Bootstrapping Hypotheses Tests

Chathurangi H. Pathiravasan, David J. Olive
Department of Mathematics, Southern Illinois University, Carbondale IL

The bootstrap is a general methodology to estimate standard error of the Test statistics. Consider testing $H_0: A\boldsymbol{\theta} = \boldsymbol{c}$ versus $H_1: A\boldsymbol{\theta} \neq \boldsymbol{c}$ where A is a known $r \times p$ matrix of rank r and c is a known $r \times 1$ vector. Let $\hat{\theta}$ be a consistent estimator of θ and make a bootstrap sample $w_i = A\hat{\theta}^* - c$ for $i = 1,2,3 \ldots, B$. Make a prediction region for the w_i and determine 0 is in the prediction region.

The percentile method, which is an interval that contains $d_B \cong k_B = \lceil \beta(1-\delta) \rceil$ of the $T_{i,n}^*$ from a bootstrap sample $T_{1,n}^*, \ldots, T_{B,n}^*$ where the statistic $T_{i,n}$ is an estimator of θ based on a sample size n.

It will be shown that this prediction region method generalizes the percentile method for $r = 1$ to $r \geq 1$. This method can be widely applied, but should be regarded as exploratory unless theory shows that the prediction region method is a large sample test.

Moreover, this prediction region method will be compared to the Efron (2014) confidence interval for variable selection and used to bootstrap a correlation matrix. Indeed, the prediction region method can also be justified as a special case of the percentile method where the test statistic is the squared Mahalanobis distance ${D_i}^{2*} = (T_i^* - \overline{T^*})^T [S_T^*]^{-1} (T_i^* - \overline{T^*})$ where $w_i = T_i^*$, and $\overline{T^*}$ and S_T^* are the sample mean and sample covariance matrix of $T_1^*, \ldots, T_B^*$.