The Bootstrap[1]

STA442/2101 Fall 2013

[1] See last slide for copyright information.

Background Reading

- Davison's *Statistical models* has almost nothing.
- The best we can do for now is the Wikipedia under Bootstrapping (Statistics)

Overview

Sampling distributions

- Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from some distribution F.
- $T = T(\mathbf{X})$ is a statistic (could be a vector of statistics).
- Need to know about the distribution of T.
- Sometimes it's not easy, even asymptotically.

Sampling distribution of T: The elementary version

- Sample repeatedly from this population (pretend).
- For each sample, calculate T.
- Make a relative frequency histogram of the T values you observe.
- As the number of samples becomes very large, the histogram approximates the distribution of T.

What is a bootstrap?
Pull yourself up by your bootstraps

This photograph was taken by Tarquin. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. For more information, see the entry at the wikimedia site.

The (statistical) Bootstrap

- Select a random sample from the population.
- If the sample size is large, the sample is similar to the population.
- Sample repeatedly from the sample. This is called *resampling*.
- Sample from the sample? Think of putting the sample data values in a jar . . .
- Calculate the statistic for every bootstrap sample.
- A histogram of the resulting values approximates the shape of the sampling distribution of the statistic.

Notation

- Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from some distribution F.
- $T = T(\mathbf{X})$ is a statistic (could be a vector of statistics).
- Form a "bootstrap sample" $\mathbf{X}^*$ by sampling n values from $\mathbf{X}$ *with replacement.*
- Repeat this process B times, obtaining $\mathbf{X}_1^*, \ldots, \mathbf{X}_B^*$
- Calculate the statistic for each bootstrap sample, obtaining $T_1^*, \ldots, T_B^*$

What can we do with $T_1^*, \ldots, T_B^*$?

We can do lots of interesting things, including constructing confidence intervals (sometimes).

In this course, we will use it mostly to calculate standard errors the easy way.

- Approximate $Var(T)$ with the sample variance of the T^* values.
- Or, if T is a vector, approximate the asymptotic covariance matrix with the sample covariance matrix of the T^* vectors.
- It can be a lot less work than the delta method.

Why does it work?

- Resampling from $\mathbf{X}$ is the same as simulation of a random variable whose distribution is the empirical distribution function $\widehat{F}(x)$.
- $\widehat{F}(x)$ is the proportion of sample observations less than or equal to x.
- The empirical distribution function converges almost surely to the true distribution function by the Law of Large Numbers.
- Moments and lots of other useful quantities are smooth functions of the distribution function.
- So as $n \to \infty$ and $B \to \infty$, bootstrap sample moments of $T_1^*, \ldots, T_B^*$ converge to the corresponding moments of the sampling distribution of T.

Example: Distribution-free regression

Independently for $i = 1, \ldots, n$, let

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where

- X_i and ϵ_i come from unknown distributions
- $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2$
- Moments of X_i will be denoted $E(X)$, $E(X^2)$, etc.
- X_i and ϵ_i are independent

Observable data consist of the pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$.

Estimation

Estimate β_0 and β_1 as usual by

$$\begin{aligned}\widehat{\beta}_0 &= \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2} \\ &= \frac{\sum_{i=1}^n X_i Y_i - n\overline{X}\,\overline{Y}}{\sum_{i=1}^n X_i^2 - n\overline{X}^2} \text{ and}\end{aligned}$$

$$\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_0 \overline{X}$$

- Consistency follows from the Law of Large Numbers and continuous mapping.
- Looks like $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are asymptotically normal.
- Use this to get tests and confidence intervals.

Set up the problem systematically

Denoting the "data vector" for case i by $\mathbf{D}_i$,

$$\mathbf{D}_i = \begin{pmatrix} X_i \\ X_i^2 \\ Y_i \\ X_iY_i \end{pmatrix}, \text{ and } \quad \overline{\mathbf{D}}_n = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i \\ \\ \frac{1}{n}\sum_{i=1}^n X_i^2 \\ \\ \frac{1}{n}\sum_{i=1}^n Y_i \\ \\ \frac{1}{n}\sum_{i=1}^n X_iY_i \end{pmatrix} \stackrel{a.s.}{\rightarrow} \begin{pmatrix} E(X) \\ \\ E(X^2) \\ \\ E(Y) \\ \\ E(XY) \end{pmatrix} = \boldsymbol{\mu}$$

$$\text{Then } \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} = g(\overline{\mathbf{D}}_n) \stackrel{a.s.}{\rightarrow} g(\boldsymbol{\mu}) = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

What would we do next
To use the Central Limit Theorem and Delta method?

CLT says $\sqrt{n}(\overline{\mathbf{D}}_n - \boldsymbol{\mu}) \stackrel{d}{\rightarrow} \mathbf{T} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Delta method says $\sqrt{n}(g(\overline{\mathbf{D}}_n) - g(\boldsymbol{\mu})) \stackrel{d}{\rightarrow} \mathbf{Y} \sim N(\mathbf{0}, \dot{\mathrm{g}}(\boldsymbol{\mu})\boldsymbol{\Sigma}\dot{\mathrm{g}}(\boldsymbol{\mu})')$, where

$$\mathbf{D}_i = \left(\begin{array}{c} X_i \\ X_i^2 \\ Y_i \\ X_iY_i \end{array} \right)$$

- Calculate $\boldsymbol{\Sigma}(\boldsymbol{\mu}) = V(\mathbf{D}_i)$ by hand.
- Calculate $\dot{\mathrm{g}}(\boldsymbol{\mu})$ by hand.
- Estimate $\boldsymbol{\Sigma}$ using sample moments, by computer.
- Calculate the estimated asymptotic covariance matrix $\frac{1}{n}\dot{\mathrm{g}}(\overline{\mathbf{D}}_n)\widehat{\boldsymbol{\Sigma}}_n\dot{\mathrm{g}}(\overline{\mathbf{D}}_n)'$ by computer
- Use that in confidence intervals and Wald-like tests.

It's a lot of work.

- Most problems have more than one explanatory variable
- You could easily make a "little" mistake.

Bootstrap approach: All by computer

- Earlier discussion implies $\widehat{\boldsymbol{\beta}}$ is asymptotically normal.
- All we need is a good estimated covariance matrix of the sampling distribution.
- Put data vectors $\mathbf{D}_i = (\mathbf{X}_i, Y_i)'$ in a jar[2].
- Sample n vectors with replacement, yielding $\mathbf{D}_1^*$. Fit the regression model, obtaining $\widehat{\boldsymbol{\beta}}_1^*$.
- Repeat B times. This yields $\widehat{\boldsymbol{\beta}}_1^* \ldots \widehat{\boldsymbol{\beta}}_B^*$.
- The sample covariance matrix of $\widehat{\boldsymbol{\beta}}_1^* \ldots \widehat{\boldsymbol{\beta}}_B^*$ is the estimated asymptotic covariance matrix we want.

[2] The definition of $\mathbf{D}_i$ is different now.

Remark

This is not a typical bootstrap regression.

- Usually people fit a model and then bootstrap the residuals, not the whole data vector.
- Bootstrapping the residuals applies to conditional regression (conditional on $\mathbf{X} = \mathbf{x}$).
- Our regression model is unconditional.
- The large-sample arguments are simpler in the unconditional case.

Copyright Information