Chapter 8. Statistical Inference

8.1: Confidence Intervals

Slides (Google Drive) Alex Tsun Video (YouTube)

We've talked about several ways to estimate unknown parameters, and desirable properties. But there is just one problem now: even if our estimator had all the good properties, the probability that our estimator for θ is *exactly* correct is 0, since θ is continuous (a decimal number)! We'll see how we can construct confidence intervals around our estimator, so that we can argue that $\hat{\theta}$ is *close to* θ with high probability.

8.1.1 Confidence Intervals Motivation

Confidence intervals are used in the Frequentist setting, which means the population parameters are assumed to be unknown but will always be **fixed**, not random variables. Credible intervals, on the other hand, are a Bayesian version of a Frequentist's confidence interval which is discussed in the next section 8.2.

When doing point estimation (such as MLE, MoM), the probability that our answer is correct (over the randomness in our iid samples) is 0:

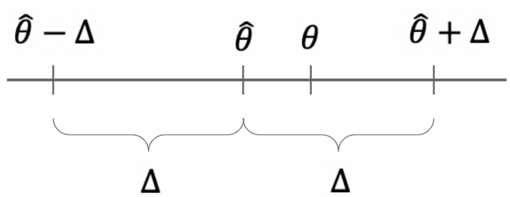
$$\mathbb{P}\left(\hat{\theta} = \theta\right) = 0$$

because θ is a real number and can take uncountably many values. Hence the probability we are ***exactly*** correct is zero, though we may be very close.

Instead, we can give an interval (often but not always centered at our point estimate $\hat{\theta}$), such that θ falls into it with high probability, like 95%.

$$\mathbb{P}\left(\theta \in \left[\hat{\theta} - \Delta, \hat{\theta} + \Delta\right]\right) = 0.95$$

The confidence interval for θ can be illustrated in the below picture. We will explain how to interpret a confidence interval at a specific confidence level soon.



Note that we can write this in any of the following three equivalent ways, as they all represent the probability that $\hat{\theta}$ and θ differ by no more than some amount Δ:

$$\mathbb{P}\left(\theta \in \left[\hat{\theta} - \Delta, \hat{\theta} + \Delta\right]\right) = \mathbb{P}\left(\left|\hat{\theta} - \theta\right| \leq \Delta\right) = \mathbb{P}\left(\hat{\theta} \in [\theta - \Delta, \theta + \Delta]\right) = 0.95$$

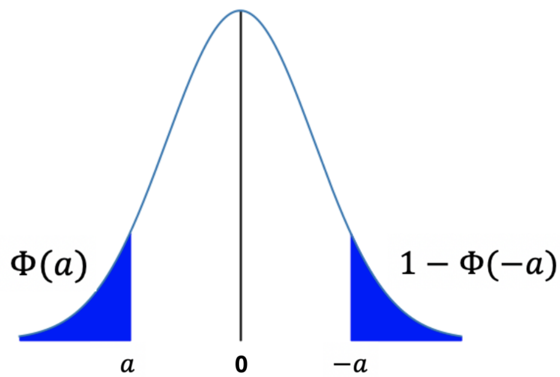
Note the first and third equivalent statements especially (swapping $\hat{\theta}$ and θ).

8.1.2 Review: The Standard Normal CDF

Before we construct confidence intervals, we need to review the standard normal CDF. It turns out, the Normal distribution frequently appears since our estimators are usually the sample mean (at least for our common distributions), and the Central Limit Theorem applies!

We have learned about the CDF of normal distribution. If $Z \sim \mathcal{N}(0,1)$, we denote the CDF $\Phi(a) = F_Z(a) = P(Z \leq a)$, since it's so commonly used. There is no closed-form formula, so one way to find a z-score associated with a percentage is to look up in a z-table.

Note: $\Phi(a) = 1 - \Phi(-a)$ by symmetry.



