Bootstrap Methods: Another Look at the Jackknife

Based on Brad Efron's 1979 Paper

Ziang Song

March 19, 2024

Question: How to estimate the bias and variance of a statistics of interest?

Introduction to Bootstrap

Question: How to estimate the bias and variance of a statistics of interest?

Prior works:

- Quenouille-Tukey jackknife (1949,1958)
- The infinitesimal jackknife (Jaeckel, L. (1972))

Description of the Jackknife

Suppose θ is some parameter of interest, and $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ is our estimate of θ based on samples.

- Jackknife is a resampling technique for estimating the bias and variance.

Description of the Jackknife

Suppose θ is some parameter of interest, and $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ is our estimate of θ based on samples.

- Jackknife is a resampling technique for estimating the bias and variance.
- Procedure: for each i from 1 to n, compute $\hat{\theta}_{(-i)}$, the estimate of θ excluding the i-th observation.

Description of the Jackknife

Suppose θ is some parameter of interest, and $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ is our estimate of θ based on samples.

- Jackknife is a resampling technique for estimating the bias and variance.
- Procedure: for each i from 1 to n, compute $\hat{\theta}_{(-i)}$, the estimate of θ excluding the i-th observation.
- Bias and Variance Estimation:
  - Bias: $\text{Bias}_{\text{jack}}(\hat{\theta}) = (n-1)\left(\hat{\theta}_{\text{jack}} - \hat{\theta}\right)$
  - Variance: $\text{Var}_{\text{jack}}(\hat{\theta}) = \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{(-i)} - \hat{\theta}_{\text{jack}}\right)^2$

where $\hat{\theta}_{\text{jack}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(-i)}$, the average of all jackknife estimates.

Limitations of the Jackknife

- The Jackknife may not perform well in certain scenarios.
- It lacks theoretical basis.

Introduction to Bootstrap

Bootstrap is a resampling technique for estimating the distribution of a statistic.

Introduction to Bootstrap

Bootstrap is a resampling technique for estimating the distribution of a statistic.

- **Basic Idea:** Use the empirical distribution to approximate the true distribution.

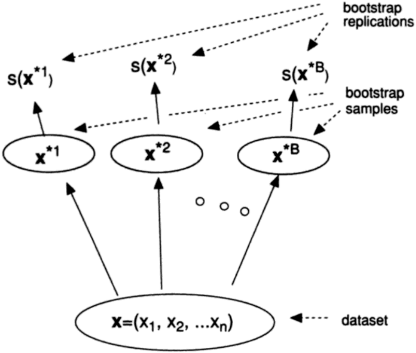


Figure: Bootstrap process (Efron, B., & Tibshirani, R. J. (1994))

Bootstrap Methods - Overview

Notations:

- $\mathbf{X} = (X_1, X_2, \ldots, X_n)$: Original sample.
- $\mathbf{x} = (x_1, x_2, \ldots, x_n)$: Observed values.
- $\hat{F}$: Empirical distribution based on the observed sample.
- $R(\mathbf{X}, F)$: Statistic of interest as a function of $\mathbf{X}$ and F.
- R^*: Bootstrap analogue of R.

Bootstrap Methods - Overview

Notations:

- $\mathbf{X} = (X_1, X_2, \ldots, X_n)$: Original sample.
- $\mathbf{x} = (x_1, x_2, \ldots, x_n)$: Observed values.
- $\hat{F}$: Empirical distribution based on the observed sample.
- $R(\mathbf{X}, F)$: Statistic of interest as a function of $\mathbf{X}$ and F.
- R^*: Bootstrap analogue of R.

Methodology:

1. Construct the empirical distribution $\hat{F}$ by placing mass $\frac{1}{n}$ at each data point $x_1, x_2, \ldots, x_n$.

Bootstrap Methods - Overview

Notations:

- $\mathbf{X} = (X_1, X_2, \ldots, X_n)$: Original sample.
- $\mathbf{x} = (x_1, x_2, \ldots, x_n)$: Observed values.
- $\hat{F}$: Empirical distribution based on the observed sample.
- $R(\mathbf{X}, F)$: Statistic of interest as a function of $\mathbf{X}$ and F.
- R^*: Bootstrap analogue of R.

