Interval estimation and confidence intervals

Consider a sample $Y_1,\ldots,Y_n$ from a probability distribution which depends on one parameter θ (which is fixed but unknown).

Suppose that we can find two functions of $Y_1,\ldots,Y_n$, say

$L = g(Y_1,\ldots,Y_n)$

$U = h(Y_1,\ldots,Y_n)$,

such that

$P(L \le \theta \le U) = 1-\alpha$.

Then [*L*, *U*] is a $100(1-\alpha)$% *confidence interval (CI)* for θ.

We call: *L* the *lower bound (LB)*,
U the *upper bound (UB)*, and
$1-\alpha$ the *coverage coefficient*,
of the CI.

Example 4 Suppose that 6.2 is a number chosen randomly between 0 and *c*. Find an 80% confidence interval for *c*.

Let $Y \sim U(0,c)$. (Since $n = 1$, we write Y instead of Y_1.)

Then $X = Y/c \sim U(0,1)$. (Proof: $f(x) = f(y)\left|\frac{dy}{dx}\right| = \frac{1}{c}|c| = 1$, $0 < x < 1$.)

We call *X* a *pivotal quantity*. (This means that *X* is a function of *Y* and *c* whose distribution does not depend on *c*.)

So $0.8 = P(0.1 \le X \le 0.9)$ (draw a picture of $f(x)$, with the area bewteen 0.1 and 0.9 shaded)

$= P(0.1 \le Y/c \le 0.9)$

$= P(0.1 \le Y/c, Y/c \le 0.9)$

$= P(c \le Y/0.1, Y/0.9 \le c)$

$= P(10Y/9 \le c \le 10Y)$.

So $[10Y/9, 10Y]$ is an 80% CI for *c*.

Here: $100(1-\alpha) = 80 \Rightarrow \alpha = 0.2$ (the *non*-coverage coefficient)

$L = g(Y) = 10Y/9$ (the lower bound)

$U = h(Y) = 10Y$ (the upper bound).

Now, the realised value of Y is $y = 6.2$.
So the realised value of $L = 10Y/9$ is $l = 10y/9 = 10(6.2)/9 = 6.89$,
and the realised value of $U = 10Y$ is $u = 10y = 10(6.2) = 62$.

Therefore an 80% confidence interval is $[l, u] = [10y/9, 10y] = [6.89, 62]$.

Notes

The term 'confidence interval' refers to both $[10Y/9, 10Y]$ and $[10y/9, 10y]$; the former is a random variable and the latter is the realised value of that random variable.

We could also write

$$0.8 = P(0.1 < X < 0.9) = \ldots = P(10Y/9 < c < 10Y);$$

and so another 80% CI for c is $(10Y/9, 10Y)$ (or $(10y/9, 10y)$). But since Y is a continuous random variable and c lies on a 'continuum' of values (from 0 to infinity), it doesn't make much difference whether our final answer is [6.89, 62] or (6.89, 62). However, in some cases, whether or not to include endpoints in a CI can be important.

Another way to derive the CI, one which does not involve a pivotal quantity, is:

$0.8 = P(0.1c < Y < 0.9c)$ (draw a picture of $f(y)$, with the area between $0.1c$ and $0.9c$ shaded)
$= P(c < Y/0.1, Y/0.9 < c)$
$= P(10Y/9 < c < 10Y)$, etc.

As an exercise for further illustration, draw (or imagine) a graph of the functions $y = 0.1c$ and $y = 0.9c$ in the first quadrant of the c-y plane. (These functions may also be expressed as $c = 10y$ and $c = 10y/9$, respectively.)

Then consider the fact that if c is equal to any particular value, call it k, Y lies between $0.1k$ and $0.9k$ with probability 0.8. (Draw the vertical line $c = k$ and mark the points $(k,0.1k)$ and $(k,0.9k)$.) This is true for all $k > 0$.

It is a consequence of this fact that if Y takes on any particular value, call it t, we can feel 80% confident that c lies between $10t/9$ and $10t$. (Draw the horizontal line $y = t$ and mark the points $(10t/9,t)$ and $(10t,t)$.)

In our example, can we say that c lies between 6.79 and 62 with probability 80%?

No. The event "6.89 < c < 62" does not involve any random variables. It is either true or false, and so its probability must be either 1 or 0, respectively, and not 0.8. Thus the statement $P(6.89 < c < 62) = 0.8$ is wrong.

The way to interpret '80% confident' is as follows. If we were to sample another number, e.g. 5.4, from the same $U(0,c)$ distribution, then we'd get another CI, e.g. (10(5.4)/9, 10(5.4)) = (6, 54).