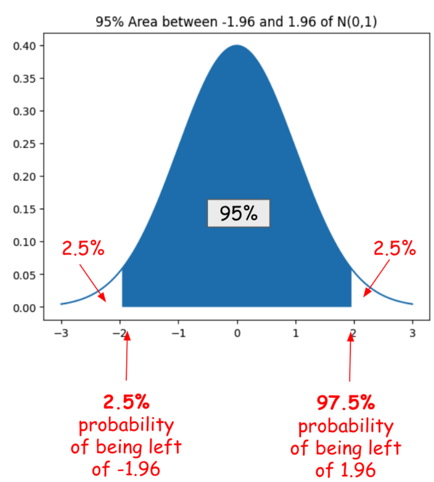
Suppose we want a (centered) interval, where the probability of being in that interval is 95%.

Left bound: the probability of being less than the left bound is 2.5%.

Right bound: the probability of being greater than the right bound is 2.5%. Thus, the probability of being less than the right bound should be 97.5%.

Note the following two equivalent statements that say that $\mathbb{P}(Z \leq 1.96) = 0.975$ (where Φ^{-1} is the inverse CDF of the standard normal):

$$\Phi(1.96) = 0.975$$
$$\Phi^{-1}(0.975) = 1.96$$



8.1.3 Confidence Intervals

Let's start by doing an example.

Example(s)

Suppose $x_1,...x_n$ are iid samples from Poi(θ) where θ is unknown. Our MLE and MoM estimates agreed at the sample mean: $\hat{\theta} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Create an interval centered at $\hat{\theta}$ which contains θ with probability 95%.

Solution Recall that if $W \sim \text{Poi}(\theta)$, then $\mathbb{E}[W] = \mathsf{Var}(W) = \theta$, and so our estimator (the sample mean) $\hat{\theta} = \bar{x}$ has $\mathbb{E}\left[\hat{\theta}\right] = \theta$ and $\text{Var}\left(\hat{\theta}\right) = \frac{\text{Var}(x_i)}{n} = \frac{\theta}{n}$. Thus, by the Central Limit Theorem, $\hat{\theta}$ is approximately Normally distributed:

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i \approx \mathcal{N}\left(\theta, \frac{\theta}{n}\right)$$

If we standardize, we get that

$$\frac{\hat{\theta} - \theta}{\sqrt{\theta/n}} \approx \mathcal{N}(0,1)$$

To construct our 95% confidence interval, we want $\mathbb{P}\left(\theta \in \left[\hat{\theta} - \Delta, \hat{\theta} + \Delta\right]\right) = 0.95$

$$\begin{aligned}
\mathbb{P}\left(\theta \in \left[\hat{\theta} - \Delta, \hat{\theta} + \Delta\right]\right) &= \mathbb{P}\left(\theta - \Delta \le \hat{\theta} \le \theta + \Delta\right) && \text{[one of 3 equivalent statements]} \\
&= \mathbb{P}\left(-\Delta \le \hat{\theta} - \theta \le \Delta\right) \\
&= \mathbb{P}\left(-\frac{\Delta}{\sqrt{\theta/n}} \le \frac{\hat{\theta} - \theta}{\sqrt{\theta/n}} \le \frac{\Delta}{\sqrt{\theta/n}}\right) \\
&= \mathbb{P}\left(-\frac{\Delta}{\sqrt{\theta/n}} \le Z \le \frac{\Delta}{\sqrt{\theta/n}}\right) && \text{[CLT]} \\
&= 0.95
\end{aligned}$$

Because $\frac{\Delta}{\sqrt{\theta/n}}$ represents the right bound, and the probability of being less than the right bound is 97.5% for a 95% interval (see the above picture again). Thus:

$$\frac{\Delta}{\sqrt{\theta/n}} = \Phi^{-1}(0.975) = 1.96 \implies \Delta = 1.96\sqrt{\frac{\theta}{n}}$$

Since we don't know θ, we plug in our estimator $\hat{\theta}$, and get

$$[\hat{\theta} - \Delta, \hat{\theta} + \Delta] = \left[\hat{\theta} - 1.96\sqrt{\frac{\hat{\theta}}{n}}, \hat{\theta} + 1.96\sqrt{\frac{\hat{\theta}}{n}}\right]$$

That is, since $\hat{\theta}$ is normally distributed with mean θ, we just need to find the Δ so that $\hat{\theta} \pm \Delta$ contains 95% of the area in a Normal distribution. The way to do so is to find $\Phi^{-1}(0.975) = 1.96$, and go $\pm$ 1.96 standard deviations of $\hat{\theta}$ in each direction! $\square$

Definition 8.1.1: Confidence Interval

Suppose you have iid samples $x_1,...,x_n$ from some distribution with unknown parameter θ, and you have some estimator $\hat{\theta}$ for θ.

