Bootstrap (Part 4)

Christof Seiler

Stanford University, Spring 2016, Stats 205
Overview

- **So far:**
  - Nonparametric bootstrap on the rows (e.g. regression, PCA with random rows and columns)
  - Nonparametric bootstrap on the residuals (e.g. regression)
  - Parametric bootstrap (e.g. PCA with fixed rows and columns)
  - Studentized bootstrap

- **Today:**
  - Bias-Corrected-accelerated (BCa) bootstrap
  - From BCa to ABC
Motivation

- Correlation coefficient of bivariate normal with $\rho = 0.577$

```r
sigma = matrix(nrow = 2, ncol = 2)
diag(sigma) = 1
rho = 0.577
sigma[1,2] = sigma[2,1] = rho
sigma
```

```
  [,1] [,2]
[1,] 1.000 0.577
[2,] 0.577 1.000
```

- Distribution of sample correlation coefficient ($n = 10$)
- Compare: Percentile, Studentized, and Bias-Corrected-Accelerated (BCa) bootstrap
bias = rho - \texttt{mean}(\text{corhat}); \text{ bias}

## [1] 0.0217078
Motivation

Percentile Bootstrap
Motivation

- Studentized bootstrap with variance stabilization fails due to numerical problems
Motivation

Studentized Bootstrap Without Variance Stabilization
Motivation

Percentile Bootstrap

Studentized Bootstrap Without Variance Stabilization

BCa Bootstrap
BCa Bootstrap

- The bias-corrected bootstrap is similar to the percentile bootstrap.
- Recall the percentile bootstrap:
  - Take bootstrap samples
    \[
    \hat{\theta}^*, \ldots, \hat{\theta}^B
    \]
  - Order them
    \[
    \hat{\theta}^{(*1)}, \ldots, \hat{\theta}^{(*B)}
    \]
  - Define interval as
    \[
    (\hat{\theta}^{(*B\alpha)}, \hat{\theta}^{(*B(1-\alpha))})
    \]
    (assuming that \(B\alpha\) and \(B(1 - \alpha)\) are integers)
BCa Bootstrap

- Assume that there is an monotone increasing transformation $g$ such that
  \[ \phi = g(\theta) \quad \text{and} \quad \hat{\phi} = g(\hat{\theta}) \]

- The BCa bootstrap is based on this model
  \[ \frac{\hat{\phi} - \phi}{\sigma_{\phi}} \sim N(-z_0, 1) \quad \text{with} \quad \sigma_{\phi} = 1 + a_{\phi} \]

- Which is a generalization of the usual normal approximation
  \[ \frac{\hat{\theta} - \theta}{\sigma} \sim N(0, 1) \]
BCa Bootstrap

- $\hat{z}_0$ is the bias estimate
- $\hat{z}_0$ measures discrepancy between the median of $\hat{\theta}^*$ and $\hat{\theta}$
- It is estimated with

$$\hat{z}_0 = \Phi^{-1} \left( \frac{\#\{\hat{\theta}^{*b} < \hat{\theta}\}}{B} \right)$$

- We obtain $\hat{z}_0 = 0$ if half of the $\hat{\theta}^{*b}$ values are less than or equal to $\hat{\theta}$
BCa Bootstrap

- \( \hat{a} \) is the skewness estimate
- \( \hat{a} \) measures the rate of change of the standard error of \( \hat{\theta} \) with respect to the true parameter \( \theta \)
- It is estimated using the Jackknife
  - Delete \( i \)th observation in original sample denote new sample by \( \hat{\theta}(i) \) and estimate
    \[
    \hat{\theta}(\cdot) = \sum_{i=1}^{n} \frac{\hat{\theta}(i)}{n}
    \]
  - Then
    \[
    \hat{a} = \frac{\sum_{i=1}^{n}(\hat{\theta}(\cdot) - \hat{\theta}(i))^3}{6\left\{\sum_{i=1}^{n}(\hat{\theta}(\cdot) - \hat{\theta}(i))^2\right\}^{3/2}}
    \]
The bias-corrected version makes two additional corrections to the percentile version.

By redefining lower $\alpha_1$ and upper $\alpha_2$ levels as

$$\alpha_1 = \Phi \left( \hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})} \right)$$

$$\alpha_2 = \Phi \left( \hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})} \right)$$

with $z^{(\alpha)}$ being the 100$\alpha$ percentile of standard normal and $\Phi$ normal CDF.

