ESTIMATION (Chapter 8)

Statistical inference can be broken down into two broad categories:
estimation (Chapters 8 and 9) and *hypothesis testing* (Chapter 10).

There are basically two types of estimation:
point estimation and *interval estimation*.

Point estimation

Example 1 We have a bent coin and are interested in p,
the probability of heads coming up on a single toss.
How can we estimate p?

First we need some *data* (observable random variable or variables).

Eg, we toss the coin n times and observe the number of heads that comes up.

The data is then that number, and we may call it Y.

(*Note*: Technically, "data" is a plural term used to represent at least two numbers, and its singular form is "datum". So we should really call Y the *datum*. However, it is conventional to use "data" for both singular and plural cases, as reflected here.)

We next need a *model* for the data, eg $Y \sim \text{Bin}(n,p)$.

Here, p may be called the *target parameter* or *estimand*.

We now need to choose an *estimator* of p.

This may be any *statistic*.

(*Note*: Recall that a statistic is defined to be a function of the observable random variables in a sample and known constants, in this case a function of Y and n.)

Eg, let the estimator be $X = Y/n$ (the proportion of tosses which result in a head).

Finally we need to actually carry out the experiment and do the calculations.

For example, we toss the coin $n = 10$ times and get 6 heads.
Then, the realised value of the data Y is $y = 6$,
and the realised value of our estimator X is $x = y/n = 6/10 = 0.6$.

We call x an *estimate* of p.
Because x is a single number, we may also call it a *point estimate* of p.
Likewise we may call X a *point estimator* of p.

(*Note*: Technically, "estimator" refers to a random variable and "estimate" to a constant. However, it is conventional to use these two terms interchangeably and to depend on the context to make it clear exactly what is meant by the term used.)

A common practice is to denote both the estimator and estimate of a parameter θ by $\hat{\theta}$. Thus in our example:

(a) the *estimator* of p is $\hat{p} = X = Y/n$ (a random variable)

(b) the *estimate* of p is $\hat{p} = x = y/n = 0.6$ (a constant).

(*Note:* This may be a bit confusing because the same symbol is used for a random variable and a constant. But usually the symbol's meaning is clear from the context.)

The question now arises: How *good* is $\hat{p}$ as an estimator of p?
What we need are some criteria for assessing the quality of estimators.

The *bias* of an estimator $\hat{\theta}$ of θ is
$$B(\hat{\theta}) = E\hat{\theta} - \theta .$$

If $B(\hat{\theta}) = 0$, we say that $\hat{\theta}$ is *unbiased* for θ.

In Example 1, what is the bias of $\hat{p}$?

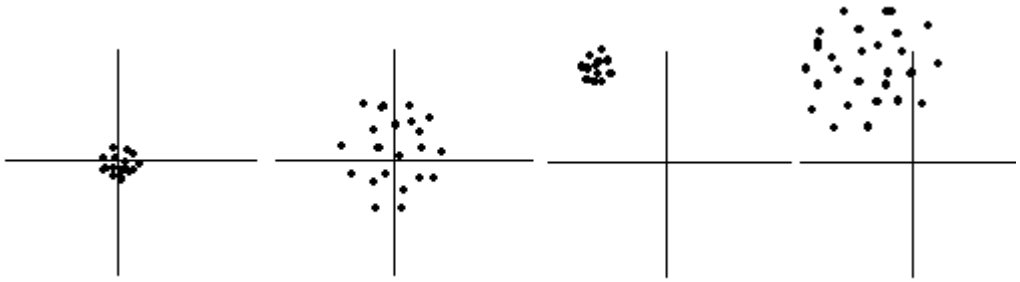
$$E\hat{p} = E\left(\frac{Y}{n}\right) = \frac{1}{n}EY = \frac{1}{n}np = p .$$

So $B(\hat{p}) = E\hat{p} - p = p - p = 0$. (Thus $\hat{p}$ is unbiased for p.)

This tells us that $\hat{p}$ is an *accurate* estimator.

But is $\hat{p}$ also *precise*?

Analogy: Target at a firing range:



accurate and precise	accurate but imprecise	innacurate but precise	innacurate and imprecise

One measure of precision is *variance* (the smaller the better).

In our example, $Var\hat{p} = Var\left(\frac{Y}{n}\right) = \frac{1}{n^2} VarY = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$.

Another measure of an estimator's quality is the mean square error.

The *mean square error (MSE)* of an estimator $\hat{\theta}$ of θ is
$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$.