A $100(1-\alpha)\%$ **confidence interval** for θ is an interval (typically but not always) centered at $\hat{\theta}$, $\left[\hat{\theta}-\Delta, \hat{\theta}+\Delta\right]$, such that the probability (over the randomness in the samples $x_1,...,x_n$) θ lies in the interval is $1-\alpha$:

$$\mathbb{P}\left(\theta \in \left[\hat{\theta}-\Delta, \hat{\theta}+\Delta\right]\right) = 1-\alpha$$

If $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean, then $\hat{\theta}$ is approximately normal by the CLT, and a $100(1-\alpha)\%$ confidence interval is given by the formula:

$$\left[\hat{\theta} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \hat{\theta} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

where $z_{1-\alpha/2} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$ and σ is the true standard deviation of a single sample (which may need to be estimated).

It is important to note that this last formula ONLY works when $\hat{\theta}$ is the sample mean (otherwise we can't use the CLT); you'll need to find some other strategy if it isn't.

If we wanted a 95% interval, then that corresponds to $\alpha = 0.05$, since $100(1-\alpha) = 95$. We were then looking up the inverse Phi table at $(1-\alpha/2) = (1-0.05/2) = 0.975$ to get our desired number of standard deviations in each direction of 1.96.

If we wanted a 98% interval, then that corresponds to $\alpha = 0.02$ since $100(1-\alpha) = 98$. We then would look up $\Phi^{-1}(0.99)$ since $1-\alpha/2 = 0.99$, because if there is to be 98% of the area in the middle, there is 1% to the left and right!

Example(s)

Construct a 99% confidence interval for θ (the unknown probability of success) in $\text{Ber}(\theta)$ given $n = 400$ iid samples $x_1,...,x_{400}$ where $\sum_{i=1}^{n} x_i = 136$ (observed 136 successes out of 400).

Solution

Recall for Bernoulli distribution $\text{Ber}(\theta)$, our MLE/MoM estimator was the sample mean:

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{136}{400} = 0.34$$

Because we want to construct a $99\% = 100(1-\alpha)\%$ confidence interval:

$$\alpha = 1 - \frac{99}{100} = 0.01$$

A 99% confidence interval would use the z-score:

$$z_{1-\alpha/2} = z_{1-0.01/2} = z_{0.995} = \Phi^{-1}(0.995) \approx 2.576$$

The population standard deviation σ is unknown, but we'll approximate it using the standard deviation of $\text{Ber}(\theta)$ as follows (since $\mathsf{Var}\left(\text{Ber}(\theta)\right) = \theta(1-\theta)$):

$$\sigma = \sqrt{\theta(1-\theta)} \approx \sqrt{\hat{\theta}(1-\hat{\theta})} = \sqrt{0.34(1-0.34)} = 0.474$$

Thus, our 99% confidence interval for θ is:

$$\left[0.34 - 2.576\frac{0.474}{\sqrt{400}}, 0.34 + 2.576\frac{0.474}{\sqrt{400}}\right] = [0.279, 0.401]$$

$\square$

8.1.4 Interpreting Confidence Intervals

How can we interpret our 99% confidence interval $[0, 279, 0.401]$ from the above example?

Incorrect: There is a 99% probability that θ falls in the confidence interval $\left[\hat{\theta} - \Delta, \hat{\theta} + \Delta\right] = [0, 279, 0.401]$

This is incorrect because there is no randomness here: θ is a fixed parameter. θ is either in the interval or out of it; there's nothing probabilistic about it.

Correct: If we repeat this process several times (getting n samples each time and constructing different confidence intervals), about 99% of the confidence intervals we construct will contain θ.

Notice the subtle difference! Alternatively, *before* you receive samples, you can say that there is a 99% probability (over the randomness in the samples) that θ will fall into our to-be-constructed confidence interval $\left[\hat{\theta} - \Delta, \hat{\theta} + \Delta\right]$. Once you plug in the numbers though, you cannot say that anymore.