Statistics 540, Bootstrap

Bootstrap Resampling

Goals .

Obtain reliable standard errors and confidence intervals for virtually any statistic, especially those lacking simple analytic form.

Setting .

Suppose that we observe the sample

$$\boldsymbol{X} = X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} F(\theta) \ ,$$

indexed by unknown parameter θ and compute the statistic

$$\hat{\theta} = \theta(X_1, \ldots, X_n) = \theta(\boldsymbol{X}) \ .$$

Denote the empirical distribution by

$$F_n(x) = \frac{\#\{X_i \leq x\}}{n} \ .$$

Approach .

Think of the parameter θ and statistic $\hat{\theta}$ as functionals,

$$\theta = \theta(F), \quad \hat{\theta}(F_n) \ .$$

The idea is to exploit the analogy

$$\theta(F) : \theta(F_n) \ :: \ \theta(F_n) : \theta(F_n^*)$$

where F_n^* denotes the empirical distribution of a sample from F_n.

Computational Procedure .

Although the bootstrap is more an approach than algorithm, it is typically implemented as a computational procedure. (Computation can be avoided in some places, such as with linear statistics.) In its nonparametric form, the idea is to simulate by sampling with replacement from F_n rather than some hypothesized distribution. For $b = 1, \ldots, B$, let

$$\boldsymbol{X}_b^* = \text{ sample from } X_1, \ldots, X_n \Rightarrow \boldsymbol{X}_b^* \stackrel{\text{iid}}{\sim} F_n \ ,$$

and then compute the bootstrap replicate

$$\hat{\theta}_b^* = \theta(\boldsymbol{X}_b^*) \ .$$

Estimate the standard error by

$$se^*(\hat{\theta}) = \sqrt{\sum_b (\hat{\theta}_b^* - \overline{\hat{\theta}_b^*})^2/(B-1)}$$

Background .

Some of the theory involves functional Taylor expansions,

$$\hat{\theta} - \theta \approx \theta'(F) \qquad \underbrace{(F_n - F)}_{\text{Brownian bridge}}$$

and

$$\hat{\theta}^* - \hat{\theta} \approx \theta'(F_n) \qquad \underbrace{(F_n^* - F_n)}_{\text{Brownian bridge}}$$

If the statistics is sufficiently smooth so that the functional derivatives $\theta'(F) \approx \theta'(F_n)$, then the procedure gets the right SE.

References .

- Bickel, P. and D. Freedman (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, **9**, 1196-1217.
- T. J. DiCicco and B. Efron (1996). Bootstrap confidence intervals. *Statistical Science*, **11**, 189-228.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1-26.
- Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.

Success ?

When it works, the bootstrap does better than the standard asymptotic normal theory. That is, its standard error is "better" than that produced the the simple delta method or jackknife approximations, and its confidence intervals have higher coverage "accuracy" than the usual normal theory interval $\hat{\theta} \pm z_\alpha SE(\hat{\theta})$ interval.