Statistics 210A Lecture 26 Notes

Daniel Raban

November 30, 2021

1 Bootstrap Confidence Intervals and Double Bootstrap

1.1 Recap: Bootstrap methods

Bootstrap is an asymptotic nonparametric method, where we use the empirical distribution as an asymptotic approximation to the true distribution. Anything we want to do with the true distribution, we substitute in the empirical distribution and call it a day.

If we have a nonparametric model $X_1, \dots, X_n \overset{\text{iid}}{\sim} P$ with "parameter" $\theta(P)$ (not necessarily 1 to 1), then we discussed the notion of a plug-in estimator $\widehat{\theta}_n(X) = \theta(\widehat{P}_n)$, where $\widehat{P}_n$ is an estimator of P. A typical choice is the **empirical distribution** $\widehat{P}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. (Note: There ar eother choices, estpeically for non-i.i.d. sampling models, e.g. time series.)

Remark 1.1. Bootstrap is often conflated with permutation tests. They are both nonparametric and involve resampling from the data, but they have very different underlying statistical logic. The permutation test is an exact, finite sample method; if you take enough permutations, you will get the exact conditional distribution of the test statistic under the null hypothesis. On the other hand, bootstrap is an approximation which only becomes accurate asymptotically.

We have seen two bootstrap algorithms so far:

- If $\widehat{\theta}_n(X)$ is any estimator we want, its standard error is

$$\mathrm{s.e.}(\widehat{\theta}_n(X)) = \sqrt{\mathrm{Var}_{X_i \overset{\text{iid}}{\sim} P}(\widehat{\theta}_n(X))},$$

  and the **bootstrap standard error** is

$$\widehat{\mathrm{s.e.}}(\widehat{\theta}_n(X)) = \sqrt{\mathrm{Var}_{X_i^* \overset{\text{iid}}{\sim} \widehat{P}_n}(\widehat{\theta}_n(X^*))}.$$

- If $\widehat{\theta}_n(X)$ is any estimator we want, its bias is

$$\mathrm{Bias}(\widehat{\theta}_n) = \mathbb{E}_{X_i \overset{\text{iid}}{\sim} P}[\widehat{\theta}_n(X)] - \theta(P),$$

and the **bootstrap bias estimator** is

$$\widehat{\text{Bias}}(\widehat{\theta}_n) = \mathbb{E}_{X_i^* \overset{\text{iid}}{\sim} \widehat{P}_n}[\widehat{\theta}_n(X^*)] - \theta(\widehat{P}_n).$$

We also have the **bias corrected bootstrap estimator**

$$\widehat{\theta}_n^{\text{BC}} = \widehat{\theta}_n - \widehat{\text{Bias}}.$$

1.2 Bootstrap confidence intervals

Suppose we want a confidence interval for $\theta(P)$. Instead of inverting a hypothesis test, we can define a random variable $R_n(X, P) = \widehat{\theta}_n(X) - \theta(P)$ for any estimator $\widehat{\theta}_n$; if we know the distribution of R_n, we can construct the confidence interval using a point estimate for R_n.

Define the CDF

$$G_{n,P}(r) = \mathbb{P}_P(\widehat{\theta}_n(X) - \theta(P) \le r).$$

The lower $\alpha/2$ quantile is

$$r_1 = G_{n,P}^{-1}(\alpha/2),$$

and the upper $\alpha/2$ quantile is

$$r_1 = G_{n,P}^{-1}(1 - \alpha/2).$$

Then

$$\begin{aligned} 1 - \alpha &= \mathbb{P}_P(r_1 \le \widehat{\theta}_n - \theta \le r_2) \\ &= \mathbb{P}_P(\theta \in [\widehat{\theta}_n - r_2, \widehat{\theta}_n - r_1]) \end{aligned}$$

The interval we get only depends on $G_{n,P}$.

If we don't know P, then we can use $\widehat{P}_n$ instead:

$$G_{n,\widehat{P}_n}(r) = \mathbb{P}_{X^* \overset{\text{iid}}{\sim} \widehat{P}_n}(\widehat{\theta}(X^*) - \theta(\widehat{P}_n) \le r).$$

This depends only on the sample X. Using this CDF in the above calculation gives us the **bootstrap confidence interval**

$$C_{n,\alpha}(X) = [\widehat{\theta}_n(X) - \widehat{r}_2, \widehat{\theta}_n(X) - \widehat{r}_1],$$

where

$$\widehat{r}_1 = G_{n,\widehat{P}_n}^{-1}(\alpha/2), \qquad \widehat{r}_2 = G_{n,\widehat{P}_n}^{-1}(1 - \alpha/2).$$

Here is the procedure in practice:

