18.650: Fundamentals of Statistics February 18, 2026

Lecture 8 — Estimators and confidence intervals

Mohammad Reza Karimi

1 Review of estimators

Let $X_1, \ldots, X_n$ be i.i.d. samples from the (unknown) probability distribution $\mathbb{P}$. We might want to estimate

- μ, the mean of $\mathbb{P}$
- σ^2, the variance of $\mathbb{P}$
- $F(t) = \mathbb{P}(X \leq t)$, the value of the CDF at some point t

We use an *estimator*—a function of the data—to estimate these quantities, e.g.,

- $\hat{\mu} = \bar{X}_n$ to estimate μ
- $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ to estimate σ^2
- $\hat{F}(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \leq t)$ to estimate $F(t)$

Properties we care about in an estimator:

- $\text{bias}(\hat{\theta}) = \mathbb{E}_\theta[\hat{\theta}] - \theta$
- standard error: $\text{se}(\hat{\theta}) = \sqrt{\mathbb{V}_\theta(\hat{\theta})}$.
- $\text{MSE}(\hat{\theta}) = \text{bias}^2(\hat{\theta}) + \text{se}^2(\hat{\theta})$.
- Consistency: $\hat{\theta} \xrightarrow{\mathbb{P}} \theta$.
- **Asymptotic normality**, which we now discuss.

2 Asymptotic Normality

To do the downstream tasks of constructing confidence intervals and testing hypotheses, we need more information: we need to know both the *rate of decay* of the fluctuations of the estimator $\hat{\theta}_n$ around θ, and the *distribution* of the fluctuations.

The "rate" is determined by the sequence a_n of increasing numbers (e.g., $a_n = \log n$, $a_n = \sqrt{n}$, $a_n = e^n$) such that $a_n(\hat{\theta}_n - \theta)$ converges to something nontrivial: it should be neither zero nor blowing up. For example, for numbers, the sequence $\frac{1}{\log n} + \frac{1}{n}$ converges to zero at rate $1/\log n$.

Thanks to the CLT, the rate of decay of fluctuations is $1/\sqrt{n}$ for most estimators $\hat{\theta}_n$, and the distribution of $\hat{\theta}_n$ is approximately normal. The formal term for such estimators is "*asymptotically normal*".

Definition 2.1: Asymptotic normality

An estimator $\hat{\theta}_n$ of θ is **asymptotically normal** if there is some $\sigma^2 > 0$ such that

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, \sigma^2) \quad \text{as} \quad n \to \infty. \tag{1}$$

The variance σ^2 is known as the **asymptotic variance**.

Remark.

A useful way to think about asymptotic normality is that

$$\hat{\theta}_n \approx \theta + \frac{\sigma}{\sqrt{n}} Z, \qquad \text{where} \quad Z \sim \mathcal{N}(0, 1).$$

By looking at this equation, we immediately see that the fluctuations decay at rate $1/\sqrt{n}$, and that the distribution of fluctuations is Gaussian.

Remark.

Our two tools for proving asymptotic normality are the CLT and the Delta Method, as the following example demonstrates.

Example.

Consider the model $\{\text{Ber}(p) \ : \ p \in (0, 1)\}$, i.e., $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p)$. We have $\mathbb{E}_p[X_i] = p$ and $\mathbb{V}_p[X_i] = p(1-p)$. Let $\hat{p} = \bar{X}_n$. Then

$$\sqrt{n}(\hat{p} - p) \rightsquigarrow \mathcal{N}(0, p(1-p))$$

by the CLT, and the asymptotic variance is $\sigma^2 = p(1-p)$.

Now, let $\hat{\theta} = (\bar{X}_n)^2$ be an estimator for $\theta = p^2$. Then $\hat{\theta}_n$ is asymptotically normal by the Delta Method, where the function g is $g(p) = p^2$. The Delta method also tells us that the asymptotic variance is $g'(p)^2\sigma^2 = (2p)^2 p(1-p) = 4p^3(1-p)$. To summarize,

$$\sqrt{n}(\hat{\theta} - \theta) \rightsquigarrow \mathcal{N}(0, 4p^3(1-p)).$$

2.1 Asymptotic normality via the standard error

If $\hat{\theta}_n$ is asymptotically normal with asymptotic variance σ^2, i.e., if $\hat{\theta}_n$ satisfies (1), then $\hat{\theta}_n \approx \mathcal{N}(\theta, \sigma^2/n)$. But if this is true, then it's also true that $\hat{\theta}_n \approx \mathcal{N}(\theta, \text{se}(\hat{\theta}_n)^2)$—in fact, this should be an even better approximation, since we have replaced the approximate variance σ^2/n by the exact variance $\text{se}(\hat{\theta}_n)^2$.

Motivated by this observation, we also give the following alternative definition of asymptotic normality, which is the one used in AoS:

Definition 2.2: Asymptotic normality, alternative definition

An estimator is asymptotically normal if

$$\frac{\hat{\theta}_n - \theta}{\text{se}(\hat{\theta}_n)} \rightsquigarrow \mathcal{N}(0, 1). \tag{2}$$

Let us compare (1) to (2). To do so, we rewrite (1) in the following equivalent form:

$$\frac{\sqrt{n}}{\sigma}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, 1). \tag{3}$$

Comparing (3) to (2), we see that (3) is simply telling us that $\text{se}(\hat{\theta}_n) \approx \sigma/\sqrt{n}$ or equivalently, that $\mathbb{V}[\hat{\theta}_n] \approx \sigma^2/n$.

It is often more difficult to compute the exact variance of an estimator than it is to derive the asymptotic variance. This is why Definition 2.1 is our preferred definition of asymptotic normality. Checking it does not require us to compute $\text{se}(\hat{\theta}_n)$.

