approaches 100% as the number of tosses increases indefinitely.

More convergence theory

Theorem

Suppose that $A \xrightarrow{p} a$ and $B \xrightarrow{p} b$ as $n \to \infty$, where a and b are constants.

Then: **(i)** $A + B \xrightarrow{p} a + b$

(ii) $AB \xrightarrow{p} ab$

(iii) $A / B \xrightarrow{p} a / b$, provided that $b \neq 0$

(iv) $g(A) \xrightarrow{p} g(a)$, provided that g is a real-valued function that is continuous at a.

Definition

Suppose that $X = X_n$ and R are random variables such that

$F_X(k) \to F_R(k)$ as $n \to \infty$ for all $k \in \Re$ at which $F_R(k)$ is continuous.

Then we say that X **converges in distribution** to R,
and write $X \xrightarrow{d} R$.

For example, $U = \dfrac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} Z \sim N(0,1)$ (central limit theorem).

Note that the above definition also applies if R has a degenerate distribution at some constant c. In that case,

$$F_R(k) = P(R \leq k) = \begin{cases} 0, & k < c \\ 1, & k \geq c. \end{cases}$$

Thus, for a constant c, $X \xrightarrow{d} c$ if

$$P(X \leq k) \to \begin{cases} 0, & k < c \\ 1, & k > c. \end{cases}$$

Three more theorems

1. Suppose that $A \xrightarrow{d} N(0,1)$ and $B \xrightarrow{p} 1$. Then $A/B \xrightarrow{d} N(0,1)$.

2. If A and B are random variables such that $A \xrightarrow{p} B$, then $A \xrightarrow{d} B$.

3. If A is a random variable and c is a constant such that $A \xrightarrow{d} c$, then $A \xrightarrow{p} c$.
(NB: This is not generally true if c is a non-degenerate random variable.)

The above theorems can be used to prove various important results, such as:

$$S^2 \xrightarrow{p} \sigma^2 \text{ as } n \to \infty$$

(i.e., the sample variance is a consistent estimator of the population variance; see Example 9.3 in text)

$$\frac{\bar{Y} - \mu}{S / \sqrt{n}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty$$

(i.e., the central limit theorem still holds when we replace the population standard deviation σ by the sample standard deviation S; see Example 9.4 in text; we used this fact in Chapter 8 to construct large sample confidence intervals for μ and p).

Methods of estimation

We will now look at two general methods for deriving point estimators:

1. the *method of moments*
2. the *method of maximum likelihood*.

1. The method of moments

Consider a random sample $Y_1, \ldots, Y_n$ from some distribution. (Thus these rv's are iid.)
Recall that $Y = Y_1$ has ***k*th raw moment** $\mu_k' = EY^k$.
We now also define the ***k*th sample moment** as $m_k' = \frac{1}{n}\sum_{i=1}^{n} y_i^k$.

Suppose that the distribution of Y involves t unknown parameters (t = 1,2,3,...).
Then the **method of moments (MOM)** involves equating:

$$\mu_1' = m_1'$$
$$\mu_2' = m_2'$$
$$\ldots\ldots\ldots\ldots$$
$$\mu_t' = m_t'.$$

The solution of these t equations leads to the **method of moments estimates (MOME's)** of the t unknown parameters.

The idea here is that the kth sample moment as a random variable, $M_k' = \frac{1}{n}\sum_{i=1}^{n} Y_i^k$, has mean μ_k' and a variance that converges to zero as n tends to infinity:

$$EM_k' = \frac{1}{n}\sum_{i=1}^{n} EY_i^k = \frac{1}{n}\sum_{i=1}^{n} \mu_k' = \mu_k'$$

$$VM_k' = \frac{1}{n^2}\sum_{i=1}^{n} VY_i^k = \frac{nVY_1^k}{n^2} = \frac{VY_1^k}{n} \to 0.$$

Thus, by the law of large numbers, M_k' is a consistent estimator of μ_k'.
That is, for each k and any $\varepsilon > 0$, $P(|M_k' - \mu_k'| > \varepsilon) \to 0$ as $n \to \infty$.
So if n is large, each m_k' should be close to μ_k'.

Note: In some cases the above definition of the MOM may require slight modification because the system of equations as indicated cannot be solved for the t unknown parameters. For example, if $\mu_1' = 0$, it may be necessary to also equate $\mu_{t+1}' = m_{t+1}'$.

Example 4 Consider a random sample of numbers from between 0 and *c*.
Find the method of moments estimate of *c*.

Here: $t = 1$, $\mu_1' = EY = c/2$, $m_1' = \bar{y}$. We now equate $\mu_1' = m_1'$.
This implies that $c/2 = \bar{y}$, and hence $c = 2\bar{y}$.
Thus the method of moments estimate of *c* is $\hat{c} = 2\bar{y}$.

Discussion

Thus also, the method of moments *estimator* of *c* is $\hat{c} = 2\bar{Y}$. This estimator is both unbiased and consistent; recall Example 3. We could also say that the *estimate* $\hat{c} = 2\bar{y}$ is unbiased and consistent, although technically this should not be said of estimates, only of estimators. But it is common for these two terms to be used interchangeably. Thus, when we say that an *estimate* is unbiased or consistent, it should be understood that we are saying this about the corresponding *estimator*.

Example 5 Suppose that $Y_1, \ldots, Y_n \sim iid\ Gam(a,b)$.
Find the method of moments estimates of *a* and *b*.

Here: $t = 2$, $\mu_1' = EY = ab$, $\mu_2' = EY^2 = VarY + (EY)^2 = ab^2 + (ab)^2$

$m_1' = \bar{y}$, $m_2' = \dfrac{1}{n}\sum_{i=1}^{n} y_i^2$.

We now equate $\mu_1' = m_1'$ and $\mu_2' = m_2'$. This implies that $ab = \bar{y}$ and $ab^2 + a^2b^2 = m_2'$. Solving these equations leads to the MOME's:

$$\hat{a} = \frac{\bar{y}^2}{m_2' - \bar{y}^2} \text{ and } \hat{b} = \frac{\bar{y}}{\hat{a}}.$$

For example, suppose that the data values are 1.3 and 2.7.
Then $\bar{y} = (1/2)(1.3 + 2.7) = 2$ and $m_2' = (1/2)(1.3^2 + 2.7^2) = 4.49$.

Therefore $\hat{a} = \dfrac{2^2}{4.49 - 2^2} = 8.1633$ and $\hat{b} = \dfrac{2}{8.1633} = 0.245$.

Exercise: $Y_1, \ldots, Y_n \sim iid\ N(a, b^2)$.
Find the MOME's of *a* and b^2.

$\mu_1' = a$, $\mu_2' = EY^2 = a^2 + b^2$, $m_1' = \bar{y}$, $m_2' = (1/n)\sum_{i=1}^{n} y_i^2$.

Equating $\mu_t' = m_t'$ for *t* = 1,2, we get: $\hat{a} = \bar{y}$

$$\hat{b}^2 = \left(\frac{n-1}{n}\right)s^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2.$$