36-705: Intermediate Statistics **Fall 2019**

Lecture 27: November 6

Lecturer: Siva Balakrishnan

Today we will continue discussing the bootstrap, and then try to understand why it works in a simple case. As we discussed in the last lecture the basic idea of the bootstrap is to construct an estimate $\widehat{P}$ of the distribution we obtain samples from. We then sample new datasets from $\widehat{P}$ instead, i.e. we construct datasets:

$$\begin{aligned} S_1^* &= \{X_{11}^*, \ldots, X_{n1}^*\} \sim \widehat{P} \\ S_2^* &= \{X_{12}^*, \ldots, X_{n2}^*\} \sim \widehat{P} \\ &\vdots \\ S_m^* &= \{X_{1m}^*, \ldots, X_{nm}^*\} \sim \widehat{P} \end{aligned}$$

for some large value m, and compute our estimator on each dataset $\{\widehat{\theta}^*(S_1^*), \ldots, \widehat{\theta}^*(S_m^*)\}$. To approximate the distribution of $\widehat{\theta}-\theta$ (which we do not have access to) we use the distribution of $\widehat{\theta}^* - \widehat{\theta}$. For the rest of today we will focus on the empirical bootstrap (i.e. we will take $\widehat{P}$ to be P_n).

As a preliminary to understanding the bootstrap let us consider a slightly simpler idea first.

27.1 Bootstrap variance estimate

To understand the idea, let us first consider the Monte-Carlo variance estimate. Suppose we had an estimator $\widehat{\theta}_n = g(X_1, \ldots, X_n)$ (this could be a complicated function), where $X_1, \ldots, X_n \sim P$ and we want to estimate $\text{Var}_P(\widehat{\theta}_n)$.

Supposing that we knew P we could try to compute the variance analytically: this might be difficult. The Monte-Carlo variance estimate would be to instead draw B samples of size n from P, i.e. we draw, $\{X_{11}, \ldots, X_{1n}\}, \ldots, \{X_{B1}, \ldots, X_{Bn}\} \sim P$, to compute our estimator on each of these samples, i.e. compute $\widehat{\theta}_n^{(1)}, \ldots, \widehat{\theta}_n^{(B)}$ and then use the sample variance, i.e.

$$\widehat{\sigma}_n^2 = \frac{1}{B}\sum_{i=1}^{B}\left(\widehat{\theta}_n^{(i)}\right)^2 - \left(\frac{1}{B}\sum_{i=1}^{B}\widehat{\theta}_n^{(i)}\right)^2.$$

By the LLN we have that $\widehat{\sigma}^2 \xrightarrow{p} \text{Var}_P(\widehat{\theta}_n)$. Unfortunately, we typically do not know P.

By now, you have already guessed the idea behind the bootstrap. The idea is to replace P in the above procedure by the empirical distribution P_n. Here is the algorithm:

Bootstrap Variance Estimator

1. Draw a bootstrap sample $X_1^*, \ldots, X_n^* \sim P_n$. Compute $\widehat{\theta}_n^* = g(X_1^*, \ldots, X_n^*)$.
2. Repeat the previous step, B times, yielding estimators $\widehat{\theta}_{n,1}^*, \ldots, \widehat{\theta}_{n,B}^*$.
3. Compute:
$$\widehat{s}^2 = \frac{1}{B} \sum_{j=1}^{B} (\widehat{\theta}_{n,j}^* - \overline{\theta})^2,$$
where $\overline{\theta} = \frac{1}{B} \sum_{j=1}^{B} \widehat{\theta}_{n,j}^*$.
4. Output $\widehat{s}^2$.

Now, let us reason about this algorithm in a very simple case. Suppose our estimator is just the sample mean, i.e.:

$$\widehat{\theta}(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

then the bootstrap is using simulation to compute $\mathrm{Var}_{P_n}(\widehat{\theta})$. Let us calculate what this is:

$$\mathrm{Var}_{P_n}(\widehat{\theta}) = \mathbb{E}_{X_1^*, \ldots, X_n^* \sim P_n} \left(\frac{1}{n} \sum_{i=1}^{n} X_i^* \right)^2 - \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^2.$$

Let us denote $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then,

$$\begin{aligned}
\mathrm{Var}_{P_n}(\widehat{\theta}) &= \mathbb{E}_{X_1^*, \ldots, X_n^* \sim P_n} \left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} X_i^* X_j^* \right) - \overline{X}^2 \\
&= \frac{1}{n} \mathbb{E}_{X^* \sim P_n} (X^*)^2 + \frac{n-1}{n} \overline{X}^2 - \overline{X}^2 \\
&= \frac{1}{n} \left[\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \overline{X}^2 \right] = \frac{\widehat{\sigma}^2}{n}.
\end{aligned}$$

So we see that, the bootstrap is using simulation to compute $\mathrm{Var}_{P_n}(\widehat{\theta})$ which for the sample mean is precisely our usual estimate of the variance of the sample mean, i.e. $\widehat{\sigma}^2/n$.

