Bootstrap & Confidence/Prediction intervals

Olivier Roustant

Mines Saint-Étienne

2017/11

Framework

Consider a model with an additive homoskedastic noise

$$y_i = g(x_i; \beta) + \varepsilon_i \qquad i = 1, \dots, n$$

with $\varepsilon_1, \dots, \varepsilon_n$ i.i.d. with a cdf F, and β is a vector of parameters.

Framework

Consider a model with an additive homoskedastic noise

$$y_i = g(x_i; \beta) + \varepsilon_i \qquad i = 1, \dots, n$$

with $\varepsilon_1, \dots, \varepsilon_n$ i.i.d. with a cdf F, and β is a vector of parameters.

Let $\hat{\beta}$ an estimator of β, and let x be a new site. We are interested in :

- The prediction mean at x : $\hat{y}(x) = g(x; \hat{\beta})$
- The prediction law at x, i.e. the law of $\hat{y}(x) + \epsilon$, where $\epsilon \sim F$ independent of the $\varepsilon_1, \dots, \varepsilon_n$.

Confidence intervals

The variability of $\hat{y}(x)$ may provide a (random) confidence interval at level α of a (deterministic) statistic of interest t for $g(x, \beta)$:

$$P(I \ni t) = 1 - \alpha$$

Confidence intervals

The variability of $\hat{y}(x)$ may provide a (random) confidence interval at level α of a (deterministic) statistic of interest t for $g(x, \beta)$:

$$P(I \ni t) = 1 - \alpha$$

Example (Gaussian linear model)

$$g(x, \hat{\beta}) = x^\top \hat{\beta} \sim \mathcal{N}\left(x^\top \beta, \sigma^2 x^\top (X^\top X)^{-1} x\right)$$

This gives a 95% confidence interval of $x^\top \beta$, the 'true' prediction at x :

$$I \approx x^\top \hat{\beta} \pm 2\hat{\sigma}\sqrt{x^\top (X^\top X)^{-1} x}$$

Confidence intervals, bootstrap estimate

- Resample the data $z_i = (x_i, y_i), i = 1, \dots, n$

$$(z_1^{\star b}, \dots, z_n^{\star b}), \qquad b = 1 \dots, B$$

Confidence intervals, bootstrap estimate

- Resample the data $z_i = (x_i, y_i), i = 1, \ldots, n$

$$(z_1^{\star b}, \ldots, z_n^{\star b}), \qquad b = 1 \ldots, B$$

- For each bootstrap sample $(z_i^{\star b})_{1 \leq i \leq n}$, refit the data

$$\hat{\beta}^{\star b}$$

Confidence intervals, bootstrap estimate

- Resample the data $z_i = (x_i, y_i), i = 1, \ldots, n$

$$(z_1^{\star b}, \ldots, z_n^{\star b}), \qquad b = 1 \ldots, B$$

- For each bootstrap sample $(z_i^{\star b})_{1 \leq i \leq n}$, refit the data

$$\hat{\beta}^{\star b}$$

- Give a 95% confidence interval by computing the empirical quantiles at level 2.5% and 97.5% of

$$g(x, \hat{\beta}^{\star b}), \qquad b = 1, \ldots, B$$

Prediction intervals

The law of $\hat{y}(x) + \epsilon$ provides a (deterministic) prediction interval I at level α, which contains the (random) prediction with probability $1 - \alpha$:

$$P(\hat{y}(x) + \epsilon \in I) = 1 - \alpha$$

Prediction intervals

The law of $\hat{y}(x) + \epsilon$ provides a (deterministic) prediction interval I at level α, which contains the (random) prediction with probability $1 - \alpha$:

$$P(\hat{y}(x) + \epsilon \in I) = 1 - \alpha$$

Example (Gaussian linear model)

$$\hat{y}(x) + \epsilon = x^\top \hat{\beta} + \epsilon \sim \mathcal{N}(x^\top \beta, \sigma^2(1 + x^\top (X^\top X)^{-1} x))$$

and $I \approx x^\top \beta \pm 2\hat{\sigma}\sqrt{1 + x^\top (X^\top X)^{-1} x}$

Prediction intervals, bootstrap estimate

- Resample the data $z_i = (x_i, y_i), i = 1, \dots, n$

$$(z_1^{\star b}, \dots, z_n^{\star b}), \qquad b = 1 \dots, B$$

- For each bootstrap sample $(z_i^{\star b})_{1 \leq i \leq n}$,
  - Refit the data : $\hat{\beta}^{\star b}$

Prediction intervals, bootstrap estimate

- Resample the data $z_i = (x_i, y_i), i = 1, \dots, n$

$$(z_1^{\star b}, \dots, z_n^{\star b}), \qquad b = 1 \dots, B$$

- For each bootstrap sample $(z_i^{\star b})_{1 \leq i \leq n}$,
  - Refit the data : $\hat{\beta}^{\star b}$
  - Sample $\varepsilon^{\star\star b}$ from the resampled residuals : $\varepsilon_i^{\star b} = y_i - g(x_i, \hat{\beta}^{\star b})$
    $\rightarrow$ *Non-parametric bootstrap on the bootstrapped residuals*

Prediction intervals, bootstrap estimate

- Resample the data $z_i = (x_i, y_i), i = 1, \dots, n$

$$(z_1^{\star b}, \dots, z_n^{\star b}), \qquad b = 1 \dots, B$$

- For each bootstrap sample $(z_i^{\star b})_{1 \le i \le n}$,
  - Refit the data : $\hat{\beta}^{\star b}$
  - Sample $\varepsilon^{\star\star b}$ from the resampled residuals : $\varepsilon_i^{\star b} = y_i - g(x_i, \hat{\beta}^{\star b})$
    $\rightarrow$ *Non-parametric bootstrap on the bootstrapped residuals*
- Give a 95% prediction interval by computing the empirical quantiles at level 2.5% and 97.5% of

$$g(x, \hat{\beta}^{\star b}) + \varepsilon^{\star\star b}, \qquad b = 1, \dots, B$$

Prediction intervals, bootstrap estimate - Normal approximation

- If the residuals $\varepsilon_1, \dots, \varepsilon_n$ are assumed $\mathcal{N}(0, \sigma^2)$, one can sample $\varepsilon^{\star\star b}$ from $\mathcal{N}(0, (\sigma^{\star b})^2)$, where $\sigma^{\star b}$ is the s.d. of $(\varepsilon_i^{\star b})_{1 \leq i \leq n}$.
  $\rightarrow$ *Parametric bootstrap on the bootstrapped residuals*

Prediction intervals, bootstrap estimate - Normal approximation

- If the residuals $\varepsilon_1, \dots, \varepsilon_n$ are assumed $\mathcal{N}(0, \sigma^2)$, one can sample $\varepsilon^{\star\star b}$ from $\mathcal{N}(0, (\sigma^{\star b})^2)$, where $\sigma^{\star b}$ is the s.d. of $(\varepsilon_i^{\star b})_{1 \leq i \leq n}$.
  $\rightarrow$ *Parametric bootstrap on the bootstrapped residuals*
- If, in addition, the law of $g(x, \hat{\beta})$ is assumed to be Gaussian, one can simply estimate the variances of $g(x, \hat{\beta})$ and ε separately

$$\hat{\sigma}^2_{g(x,.)} = \frac{1}{B-1} \sum_{b=1}^{B} \left(g(x, \hat{\beta}^{\star b}) - \overline{g(x,.)} \right)^2 \qquad \hat{\sigma}^2_{\varepsilon} = \frac{1}{n-1} \sum_{i=1}^{n} \epsilon_i^2$$

and compute the prediction intervals as

$$\overline{g(x,.)} \pm 2 \sqrt{\hat{\sigma}^2_{g(x,.)} + \hat{\sigma}^2_{\varepsilon}}$$

References

For more details, in particular about bias induced by bootstrap, one can read :

ESL T. Hastie, R. Tibshirani and J. Friedman (2009), The Elements of Statistical Learning, Springer, 2nd edition, print 10.