STAT 460/560 Class 8: Consistency of the bootstrap

Ben Bloem-Reddy

Reading: Chapter 23.1-23.2, [van98].

Today we'll establish the consistency of bootstrap methods, and in particular, bootstrap confidence intervals. Extending the empirical CDF to more general settings, for n i.i.d. observations $X_1, \dots, X_n$ taking values in $\mathbb{R}^d$, $\hat{P}_n$ is the empirical measure, defined as

$$\hat{P}_n(A) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(A) \;, \quad A \subset \mathbb{R}^d \;.$$

As with the empirical CDF, sampling from $\hat{P}_n$ is just sampling with replacement from $X_1, \dots, X_n$.

The asymptotics today may seem a little strange, because many of the statements will be something like "$S_n \rightsquigarrow T$, conditionally almost surely." What this means is that for P-almost every sequence $X_1, X_2, \dots$, the conditional distribution of S_n given $X_1, \dots, X_n$ converges to the distribution of T. Just think of this as the sequence of conditional distributions converging, but keep in mind that a conditional distribution is a random distribution (it is a function of the random variables on which it conditions), which makes the "almost surely" necessary.

Before we get to the bootstrap, we need an intermediate result, the proof of which can be found as Lemma 21.2 in [van98]. Recall that if F is a CDF, the **quantile function**, denoted F^{-1}, is the generalized inverse of F,

$$F^{-1}(p) = \inf\{x \colon F(x) \geq p\} \;.$$

If F is continuous and strictly increasing then this is just the usual inverse. The generalized part comes in when F has jumps, flat spots, etc.

Lemma 8.1 (Lemma 21.2, [van98])**.** *Let F_n be any sequence of CDFs and F another CDF. $F_n^{-1}(p) \to F^{-1}(p)$ at every p where F^{-1} is continuous if and only if $F_n(x) \to F(x)$ at every x where F is continuous.*

Note that this kind of convergence is equivalent to $X_n \rightsquigarrow X$ if F_n is the CDF of X_n and F is the CDF of X. For that reason, we write $F_n \rightsquigarrow F$ (but this is a convenient overloading of notation).

1. Consistency of interval from convergence of estimator

Recall that the bootstrap conditions on an estimate $\hat{P}_n$ of P (formed from the sample), and generates a bootstrap sample $X_1^*, \dots, X_n^*$. In practice, we'll simulate the bootstrap sample many times to estimate things like the distribution of $(\hat{\theta}_n^* - \hat{\theta}_n)/\hat{\sigma}_n^*$. For the purposes of analysis, we'll assume that we have access to the actual bootstrap distribution (i.e., what we would obtain in the limit $B \to \infty$).

To form the bootstrap interval, we will estimate quantiles of $(\hat{\theta}_n - \theta)/\sigma$ by quantiles of the bootstrap quantity $(\hat{\theta}_n^* - \hat{\theta}_n)/\sigma_n^*$. In particular, let

$$\hat{F}_n(x) := P\left(\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\sigma_n^*} \leq x \mid \hat{P}_n\right) \;, \tag{8.1}$$

with $\hat{\xi}_{n,1-\alpha} = \inf\{x\colon \hat{F}_n(x) \geq 1-\alpha\}$. This is a random number, where the randomness comes from conditioning on $\hat{P}_n$. If T is a random variable with CDF F, we say that $(\hat{\theta}_n^* - \hat{\theta}_n)/\hat{\sigma}_n^* \rightsquigarrow T$ conditionally on $X_1, X_2, \dots$ if $\hat{F}_n \rightsquigarrow F$ conditioned on $X_1, X_2, \dots$.

Using the estimated quantile, we can form an approximate interval as

$$\hat{I}_n(\alpha,\beta) = [\hat{\theta}_n - \hat{\xi}_{n,\beta}\hat{\sigma}_n, \hat{\theta}_n - \hat{\xi}_{n,1-\alpha}\hat{\sigma}_n] \ . \tag{8.2}$$

A sufficient condition for consistency of the bootstrap intervals is that the bootstrap quantities converge in distribution to the same place as the non-bootstrap quantities.