Now imagine sampling very many such numbers, so as to get the same number of corresponding CIs, e.g. (6.89, 62), (6,54), (8.89, 80),....

Then close to 80% of these CIs will contain c. This is an expression of the fact that $P(10Y/9 < c < 10Y) = 0.8$.

But for any particular value y of Y, we should never write $P(10y/9 < c < 10y) = 0.8$.

The 80% CI derived above, (10Y/9, 10Y), may also be called a *central* CI. It is also possible to construct other types of CI. The two other major types we will look at are *upper range* and *lower range* CI's. These three major types are defined as follows.

Let $I = [L, U] = [g(Y_1,...,Y_n), h(Y_1,...,Y_n)]$ be a $100(1-\alpha)$% confidence interval for θ.

(Thus $P(L \le \theta \le U) = 1-\alpha$ for all θ.)

We say that I is an *upper range confidence interval* if $P(\theta \ge L) = 1-\alpha$

(Then $P(\theta < L) = \alpha$, $U = \infty$ & we call L the $1-\alpha$ *lower confidence limit* for θ.)

We say that I is a *lower range confidence interval* if $P(\theta \le U) = 1-\alpha$.

(Then $P(\theta > U) = \alpha$, $L = -\infty$ & we call U the $1-\alpha$ *upper confidence limit* for θ.)

We say that I is a *central* confidence interval if $P(\theta < L) = P(\theta > U) = \alpha/2$.

(Analogous definitions apply for CIs of the form $I = (L, U)$.)

Thus in Example 4, [10*Y*/9, 10*Y*] is a *central* 80% CI for *c*.

Check: $P(\theta < L) = P(c < 10Y/9) = P(Y > 0.9c) = 0.1 = \alpha/2$
$P(\theta > U) = P(c > 10Y) = P(Y < 0.1c) = 0.1 = \alpha/2$.

Example 5 Refer to Example 4. Find an upper range 80% CI for *c*.

0.8 = $P(X \leq 0.8) = P(Y/c \leq 0.8) = P(5Y/4 \leq c)$.
(draw a picture of $f(x)$, with the area between 0 and 0.8 shaded).

So $L = 5Y/4$, with realised value 5(6.2)/4 = 7.75.
So an upper range 80% CI for *c* is [7.75, ∞).

What is a lower range 80% CI for *c*?

0.8 = $P(X \geq 0.2) = P(Y/c \geq 0.2) = P(c \leq 5Y)$
(draw a picture of $f(x)$, with the area bewteen 0.2 and 1 shaded).

So $U = 5Y$, with realised value 5(6.2) = 31.
So a lower range 80% CI for *c* is ($-\infty$, 31].

This is true, but we can 'refine' the answer, as follows.

First, observe that *c* cannot be negative.
So a lower range 80% CI for *c* is [0, 31].

Secondly, we also know that $0 \leq y \leq c$, and therefore $c \geq y = 6.2$.
So a lower range 80% CI for *c* is [6.2, 31].

Summary: Lower range 80% CI = [6.2, 31]
Upper range 80% CI = [7.75, ∞)
Central 80% CI = [6.89, 62].

Note that the lower range CI is the shortest of the three. This is an attractive property, and sometimes we may choose one CI formula over another because it leads to a shorter CI *on average*. However, in some cases (not considered here), choosing the shortest of two or more CIs only *after* observing the data (as is sometimes done) may cause the confidence coefficient to be altered so that it is no longer the nominal $1-\alpha$.

Example 6 Suppose that 1.2, 3.9 and 2.4 are a random sample from a normal distribution with variance 7.

Find a 95% confidence interval for the normal mean.

Let $Y_1,\ldots,Y_n \sim iid\ N(\mu,\sigma^2)$ (in our case, n = 3 and σ^2 = 7).
Let's find a $100(1-\alpha)$% CI for μ generally.

Recall that $Z = \dfrac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ (Z is a pivotal quantity).

Therefore
$$\begin{aligned}1-\alpha &= P(-z_{\alpha/2} < Z < z_{\alpha/2}) \\ &= P\left(-z_{\alpha/2} < \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \\ &= P\left(-z_{\alpha/2}\sigma/\sqrt{n} < \bar{Y}-\mu < z_{\alpha/2}\sigma/\sqrt{n}\right) \\ &= P\left(-\bar{Y}-z_{\alpha/2}\sigma/\sqrt{n} < -\mu < -\bar{Y}+z_{\alpha/2}\sigma/\sqrt{n}\right) \\ &= P\left(\bar{Y}-z_{\alpha/2}\sigma/\sqrt{n} < \mu < \bar{Y}+z_{\alpha/2}\sigma/\sqrt{n}\right).\end{aligned}$$

So a $100(1-\alpha)$% CI for μ is $\left(\bar{Y}-z_{\alpha/2}\sigma/\sqrt{n}, \bar{Y}+z_{\alpha/2}\sigma/\sqrt{n}\right)$.
This interval may also be written $\left(\bar{Y} \pm z_{\alpha/2}\sigma/\sqrt{n}\right)$.