27.2 Bootstrap Confidence Intervals

The bootstrap can also be used to obtain confidence intervals. If your estimator has a normal limit then you could just use a Wald interval with the bootstrap variance estimate, i.e. $C_n = [\widehat{\theta}_n - \widehat{s} z_{\alpha/2}, \widehat{\theta}_n + \widehat{s} z_{\alpha/2}]$.

It is often more accurate to use the distribution of the bootstrap estimates itself to construct the bootstrap confidence interval.

27.2.1 Hypothetical confidence interval

Suppose we knew the distribution of our estimator, in particular suppose we knew the distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$. Let us denote the distribution by G and denote its $\alpha/2$ and $1 - \alpha/2$ quantiles by $g_{\alpha/2}$ and $g_{1-\alpha/2}$.

Then a $1 - \alpha$ confidence interval would be:

$$C_n = \left[\widehat{\theta}_n - \frac{g_{1-\alpha/2}}{\sqrt{n}}, \widehat{\theta}_n - \frac{g_{\alpha/2}}{\sqrt{n}} \right].$$

This might seem a little strange, but this is probably because you are used to confidence intervals based on the normal distribution which has symmetric quantiles. To verify this,

$$\begin{aligned}
\mathbb{P}(\theta \in C_n) &= \mathbb{P}\left(g_{\alpha/2} \leq \sqrt{n}(\widehat{\theta}_n - \theta) \leq g_{1-\alpha/2} \right) \\
&= 1 - \alpha/2 - \alpha/2 = 1 - \alpha.
\end{aligned}$$

Again the point is that we do not know the distribution G above so we try to approximate this using the bootstrap.

27.2.2 Bootstrap confidence interval algorithm

Bootstrap Confidence Interval

1. Draw a bootstrap sample $X_1^*, \ldots, X_n^* \sim P_n$. Compute $\widehat{\theta}_n^* = g(X_1^*, \ldots, X_n^*)$.
2. Repeat the previous step, B times, yielding estimators $\widehat{\theta}_{n,1}^*, \ldots, \widehat{\theta}_{n,B}^*$.
3. Let
$$\widehat{G}(t) = \frac{1}{B} \sum_{j=1}^{B} I\Big(\sqrt{n}(\widehat{\theta}_{n,j}^* - \widehat{\theta}_n\Big) \leq t).$$
4. Let
$$C_n = \left[\widehat{\theta}_n - \frac{g_{1-\alpha/2}}{\sqrt{n}},\ \widehat{\theta}_n - \frac{g_{\alpha/2}}{\sqrt{n}}\right]$$
where $g_{\alpha/2} = \widehat{G}^{-1}(\alpha/2)$ and $g_{1-\alpha/2} = \widehat{G}^{-1}(1-\alpha/2)$.
5. Output C_n.

27.3 Justifying the Bootstrap

This part is going to be a little bit technical. Before we get into it, we should try to figure out what it means to "justify the bootstrap". Roughly, we want that the quantiles of the bootstrap distribution of our statistic should be close to the quantiles its actual distribution, i.e. suppose we define:

$$\widehat{F}_n(t) = \mathbb{P}_n(\sqrt{n}(\widehat{\theta}_n^* - \widehat{\theta}_n) \geq t | X_1, \ldots, X_n),$$

to be the CDF of the bootstrap distribution, and

$$F_n(t) = \mathbb{P}(\sqrt{n}(\widehat{\theta}_n - \theta) \geq t),$$

to be the CDF of the true sampling distribution of our statistic, then the bootstrap works if for instance:

$$\sup_t |\widehat{F}_n(t) - F_n(t)| \to 0.$$

This turns out to be true in quite a bit of generality, only requiring mild conditions (Hadamard differentiability, see Bootstrap chapter in van der Vaart), but we will prove it in the simplest case: when $\widehat{\theta}_n$ is a sample mean. In this case there are much simpler ways to construct confidence intervals (using Normal approximations) but that is not really the point.

Suppose that $X_1, \ldots, X_n \sim P$ where X_i has mean μ and variance σ^2. Suppose we want to construct a confidence interval for μ.

Let $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and define

$$F_n(t) = \mathbb{P}(\sqrt{n}(\widehat{\mu}_n - \mu) \leq t). \tag{27.1}$$

We want to show that

$$\widehat{F}_n(t) = \mathbb{P}\Big(\sqrt{n}(\widehat{\mu}_n^* - \widehat{\mu}_n) \leq t \,\Big|\, X_1, \ldots, X_n\Big)$$

is close to F_n.

Theorem 27.1 (Bootstrap Theorem) *Suppose that $\mu_3 = \mathbb{E}|X_i|^3 < \infty$. Then,*

$$\sup_t |\widehat{F}_n(t) - F_n(t)| = O_P\left(\frac{1}{\sqrt{n}}\right).$$

To prove this result, let us recall that Berry-Esseen Theorem.