1. For $b = 1, \dots, B$, let $X_1^{*b}, \dots, X_n^{*b} \overset{\text{iid}}{\sim} \widehat{P}_n$.

2. For $b = 1, \ldots, B$, let $R_n^{*b} = \widehat{\theta}_n(X^{*b}) - \theta(\widehat{P}_n)$.

3. Return $\widehat{G}_n(r) = \frac{1}{B}\sum_{k=1}^{B} \mathbb{1}_{\{R_n^{*b} \leq r\}}$

4. Invert this to recover $\widehat{r}_1$ and $\widehat{r}_2$.

This is not the only way to make a bootstrap confidence interval. Other examples of estimators we could use bootstrap with for confidence intervals are

- The **studentized root**
$$R_n(X, P) = \frac{\widehat{\theta}_n(X) - \theta(P)}{\widehat{\sigma}(X)}.$$

- The **relative error**
$$R_n(X, P) = \frac{\widehat{\theta}_n(X)}{\theta(P)}.$$

With the studentized root,

$$C_{n,\alpha} = [\widehat{\theta}_n - r_2\widehat{\sigma}, \widehat{\theta}_n - r_1\widehat{\theta}],$$

where we can estimate r_1, r_2 using a the plug-in estimator R_n.

Remark 1.2. Our first version of the bootstrap confidence interval works best when $G_{n,P}$ is not so sensitive to varying P.

1.3 Double bootstrap

Bootstrap is an approximation. Is it a good approximation? Suppose we have, for example, a bootstrap confidence interval

$$C_{n,\alpha} = [\widehat{\theta}_n(X) - \widehat{r}_2(X)\widehat{\sigma}(X), \widehat{\theta}_n(X) - \widehat{r}_1(X)\widehat{\sigma}(X)].$$

What is the probability

$$\mathbb{P}_{X_i \overset{\text{iid}}{\sim} P}(\widehat{\theta}_n(P) \in C_{n,\alpha}(X))?$$

We can use bootstrap to estimate this:

$$\mathbb{P}_{X_i^* \overset{\text{iid}}{\sim} \widehat{P}_n}(\widehat{\theta}_n(\widehat{P}_n) \in C_{n,\alpha}(X^*))?$$

Suppose we estimate that $C_{n,0.1}$ has $\approx 87\%$ coverage, but $C_{n,0.08}$ has $\approx 90\%$ coverage. Then we want the latter confidence interval. In particular, we are using bootstrap to calibrate the confidence level of the confidence interval.

Remark 1.3. We could do this bootstrap "tuning" of α using any confidence interval, not just one that was originally obtained through bootstrap.

Here is how we can implement this α "tuning" in practice:

1. For $a = 1, \ldots, A$, let $X_1^{*a}, \ldots, X_n^{*a} \overset{\text{iid}}{\sim} \widehat{P}_n$.
2. Calculate $C_{n,\alpha'}(X^{*a})$ for α' in some grid (try $\alpha' = 10\%, 9\%, 8\%$, etc.) using whatever method you are using to obtain a confidence interval (bootstrap or not).

   We can specify this in particular for the double bootstrap:

   (a) Let $\widehat{P}_n^{*a} = \frac{1}{n}\sum_{i=1}^n \delta_{X_i^*}$.

   (b) For $b = 1, \ldots, B$,

      i. Let $X_1^{**a,b}, \ldots, X_n^{**a,b} \overset{\text{iid}}{\sim} \widehat{P}_n^{*a}$.

      ii. Let
      $$R_n^{**a,b} = \frac{\widehat{\theta}_n(X^{**a,b}) - \theta(\widehat{P}_n^{*a})}{\widehat{\sigma}(X^{**a,b})}.$$

   (c) Let $G_n^{*a} = \mathrm{ecdf}(R_n^{**a,1}, \ldots, R_n^{**a,B})$.

   (d) For α' in the grid, let
   $$C_{n,\alpha'}(X^{*a}) = [\widehat{\theta}_n^* - \widehat{\sigma}^{*a}\widehat{r}_2(G_n^{*a}), \widehat{\theta}_n^* - \widehat{\sigma}^{*a}\widehat{r}_1(G_n^{*a})].$$
3. For α' in this grid, let
   $$\widehat{\mathrm{Coverage}}(\alpha') = \frac{1}{A}\sum_{a=1}^A \mathbb{1}_{\{C_{n,\alpha'}(X^{*a}) \ni \theta(\widehat{P}_n)\}}.$$
4. Take $\widehat{\alpha} = \max\{\alpha' : \widehat{\mathrm{Coverage}}(\alpha') \geq 1 - \alpha\}$, and return $C_{n,\widehat{\alpha}}(X)$.

Remark 1.4. This seems like circular logic, where this method will suffer from the same issues as the original bootstrap confidence interval. The heuristic idea is that the double bootstrap confidence interval may be less sensitive to changes in P than the original confidence interval.