Theorem 8.2. *Suppose that $(\hat{\theta}_n - \theta)/\hat{\sigma}_n \rightsquigarrow T$, and that for almost every sequence $X_1, X_2, \dots,$, $(\hat{\theta}_n^* - \hat{\theta}_n)/\sigma_n^* \rightsquigarrow T$ conditionally on $X_1, X_2, \dots$, for a random variable T with a continuous CDF. Then*

$$P(\hat{I}_n(\alpha,\beta)) \to 1-\alpha-\beta \ . \tag{8.3}$$

Before going through the proof, we have made a couple of assumptions that are not strictly necessary, but that simplify the proof. The first is that we assume that F is a continuous function. This rules out having to deal with CDFs that can differ on a null set of discontinuities (a headache that doesn't offer additional insight). The second is that we assume that convergence in distribution $(\hat{\theta}_n^* - \hat{\theta}_n)/\sigma_n^* \rightsquigarrow T$ holds almost surely, conditionally on $X_1, X_2, \dots$. This can be weakened to convergence in probability by arguing along subsequences but again, that doesn't offer additional insight.

Proof. By assumption, $\hat{F}_n \rightsquigarrow F$, where F is the CDF of T. By Lemma 8.1, $F_n^{-1}(p) \to F^{-1}(p)$ at every continuity point of F^{-1}. Choose α so that $1-\alpha$ is such a continuity point. Then $\hat{\xi}_{n,1-\alpha} = F_n^{-1}(1-\alpha)$ converges almost surely to $F^{-1}(1-\alpha)$. Hence, by Slutsky's lemma,

$$(\hat{\theta}_n - \theta)/\hat{\sigma}_n - \hat{\xi}_{n,1-\alpha} \rightsquigarrow T - F^{-1}(1-\alpha) \ .$$

Therefore,

$$P(\theta \geq \hat{\theta}_n - \hat{\sigma}_n\hat{\xi}_{n,1-\alpha}) = P\left(\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \leq \hat{\xi}_{n,1-\alpha}\right) \to P(T \leq F^{-1}(1-\alpha)) = 1-\alpha \ .$$

This must hold for all but at most countably many values of α. Both the LHS and the RHS of the previous math display are monotone functions of α and the RHS is continuous in α, the convergence must hold for every α. □

With this in hand, showing that bootstrap confidence intervals are consistent amounts to showing that our statistic of interest converges in distribution, and that the bootstrap version of our statistic of interest converges in distribution to the same place.

Theorem 8.3. *Let $X_1, X_2, \dots$ be i.i.d. random variables in $\mathbb{R}$ with mean μ and variance σ^2. Then conditionally on $X_1, X_2, \dots$, for almost every sequence $X_1, X_2, \dots$,*

$$\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \rightsquigarrow \mathcal{N}(0, \sigma^2) \ .$$

Activity 8.1. Show that

$$E[X_i^*|\hat{P}_n] = \bar{X}_n \ , \quad \text{and} \quad E[(X_i^* - \bar{X}_n)^2|\hat{P}_n] = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}_n^2 \ .$$

Solution: Conditioning on $\hat{P}_n$ is the same as conditioning on $X_1, \dots, X_n$ but ignoring the order of the observations. Since, conditioned on $\hat{P}_n$, X_i^* is generated by sampling uniformly at random from

$X_1, \ldots, X_n$, there is the general formula

$$E[f(X_i^*)|\hat{P}_n] = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \ .$$

Another way to view this is that $\hat{P}_n$ is a discrete probability measure that puts probability $1/n$ at each of $X_1, \ldots, X_n$. Applying the general formula to $E[X_i^*|\hat{P}_n]$ and to $E[(X_i^* - \bar{X}_n)^2|\hat{P}_n]$ yields the stated results.

By the (strong) law of large numbers, these converge almost surely to μ and to σ^2, respectively. In other words, for P-almost every sequence $X_1, X_2, \ldots$, $\bar{X}_n \to \mu$ and similarly for the variance.

Showing the asymptotic normality of $\bar{X}_n^*$ requires a strong version of the CLT that is suitable for the case in which the X_i^*'s are sampled from a different distribution for each n. In particular, the Lindeberg–Feller CLT applies to *triangular arrays*, and although van der Vaart [van98, Prop. 2.27] calls it "the simplest extension of the classical central limit theorem," it is extremely useful. An infinite triangular array is formed by taking a vector Y_n of length k_n, putting it in the n-th row, and appending infinitely many zeros. In the theorem, the entries in each row are assumed to be independent, but there may be dependence between rows. It can also be extended to the situation in which each entry of the array is actually a vector of fixed length [see van98, Prop. 2.27].