When $\hat{a}$ and $\hat{z}_0$ are equal to zero then $\alpha_1 = \alpha$ and $\alpha_2 = 1 - \alpha$.

The interval is then given by

$$\left( \hat{\theta}(\star B\alpha_1), \hat{\theta}(\star B\alpha_2) \right)$$

(assuming that $B\alpha_1$ and $B\alpha_2$ are integers)
BCa Bootstrap

- Same asymptotic accuracy as the studentized bootstrap
- Can handle out of range problem as well
- Efron (1987) for detailed justification of this model
BCa Bootstrap in R

```r
library(bootstrap)
xdata = matrix(rnorm(30), ncol=2); n = 15
theta = function(x, xdata) {
  cor(xdata[x,1], xdata[x,2])
}
results = bcanon(1:n, 100, theta, xdata,
  alpha=c(0.025, 0.975))
results$confpoints

## alpha   bca point
## [1,] 0.025 -0.39659
## [2,] 0.975  0.69326
```
### Properties of Different Bootstrap Methods

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th>Percentile</th>
<th>Studentized*</th>
<th>BCa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptotic Accuracy</td>
<td>$O(\sqrt{n})$</td>
<td>$O(\sqrt{n})$</td>
<td>$O(1/n)$</td>
<td>$O(1/n)$</td>
</tr>
<tr>
<td>Range-Preserving</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Transformation-Invariant</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Bias-Correcting</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Skeweness-Correcting</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>$\hat{\sigma}, \hat{\sigma}^*$ required</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Analytic constant or variance stabilizing tranformation required</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* with variance stabilization
Properties of Different Bootstrap Methods

For nonparametric bootstrap:

Are the standard deviations $\hat{\sigma}$, $\hat{\sigma}^*$ available?

Yes

Is there a correlation between $\hat{\sigma}^*$ and $\hat{\theta}^*$?

Yes

Variance stabilised bootstrap-t or BCa

No

Bootstrap-t or BCa

No

BCa

Source: Carpenter and Bithell (2000)
Many More Topics

- Using the bootstrap for better confidence in model selection (Efron 2014)
- Using the jackknife and the infinitesimal jackknife for confidence intervals in random forests prediction or classification (Wager, Hastie, and Efron 2014)
Approximate Bayesian Computation (ABC)

- **Goal:** We wish to sample from the posterior distribution \( p(\theta|D) \) given data \( D \)

\[
p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}
\]

- **Setting:**
  - The likelihood \( p(D|\theta) \) is hard to evaluate or expensive to compute (e.g. missing normalizing constant)
  - Easy to sample from likelihood \( p(D|\theta) \)
  - Easy to sample from prior \( p(\theta) \)

- **Examples:**
  - Population genetics (latent variables)
  - Ecology, epidemiology, systems biology (models based on differential equations)
Approximate Bayesian Computation (ABC)

- **Sampling algorithm (with data \( D = \{y_1, \ldots, y_n\} \)):**
  1. Sample \( \theta_i \sim p(\theta) \)
  2. Sample \( x_i \sim p(x|\theta_i) \)
  3. Reject \( \theta_i \) if \( x_i \neq y_j \) for \( j = 1, \ldots, n \)

- **ABC sampling (define statistics \( \mu \), distance \( \rho \), and tolerance \( \epsilon \)):**
  1. Sample \( \theta_i \sim p(\theta) \)
  2. Sample \( \hat{D}_i = \{x_1, \ldots, x_k\} \sim p(x|\theta_i) \)
  3. Reject \( \theta_i \) if \( \rho(\mu(\hat{D}_i), \mu(D)) > \epsilon \)
1. Compute summary statistic $\mu$ from observational data.

2. Given a certain model, perform $n$ simulations, each with a parameter drawn from the prior distribution.

3. Compute summary statistic $\mu_i$ for each simulation.

4. Based on a distance $\rho(\cdot, \cdot)$ and a tolerance $\epsilon$, decide for each simulation whether its summary statistic is sufficiently close to that of the observed data.

5. Approximate the posterior distribution of $\theta$ from the distribution of parameter values $\theta$, associated with accepted simulations.
References

- Hall (1992). The Bootstrap and Edgeworth Expansion
- Efron and Tibshirani (1994). An Introduction to the Bootstrap
- Marin, Pudlo, Robert, and Ryder (2012). Approximate Bayesian Computational Methods
- Efron (2014). Estimation and Accuracy after Model Selection