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

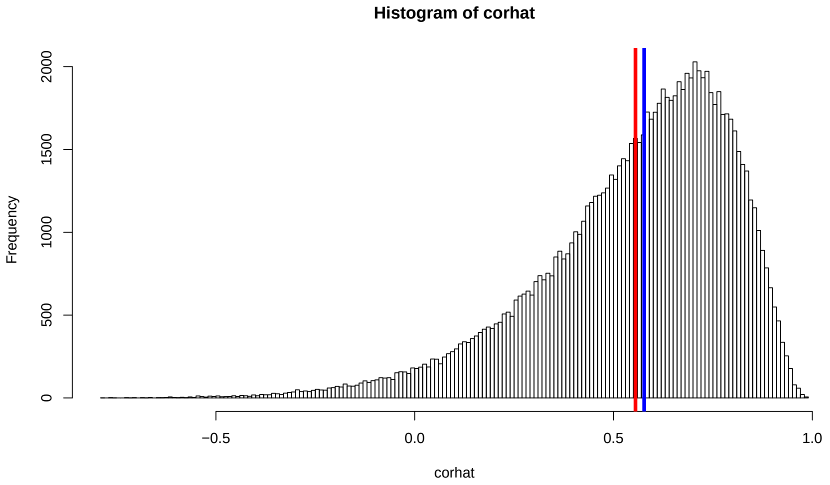
- Correlation coefficient of bivariate normal with $\rho = 0.577$

```
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

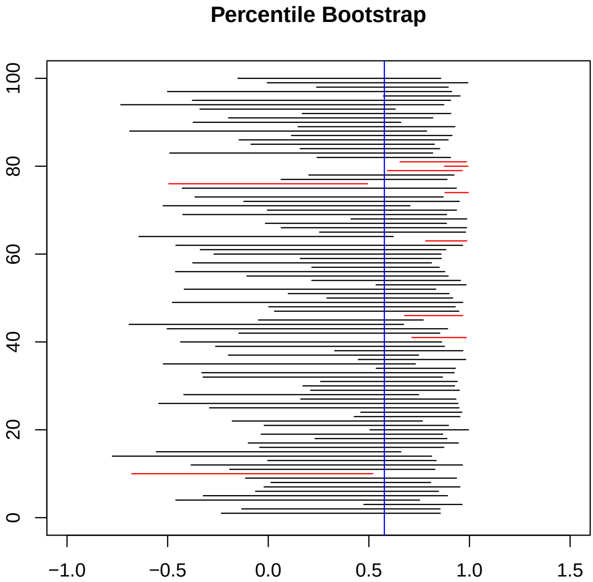
Motivation




```
bias = rho - mean(corhat); bias
```

```
## [1] 0.0217078
```

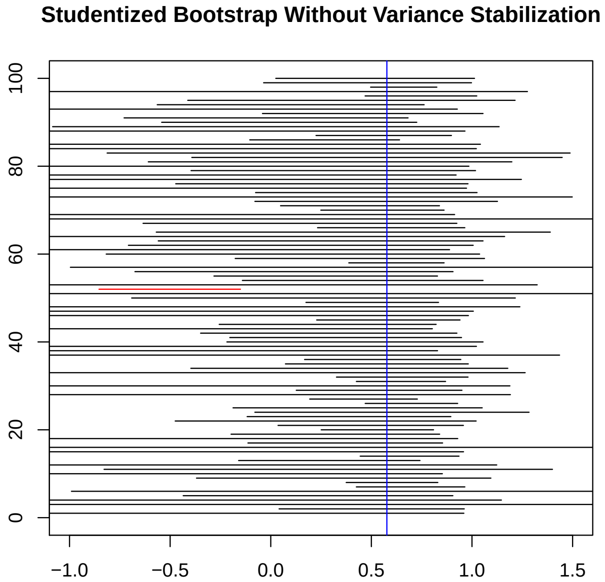
Motivation



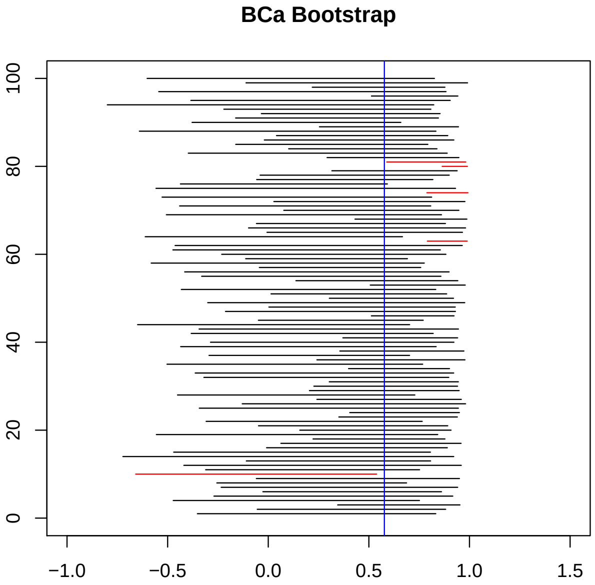
Motivation

- Studentized bootstrap with variance stabilization fails due to numerical problems

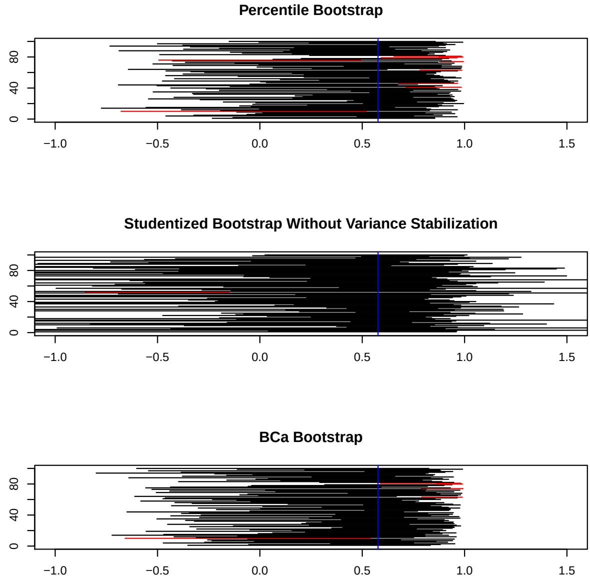
Motivation


Studentized Bootstrap Without Variance Stabilization
0
20
40
60
80
100
−1.0
−0.5
0.0
0.5
1.0
1.5

Motivation



Motivation


Percentile Bootstrap
0
40
80
−1.0
−0.5
0.0
0.5
1.0
1.5
Studentized Bootstrap Without Variance Stabilization
0
40
80
−1.0
−0.5
0.0
0.5
1.0
1.5
BCa Bootstrap
0
40
80
−1.0
−0.5
0.0
0.5
1.0
1.5

BCa Bootstrap

- The bias-corrected bootstrap is similar to the percentile bootstrap
- Recall the percentile bootstrap:
- Take bootstrap samples

$$\hat{\theta}^{*1}, \ldots, \hat{\theta}^{*B}$$

- 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 θ
- It is estimated using the Jackknife
  - Delete ith 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

$$\hat{a} = \frac{\sum_{i=1}^{n} (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^3}{6\{\sum_{i=1}^{n} (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2\}^{3/2}}$$

BCa Bootstrap

