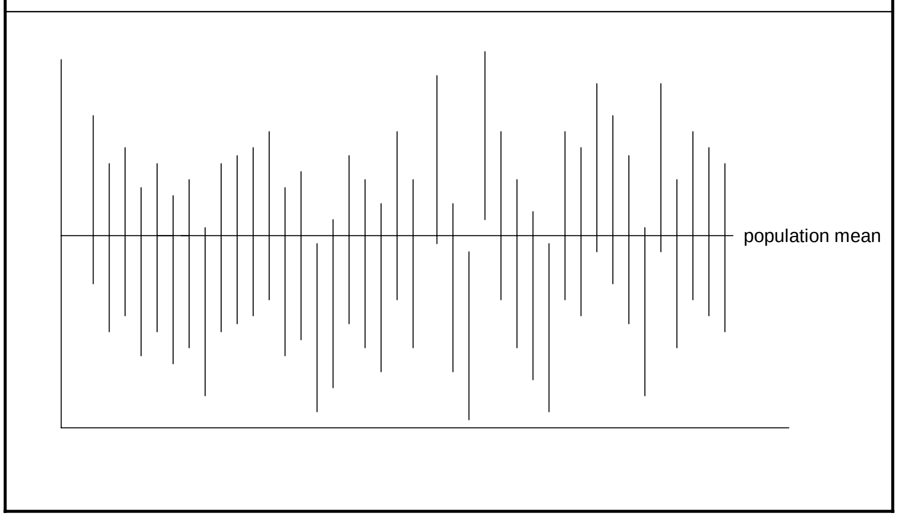
What does a confidence interval mean?

Elliott Sober

Suppose you sample organisms from a population and find that the average height in the sample is 6 units. If so, the maximum likelihood estimate of the average height in the whole population is 6 units. But how confident should you be that the population mean is around 6 feet? It may *seem* that drawing a 95% confidence interval around 6 gives you an answer to this question. In fact, it does not.

When you say that [5,7] is a 95% confidence interval, this does NOT mean that there is a 95% probability that the true value of the parameter being estimated falls between 5 and 7. Confidence intervals are part of frequentism, not Bayesianism. What, then, does a confidence interval mean? Consider the following figure; it depicts different 95% CI's constructed from different data sets drawn from the same population. (I assume that the population is normal and that the standard deviation is known.)



Each line represents a 95% confidence interval for a different data set drawn from the same population.

The crucial point is that when we say that "a 95% CI has a probability of 0.95 of containing the true value of the parameter Θ," this is <u>conditional</u> on the parameter's having the value it does. That is, what is true is that

(T) For any x, Pr(a 95% CI constructed from data drawn from the population will contain x | Θ = x) = 0.95.

This is very different from the following claim about an <u>unconditional</u> probability:

(F) For any x, Pr(a 95% CI constructed from data drawn from the population will contain x, where Θ = x) = 0.95.

The difference between the true principle (T) and the false principle (F) is the same as the difference between $Pr(X \mid Y)$ and $Pr(X\&Y)$. The sentence quoted at the start of this paragraph is false or misleading as stated. Notice that the figure could not be drawn without first deciding what the true value of Θ, the population mean, is.

The probability of 0.95 in (T) indicates that you should expect that about 95% of the CI's constructed from different data sets will include the population mean and about 5% will fail to do so. The figure shows the CI's drawn for 40 data sets. There are 4 failures here, which is a few more than the expected value (2), but of course that sort of thing can happen.

One further comment: (T) talks about the probability that "a" 95% CI will contain the number x. Why the indefinite article? Two points are relevant. First, there are many 95% CI's that might be constructed from a single data set. Second, the claim is not that <u>at least one</u> of those 95% CI's will have a probability of 0.95 of containing the number x. Rather, (T) should be understood to assert that <u>each</u> of those 95% CI's has a probability of 0.95 of containing the number x. To simplify our discussion, I'm going to assume that there is just one 95% CI that an investigator will construct from a given data set. This involves rewriting (T), slightly, as follows:

(T*) For any x, Pr(the 95% CI constructed from data drawn from the population will contain x $\mid \Theta = x) = 0.95$.

* * *

The true principle (T*) is a universal generalization, and the principle from deductive logic called universal instantiation (from "all x's are F," infer that the individual i is F) applies to it. That is, it follows from (T*) that

Pr(the 95% CI constructed from data drawn from the population will contain 14 $\mid \Theta = 14) = 0.95$.

