Nonparametric Inference in Regression Analysis

University of California, San Diego
Instructor: Armin Schwartzman

why we have M-estimators? to be more robust to outliers

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 is 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.)

Student confidence interval for the mean

- Suppose we have a sample $X_1, \dots, X_n$ IID from some distribution on $\mathbb{R}$.
- Suppose the underlying distribution has a well-defined mean μ and that we want to compute a $(1-\alpha)$-confidence interval for μ.
- First assume that the distribution is normal $\mathcal{N}(\mu, \sigma^2)$, with variance σ^2 unknown, as is often the case.
- The (two-sided) Student $(1-\alpha)$-confidence interval for θ is

$$\bar{X} \pm t_{n-1}^{(\alpha/2)} \frac{S}{\sqrt{n}}$$

  where $\bar{X}$ is the sample mean, S is the sample standard deviation, and $t_m^{(\alpha)}$ is the α-quantile of the t-distribution with m degrees of freedom

- This interval hinges on the fact that

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has the t-distribution with $n-1$ degrees of freedom when the sample is normal.

Indeed, for any $a < b$,

$$\mathbb{P}\left(\bar{X} + aS/\sqrt{n} \le \mu \le \bar{X} + bS/\sqrt{n}\right) = \mathbb{P}\left(-b \le T \le -a\right)$$

- The confidence level is exact if the population is indeed normal.

It is asymptotically correct if the population has finite variance, because of the Central Limit Theorem (CLT).

In practice, it is approximately correct if the sample is large enough and the underlying distribution is not too asymmetric or heavy-tailed.

Montecarlo confidence interval for the mean

- If we could simulate new observations from the same distribution (denoted F), then we could estimate the distribution of this t-ratio.
- Let B be a large integer.

1. For $b = 1, \ldots, B$, do the following:
   (a) Generate $X_1^{*b}, \ldots, X_n^{*b}$ iid from F.
   (b) Compute

$$T_b = \frac{\bar{X}_b - \mu}{S_b/\sqrt{n}}, \quad \text{where}$$

$$\bar{X}_b = \frac{1}{n}\sum_{i=1}^{n} X_i^{*b}, \quad S_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^{*b} - \bar{X}_b)^2$$

2. Compute $t_{\mathrm{MC}}^{(\alpha/2)}$, the α-quantile of $\{T_b : b = 1, \ldots, B\}$.

- A Montecarlo $(1-\alpha)$-CI for $\mu = \mathrm{mean}(F)$ is

$$\left[\bar{X} + t_{\mathrm{MC}}^{(\alpha/2)}\frac{S}{\sqrt{n}} \quad , \quad \bar{X} + t_{\mathrm{MC}}^{(1-\alpha/2)}\frac{S}{\sqrt{n}}\right]$$

The nonparametric bootstrap

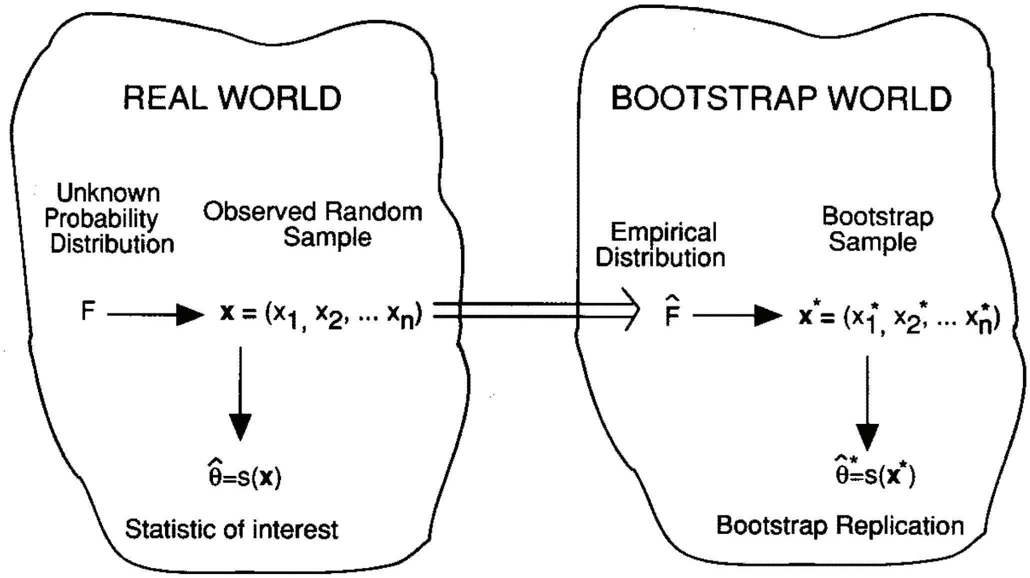

Figure 8.1. *A schematic diagram of the bootstrap as it applies to one-sample problems. In the real world, the unknown probability distribution F gives the data $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ by random sampling; from $\mathbf{x}$ we calculate the statistic of interest $\hat{\theta} = s(\mathbf{x})$. In the bootstrap world, $\hat{F}$ generates $\mathbf{x}^*$ by random sampling, giving $\hat{\theta}^* = s(\mathbf{x}^*)$. There is only one observed value of $\hat{\theta}$, but we can generate as many bootstrap replications $\hat{\theta}^*$ as affordable. The crucial step in the bootstrap process is "$\Longrightarrow$", the process by which we construct from $\mathbf{x}$ an estimate $\hat{F}$ of the unknown population F.*

36

The nonparametric bootstrap interval for the mean

- The following procedure is nonparametric — it does not assume a particular parametric model for the distribution of the data.

