Subset Bootstrap

ZHANG Yan

April 12, 2021

Original bootstrap:
1. draw $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ from F, and obtain $\hat{\theta} = \hat{\theta}(\boldsymbol{X})$;
2. resample $\boldsymbol{X}^{*b} = (X_1^{*b}, X_2^{*b}, \ldots, X_n^{*b})$ from $\boldsymbol{X}$, $b = 1, \ldots, B$;
3. obtain $\hat{\boldsymbol{\theta}}^* = (\hat{\theta}^{*1}, \hat{\theta}^{*2}, \ldots, \hat{\theta}^{*B})$, where $\hat{\theta}^{*b} = \hat{\theta}(\boldsymbol{X}^{*b})$;
4. estimate the $1-\alpha$ confidence interval: $[2\hat{\theta} - \hat{\boldsymbol{\theta}}^*_{(1-\alpha/2)}, 2\hat{\theta} - \hat{\boldsymbol{\theta}}^*_{(\alpha/2)}]$.

Subset bootstrap: (assume the bias of $\hat{\theta}$ is asymptotically negligible, and $STD(\hat{\theta}) \propto n^{-\beta}$, normally $\beta = 1/2$)
1. draw $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ from F, and obtain $\hat{\theta} = \hat{\theta}(\boldsymbol{X})$;
2. set $0 < \gamma < 1$ so that γn is an integer, resample (with or without replacement) $\boldsymbol{X}_\gamma^{*b} = (X_1^{*b}, X_2^{*b}, \ldots, X_{\gamma n}^{*b})$ from $\boldsymbol{X}$, $b = 1, \ldots, B$;
3. obtain $\hat{\boldsymbol{\theta}}_\gamma^* = (\hat{\theta}_\gamma^{*1}, \hat{\theta}_\gamma^{*2}, \ldots, \hat{\theta}_\gamma^{*B})$, where $\hat{\theta}_\gamma^{*b} = \hat{\theta}(\boldsymbol{X}_\gamma^{*b})$;
4. scale $\hat{\boldsymbol{\theta}}_\gamma^*$ to be $\hat{\boldsymbol{\theta}}^* = \gamma^\beta(\hat{\boldsymbol{\theta}}_\gamma^* - \hat{\theta}) + \hat{\theta}$;
5. estimate the $1-\alpha$ confidence interval: $[2\hat{\theta} - \hat{\boldsymbol{\theta}}^*_{(1-\alpha/2)}, 2\hat{\theta} - \hat{\boldsymbol{\theta}}^*_{(\alpha/2)}]$.

Experiment:
Here I set $n = 2000$, $B = 1000$, $\gamma = 0.1$ and $\alpha = 0.1$ with 100 replicates to estimate the probabilities that the confidence intervals contain the target parameters.

	normal (one mode)	gamma	normal (two mode)
mean	0.87/0.86/0.84	0.87/0.88/0.86	0.91/0.92/0.9
median	0.89/0.88/0.88	0.88/0.89/0.88	**0.81/0.7/0.68**
std	0.93/0.93/0.92	0.93/0.92/0.93	0.86/0.87/0.85
variance	0.91/0.91/0.89	0.91/0.92/0.91	0.9/0.88/0.87
0.05 quantile	0.86/0.85/0.85	0.86/0.91/0.88	0.86/0.88/0.85

Table 1: Original bootstrap/Subset with replacement/Subset without replacement

5×3 scenarios were tried (5 different estimators and 3 different distributions). We can see that the subset bootstrap has the similar property to the original bootstrap in most occasions except the bold one. And if I increase the sample size n, subset bootstrap will also work for the bold scenario.

Besides, one can also easily detect that the interval based on subset without replacement always tend to be smaller than that with replacement and the original bootstrap, which may be adjusted by scaling the bootstrap sample somehow.