

Presenting groups: 7 and 8, Date: 9.6.2021

The aim of this exercise is to compare coverage probabilities of analytical confidence intervals and bootstrapped confidence intervals.

Exercise 1:

Consider the following data generating process in which we have $n = 200$ observations and one covariate X that enters the DGP nonlinearly, additional to a constant. Specifically, we assume that we observe n realisations of a simple linear model:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \beta_4 x_i^3 + \beta_5 x_i^4 + \varepsilon_i,$$

where $X \sim \mathcal{N}(0, 1)$, $\beta = (1 \quad 1.5 \quad -1.5 \quad 1.5 \quad 0.5)$ and the errors are drawn from a normal distribution $\varepsilon \sim \mathcal{N}(0, 1)$.

Suppose we have computed the fit at a particular value, x_0 :

$$\hat{f}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0 + \hat{\beta}_2 x_0^2 + \hat{\beta}_3 x_0^3 + \hat{\beta}_4 x_0^4. \quad (1)$$

The variance $\text{Var} \hat{f}(x_0)$ (at a particular point) can be calculated as follows:

Denote by $\hat{\mathbf{C}}$ the 5×5 variance-covariance matrix of the estimator $\hat{\beta}$ (that is calculated by the OLS, for example). Denote by $\ell_0 = (1, x_0, x_0^2, x_0^3, x_0^4)^\top$ a data vector. Then we can calculate the necessary value by $\text{Var}[\hat{f}(x_0)] = \ell_0^\top \hat{\mathbf{C}} \ell_0$.

To retrieve the point-wise standard errors, we take the square root of the point-wise variances. Twice the standard error is a good approximation of normal confidence interval with $\alpha = 5\%$ (or you may use the normal quantile). This computation is repeated at each reference point along the range of X .

- Calculate the analytical standard errors for the polynomial specification above (as presented in the beginning of the exercise) and use these to calculate the approximate confidence intervals as $2 * SE$ for each value of X you consider.
- Calculate the naive bootstrap confidence intervals (i.e. using the naive bootstrap quantiles) for $B = 200$ bootstrap draws from the original data, such that the nominal coverage for the two methods is the same.
- Calculate the coverage probability at four different values of X , chosen by you.
- Calculate the interval length at four different values of X .

Exercise 2:

Evaluate the two types of confidence intervals above along two dimensions: interval length and coverage probability.

- Calculate both for a small simulation study of 100 repetitions.
- How could you change the data-generating process to give a competitive advantage to the bootstrap? Suggest two changes and check each in a simulation study.

Hint: you can change the underlying distribution as well, not only the sample size or the parameter values.