  The idea is to use resampling to estimate the distribution of the t-ratio.

- Define the sample (aka empirical) distribution as the uniform distribution over the sample denoted by $\hat{F}$.

> Generating an iid sample of size k from the empirical distribution is done by sampling with replacement k times from the data
> $$\{X_1, \ldots, X_n\}$$

why we sample with replacement?
-> to make sure the sample size is the same

Note that even if all the observations $X_1, \ldots, X_n$ are distinct, a sample from the empirical distribution may contain many repeats and may not include all the observations.

The nonparametric bootstrap interval for the mean

- Let B be a large integer.

  1. For $b = 1, \ldots, B$, do the following:
     (a) Generate $X_1^{*b}, \ldots, X_n^{*b}$ iid from $\hat{F}$.
     (b) Compute the corresponding t-ratio

$$T_b = \frac{\bar{X}_b - \bar{X}}{S_b/\sqrt{n}}, \quad \text{where}$$

$$\bar{X}_b = \frac{1}{n}\sum_{i=1}^{n} X_i^{*b}, \quad S_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^{*b} - \bar{X}_b)^2$$

  2. Compute $t_{\text{boot}}^{(\alpha)}$, the α-quantile of $\{T_b : b = 1, \ldots, B\}$

- A bootstrap $(1-\alpha)$-CI for $\mu = \text{mean}(F)$ is

$$\left[\bar{X} + t_{\text{boot}}^{(\alpha/2)} \frac{S}{\sqrt{n}} \quad , \quad \bar{X} + t_{\text{boot}}^{(1-\alpha/2)} \frac{S}{\sqrt{n}}\right]$$

  Note that the confidence level is not exact.

The bootstrap variance estimate

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

- We can compute it by integration if F (or its density) is known in closed form.

- We can compute it by Monte Carlo integration if we can simulate from F.

  Let B be a large integer.

  1. For $b = 1, \ldots, B$:
     (a) Sample $X_1^{\bullet b}, \ldots, X_n^{\bullet b}$ IID from F.
     (b) Compute $D_b = \Lambda(X_1^{\bullet b}, \ldots, X_n^{\bullet b})$.
  2. Compute the sample mean and variance

$$\bar{D} = \frac{1}{B}\sum_{b=1}^{B} D_b, \qquad \widehat{\text{SE}}_{\text{MC}}^2 = \frac{1}{B-1}\sum_{b=1}^{B}\left(D_b - \bar{D}\right)^2$$

  (MC = Monte Carlo)

- We can estimate it by the nonparametric bootstrap.

  The procedure is the same as above except that we sample from $\hat{F}$ (the sample distribution) instead of F (the population distribution).

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

  Let $\widehat{\mathrm{SE}}_{\mathrm{boot}}$ denote the bootstrap variance estimate.

Bootstrap confidence intervals

- Consider a functional $\mathcal{A}$ and let $\theta = \mathcal{A}(F)$.

  For example, $\mathcal{A}(F) = \mathrm{median}(F)$ or $\mathcal{A}(F) = \mathrm{MAD}(F)$, etc.

- Suppose we want a $(1-\alpha)$-confidence interval for θ.
- Define $\hat{\theta} = \mathcal{A}(\hat{F})$, which is the plug-in estimate for θ.
- The bootstrap procedure is is based on generating many bootstrap samples and computing the statistic of interest on each sample.

  Let B be a large integer. For $b = 1, \ldots, B$, do the following:

  1. Generate $X_1^{*b}, \ldots, X_n^{*b}$ iid from $\hat{F}$.
  2. Compute
  $$\hat{\theta}_{*b} = \mathcal{A}(\hat{F}_{*b})$$
  where $\hat{F}_{*b}$ is the sample distribution of $X_1^{*b}, \ldots, X_n^{*b}$.

  Let $\hat{\theta}_{(*b)}$ denote the b-th largest bootstrap statistic, so that

  $$\hat{\theta}_{(*1)} \leq \cdots \leq \hat{\theta}_{(*B)}$$

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. (This interval is very closely related to the next one, which requires fewer assumptions.)

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 $\widehat{f}$.

Bootstrap inference for regression

why you apply this method
instead of that method?

case 2: x is not re-sampled

if x is observed, it's
changed
if x is designed, it's fixed

- 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 $\mathbf{X}$.)