In our case: $100(1-\alpha)$ = 95 $\Rightarrow \alpha = 0.05$
$z_{\alpha/2} = z_{0.025} = 1.96$
n = 3
$\bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i = \dfrac{1}{3}(1.2+3.9+2.4)$ = 2.5
σ^2 = 7.

So the 95% CI for μ is
$$\left(\bar{y} \pm z_{\alpha/2}\sigma/\sqrt{n}\right) = \left(2.5 \pm 1.96\sqrt{7}/\sqrt{3}\right) = (2.5 \pm 3.0) = (-0.5, 5.5).$$

Example 7 Suppose that 1.2, 3.9 and 2.4 are a random sample from a normal distribution with *unknown* variance.

Find a 95% confidence interval for the normal mean.

Let $Y_1,\ldots,Y_n \sim iid\ N(\mu,\sigma^2)$.

Recall that $T=\dfrac{\overline{Y}-\mu}{S/\sqrt{n}} \sim t(n-1)$ (T is a pivotal quantity).

Therefore $1-\alpha = P(-t<T<t)$ where $t = t_{\alpha/2}(n-1)$

$$= P\left(-t<\frac{\overline{Y}-\mu}{S/\sqrt{n}}<t\right)$$

$$= P\left(\overline{Y}-tS/\sqrt{n}<\mu<\overline{Y}+tS/\sqrt{n}\right).$$

So a $100(1-\alpha)$% CI for μ is

$$\left(\overline{Y}\pm t_{\alpha/2}(n-1)S/\sqrt{n}\right).$$

In our case: $100(1-\alpha) = 95 \Rightarrow \alpha = 0.05$

$n = 3$

$t_{\alpha/2}(n-1) = t_{0.025}(2) = 4.303$

$\overline{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i = \dfrac{1}{3}(1.2+3.9+2.4) = 2.5$ [as before]

$s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(y_i-\overline{y})^2$

$= \dfrac{1}{2}\{(1.2-2.5)^2 + (3.9-2.5)^2 + (2.4-2.5)^2\} = 1.83.$

So the 95% CI for μ is

$$\left(\overline{y}\pm t_{\alpha/2}(n-1)s/\sqrt{n}\right) = \left(2.5\pm 4.303\sqrt{1.83}/\sqrt{3}\right) = (2.5\pm 3.36) = (-0.86,\ 5.86).$$

Note that this interval is wider than the CI in Example 5, (–0.5, 5.5). This corresponds to the fact that σ is now *unknown* and effectively needs to be estimated from the sample, by *s*. This illustrates the fact that when information is *decreased*, CI's tend to become wider (which makes sense because there is greater uncertainty). However, This is not always the case, and sometimes a decrease in information can, by chance, lead to an interval (with the same confidence coefficient) which is much narrower.

Example 8 200 people were randomly sampled from the population of Australia, and their heights were measured. The sample mean was 1.673 and the sample standard deviation was 0.310.

Find a 95% confidence interval for the average height of all Australians.

Let Y_i be the *i*th height and assume that $Y_1,\ldots,Y_n \sim iid\,(\mu,\sigma^2)$.

(Note that we are not making any *distributional* assumptions.)

Now, by the central limit theorem, $\dfrac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \dot{\sim} N(0,1)$ (since n = 200 is large).

But σ is unknown, and so we cannot make use of this pivotal quantity.

However, it can be shown that $Z = \dfrac{\overline{Y}-\mu}{S/\sqrt{n}} \dot{\sim} N(0,1)$ also (see Chapter 9).

Using the same logic as in the last few examples, we find that
a $100(1-\alpha)$% CI for μ is $\left(\overline{Y} \pm z_{\alpha/2} S/\sqrt{n}\right)$.

(Note that this is only an *approximate* CI, meaning that the probability of it containing μ is only *approximately* $1-\alpha$. However, the closeness of the approximation will be very good if n is large.)

In our case, this 95% CI for the average height of all Australians works out as
$\left(\overline{y} \pm z_{\alpha/2} s/\sqrt{n}\right) = (1.673 \pm 1.96(0.310)/\sqrt{200}) = (1.673 \pm 0.043) = (1.630, 1.716)$.

(Note that if σ were known, we would use it in place of S.
Thus we would instead report the interval $\left(\overline{y} \pm z_{\alpha/2}\sigma/\sqrt{n}\right)$.)