It turns out that $MSE(\hat{\theta}) = Var\hat{\theta} + B(\hat{\theta})^2$. (Prove this as an exercise.)

Thus the MSE is a combined measure of accuracy and precision.
(It will be small only if both the mean and variance are small.)

In our example, $MSE(\hat{p}) = Var\hat{p} + B(\hat{p})^2 = \frac{p(1-p)}{n} + 0^2 = \frac{p(1-p)}{n}$.

(*Note*: The MSE and variance are the same for all unbiased estimators.)

In our example, $\hat{p} = Y/n$ seemed like an 'obvious' estimator of p.
But in many situations there will be several plausible estimators to consider.
One must then decide which one is the 'best'.
This will typically be done by comparing the bias, variance and MSE
of the candidate estimators.

Example 2 Two numbers are randomly chosen between 0 and c.
They are $x = 3.6$ and $y = 5.4$.

Consider the two following estimates of c:

$u = x + y = 3.6 + 5.4 = 9.0$

$v = \max(x,y) = \max(3.6,5.4) = 5.4$.

Which estimate should we choose, and why?

(Why focus on these two estimates? This question will be answered in Chapter 9.)

Let $X, Y \sim$ iid U(0,c).
Then the two estimators we wish to compare are:

(a) $U = X + Y$

(b) $V = \max(X,Y)$.

Let's now find the mean, bias, variance and MSE of each estimator.

(a) $EU = E(X + Y) = EX + EY = c/2 + c/2 = c$.
So $B(U) = EU - c = c - c = 0$. (So U is unbiased for c.)

$VarU = Var(X + Y) = VarX + VarY = c^2/12 + c^2/12 = c^2/6$.
$MSE(U) = VarU = c^2/6$.

(b) What is the distribution of V? Note that V is an order statistic.

$$F(v) = P(V < v) = P(X < v, Y < v) = P(X < v)P(Y < v) = (v/c)^2, \quad 0 < v < c.$$
$$f(v) = F'(v) = 2v/c^2, \quad 0 < v < c.$$

So: $$EV = \int_0^c v\frac{2v}{c^2}dv = \frac{2c}{3}$$

$$B(V) = EV - c = \frac{2c}{3} - c = -\frac{c}{3} \qquad (V \text{ is biased})$$

$$EV^2 = \int_0^c v^2\frac{2v}{c^2}dv = \frac{c^2}{2}, \qquad VarV = \frac{c^2}{2} - \left(\frac{2c}{3}\right)^2 = \frac{c^2}{18}$$

$$MSE(V) = VarV + B(V)^2 = \frac{c^2}{18} + \left(-\frac{c}{3}\right)^2 = \frac{c^2}{6}.$$

Summary

estimator of c	mean	bias	variance	MSE
$U = X + Y$	c	0	$c^2/6$	$c^2/6$
$V = \max(X,Y)$	$2c/3$	$-c/3$	$c^2/18$	$c^2/6$

We see that U is:
- more accurate than V (since $|0| < |-c/3|$)
- less precise than V (since $c^2/6 > c^2/18$)
- overall about as good as V (since $c^2/6 = c^2/6$).

Since U is unbiased and V is not, we choose U as the better of the two estimators.
So we should choose $u = 9.0$ as the better estimate of c.

Does this mean that the estimate $v = 5.4$ is useless? No. We can make use of v as a starting point (or basis) for constructing an estimate that is better than u.

Observe that $EV = 2c/3$. This means that $E(3V/2) = c$.
So we define $W = 3V/2$. This is another (third) estimator of c to be considered.

Then: $EW = c$, $B(W) = EW - c = c - c = 0$ (W is unbiased)

$$VarW = \left(\frac{3}{2}\right)^2 VarV = \frac{9}{4} \times \frac{c^2}{18} = \frac{c^2}{8}$$

$MSE(W) = VarW = c^2/8$.

New summary

estimator of c	mean	bias	variance	MSE
$U = X + Y$	c	0	$c^2/6$	$c^2/6$
$V = \max(X,Y)$	$2c/3$	$-c/3$	$c^2/18$	$c^2/6$
$W = 1.5\max(X,Y)$	c	0	$c^2/8$	$c^2/8$

We see that W is:
- just as accurate as U (both estimators are unbiased)
- more precise than U (since $c^2/8 < c^2/6$)
- overall better than U (since $c^2/8 < c^2/6$).