- The bias-corrected version makes two additional corrections to the percentile version
- By redefining lower α_1 and upper α_2 levels as

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})}\right) \qquad \alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})}\right)$$

  with $z^{(\alpha)}$ being the 100α percentile of standard normal and Φ 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

$$(\hat{\theta}^{(*B\alpha_1)}, \hat{\theta}^{(*B\alpha_2)})$$

  (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

```
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
```

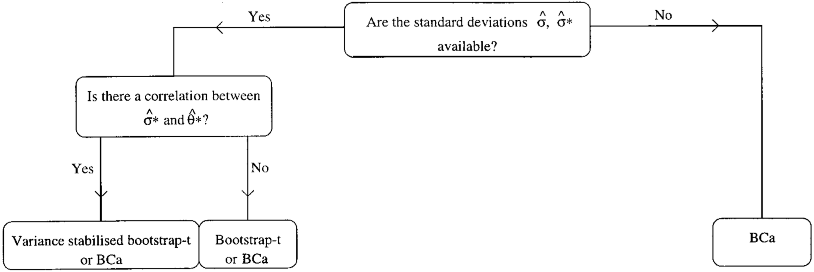
Properties of Different Boostrap Methods

	Standard	Percentile	Studentized*	BCa
Asymptotic Acurracy	$O(\sqrt{n})$	$O(\sqrt{n})$	$O(1/n)$	$O(1/n)$
Range-Preserving	No	Yes	No	Yes
Transformation-Invariant	No	Yes	No	Yes
Bias-Correcting	No	No	No	Yes
Skeweness-Correcting	No	Yes	Yes	Yes
$\hat{\sigma}, \hat{\sigma}^*$ required	No	No	Yes	No
Analytic constant or variance stabilizing tranformation required	No	No	Yes	Yes

* with variance stabilization

Properties of Different Boostrap Methods

For nonparametric boostrap:



Source: Carpenter and Bithell (2000)

Many More Topics

- Using the boostrap 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)

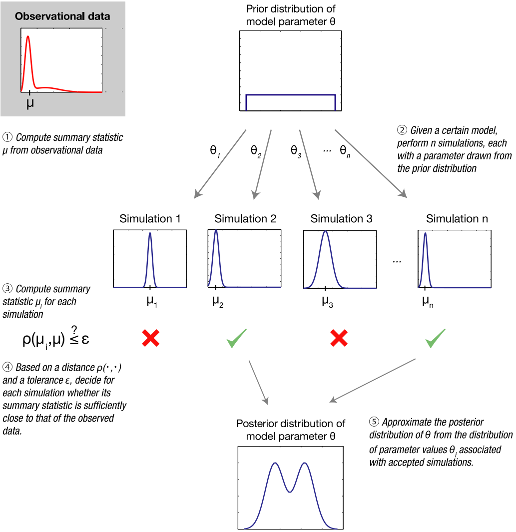
Approximate Bayesian Computation (ABC)

- 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 θ_i if

$$x_i \neq y_j \text{ for } j = 1, \ldots, n$$