Methodology:

1. Construct the empirical distribution $\hat{F}$ by placing mass $\frac{1}{n}$ at each data point $x_1, x_2, \ldots, x_n$.
2. Draw bootstrap samples $\mathbf{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)$ from $\hat{F}$.

Bootstrap Methods - Overview

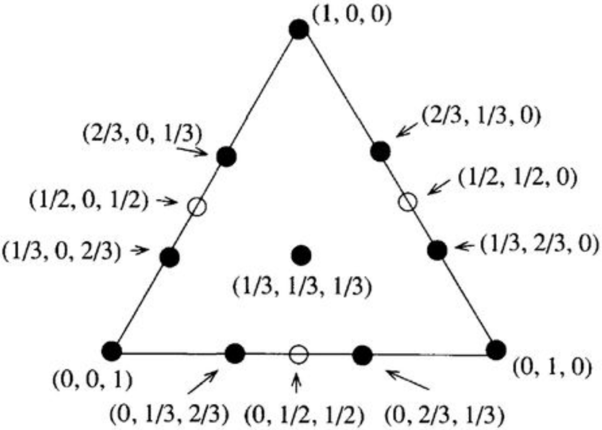
Notations:

- $\mathbf{X} = (X_1, X_2, \ldots, X_n)$: Original sample.
- $\mathbf{x} = (x_1, x_2, \ldots, x_n)$: Observed values.
- $\hat{F}$: Empirical distribution based on the observed sample.
- $R(\mathbf{X}, F)$: Statistic of interest as a function of $\mathbf{X}$ and F.
- R^*: Bootstrap analogue of R.

Methodology:

1. Construct the empirical distribution $\hat{F}$ by placing mass $\frac{1}{n}$ at each data point $x_1, x_2, \ldots, x_n$.
2. Draw bootstrap samples $\mathbf{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)$ from $\hat{F}$.
3. Approximate the sampling distribution of $R(\mathbf{X}, F)$ by the bootstrap distribution of $R^* = R(\mathbf{X}^*, \hat{F})$.

Simplex for $n = 3$. The solid points indicates the support points of the bootstrap distribution while the open circles are the jackknife points.



A simple example

- Consider a probability distribution F putting all of its mass at zero or one, and let $\theta(F) = \mathrm{Prob}_F(X = 1)$. The statistic of interest is

$$R(\mathbf{X}, F) = \bar{X} - \theta(F), \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

A simple example

- Consider a probability distribution F putting all of its mass at zero or one, and let $\theta(F) = \mathrm{Prob}_F(X = 1)$. The statistic of interest is

$$R(\mathbf{X}, F) = \bar{X} - \theta(F), \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- Having observed $\mathbf{X} = \mathbf{x}$, the bootstrap sample $\mathbf{X}^*$ has each component $X_i^* \sim_{iid} Bernoulli(\bar{x})$.

A simple example

- Consider a probability distribution F putting all of its mass at zero or one, and let $\theta(F) = \mathrm{Prob}_F(X = 1)$. The statistic of interest is

$$R(\mathbf{X}, F) = \bar{X} - \theta(F), \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- Having observed $\mathbf{X} = \mathbf{x}$, the bootstrap sample $\mathbf{X}^*$ has each component $X_i^* \sim_{iid} Bernoulli(\bar{x})$.
- $R^* = R(\mathbf{X}^*, \hat{F}) = \bar{X}^* - \bar{x}$ has mean and variance

$$\mathbb{E}_* R^* = 0, \quad \mathrm{Var}_* R^* = \frac{\bar{x}(1-\bar{x})}{n}.$$

