The Bayesian Bootstrap (1981) by Donald B. Rubin

TCD Stats Department Reading Group

September 2, 2020

The Bootstrap

Suppose $\hat{\phi}(x)$ is chosen to estimate a parameter ϕ of the distribution of X. Given a sample $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, we can estimate the sampling distribution of $\hat{\phi}(\boldsymbol{x})$ using the bootstrap method (Efron, 1979):

For each bootsrap iteration b:

1. Resample from $\boldsymbol{x}$ with replacement, call it $\boldsymbol{x}^b$.
2. Calculate $\hat{\phi}(\boldsymbol{x}^b)$.
3. Repeat this procedure B times to get $\hat{\phi}(\boldsymbol{x}^1), \ldots, \hat{\phi}(\boldsymbol{x}^B)$.

$\{\hat{\phi}(\boldsymbol{x}^1), \ldots, \hat{\phi}(\boldsymbol{x}^B)\}$ forms the emperical bootstrap distribution.

The Bootstrap: Alternate Perspective

- Let $\boldsymbol{p}^b = \{p_1^b, \dots p_n^b\}$ be the proportion of times x_i is drawn in bootstrap sample b, $p_i^b \in \{0/n, 1/n, \dots, n/n\}$, subject to $\sum_i^n p_i^b = 1$.

- We can think of bootstrapping as simulating these $\boldsymbol{p}^b$ and assigning them as weights to the original data.

- For example, the bootstrap sample mean $\frac{1}{n}\sum_i^n x_i^b$ can be equivalently expressed as $\sum_i^n p_i^b x_i$.

The Bayesian Bootstrap

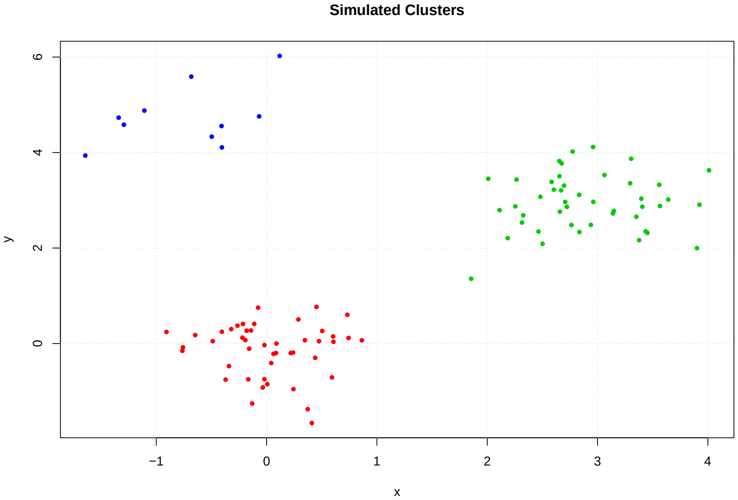
The bayesian bootstrap (Rubin, 1981) proceeds along the above lines. Let $\boldsymbol{d} = \{d_1, \ldots, d_K\}$ be the number of times each distinct value of $x_1, \ldots, x_K$ are observed.

For each bootstrap iteration b:

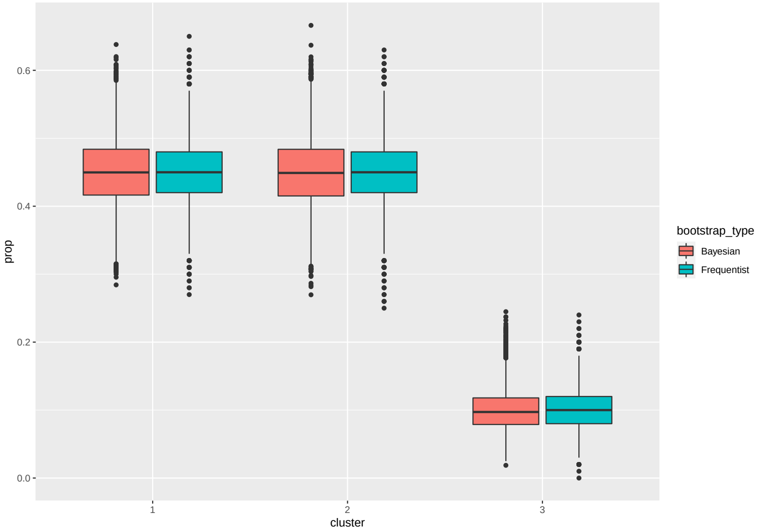
1. Sample $\boldsymbol{p}^b \sim \text{Dirichlet}(\boldsymbol{d})$.
2. Assume $P(X = x_k) = p_k^b$ and calculate ϕ^b under this distribution.
3. Repeat this procedure B times to get $\phi^1, \ldots, \phi^B$.

Simulated Illustration

How does the frequentist and bayesian methods differ for the following data?



Simulated Illustration



Bayesian Bootstrap Theory

Let $\boldsymbol{d}$ and $\boldsymbol{p}$ be defined as before. Each iteration of the bayesian bootstrap samples from the following posterior distribution:

$$\pi(\boldsymbol{p}|\boldsymbol{d}) \propto \mathcal{L}(\boldsymbol{d}|\boldsymbol{p})\pi(\boldsymbol{p})$$

where

$$\begin{aligned} \mathcal{L}(\boldsymbol{d}|\boldsymbol{p}) &= \text{Multinomial}(K, p_1, \ldots, p_K) \\ \pi(\boldsymbol{p}) &= \text{Dirichlet}(\mathbf{0}) \end{aligned}$$

which implies

$$\pi(\boldsymbol{p}|\boldsymbol{d}) = \text{Dirichlet}(\boldsymbol{d})$$

Comparison with Frequentist Bootstrap

In contrast, the frequentist method utilises the likelihood:

$$n\boldsymbol{p} \sim \text{Multinomial}(n, 1/n, \ldots, 1/n)$$

Critisisms Against Bayesian Bootstrap

Rubin took issue with the following characteristics of the bootstrap:

- The assumption that the sample cdf essentially replicates the population cdf.
- The assumption that the drawn proportions are independent
- The unusual "Dirichlet($\mathbf{0}$)" prior.

Efron and Rubin Outlooks

Efron (paraphrasing Tukey)

"[The bootsrap] can blow the head off any problem if the statistician can stand the resulting mess."

Rubin

"Although the bootstrap may be useful in many particular contexts, there are no general data analytic panaceas that allow us to pull ourselves up by our bootstraps."

Questions For The Group

- Is the bayesian bootstrap redundant in a world where we have MCMC?
- Why do you think the frequentist bootstrap appears to be the more popular method?
- Rubin takes a somewhat harsh tone towards the bootstrap assumptions. Do you agree with his critisicms?

References

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics*, 7(1):1–26.

Rubin, D. B. (1981). The Bayesian Bootstrap. *Annals of Statistics*, 9(1):130–134.