18.650: Fundamentals of Statistics October 13, 2023

Lecture 13 — Bootstrap properties and confidence intervals

Prof. Philippe Rigollet *Scribe: Anya Katsevich*

Recap of lecture 12: we have one true sample $X_1, \dots, X_n$. A bootstrap sample is

$$X_1^*, \dots, X_m^* \overset{\text{i.i.d.}}{\sim} \text{Unif}(X_1, \dots, X_n).$$

To get a sense of the distribution of an estimator $\hat{\theta}$, we can create B bootstrap samples, and use each one to get a new value of $\hat{\theta}$:

$$\begin{aligned}
X_{1:n}^{*(1)} &= \{X_1^{*(1)}, \dots, X_n^{*(1)}\} \to \hat{\theta}^{(1)} \\
X_{1:n}^{*(2)} &= \{X_1^{*(2)}, \dots, X_n^{(2)}\} \to \hat{\theta}^{(2)} \\
&\vdots \\
X_{1:n}^{*(B)} &= \{X_1^{*(B)}, \dots, X_n^{*(B)}\} \to \hat{\theta}^{(B)},
\end{aligned} \tag{1}$$

where

$$X_i^{(b)} \overset{\text{i.i.d.}}{\sim} \hat{\mathbb{P}} = \text{Unif}(X_1, \dots, X_n), \quad i = 1, \dots, n, \quad b = 1, \dots, B. \tag{2}$$

We can then draw a histogram of the $\hat{\theta}^{(b)}$, or compute the sample variance:

$$v_{\text{boot}} = \frac{1}{B} \sum_{b=1}^{B} \left(\hat{\theta}^{(b)} - \frac{1}{B} \sum_{c=1}^{B} \hat{\theta}^{(c)} \right)^2 \tag{3}$$

1 Properties of the bootstrap sample.

Let $X_1^*, \dots, X_m^*$ be a bootstrap sample of $X_1, \dots, X_n$, and suppose

$$\mu = \mathbb{E}[X_1], \sigma^2 = \mathbb{V}[X_1].$$

Then the bootstrap sample satisfies the following properties.

1. $X_i^* \overset{\text{i.i.d.}}{\sim} \text{Unif}(X_1, \dots, X_n)$ conditionally on $\{X_1, \dots, X_n\}$. This means

$$\mathbb{P}(X_i^* = X_j \mid X_1, \dots, X_n) = 1/n \quad \forall j = 1, \dots, n.$$

2. The conditional expectation of X_i^* is the sample mean:

$$\mathbb{E}[X_i^* \mid X_1, \dots, X_n] = \sum_{j=1}^{n} X_j \mathbb{P}(X_i^* = X_j \mid X_1, \dots, X_n) = \frac{1}{n} \sum_{j=1}^{n} X_j = \bar{X}_n.$$

3. The unconditional expectation of X_i^* is μ, by the tower property of conditional expectations:

$$\mathbb{E}[X_i^*] = \mathbb{E}\left[\mathbb{E}\left[X_i^* \mid X_1, \ldots, X_n\right]\right] = \mathbb{E}[\bar{X}_n] = \mu.$$

4. The conditional variance of X_i^* is the sample variance:

$$\begin{aligned}\mathbb{V}[X_i^* \mid X_1, \ldots, X_n] &= \mathbb{E}[(X_i^*)^2 \mid X_1, \ldots, X_n] - \mathbb{E}[X_i^* \mid X_1, \ldots, X_n]^2 \\ &= \frac{1}{n}\sum_{j=1}^{n} X_j^2 - (\bar{X}_n)^2 = \hat{\sigma}_n^2.\end{aligned}$$

