The Bootstrap Method for Estimating MSE

Bootstrap methods are used throughout the field of statistics, but we will introduce the bootstrap as a way to address the question: How close is an estimator $\widehat{\theta}$ to its estimand θ (the parameter that it is estimating)? We have seen that for unbiased estimators, we often construct confidence intervals of the form:

$$\text{estimator} \ \pm \ \text{(multiplier)(standard error of estimator)}$$

However, many estimators are biased, so we instead focus on the mean-squared error of an estimator $\widehat{\theta}$, defined as:

$$MSE(\widehat{\theta}) = E(\widehat{\theta} - \theta)^2 = var(\widehat{\theta}) + [E(\widehat{\theta}) - \theta]^2 = var(\widehat{\theta}) + [bias(\widehat{\theta})]^2.$$

We can use the mean-squared error of an estimator $\widehat{\theta}$ to see how close it is to θ with the Tchebyshev-Markov inequality:

$$P(\,|\widehat{\theta} - \theta| \ \leq \ k\sqrt{MSE(\widehat{\theta})}\,) \geq 1 - 1/k^2$$

If we use $k = 2$ in this expression we get

$$P(\,|\widehat{\theta} - \theta| \ \leq \ 2\sqrt{MSE(\widehat{\theta})}\,) \geq 3/4\ .$$

This last expression motivates the definition of the term

$$\text{margin of error} = 2\sqrt{MSE(\widehat{\theta})},$$

which is often used as a measure of the accuracy of an estimator $\widehat{\theta}$.

How do we find the MSE of an estimator?

1) From theory: for some statistics we can derive their MSE, for example:

$$E(\overline{x}) = \mu,\ var(\overline{x}) = \sigma^2/n$$

2) From simulation: for known F, with sample size n, (F_n) and an estimator $\widehat{\theta}$ of interest we can examine the distribution of $\widehat{\theta}$ about θ via simulation. We do this by taking a large number, REP, of samples from

F_n , and for each of these samples we can calculate the estimator $\widehat{\theta}_i$, $i = 1, \cdots, REP$. Now we can estimate the MSE of $\widehat{\theta}$ by calculating:

$$\widehat{MSE} = \frac{1}{REP} \sum_{i=1}^{REP} (\widehat{\theta}_i - \theta)^2.$$

These approaches are fine for when we can derive expressions for $E(\widehat{\theta})$ and $var(\widehat{\theta})$ or if we know F_n, but what can we do if we cannot use these methods? The key idea of the bootstrap is that we do something similar to the simulation method above, but since we do not know F_n, we estimate it via the sample ($\widehat{F}_n$) by taking bootstrap samples of size n (random samples with replacement) from $\widehat{F}_n$. Thus, as shown on page 251 of the text, to obtain a bootstrap estimate of the MSE of $\widehat{\theta}$, we take a large number, REP, of bootstrap samples from $\widehat{F}_n$ (random samples with replacement), and for each of these samples we calculate the estimator $\widehat{\theta}_{b,i}$, $i = 1, \cdots, REP$. Then we calculate the bootstrap estimate of MSE as:

$$\widehat{MSE} = \frac{1}{REP} \sum_{i=1}^{REP} (\widehat{\theta}_{b,i} - \widehat{\theta})^2.$$

Other bootstrap calculations

Analogous to the way we estimate MSE from bootstrap calculations, we can also estimate other quantities of interest. Thus the expected value and bias of $\widehat{\theta}$ can be estimated as:

$$\widehat{E} = \frac{1}{REP} \sum_{i=1}^{REP} \widehat{\theta}_{b,i} \text{ and } \widehat{B} = \widehat{E} - \widehat{\theta},$$

and the variance of $\widehat{\theta}$ can be estimated by:

$$var = \frac{1}{REP} \sum_{i=1}^{REP} (\widehat{\theta}_{b,i} - \widehat{E})^2.$$

Nonparametric versus parametric bootstrap estimates