We conclude that W is the best of the three estimators.
So our best estimate of c is $w = 1.5\max(3.6,5.4) = 8.1$, our second best estimate is $u = 3.6 + 5.4 = 9.0$, and our third choice would be $v = \max(3.6,5.4) = 5.4$.

Note: Ideally, we should choose which estimator to use (*U*, *V* or *W*) *before* observing the data (*X* and *Y*). If we observe $x = 3.6$ and $y = 5.4$ *first*, and *then* make a decision on whether to estimate c by $w = 8.1$, $u = 9.0$ or $v = 5.4$, we could be accused of 'biasing' the results (e.g. subconsciously favouring one of the numbers). However, in practice we see the numbers (data) first and only then think about how to use them. So we should be aware of such biases that might arise and try to be as 'impartial' as possible.

Two important results for estimation

Suppose that: $Y_1, \ldots, Y_n \sim iid\ (\mu, \sigma^2)$
(the Y_i are *iid*, not necessarily normal,
with common mean $EY_i = \mu$, $VarY_i = \sigma^2$)

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad \text{(sample mean)}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (Y_i - \bar{Y})^2 \qquad \text{(sample variance).}$$

Then: **(a)** $E\bar{Y} = \mu$ (population mean)

(b) $ES^2 = \sigma^2$ (population variance).

Proof of (a): $E\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} EY_i = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu$.

Proof of (b): $S^2 = \frac{1}{n-1}\sum_{i=1}^{n} D_i^2$, where $D_i = Y_i - \bar{Y}$.

Now $ED_i^2 = VarD_i$ (since $ED_i = EY_i - E\bar{Y} = \mu - \mu = 0$)

$= VarY_i + Var\bar{Y} - 2Cov(Y_i, \bar{Y})$ (see *Note* 1 below)

$= \sigma^2 + \frac{\sigma^2}{n} - 2 \times \frac{\sigma^2}{n}$

$= \frac{n-1}{n}\sigma^2$.

Therefore $ES^2 = \frac{1}{n-1}\sum_{i=1}^{n} ED_i^2 = \frac{1}{\cancel{n-1}}\sum_{i=1}^{n} \frac{\cancel{n-1}}{n}\sigma^2 = \frac{1}{n} n\sigma^2 = \sigma^2$.

Note 1: $Cov(Y_i, \bar{Y}) = Cov(Y_1, \bar{Y}) = Cov\left(Y_1, \frac{1}{n}(Y_1 + \ldots + Y_n)\right)$

$$= \frac{1}{n}\{Cov(Y_1, Y_1) + Cov(Y_1, Y_2) + \ldots + Cov(Y_1, Y_n)\}$$

$$= \frac{1}{n}\{VarY_1 + 0 + \ldots + 0\} = \frac{\sigma^2}{n}.$$

Note 2: The results (a) and (b) may be true even if the Y_i are not iid. A sufficient condition for (a) and (b) is that the Y_i are *uncorrelated* (having 0 correlation, or equivalently, 0 covariance) with common mean and common variance.

Example 3 A bottling machine dispenses volumes independently with mean μ and variance σ^2.

n = 3 bottles are randomly sampled from the output of the machine, and their volumes are 1.9, 1.4, 1.8 (litres).

Find unbiased estimates of μ and σ^2.

Unbiased estimates are $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{3}(1.9 + 1.4 + 1.8) = 1.7$

and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{3-1}\{(1.9-1.7)^2 + (1.4-1.7)^2 + (1.8-1.7)^2\} = 0.07.$

Equivalently, $s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right) = \frac{1}{3-1}\{[1.9^2 + 1.4^2 + 1.8^2] - 3(1.7)^2\} = 0.07$.

Note: Suppose we sample 3 bottles independently a huge number of times (e.g. 100,000, making a total of 300,000 sampled bottles) and each time compute $\bar{y}$ and s^2. Then the average of the resulting 100,000 $\bar{y}$ values would be very close to μ, and the average of the 100,000 s^2 values would be very close to σ^2. This statement is an expression of the fact that $\bar{y}$ and s^2 are *unbiased* estimates of μ and σ^2.

Suppose we changed n = 3 to a very large number (e.g. 100,000). Then we would feel 'confident' that $\bar{y}$ is 'very close' to μ, and that s^2 is 'very close' to σ^2. This intuitively obvious statement is an expression of the fact that $\bar{y}$ and s^2 are *consistent* estimates of μ and σ^2. The property of consistency will be formally defined later.