Lab 8

Bayesian Bootstrap

Why do we need to do Bootstrap

Recall from lab 7 that we want to calculate the ATE:

$$\Psi := ATE = E_Z\left[E_{Y|Z,X}\left[Y_i \mid X_i = 1, Z_i\right] - E_{Y|Z,X}\left[Y_i \mid X_i = 0, Z_i\right]\right]$$

Note that here X is the treatment variables, where Z is the confounders. We know from the lecture and previous labs that under a given model, we can express $E_{Y|Z,X}\left[Y_i \mid X_i, Z_i\right]$ in a regression function, whether that is parametric or not. For simplicity we will consider this parametric model:

$$E_{Y|Z,X}\left[Y_i \mid X_i, Z_i\right] = \beta_0 + \beta_1 X_i + \beta_1 X_i Z_i$$

we will see that

$$E_{Y|Z,X}\left[Y_i \mid X_i = 1, Z_i\right] = \beta_0 + \beta_1 + \beta_2 Z_i$$

and

$$E_{Y|Z,X}\left[Y_i \mid X_i = 0, Z_i\right] = \beta_0$$

Thus

$$\Psi = ATE = \beta_1 + \beta_2 E_Z[Z]$$

Now if we place priors on $\beta_0, \beta_1, \beta_2$ we can get posterior draws (say, size M) for these parameters : $\{\beta_0^{(m)}, \beta_1^{(m)}, \beta_2^{(m)}\}_{m=1}^{M}$. If we know the distribution of the confounder Z, than we can conduct inference of ATE by generating a series of posterior draws of $\Psi^{(m)} = \beta_1^{(m)} + \beta_2^{(m)} E_Z[Z]$. The problem however is that in regression models we tend to treat the confounders as fixed and known and usually do not specifiy a distribution in modeling stage. The requires us to use a nonparametric method (nonparametric in that we do not compute $E[Z]$ based on parametric assumptions of the distribution of Z).

- **Remark**: the process of 'integrating over Z' is called **standardization** in causal inference. Bayesian bootstrapping is one way of doing standardization when we conduct Bayesian analysis for ATE estimation.
- **Remark 2**: Essentially, you can think about Bootstrapping as a way to 'calculate' things like expectation (in our case, $E[Z]$), where the only piece of information you have about Z is the observed data itself (i.e. a bunch of individual observations, $Z_1, ..., Z_n$).

How to do Bayesian Bootstrap

Say now we have a vector of confounder variable Z: $Z_1, ..., Z_n$, and we want to estimate $E[Z]$ through Bootstrapping. There are generally two steps in the Bootstrapping procedure:

- Generate a bunch of 'pseudo-datasets' $\mathbf{Z}'_1, .., \mathbf{Z}'_m$ (say, of size M) based on $Z_1, ..., Z_n$.

- For each of these pseudo datasets, we calculate the average of the pseudo-dataset:

$$\bar{Z'}_1, ..., \bar{Z'}_m$$

(We will learn how to generate these pseudo-datasets in a minute.)

We can use the set of pseudo-dataset averages for inference of $E[Z]$, (say, we can calculate sample quatiles as a credible/confidence intervals of $E[Z]$, or calculate the average of the averages as a point estimate for $E[Z]$).

We then see how to understand this under the Bayesian framework.

- Generating pseudo-datasets: draw weights from a uniform Dirichlet distribution with the same dimension as the number of data points. If the data is $Z_1, ..., Z_n$, we sample a vector of weights p from a Dirichlet distribution of $Dir(1, 1, ..., 1)$ where there are n number ones.
- What is a uniform Dirichlet distribution? For a uniform Dirichlet distribution of dimension K, the samples drawn from this distribution will be evenly distributed on a standard simplex of dimension $K - 1$:

$$\sum_{i=1}^{K} x_i = 1 \text{ and } x_i \geq 0 \text{ for all } i \in \{1, \dots, K\}$$

You can see a vector $(x_1, ..., x_K)$ like this will correspond to a discrete distribution of K different outcomes. This means that all possible distributions are likely to be drawn from this uniform Dirichlet distribution. - Now we have a vector of weights $(p_1, ..., p_n)$, we then draw a pseudo dataset of size n' (notice that n' can be different that n), consisting of n' i.i.d samples from the discrete distribution associated with $(p_1, ..., p_n)$. i.e.

$$P(Z'_i = k) = p_k, k = 1, ..., K$$

where Z'_i is the i-th element in the pseudo-dataset. - Finally, we calculate the mean from this pseudodataset. This is the first pseudo dataset so we denote it as $\overline{Z'}^{(1)}$. We repeat it until we have M pseudo datasets, and we have

$$\bar{Z'}^{(1)}, ..., \bar{Z'}^{(m)}$$

. The mean of this will serve as an estimator for $E[Z]$.

Okay this is super complicated – do we have an easy way out?

- Luckily if the population mean is what we care about, we do not need to sample the pseudo dataset – all we need are the weights. (Why?)
- for each $b = 1, ..., B$, we just need to sample weights from the uniform dirichlet distribution:

$$(p_1^{(m)}, p_n^{(m)}) \sim Dir(1, ..., 1)$$

- Then weight our data by the weights

$$Z'^{(m)} = \sum_{i=1}^{n} p_i^{(m)} Z_i$$

The rest is the same as the first procedure.

Estimating ATE: putting it back together

Now we put this back into our picture of estimating ATE. Suppose we have M posterior samples, $\{\beta_0^{(m)}, \beta_1^{(m)}, \beta_2^{(m)}\}_{m=1}^M$, we now replace $E_Z[Z]$ by each of the $\bar{Z}'^{(1)}, ..., \bar{Z}'^{(m)}$ for $m = 1, ..., M$. This gives us a set of 'estimated' posterior draws of ATE:

$$\Psi^{(m)} = \beta_1^{(m)} + \beta_2^{(m)} \bar{Z}'^{(m)}$$

This allow us to do inference for ATE: posterior mean, credible intervals, etc…

Cracking the code

Now we take a look at the example code of Bayesian Bootstrapping, so that you can adapt it for your own model!

- mu_a1, mu_a0: estimates for $E[Y|X = 1, Z]$ and $E[Y|X = 0, Z]$
- mu_a1, mu_a0 are matrices: the number of columns represents the number of posterior draws, why the number of rows represents the number of observations in the dataset (row of the dataset)
- we repeat this procedure for M times, since we have M draws
- for each iteration, we calculate $ATE^{(m)}$ using the formula we have above
- the function returns psi_post, the vector of posterior draws of ATE

```
bayes_boot = function(mu_a1, mu_a0) {
  n = nrow(mu_a1)
  M = ncol(mu_a1)
  psi_post = numeric(M)
  for (m in 1:M) {
    bb_weights = rdirichlet(1, rep(1, n))
    psi_post[m] = sum(bb_weights * (mu_a1[, m] - mu_a0[, m]))
  }
  return(psi_post)
}
```