- ABC sampling (define statistics μ, distance ρ, and tolerance ϵ):
  1. Sample $\theta_i \sim p(\theta)$
  2. Sample $\hat{D}_i = \{x_1, \ldots, x_k\} \sim p(x|\theta_i)$
  3. Reject θ_i if

$$\rho(\mu(\hat{D}_i), \mu(D)) > \epsilon$$


Observational data
μ
Prior distribution of model parameter θ
① Compute summary statistic μ from observational data
② Given a certain model, perform n simulations, each with a parameter drawn from the prior distribution
θ1 θ2 θ3 ... θn
Simulation 1
Simulation 2
Simulation 3
Simulation n
...
③ Compute summary statistic μi for each simulation
μ1 μ2 μ3 μn
ρ(μi,μ) ≤? ε
④ Based on a distance ρ(·,·) and a tolerance ε, decide for each simulation whether its summary statistic is sufficiently close to that of the observed data.
Posterior distribution of model parameter θ
⑤ Approximate the posterior distribution of θ from the distribution of parameter values θi associated with accepted simulations.

References

- Efron (1987). Better Bootstrap Confidence Intervals
- Hall (1992). The Bootstrap and Edgeworth Expansion
- Efron and Tibshirani (1994). An Introduction to the Bootstrap
- Carpenter and Bithell (2000). Bootstrap Conidence Intervals: When, Which, What? A Practical Guide for Medical Statisticians
- Marin, Pudlo, Robert, and Ryder (2012). Approximate Bayesian Computational Methods
- Efron (2014). Estimation and Accuracy after Model Selection
- Wager, Hastie, and Efron (2014). Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife