Nonparametric Inference in Regression Analysis

University of California, San Diego
Instructor: Ery Arias-Castro

http://math.ucsd.edu/~eariasca/teaching.html

Inference for regression

- Inference when using least squares to fit a linear model is usually done under the standard assumptions.
- M-estimators are asymptotically normal under some conditions (satisfied here) and so is the LTS estimator. Various estimates for the asymptotic covariance matrix are available. This asymptotic theory can be turned into inference that this valid for large samples... But how large?
- Computer-intensive methods are based on resampling the data (bootstrap, jackknife, permutation) and do not make parametric assumptions on the distribution of the errors (such as normality).

The bootstrap

- The bootstrap is a resampling method that can be used to estimate the variance of an estimator, to produce a confidence interval for a parameter of interest, or to estimate a P-value, and more.

  (We follow here the book All of Nonparametric Statistics, by L. Wasserman.)

The bootstrap variance estimate

- Suppose we have a sample $X_1, \dots, X_n$ IID F and want to estimate the variance of a statistic $T_n = T(X_1, \dots, X_n)$. We have several options, depending on what information we have access to:

1. Compute it by integration if F (or its density) is known in closed form.
2. Compute it by Monte Carlo simulations if we can simulate from F.

   Let B be a large integer. For $b = 1, \dots, B$, sample $X_1^{\bullet b}, \dots, X_n^{\bullet b}$ IID from F and compute $T_n^{\bullet b} = T(X_1^{\bullet b}, \dots, X_n^{\bullet b})$. Compute the sample mean and variance

$$\bar{T}_n = \frac{1}{B} \sum_{b=1}^{B} T_n^{\bullet b}, \qquad \widehat{\mathrm{SE}}_{\mathrm{MC}}^2 = \frac{1}{B-1} \sum_{b=1}^{B} \left(T_n^{\bullet b} - \bar{T}_n\right)^2$$

   (MC = Monte Carlo)

3. Estimate it by the nonparametric bootstrap. Same as above except that we sample $X_1^{*b}, \dots, X_n^{*b}$ IID from $\hat{F}$, the empirical distribution based on the original sample. Let the result be denoted $\widehat{\mathrm{SE}}_{\mathrm{boot}}^2$.

$$\boxed{\text{The nonparametric bootstrap does as if the sample were the population.}}$$

Nonparametric bootstrap sampling

- Recall that empirical distribution is the discrete distribution with support $\{X_1, \ldots, X_n\}$ that puts mass $1/n$ on each X_i.

  If $X_1, \ldots, X_n$ are real, then their empirical distribution function is

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{X_i \leq x\}$$

- Generating an IID sample of size k from $\hat{F}$ is done by sampling with replacement k times from the original sample $\{X_1, \ldots, X_n\}$.

  Thus, $X_1^{*b}, \ldots, X_n^{*b}$ contains many repeats and does not include all the observations (in fact, on average, about $1/3$ are missing).

Bootstrap confidence intervals

- Consider a functional T and let $\theta = T(F)$.

  For example, $T(F) = \mathrm{median}(F)$.

- Suppose we want a $(1-\alpha)$-confidence interval for θ.

- Define $\hat{\theta} = T(\hat{F})$, which we use as an estimate for θ.

Bootstrap normal confidence interval

- The bootstrap normal confidence interval is

$$\hat{\theta} \pm z_{\alpha/2}\,\widehat{\mathrm{SE}}_{\mathrm{boot}}$$

  where $\widehat{\mathrm{SE}}_{\mathrm{boot}}$ is (obviously) the bootstrap estimate for standard error for the statistic $\hat{\theta}$.

- This interval is only accurate if $\hat{\theta}$ is approximately unbiased for θ and approximately normal.

Bootstrap percentile confidence interval

- The bootstrap percentile confidence interval is

$$\left(\hat{\theta}^{*(B\alpha/2)},\ \hat{\theta}^{*(B(1-\alpha/2))}\right)$$

  where $\hat{\theta}^{*(b)}$ denotes the bth largest $\hat{\theta}^{*b}, b = 1, \ldots, B$.

- To be valid, this method requires some special assumptions.

  (Think of a useless statistic like $\hat{\theta} \equiv 1$.)

Bootstrap pivotal confidence interval

- The bootstrap pivotal confidence interval is

$$\left(2\hat{\theta} - \hat{\theta}^{*(B(1-\alpha/2))},\ 2\hat{\theta} - \hat{\theta}^{*(B\alpha/2)}\right)$$

- This is justified by considering the pivot $S = \hat{\theta} - \theta$.
  - If $H(s) = \mathbb{P}(S \le s)$ and $s_\alpha = H^{-1}(\alpha)$, then

    $$\mathbb{P}(s_{\alpha/2} \le \hat{\theta} - \theta \le s_{1-\alpha/2}) = 1 - \alpha$$

    equivalently, $\theta \in \left[\hat{\theta} - s_{1-\alpha/2}, \hat{\theta} - s_{\alpha/2}\right]$ with probability $1-\alpha$.
  - We estimate H by the bootstrap

    $$\hat{H}(s) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\{S^{*b} \le s\}$$

    where $S^{*b} = \hat{\theta}^{*b} - \hat{\theta}$.

Bootstrap Studentized pivotal confidence interval

- Let B and C be two large integers.

  For $b = 1, \ldots, B$, do the following:

  1. Generate $X_1^{*b}, \ldots, X_n^{*b}$ from $\hat{F}$.
     Let $\hat{F}^{*b}$ denote the corresponding empirical distribution.
  2. Compute $\hat{\theta}^{*b} = T(\hat{F}^{*b})$.
  3. For $c = 1, \ldots, C$, do the following: (2nd bootstrap loop)
     (a) Generate $X_1^{*(b,c)}, \ldots, X_n^{*(b,c)}$ from $\hat{F}^{*b}$.
     Let $\hat{F}^{*(b,c)}$ denote the corresponding empirical distribution.
     (b) Compute $\hat{\theta}^{*(b,c)} = T(\hat{F}^{*(b,c)})$.
  4. Compute

     $$\bar{\theta}^{*b} = \frac{1}{C}\sum_{c=1}^{C} \hat{\theta}^{*(b,c)}, \qquad \widehat{\mathrm{SE}}_{*b}^2 = \frac{1}{C-1}\sum_{c=1}^{C} \left(\hat{\theta}^{*(b,c)} - \bar{\theta}^{*b}\right)^2$$

  5. Compute

     $$Z^{*b} = \frac{\hat{\theta}^{*b} - \hat{\theta}}{\widehat{\mathrm{SE}}_{*b}}$$

- Note that $\bar{\theta}^{*b}$ is different from $\hat{\theta}^{*b}$.
- The bootstrap Studentized pivotal confidence interval is

$$(\hat{\theta} - z^*_{1-\alpha/2}\widehat{\mathrm{SE}}_{\mathrm{boot}},\ \hat{\theta} - z^*_{\alpha/2}\widehat{\mathrm{SE}}_{\mathrm{boot}})$$

  where $z^*_\alpha = Z^{*(B\alpha)}$.
- The rationale is to do as in the bootstrap pivot confidence interval, where instead of S we use as pivot

$$Z = (\hat{\theta} - \theta)/\widehat{\mathrm{SE}}_{\mathrm{boot}}(\hat{\theta})$$

  The standard deviation bootstrap estimate requires a bootstrap loop, and this is carried out for each bootstrap sample, giving rise to a double loop!

Bootstrap P-values

- Suppose we want to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.
- We can simply build a confidence interval using one of the aforementioned methods. If $\hat{I}_{1-\alpha}$ is a bootstrap $(1-\alpha)$-confidence interval, then

$$\text{P-value} = \sup\{\alpha : \theta_0 \notin \hat{I}_{1-\alpha}\}$$

- We can perform one-sided testing by considering appropriate one-sided confidence intervals.

Standardized jackknife influence

- We want to gage how much each observation affects the bootstrap distribution of a given statistic $\hat{\theta} = T(X_1, \ldots, X_n)$.
- For $i \in \{1, \ldots, n\}$, let

$$\bar{\theta}^*_{(i)} = \mathrm{Ave}\big(\hat{\theta}^{*b}, i \notin I^{*b}\big)$$

  the average of the bootstrapped statistics over the (bootstrap) samples that do not contain i. (I^{*b} indexes the observations in the b-th bootstrap sample.)
- Define

$$\bar{\theta}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b}$$

  which can also be expressed as a weighted average of $\bar{\theta}^*_{(i)}, i = 1, \ldots, n$.
- The standardized jackknife influence of observation i is defined as

$$J_{(i)} = \frac{u_{(i)}}{\sqrt{\frac{1}{n-1}\sum_i u^2_{(i)}}}$$

  where

$$u_{(i)} = (n-1)\big(\bar{\theta}^* - \bar{\theta}^*_{(i)}\big)$$

Jackknife-after-bootstrap plot

- This plot shows the sensitivity of $\hat{\theta}$ and of the percentiles of its bootstrapped distribution to deletion of individual observations, and thus can be seen as a diagnostic plot for discovering outliers.
  - Each line traces a given quantile of the distribution over the bootstrap samples that do not contain a given observation.
  - On the horizontal axis the observations are ordered according to their standardized jackknife influence.
- Fix $\alpha \in [0, 1]$ and let $\gamma_\alpha(z_1, \ldots, z_m)$ denote the α-quantile of a sample $z_1, \ldots, z_m \in \mathbb{R}$. Then let

$$\gamma_{\alpha,(i)} = \gamma_\alpha(\hat{\theta}^{*b}, i \notin I^{*b})$$

  the α-quantile of the bootstrap distribution over samples that do not contain observation i.
- After ordering the $J_{(i)}$'s, we then plot $\gamma_{\alpha,(i)}$ versus $J_{(i)}$.
- This is done for several α's.

Bootstrap inference for regression

- Suppose we have an additive error model

$$y = f(\mathbf{x}) + \varepsilon$$

  where we assume that $\mathbb{E}(\varepsilon|\mathbf{x}) = 0$, so that $f(\mathbf{x}) = \mathbb{E}(y|\mathbf{x})$.
- We observe a sample from this model $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ and want to learn about an estimator $\widehat{f}$.

- Bootstrapping cases (random-x bootstrap)

$$(\mathbf{x}_1^{*b}, y_1^{*b}), \ldots, (\mathbf{x}_n^{*b}, y_n^{*b})$$

  are sampled with replacement from the sample.
- Bootstrapping residuals (fixed-x bootstrap)

$$(\mathbf{x}_1, y_1^{*b}), \ldots, (\mathbf{x}_n, y_n^{*b})$$

  where

$$y_i^{*b} = \widehat{f}(\mathbf{x}_i) + e_i^{*b}$$

  $e_1^{*b}, \ldots, e_n^{*b}$ are sampled with replacement from the residuals $e_1, \ldots, e_n$:

$$e_i = y_i - \widehat{f}(\mathbf{x}_i)$$

  This method implicitly assumes that the errors $\varepsilon_1, \ldots, \varepsilon_n$ are IID and independent of the $\mathbf{x}$'s. This is *not* the case if $\hat{f}$ is biased for f, or if the errors have different variances, or if there are outliers.

Example: confidence interval for $f(\mathbf{x}_0)$

- Fix $\mathbf{x}_0$. We can obtain a confidence interval for $\theta = f(\mathbf{x}_0)$ by simply computing a bootstrap confidence interval based on $\hat{\theta} = \widehat{f}(\mathbf{x}_0)$.
- In more detail, the estimator is computed on b-th bootstrap sample, resulting in $\widehat{f}^{*b}$, and a confidence interval is computed based on $\hat{\theta}^{*b} = \widehat{f}^{*b}(\mathbf{x}_0)$, $b = 1, \ldots, B$.

Example: inference for the coefficients of a linear model

- Suppose we fit a linear model $f(\mathbf{x}) = \boldsymbol{\beta}^\top \mathbf{x}$ by a certain method, for example, by M-estimation using Huber's function.

  Let $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_p)$ denote the resulting coefficient estimate.
- We can obtain a confidence interval for $\theta = \beta_j$ by simply computing a bootstrap confidence interval based on $\hat{\theta} = \widehat{\beta}_j$.

  And testing $\beta_j = 0$, for example, can be done via these confidence intervals as we saw earlier.
- **Remark.** If $\widehat{\boldsymbol{\beta}}$ is computed by least squares, a bootstrap Studentized pivot can be obtained with a single bootstrap loop. Indeed, we can directly use the analytic expression to estimate the variance of $\widehat{\beta}_j$:

$$\widehat{\mathrm{SE}}^2_{*b} = \widehat{\sigma}^2_{*b} (\mathbf{X}_{*b}^\top \mathbf{X}_{*b})^{-1}_{jj}$$

  (This assumes that we resample the observations. If we resample the residuals, the design matrix remains $\mathbf{X}$.)