3 Confidence intervals

A confidence interval is an interval in which θ lies with some probability.

Definition 3.1: confidence interval

A $(1 - \alpha)$ confidence interval (CI) for θ is a *random* interval of the form $C_n = (A_n, B_n)$ such that

$$\mathbb{P}_\theta(\theta \in (A_n, B_n)) = \mathbb{P}_\theta(A_n < \theta < B_n) \geq 1 - \alpha \quad \forall \theta.$$

Here, $1 - \alpha$ is called the *coverage* of the CI.

Note that the randomness comes from A_n and B_n, *not* from θ.

Interpretation. If for a given dataset we get $A_n = 2, B_n = 4$ for our 95% confidence interval, then it is *not* true that $\mathbb{P}_\theta(\theta \in (2,4)) \geq 0.95$. That probability is either 0 or 1, because either θ lies in $(2,4)$ or it doesn't. The way to interpret a confidence interval is that if we repeat an experiment a 100 times (collecting new data each time), and each time we construct a confidence interval based on the data, then we can expect that 5 out of those confidence intervals "miss" (don't contain θ), but the other 95 succeed at trapping θ inside.

Remark.

> If (A_n, B_n) is a $1-\alpha$ confidence interval, then so is $(A_n - 10, B_n + 10)$, because widening the interval will only increase the probability that θ lies in it. But this is giving us less precise information. So we shoot for the narrowest interval at the given confidence level.

Remark.

> In higher dimensions, we construct confidence *sets* rather than confidence intervals. The shape of the confidence set can be pretty much anything.

3.1 Constructing a CI

Typically our CI looks like

$$\hat{\theta} - c_\alpha < \theta < \hat{\theta} + c_\alpha,$$

and we just need to figure out how big to take c_α to ensure this is a $1-\alpha$ CI, i.e., to ensure

$$\mathbb{P}_\theta\left(\hat{\theta} - c_\alpha < \theta < \hat{\theta} + c_\alpha\right) \geq 1 - \alpha \tag{4}$$

To do so, we use asymptotic normality. Suppose $\hat{\theta}$ is asymptotically normal, so that $\hat{\theta} \approx \mathcal{N}(\theta, \sigma^2/n)$, where σ^2 is the asymptotic variance. Then we can rearrange the probability (4) as follows:

$$\begin{aligned}
\mathbb{P}_\theta(\hat{\theta} - c_\alpha < \theta < \hat{\theta} + c_\alpha) &= \mathbb{P}_\theta(-c_\alpha \leq \hat{\theta} - \theta \leq c_\alpha) \\
&= \mathbb{P}_\theta\left(-\frac{c_\alpha}{\sigma/\sqrt{n}} \leq \frac{\hat{\theta} - \theta}{\sigma/\sqrt{n}} \leq \frac{c_\alpha}{\sigma/\sqrt{n}}\right) \\
&\approx \mathbb{P}\left(-\frac{c_\alpha}{\sigma/\sqrt{n}} \leq Z \leq \frac{c_\alpha}{\sigma/\sqrt{n}}\right), \qquad Z \sim \mathcal{N}(0,1).
\end{aligned}$$

To get this probability to be $1-\alpha$, we set $c_\alpha/(\sigma/\sqrt{n}) = z_{\alpha/2}$, which is the point such that $\mathbb{P}(Z \geq z_{\alpha/2}) = \mathbb{P}(Z \leq -z_{\alpha/2}) = \alpha/2$, and hence $\mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$. Solving for c_α gives

$$c_\alpha = z_{\alpha/2}\, \sigma/\sqrt{n}.$$

In other words, our $1-\alpha$ confidence interval is

$$\left(\hat{\theta} - z_{\alpha/2}\,\sigma/\sqrt{n}, \hat{\theta} + z_{\alpha/2}\,\sigma/\sqrt{n}\right). \tag{5}$$

However, σ often depends on the unknown parameter θ! So instead we use an estimate for σ. Let's see an example of this in the context of Bernoulli random variables.

Example.

Suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p)$. We want to construct a confidence interval for p, using our estimator $\hat{p} = \bar{X}_n$. Since $\mathbb{E}[X_i] = p$, $\mathbb{V}[X_i] = p(1-p)$, the CLT tells us that

$$\sqrt{n}(\bar{X}_n - p) \rightsquigarrow \mathcal{N}(0, p(1-p)).$$

Therefore the asymptotic variance is $\sigma^2 = p(1-p)$, so (5) suggests to take the CI to be

$$\left(\bar{X}_n - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}, \bar{X}_n + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right).$$

But we don't know p, so this is not a CI we can actually compute from our data. Instead, there are two options. First, we can estimate p by $\bar{X}_n$. This gives the valid CI

$$\left(\bar{X}_n - z_{\alpha/2}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + z_{\alpha/2}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\right).$$

Second, we can notice that $p(1-p)$ is maximized at $p = 1/2$, so that $p(1-p) \leq \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ for all $p \in [0,1]$. We can therefore get the following more *conservative* CI (for $p \neq 1/2$, it will be slightly wider than it has to be):

$$\left(\bar{X}_n - z_{\alpha/2}\sqrt{\frac{1}{4n}}, \bar{X}_n + z_{\alpha/2}\sqrt{\frac{1}{4n}}\right).$$

Formally, the CI we have constructed is only valid *asymptotically.*

Definition 3.2: Asymptotic coverage

If

$$\lim_{n\to\infty} \mathbb{P}_\theta(\theta \in C_n) \geq 1 - \alpha$$

then we say that C_n has asymptotic coverage $1-\alpha$.