Statistica Sinica Preprint No: SS-2016-0546R1

Title	Calibrated Percentile Double Bootstrap For Robust Linear Regression Inference
Manuscript ID	SS-2016-0546R1
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0546
Complete List of Authors	Kai Zhang Daniel McCarthy Lawrence Brown Richard Berk Andreas Bujas Edward George and Linda Zhao
Corresponding Author	Kai Zhang
E-mail	zhangk@email.unc.edu
Notice: Accepted version subject to English editing.	

the regression setting in Section 2.4. We assume that we observe $Z_1, \ldots, Z_m \overset{iid}{\sim} F$ for some distribution F. Let $\theta = \theta(F)$ be a parameter of our interest. We will estimate θ through the empirical distribution $\hat{F}(z) = \frac{1}{m}\sum_{i=1}^{m} I(Z_i \leq z)$. The estimator is denoted by $\hat{\theta} = \theta(\hat{F}) = \theta(Z_1, \ldots, Z_m)$. The construction of the confidence interval is illustrated in Figure 1 and is described as follows.

1. For chosen bootstrap sample size B_1, obtain bootstrap samples $(\mathbf{Z}_1^*, \ldots, \mathbf{Z}_{B_1}^*)$. Each $\mathbf{Z}_j^*$ consists of m i.i.d. samples with replacement from $\hat{F}$. For chosen bootstrap sample size B_2, obtain double bootstrap samples corresponding to all bootstrap samples, $(\mathbf{Z}_{1,1}^{**}, \ldots, \mathbf{Z}_{1,B_2}^{**}, \mathbf{Z}_{2,1}^{**}, \ldots, \mathbf{Z}_{2,B_2}^{**}, \ldots, \mathbf{Z}_{B_1,1}^{**}, \ldots, \mathbf{Z}_{B_1,B_2}^{**})$ in the same manner as in the first-level bootstrap. Denote the empirical distributions by $\hat{F}_j^*$'s, $j = 1, \ldots, B_1$, and $\hat{F}_{j,k}^{**}$'s, $j = 1, \ldots, B_1, k = 1, \ldots, B_2$, respectively.

2. Obtain parameter estimates corresponding to the observed sample, $\hat{\theta} = \theta(\hat{F})$, all bootstrap samples, $(\hat{\theta}_1^*, \ldots, \hat{\theta}_{B_1}^*)$ with $\hat{\theta}_j^* = \theta(\hat{F}_j^*)$ and all double bootstrap samples corresponding to all bootstrap samples, $(\hat{\theta}_{1,1}^{**}, \ldots, \hat{\theta}_{1,B_2}^{**}, \hat{\theta}_{2,1}^{**}, \ldots, \hat{\theta}_{2,B_2}^{**}, \ldots, \hat{\theta}_{B_1,1}^{**}, \ldots, \hat{\theta}_{B_1,B_2}^{**})$ with $\hat{\theta}_{j,k}^{**} = \theta(\hat{F}_{j,k}^{**})$.

3. Form B_1 double bootstrap histograms $\hat{\boldsymbol{\theta}}_1^{**}, \ldots, \hat{\boldsymbol{\theta}}_{B_1}^{**}$, where each histogram $\hat{\boldsymbol{\theta}}_j^{**}$ is comprised of all B_2 double bootstrap estimates $(\hat{\theta}_{j,1}^{**}, \ldots, \hat{\theta}_{j,B_2}^{**})$ corresponding to the jth bootstrap sample and estimate, $\mathbf{Z}_j^*$ and $\hat{\theta}_j^*$, respectively, $j \in \{1, 2, \ldots, B_1\}$.

4. Find the largest $\hat{\lambda}$ such that $1/2 \leq \hat{\lambda} < 1$ and that $\hat{\theta}$ lies in the $1 - \hat{\lambda}$ percentile and the $\hat{\lambda}$ percentile of the histograms $1 - \alpha$ proportion of the time.

5. $\hat{\theta}$ lies between the $(1 - \hat{\lambda}, \hat{\lambda})$ percentiles of the second-level bootstrap distributions $1 - \alpha$ proportion of the time. Therefore our `perc-cal` $(1 - \alpha)$ interval for θ is equal to the $(1 - \hat{\lambda}, \hat{\lambda})$ percentiles of the first-level bootstrap distribution, $[\hat{\boldsymbol{\theta}}^*_{(1-\hat{\lambda})}, \hat{\boldsymbol{\theta}}^*_{(\hat{\lambda})}]$.

For a $(1 - \alpha)$ left-sided `perc-cal` confidence interval for θ, the only change in the procedure is in Step 4, where one uses the histograms to find the smallest $\hat{\lambda}$ such that $\hat{\theta}$ lies below the $\hat{\lambda}$ percentile of the histograms $1-\alpha$ percent of the time. In what follows, we shall refer the two-sided `perc-cal` interval as $\mathcal{I}_2 = [\hat{\boldsymbol{\theta}}^*_{(1-\hat{\lambda})}, \hat{\boldsymbol{\theta}}^*_{(\hat{\lambda})}]$ and the one-sided `perc-cal` interval as $\mathcal{I}_1 = (\infty, \hat{\boldsymbol{\theta}}^*_{(\hat{\lambda})}]$.