And that

Pr(the 95% CI constructed from data drawn from the population will contain 6 $\mid \Theta = 6) = 0.95$.

Now suppose your data lead you to construct the 95% CI whose bounds are [5,7]. Does it follow from the above two displayed statements that

Pr([5,7] contains 14 $\mid \Theta = 14) = 0.95$.

and that

Pr([5,7] contains 6 $\mid \Theta = 6) = 0.95$?

These had better <u>not</u> follow, since these last two probability statements are false; the correct value for the first is 0, while the correct value for the second is unity.

In fact, the following argument is fallacious (see Howson and Urbach, pp. 239-240):

Pr(the 95% CI constructed from the data contains 6 | Θ = 6) = 0.95.
[5,7] = the 95% CI constructed from the data.

Pr([5,7] contains 6 | Θ = 6) = 0.95.

The logical flaw in this argument has nothing special to do with CI's. The same mistake occurs in the following argument about the throwing of a fair die:

Pr(the result of the next throw is an even number | the die is thrown) = ½ .
4 = the result of the next throw.

Pr(4 is an even number | the die is thrown) = ½ .

The premises are true but the conclusion is false; the proposition that 4 is an even number has a probability of unity, no matter what you conditionalize on. The general lesson here is that probability statements contain "opaque contexts" (a term from the philosophy of language); this means that substituting co-referring terms is not logically valid. Here are two examples from outside probability:

It's a contingent fact that the number of planets > 7.
The number of planets = 9.

It's a contingent fact that 9>7.

Lois Lane believes that Superman can fly.
Superman = Clark Kent.

Lois Lane believes that Clark Kent can fly.

Both these arguments are fallacious; modal statements ("it's contingent that p," "it's necessary that q") contain opaque contexts, and so do statements that involve propositional attitudes about what agents believe and desire.* Probability is another modality, so it is no surprise that the probability operator Pr(- | -) contains opaque contexts. Watch out for substitutions!

* * *

* As an exercise, show that "S desires ____" contains an opaque context.

Confidence intervals are regarded as interval estimates. For example, the statement that [5,7] is a 95% CI is usually taken to mean that [5,7] is a "good" interval estimate of Θ. This might lead you to think that there is a probability statement that has the proposition "Θ falls in the interval [5,7]" in it somewhere, and which assigns to the probability in question a value of 0.95. Let us consider some options concerning what this probability statement might be:

(1) Pr(Θ is between 5 and 7 | the sample mean is between 5 and 7) = 0.95.

(2) Pr(the sample mean is between 5 and 7 | Θ is between 5 and 7) = 0.95.

(3) Pr(the sample mean is 6 | Θ is between 5 and 7) = 0.95.

(4) Pr(Θ is between 5 and 7 | Θ = 6) = 0.95.

None of these statements correctly describes what it means for [5,7] to be a 95% CI for Θ. Statement (1) assigns a posterior probability to the interval estimate of Θ. But confidence intervals are not Bayesian and principle (T) doesn't assign a posterior probability to any interval estimate. Statement (2) improves matters; at least the claim about the unknown parameter Θ is on the right side of the conditional probability sign. But (2) isn't what it means to say that [5,7] is a 95% CI. Notice that the hypothesis described in (2) – that Θ is between 5 and 7 -- is composite; there are many ways that Θ can be between 5 and 7. The same is true of statement (3); however, the probability of obtaining a sample whose mean value is exactly 6 is zero if the population is normal and has a nonzero standard deviation. So (3) is false. And so is (4); as already noted, the true value of the conditional probability it describes is unity, not 0.95.

Proposition (T) describes what it means to construct a 95% CI. It says that if you apply this procedure of construction to different data sets drawn from the same population (as the figure depicts), you can expect to construct intervals that include the true population mean about 95% of the time. When you use this procedure on the single data set at hand, you're using a procedure that has a probability of 95% of producing an interval that includes the true value. But this statement about *the procedure* you're using doesn't allow you to make a probability statement about *the confidence interval* you construct. The interval [5,7] either includes Θ, or it doesn't. All that can be said of the interval [5,7] is that it was constructed by a procedure that is described by (T). The probability claim can be made about the procedure of constructing a confidence interval, not about the confidence interval you construct.