Residuals Scaling for Bootstrap Prediction Interval in Regression

Sha Taole  
Bachelor of Science (Statistics)

Problem Statement

i.i.d. data pairs \((x_i, y_i)\) \(i = 1, 2, ..., N\) of \((X, Y)\)

Modelling assumption: \(y = f(x) + \epsilon\)

Unknown function: \(f(x) = E(Y | X = x)\)

Noise: \(\epsilon \sim F\), mean 0, variance \(\sigma^2\)

Regression estimator at a future input: \(\hat{f}(x^*) = \hat{E}(Y | X = x^*)\)

Target prediction interval: \(I(x^*) = [l(x^*), u(x^*)]\)

\(\beta\)-expectation: \(Pr(y^* \in I(x^*)) = \beta\)

Bootstrap Prediction Interval

Decompose the output: \(y^* = \hat{f}(x^*) + f(x^*) - \hat{f}(x^*) + \epsilon^*\)

The unknown target: \((R^* = f(x^*) - \hat{f}(x^*) + \epsilon^*) \sim G^*\)

Approximation: \((\hat{R}_b^* = \hat{f}(x^*) - \hat{f}_b(x^*) + \hat{e}_b^*), \text{empirically } G^*\)

\(b\) denotes \(b^{th}\) Bootstrap replicate

Bootstrap prediction interval:

\[I_b^*(x^*) = \left[\hat{f}(x^*) + G^{*^{-1}}\left(1 - \frac{1}{2}\beta\right), \hat{f}(x^*) + G^{*^{-1}}\left(1 - \frac{1}{2}\beta\right)\right]\]

The idea back to Stine (1985) in a linear regression setting

Generate Bootstrap Replicates

Sample residuals: \(\hat{e}_i = y_i - \hat{f}(x_i), \text{empirically } F_N\)

Generate \(y_{b,i} = \hat{f}(x_i) + \hat{e}_i, \hat{e}_i \sim F_N\)

\(\hat{f}_b(x^*)\) is dependent on \((x_i, y_{b,i}) i = 1, 2, ..., N\)

Noise term: \(\hat{e}^2_b \sim F_N\)

Asymptotic Validity

Assumption: as \(N \to \infty, \hat{f}(x) - f(x) \xrightarrow{P} 0, \hat{f}_b(x) - f(x) \xrightarrow{P} 0\)

Asymptotic \(\beta\)-expectation: \(Pr\left(y^* \in I_b^*(x^*)\right) \xrightarrow{P} \beta\)

Overfitting the observations \(\to\) Finite Sample Undercoverage

Scaling the Residuals

Definition (Linear Estimator):

\[\hat{f}(x^*) = \sum_{i=1}^{N} y_i w_i(x^*)\]

\(w_i(x^*)\) depends only on \(N, i, x^*, x_i, x_2, ..., x_N\)

In-sample matrix form:

\[\hat{f}(X) = WY, \quad W = \begin{bmatrix} w_1(x_1) & \ldots & w_N(x_1) \\ \vdots & \ddots & \vdots \\ w_1(x_N) & \ldots & w_N(x_N) \end{bmatrix}\]

In-sample residuals:

\[\hat{e} = (I - W)Y, \quad Cov(e) = (I - W)\sigma^2 (I - W)^T\]

Define \(s_i\):

\[s_i = \left(\frac{1}{1-w(x_i)^2} + \frac{1}{\sum_{j=i+1}^{N}w(x_i)^2}\right)^{\frac{1}{2}}\]

By the Lindeberg-Feller central limit theorem:

\[\frac{f(x^*) - \hat{f}(x^*)}{\|W^*(x^*)\|^2} \to N(0, \sigma^2), \quad \frac{f(x^*) - \hat{f}_b(x^*)}{\|W^*(x^*)\|^2} \to N(0, Variance(s_i \hat{e}_i) = s^2^\sigma^2\]

Denote a diagonal matrix \(S\) whose \((i, i)^{th}\) entry is \(s_i^\sigma^2\):

\[E\left[\frac{\hat{e}^T (I - W)^T S^2 S (I - W)\hat{e}}{N}\right] = \sigma^2 + \frac{\sigma^2}{\|W^*(x^*)\|^2}\]

Handing the Heteroscedasticity

Alternative modelling assumption: \(y = f(x) + \epsilon | x = x\)

Noise: \(\epsilon \sim F, \text{mean 0, variance } \sigma^2\)

Heteroscedasticity scale: \(h(x)\)

• \(h(x) = 1\) backs to the homoscedastic situation

• \(h(x) = E(\epsilon | x = x)\) forwards to the heteroscedastic situation

Regressing \(\hat{e}_i / s_i \hat{e}_i\) on \(X\) to have the \(h(x)\) (Lei et al., 2018)

Additional Bootstrap Alternative for Scaling

What if the regression estimator is black-box to the user?

Sample residuals:

\(\hat{e}_i = y_i - \hat{f}(x_i), \text{empirically } F_N\) with \(\text{sample variance } \sigma^2\)

Generate \(y_{a,i} = \hat{f}(x_i) + \hat{e}_i, \hat{e}_i \sim F_N\), \(a = 1, 2, ..., A\)

\(\hat{f}_a(x^*)\) is dependent on \((x_i, y_{a,i}) i = 1, 2, ..., N\)

Bootstrap sample residuals:

\[\hat{e}_{a,i} = y_{a,i} - \hat{f}_a(x_i), \quad \hat{\sigma}_a^2 = \frac{\sum_{i=1}^{A} e_{a,i}^2}{A}\]

Alternative scaling factors:

\[s_i = \frac{\sigma^2}{\hat{\sigma}_a^2} \frac{1}{2}\]

References (a selected list)