Theorem 27.2 (Berry-Esseen Theorem) *Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^2. Let $\mu_3 = \mathbb{E}[|X_i - \mu|^3] < \infty$. Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ be the sample mean and let Φ be the cdf of a $N(0,1)$ random variable. Let $Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$. Then*

$$\sup_z \Big|\mathbb{P}(Z_n \leq z) - \Phi(z)\Big| \leq \frac{33}{4}\frac{\mu_3}{\sigma^3\sqrt{n}}. \tag{27.2}$$

Proof of the Bootstrap Theorem. Let $\Phi_\sigma(t)$ denote the cdf of a Normal with mean 0 and variance σ^2. Let $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \widehat{\mu}_n)^2$. Thus, $\widehat{\sigma}^2 = \mathrm{Var}(\sqrt{n}(\widehat{\mu}_n^* - \widehat{\mu}_n)|X_1, \ldots, X_n)$. Now, by the triangle inequality,

$$\begin{aligned}\sup_t |\widehat{F}_n(t) - F_n(t)| &\leq \sup_t |F_n(t) - \Phi_\sigma(t)| + \sup_t |\Phi_\sigma(t) - \Phi_{\widehat{\sigma}}(t)| + \sup_t |\widehat{F}_n(t) - \Phi_{\widehat{\sigma}}(t)| \\ &= \mathrm{I} + \mathrm{II} + \mathrm{III}.\end{aligned}$$

Let $Z \sim N(0,1)$. Then, $\sigma Z \sim N(0, \sigma^2)$ and from the Berry-Esseen theorem,

$$\begin{aligned}\mathrm{I} = \sup_t |F_n(t) - \Phi_\sigma(t)| &= \sup_t \left|\mathbb{P}\left(\sqrt{n}(\widehat{\mu}_n - \mu) \leq t\right) - \mathbb{P}\left(\sigma Z \leq t\right)\right| \\ &= \sup_t \left|\mathbb{P}\left(\frac{\sqrt{n}(\widehat{\mu}_n - \mu)}{\sigma} \leq \frac{t}{\sigma}\right) - \mathbb{P}\left(Z \leq \frac{t}{\sigma}\right)\right| \leq \frac{33}{4}\frac{\mu_3}{\sigma^3\sqrt{n}}.\end{aligned}$$

Using the same argument on the third term, we have that

$$\mathrm{III} = \sup_t |\widehat{F}_n(t) - \Phi_{\widehat{\sigma}}(t)| \leq \frac{33}{4}\frac{\widehat{\mu}_3}{\widehat{\sigma}^3\sqrt{n}}$$

$$\begin{tikzcd}
F_n \arrow[r, "O(1/\sqrt{n})"] \arrow[d, dashed] & L \arrow[d, "O_P(1/\sqrt{n})"] \\
\widehat{F}_n \arrow[r, "O_P(1/\sqrt{n})"'] \arrow[d, "O(1/\sqrt{B})"] & \widehat{L} \\
\overline{F} &
\end{tikzcd}$$

Figure 27.1: *The distribution $F_n(t) = \mathbb{P}(\sqrt{n}(\widehat{\theta}_n - \theta) \leq t)$ is close to some limit distribution L. Similarly, the bootstrap distribution $\widehat{F}_n(t) = \mathbb{P}(\sqrt{n}(\widehat{\theta}^*_n - \widehat{\theta}_n) \leq t | X_1, \ldots, X_n)$ is close to some limit distribution $\widehat{L}$. Since $\widehat{L}$ and L are close, it follows that F_n and $\widehat{F}_n$ are close. In practice, we approximate $\widehat{F}_n$ with its Monte Carlo version $\overline{F}$ which we can make as close to $\widehat{F}_n$ as we like by taking B large.*

where $\widehat{\mu}_3 = \frac{1}{n}\sum_{i=1} |X_i - \widehat{\mu}_n|^3$ is the empirical third moment. By the strong law of large numbers, $\widehat{\mu}_3$ converges almost surely to μ_3 and $\widehat{\sigma}$ converges almost surely to σ. So, almost surely, for all large n, $\widehat{\mu}_3 \leq 2\mu_3$ and $\widehat{\sigma} \geq (1/2)\sigma$ and III $\leq \frac{33}{4}\frac{4\mu_3}{\sqrt{n}}$. From the fact that $\widehat{\sigma} - \sigma = O_P(\sqrt{1/n})$ it may be shown that $\text{II} = \sup_t |\Phi_\sigma(t) - \Phi_{\widehat{\sigma}}(t)| = O_P(\sqrt{1/n})$. (This may be seen by Taylor expanding $\Phi_{\widehat{\sigma}}(t)$ around σ.) This completes the proof. □

So far we have focused on the mean. Similar theorems may be proved for more general parameters. The details are complex so we will not discuss them here.

27.4 Failure of the Bootstrap

As usual when we need a counterexample we try the uniform distribution. Suppose that $X_1, \ldots, X_n \sim U[0, \theta]$ and we try to bootstrap the MLE to construct a confidence interval for θ.

The natural bootstrap confidence interval would have no coverage, even asymptotically, because on each of the bootstrap samples, as well as on the original sample we underestimate θ.