5. The unconditional variance is σ^2, using the law of total variance:

$$\begin{aligned}\mathbb{V}[X_i^*] &= \mathbb{E}\left[\mathbb{V}\left[X_i^* \mid X_1, \ldots, X_n\right]\right] + \mathbb{V}\left[\mathbb{E}\left[X_i^* \mid X_1, \ldots, X_n\right]\right] \\ &= \mathbb{E}[\hat{\sigma}_n^2] + \mathbb{V}[\bar{X}_n] = \frac{n-1}{n}\sigma^2 + \frac{1}{n}\sigma^2 = \sigma^2.\end{aligned}$$

2 Bootstrap confidence intervals

2.1 Method 1: the normal interval (simplest)

If $\hat{\theta}$ is approximately normal, then we can construct CI's as before, but now using $\sqrt{v_{\text{boot}}}$ as our estimate of the standard error. In other words,

$$[\hat{\theta} \pm z_{\alpha/2}\sqrt{v_{\text{boot}}}]$$

is a $1-\alpha$ CI. For example, $z_{\alpha/2} = 1.96$ if $\alpha = 5\%$.

But what if $\hat{\theta}$ is not Gaussian?

2.2 Method 2: pivotal interval (preferred)

TLDR: see Definition 2.1.

For maximal clarity, in this section we'll use the following notation: let θ be the ground truth (a number), $\hat{\theta}$ be a *random variable*, and $\hat{\theta}^{(0)}$ be a draw from $\hat{\theta}$, i.e. $\hat{\theta}^{(0)}$ is also just a number. To explain the intuition for the pivotal interval, we start with the following

Thought experiment: say we know both θ and the cdf F of $\hat{\theta}$.

To construct a $(1-\alpha)$ confidence interval for θ, we find t such that

$$\mathbb{P}(\theta - t \le \hat{\theta} \le \theta + t) \ge 1 - \alpha.$$

The narrowest confidence interval takes t to be the smallest possible; specifically, we define

$$t_\alpha = \text{smallest } t \text{ such that } F(\theta + t) - F(\theta - t) \geq 1 - \alpha. \tag{4}$$

Next, note that

$$\mathbb{P}(\theta - t_\alpha \leq \hat{\theta} \leq \theta + t_\alpha) = \mathbb{P}(\hat{\theta} - t_\alpha \leq \theta \leq \hat{\theta} + t_\alpha).$$

So the *random* interval

$$(\hat{\theta} - t_\alpha, \hat{\theta} + t_\alpha)$$

is a $1 - \alpha$ confidence interval for θ, meaning $1 - \alpha$ proportion of the times you draw $\hat{\theta}$ out of a hat, the point θ will be covered by the interval $(\hat{\theta} - t_\alpha, \hat{\theta} + t_\alpha)$.

In particular, for our single draw $\hat{\theta}^{(0)}$ of $\hat{\theta}$, we get a single draw of the confidence interval, namely $(\hat{\theta}^{(0)} - t_\alpha, \hat{\theta}^{(0)} + t_\alpha)$. Now, note from (4) that t_α depends on both F and θ! Reminding ourselves of this dependence, we write the confidence interval we have produced as follows:

$$\text{CI} = (\hat{\theta}^{(0)} - t_\alpha(F, \theta),\ \hat{\theta}^{(0)} + t_\alpha(F, \theta)). \tag{5}$$

Step 1: removing the assumption of knowing θ.

We don't know θ, but we have observed a single sample $\hat{\theta}^{(0)}$ of θ. If the random variable $\hat{\theta}$ is a good estimator for θ (say, it's unbiased, i.e. $\mathbb{E}[\hat{\theta}] = \theta$ and has small variance), then we can reasonably expect that $\hat{\theta}^{(0)}$ is close to θ. So we simply replace θ by $\hat{\theta}^{(0)}$ in (5), which gives

$$\text{CI} = (\hat{\theta}^{(0)} - t_\alpha(F, \hat{\theta}^{(0)}),\ \hat{\theta}^{(0)} + t_\alpha(F, \hat{\theta}^{(0)})) \tag{6}$$

In other words, we redefine t_α to be

$$t_\alpha = \text{smallest } t \text{ such that } F(\hat{\theta}^{(0)} + t) - F(\hat{\theta}^{(0)} - t) \geq 1 - \alpha. \tag{7}$$

Step 2: removing the assumption of knowing F.