Theorem 8.4 (Lindeberg–Feller CLT)**.** *For each $n \geq 1$, let $(Y_{n,1}, \ldots, Y_{n,k_n})$ be a vector of independent random variables with finite variances and for every $\epsilon > 0$,*

$$\sum_{i=1}^{k_n} E[\|Y_{n,i}\|^2 \mathbf{I}\{\|Y_{n,i}\| > \epsilon\}] \to 0 \quad \textit{as } n \to \infty \ .$$

Moreover, assume that $\sum_{i=1}^{k_n} \mathit{Var}(Y_{n,i}) \to \sigma^2$. Then as $n \to \infty$,

$$\sum_{i=1}^{k_n} (Y_{n,i} - E[Y_{n,i}]) \rightsquigarrow \mathcal{N}(0, \sigma^2) \ .$$

Back to our bootstrap analysis, $\sqrt{n}\bar{X}_n^* = \sum_{i=1}^n \frac{1}{\sqrt{n}} X_{n,i}^* = \sum_{i=1}^n Y_{n,i}$, where we have defined $Y_{n,i} = \frac{1}{\sqrt{n}} X_{n,i}^*$, and $X_{n,i}^*$ is the i-th bootstrap sample from $\hat{P}_n$. We just need to check the conditions of the Lindeberg–Feller CLT. To that end, fix arbitrary $\epsilon > 0$, and consider

$$\begin{aligned}
\sum_{i=1}^{n} E\left[\|Y_{n,i}\|^2 \mathbf{I}\{\|Y_{n,i}\| > \epsilon\} \mid \hat{P}_n\right] &= \sum_{i=1}^{n} E\left[\frac{1}{n}\|X_{n,i}^*\|^2 \mathbf{I}\{\|X_{n,i}^*\| > \sqrt{n}\epsilon\} \mid \hat{P}_n\right] \\
&= \frac{n}{n}\frac{1}{n}\sum_{i=1}^{n} \|X_{n,i}\|^2 \mathbf{I}\{\|X_{n,i}\| > \sqrt{n}\epsilon\} \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \|X_{n,i}\|^2 \mathbf{I}\{\|X_{n,i}\| > M\} \xrightarrow{\text{a.s.}} E[\|X_{n,i}\|^2 \mathbf{I}\{\|X_{n,i}\| > M\}] \ ,
\end{aligned}$$

for $M \leq \sqrt{n}\epsilon$, and where the convergence follows from the strong law of large numbers.

Now, we assumed that $\mathrm{Var}[X_i] = \sigma^2$, which implies that

$$\lim_{M \to \infty} E[\|X_{n,i}\|^2 \mathbf{I}\{\|X_{n,i}\| > M\}] = 0 \ . \tag{8.4}$$

(See the optional section on integrability below for details.) Hence, for each $\eta > 0$, there is some M_η such that

$$E[\|X_{n,i}\|^2 \mathbf{I}\{\|X_{n,i}\| > M_\eta\}] < \eta \ ,$$

which implies that for n large enough,

$$\sum_{i=1}^{n} E\left[\|Y_{n,i}\|^2 \mathbf{I}\{\|Y_{n,i}\| > \epsilon\} \mid \hat{P}_n\right] < \eta \,,$$

for P-almost every sequence $X_1, X_2, \ldots$. That shows that the first condition of the Lindeberg–Feller CLT is satisfied for P-almost every sequence.

Finally, using the activity above and the strong law of large numbers,

$$\begin{aligned}
\sum_{i=1}^{k_n} \mathrm{Var}(Y_{n,i}) &= \frac{1}{n}\sum_{i=1}^{n} E[(X_i^* - \bar{X}_n)^2 | \hat{P}_n] \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2 \xrightarrow{\text{a.s.}} \sigma^2 \,.
\end{aligned}$$

Hence, the conclusion of the Lindeberg–Feller CLT applied here is that

$$\sum_{i=1}^{k_n} (Y_{n,i} - E[Y_{n,i} \mid \hat{P}_n]) = \sqrt{n}(\bar{X}_n^* - \bar{X}_n) \rightsquigarrow \mathcal{N}(0, \sigma^2)$$

holds for P-almost every sequence $X_1, X_2, \ldots$.

Theorem 23.5 in [van98] extends this substantially by proving the validity of the delta method for the bootstrap.

Integrability (optional)

A random variable X is *integrable* if $\mathbb{E}[\|X\|] < \infty$. An equivalent condition [see, e.g., Çin11, Lemma 3.10] is that

$$\lim_{b\to\infty} \mathbb{E}[\|X\| \mathbf{I}\{\|X\| > b\}] = 0 \,.$$

Applied to the situation above, (8.4) holds because $\mathbb{E}[X_i^2] = \mu^2 + \sigma^2 < \infty$ (i.e., it is integrable).

References

[Çin11] E. Çinlar. *Probability and Stochastics*. Springer New York, 2011.

[van98] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.