

A.11 Another look at confidence intervals

A.11.1 Why are confidence intervals often misinterpreted?

The concept of a confidence interval is often misunderstood and misinterpreted. A common misinterpretation is that a (realised) confidence interval ‘contains the parameter value with a certain *probability*’. **This is incorrect**, although it is easy to see why this mistake is made: the problem is the use (explicit or implicit) of the word ‘realised’, which means we have an actual value for the confidence interval, e.g. $(-1.57, 1.92)$.

A.11.2 Random variables vs realisations

In Section 7.1.1, the difference between a random variable X and its realisation x are re-emphasised. A random variable X is a function, while x is a specific value in \mathbb{R} .

Similarly, given multiple random variables X_1, X_2, \dots, X_n and their realisations x_1, x_2, \dots, x_n , respectively, we have the statistic \bar{X} , known as the sample mean, and its realisation, \bar{x} , which is also known as the sample mean. However, \bar{X} is a random variable, while \bar{x} is a real number. This is true for all statistics; a statistic is a random variable defined in terms of other random variables, but it will have a realisation or value once those random variables are observed.

The same is true for estimators. In fact, \bar{X} is also an estimator, perhaps of the true mean θ of the (i.i.d.) random variables X_1, X_2, \dots, X_n , while \bar{x} is then an estimate of θ . Here, \bar{X} is an example of a point estimator, while \bar{x} is a point estimate.

We have also seen in Section 1.5.2 the concept of interval estimators, which is a pair of estimators $L(X_1, X_2, \dots, X_n)$ and $U(X_1, X_2, \dots, X_n)$, with the property $L(\mathbf{X}) \leq U(\mathbf{X})$, and which will have realisations $L(x_1, x_2, \dots, x_n)$ and $U(x_1, x_2, \dots, x_n)$.

A.11.3 What is a confidence interval?

Let’s go over the motivation behind the creation of a confidence interval one more time.

There is a distribution F_X with a parameter θ . Following the **frequentist** approach to statistics, we assume that θ has a **fixed**, true value, but this value is unknown. We wish to estimate the value of θ .

We record data x_1, x_2, \dots, x_n , which we assume are observations of random variables X_1, X_2, \dots, X_n following distribution F_X . A common assumption is that the random variables are independent, in order to allow us to use certain theoretical results. We may decide to make additional assumptions about F_X , e.g. regarding the value of other parameters, or the family of distributions to which it belongs, to allow us to use further theoretical results.

Given a significance threshold $\alpha \in (0, 1)$, we can use our assumptions and theoretical results to define an interval estimator, which is a pair of estimators $L(\mathbf{X}) = L(X_1, X_2, \dots, X_n)$ and $U(\mathbf{X}) = U(X_1, X_2, \dots, X_n)$, with the property $L(\mathbf{X}) \leq U(\mathbf{X})$ and

$$P(\theta \in [L(\mathbf{X}), U(\mathbf{X})]) \geq 1 - \alpha. \quad (\text{A.8})$$

We then call $[L(\mathbf{X}), U(\mathbf{X})]$ a $1 - \alpha$ **confidence interval**.

So far, everything is theoretical and the probabilistic statement in Equation (A.8) is valid. However, most of the time when people are referring to confidence intervals, they are referring to the **realisation**, and once we compute a realisation of the confidence interval, using $\mathbf{x} = (x_1, x_2, \dots, x_n)$, we need to be more careful how we interpret the realised confidence interval $[L(\mathbf{x}), U(\mathbf{x})]$ which is now a pair of real numbers.

We cannot say that the interval $[L(\mathbf{x}), U(\mathbf{x})]$ contains θ with a certain probability, but we can say that the interval $[L(\mathbf{x}), U(\mathbf{x})]$ contains θ with **confidence** $1 - \alpha$. Let's look at an example to see why this is the case.

A.11.4 Example: heights of a group of students

Suppose a lecturer wants to estimate θ , the mean height of the cohort of Year 1 students (in cm). The assumption is that the heights of the students follow a distribution F_X , and the mean of this distribution is θ , and the height of each student is an independent random variable X following F_X .

The lecturer decides to use a sample of n student heights to construct a confidence interval, using the sample mean \bar{X} . For simplicity, the lecturer decides to assume that the standard deviation is less or equal to 12cm, i.e. $\sigma \leq 12$. Furthermore, it is assumed that the heights of the students follow a normal distribution, i.e. the height of a student is a random variable X , with $X \sim N(\theta, \sigma^2)$. For any significance threshold α , one can derive

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

Choosing $\alpha = 0.01$, this becomes a 99% confidence interval, i.e.

$$P\left(\bar{X} - 2.576 \frac{12}{\sqrt{n}} < \theta < \bar{X} + 2.576 \frac{12}{\sqrt{n}}\right) = 0.99.$$

As discussed above, since we are still dealing with random variables, this probabilistic statement holds. Having constructed this interval theoretically, the lecturer decides to obtain some data.

A.11.5 First observed sample

Before the start of the next class, the lecturer decides to randomly measure the heights of 10 students (in cm), to obtain a realisation of the constructed confidence interval. The following values are recorded:

$$\{155.5, 163.5, 170.1, 153.2, 169.3, 167.4, 168.0, 180.4, 152.4, 182.2\}.$$

One can compute $\bar{x} = 166.2$, and so the realised 99% confidence interval is

$$\left(\bar{x} - 2.576 \frac{12}{\sqrt{n}}, \bar{x} + 2.576 \frac{12}{\sqrt{n}}\right) = \left(166.2 - 2.576 \frac{12}{\sqrt{10}}, 166.2 + 2.576 \frac{12}{\sqrt{10}}\right) = (156.4, 176.0).$$

One then can say that the interval $(156.4, 176.0)$ contains the value of θ with **confidence** 0.99 (or 99%).

However, it does not make sense to say that the interval contains θ with *probability* 0.99! The value of θ , while unknown, is assumed to be a fixed, true value. This value is either in the interval, or not. Imagine the value of θ were 168. Then it does not make sense to say that 168 is in the interval $(156.4, 176.0)$ with probability 0.99. In fact, the unknown value θ is contained in a realised confidence interval with probability 0 or 1: it is either in the interval, or not.

A.11.6 Second observed sample

Suppose the lecturer decides to record some difference measurements at the next lecture, and again measures the heights of 10 randomly selected students to obtain the sample:

$$\{160.9, 187.6, 155.9, 171.8, 191.5, 163.8, 165.3, 163.4, 167.6, 172.7\}$$

The sample mean can be computed as $\bar{x} = 170.1$, and this results in the (realised) 99% confidence interval:

$$\left(170.1 - 2.576 \frac{12}{\sqrt{10}}, 170.1 + 2.576 \frac{12}{\sqrt{10}}\right) = (160.32, 179.9).$$

So, one then can again say that the interval $(160.32, 179.9)$ contains the value of θ with **confidence** 0.99 (or 99%).

Note how this also shows the problem with saying a realised confidence interval contains θ with probability 0.99; the two intervals $(156.4, 176.0)$ and $(160.32, 179.9)$ could not both contain θ with probability 0.99!

A.11.7 Infinitely many samples

So, one may be wondering, does a realised confidence interval actually contain the true value of θ , or not? The truth is, for a given (realised) confidence interval, we don't know. However, this is where the level of confidence comes in: if the lecturer repeated this experiment at the start of every class, and repeated this a large number of times, then approximately **99% of the realised confidence intervals** would contain θ .

In other words, if the experiment could be repeated infinitely many times, then the proportion of confidence intervals actually containing θ would be $1 - \alpha$, for the chosen significance threshold α .

This is where the name 'frequentist' comes from; probabilities are related to the frequency (or proportion) of results from repeated sampling of data.

Finally, note that although in this example we were interested in the mean height of the students, we could easily have used another measurable characteristic such as weight or shoe size.