This is where our bootstrap samples come in. We have drawn B bootstrap samples $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}$. We can use these samples to construct an empirical CDF:

$$\hat{F}(s) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}(\hat{\theta}^{(b)} \leq s).$$

We now simply use $\hat{F}$ in place of F in the definition (7) of t_α:

$$t_\alpha = \text{smallest } t \text{ such that } \hat{F}(\hat{\theta}^{(0)} + t) - \hat{F}(\hat{\theta}^{(0)} - t) \geq 1 - \alpha. \tag{8}$$

But note that

$$\hat{F}\left(\hat{\theta}^{(0)}+t\right)-\hat{F}\left(\hat{\theta}^{(0)}-t\right)=\frac{1}{B}\sum_{b=1}^{B}\mathbb{1}\left(\hat{\theta}^{(0)}-t\leq\hat{\theta}^{(b)}\leq\hat{\theta}^{(0)}+t\right), \tag{9}$$

i.e. the proportion of bootstrap samples that lie between $\hat{\theta}^{(0)}-t$ and $\hat{\theta}^{(0)}+t$. To summarize, the t_α used in the pivotal confidence interval is given by

$$t_\alpha = \text{smallest } t \text{ such that } \frac{1}{B}\sum_{b=1}^{B}\mathbb{1}\left(\hat{\theta}^{(0)}-t\leq\hat{\theta}^{(b)}\leq\hat{\theta}^{(0)}+t\right)\geq 1-\alpha. \tag{10}$$

Definition 2.1: Pivotal confidence interval

Given a sample $\hat{\theta}^{(0)}$ of the random variable $\hat{\theta}$, and B bootstrap samples $\hat{\theta}^{(1)},\dots,\hat{\theta}^{(B)}$, the $(1-\alpha)$ pivotal CI is given by

$$(\theta^{(0)}-t_\alpha,\ \theta^{(0)}+t_\alpha), \tag{11}$$

where t_α is as defined in (10).

Exercise: check that the CI in Definition 2.1 is equivalent to (8.6) in the book.

2.3 Method 3: percentile interval

The percentile CI is

$$\text{CI} = (q^*_{\alpha/2}, q^*_{1-\alpha/2}),$$

where $q^*_{\alpha/2}$ and $q^*_{1-\alpha/2}$ are quantiles of the empirical bootstrap distribution, i.e. $q^*_{\alpha/2}$ is the point such that

$$\frac{1}{B}\sum_{b=1}^{B}\mathbb{1}(\theta^{(b)}\leq q^*_{\alpha/2})=\alpha/2,$$

and similarly for $q^*_{1-\alpha/2}$. Note that the original, true sample $\hat{\theta}^{(0)}$ does not get used at all!

Remark.

The jackknife is a precursor to the bootstrap, due to Quenouille in 1949. The idea is the following: given data $X_1, \ldots, X_n$, define

$$\hat{\theta}^{(-i)} = \hat{\theta}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n),$$

removing X_i. We have n such values: $\hat{\theta}^{(-1)}, \ldots, \hat{\theta}^{(-n)}$. We let

$$\bar{\hat{\theta}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}^{(-i)},$$

and we take

$$v_{\text{jack}} = \frac{n-1}{n} \sum_{i=1}^{n} \left(\hat{\theta}^{(-i)} - \bar{\hat{\theta}} \right)^2$$

as our variance estimator.

Exercise: Suppose $\hat{\theta} = \bar{X}_n$, for which we know the true variance is $\mathbb{V}[\hat{\theta}] = \sigma^2/n$. Show that $\bar{\hat{\theta}} = \bar{X}_n$ as well, and that

$$v_{\text{jack}} = \frac{\hat{\sigma}^2}{n},$$

where

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

is the debiased sample variance.