A simple example

- Consider a probability distribution F putting all of its mass at zero or one, and let $\theta(F) = \mathrm{Prob}_F(X = 1)$. The statistic of interest is

$$R(\mathbf{X}, F) = \bar{X} - \theta(F), \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- Having observed $\mathbf{X} = \mathbf{x}$, the bootstrap sample $\mathbf{X}^*$ has each component $X_i^* \sim_{iid} Bernoulli(\bar{x})$.
- $R^* = R(\mathbf{X}^*, \hat{F}) = \bar{X}^* - \bar{x}$ has mean and variance

$$\mathbb{E}_* R^* = 0, \quad \mathrm{Var}_* R^* = \frac{\bar{x}(1-\bar{x})}{n}.$$

- This suggest $\bar{X}$ is unbiased for θ, with variance approximately equal to $\bar{x}(1-\bar{x})/n$.

Calculation of the bootstrap distribution

- Method 1. Direct theoretical calculation, as in the last example.
- Method 2. Monte Carlo approximation.
- Method 3. Taylor series expansion methods.

Choices of $\hat{F}$

Suppose we wish to estimate the median of F using the sample median. Let $\theta(F)$ indicate the median of F, and let $t(X)$ be the sample median. Assume $n = 2m - 1$. We can use different $\hat{F}$:

Choices of $\hat{F}$

Suppose we wish to estimate the median of F using the sample median. Let $\theta(F)$ indicate the median of F, and let $t(X)$ be the sample median. Assume $n = 2m - 1$. We can use different $\hat{F}$:

- Empirical distribution of $x_1, \cdots, x_n$.

Choices of $\hat{F}$

Suppose we wish to estimate the median of F using the sample median. Let $\theta(F)$ indicate the median of F, and let $t(X)$ be the sample median. Assume $n = 2m - 1$. We can use different $\hat{F}$:

- Empirical distribution of $x_1, \cdots, x_n$.
- If we know that F is symmetric, we can replace $\hat{F}$ by

$$\hat{F}_{\mathrm{SYM}} : \text{probability mass } \frac{1}{2n-1} \text{ at } x_{(1)}, \cdots, x_{(n)},$$
$$2x_{(m)} - x_{(1)}, \cdots, 2x_{(m)} - x_{(n)}$$

Choices of $\hat{F}$

Suppose we wish to estimate the median of F using the sample median. Let $\theta(F)$ indicate the median of F, and let $t(X)$ be the sample median. Assume $n = 2m - 1$. We can use different $\hat{F}$:

- Empirical distribution of $x_1, \cdots, x_n$.
- If we know that F is symmetric, we can replace $\hat{F}$ by

$$\hat{F}_{\mathrm{SYM}} : \text{probability mass } \frac{1}{2n-1} \text{ at } x_{(1)}, \cdots, x_{(n)},$$
$$2x_{(m)} - x_{(1)}, \cdots, 2x_{(m)} - x_{(n)}$$

- Smoothed Bootstrap: take $X_i^* = \bar{x} + c(x_{I_i} - \bar{x} + \hat{\sigma} Z_i)$ where $I_i \sim_{iid}$ Unif$([n])$ and Z_i are sampled from some fixed distribution.

Choices of $\hat{F}$

TABLE 1*

Trial #	Unsmoothed Bootstrap (3.6)	(3.10)	Smoothed Bootstrap (3.11): Z_i uniform dist. on $[-d/2, d/2]$, $d = 0$	$d = .25$	$d = .5$	$d = 1$	Z_i triangular dist., $\sigma_Z^2 = 1/12$
1	1.07	1.18	1.09	1.10	1.12	1.11	1.16
2	.96	.74	1.10	1.10	1.08	1.09	1.15
3	1.22	.74	1.36	1.35	1.33	1.43	1.52
4	1.38	1.51	1.44	1.41	1.38	1.28	1.30
5	1.00	.83	1.03	1.05	1.09	1.14	1.17
6	1.13	1.21	1.27	1.26	1.23	1.20	1.26
7	1.07	.98	1.01	.94	.83	.79	.92
8	1.51	1.40	1.40	1.45	1.47	1.51	1.50
9	.56	.64	.69	.71	.74	.80	.81
10	1.05	.86	1.14	1.17	1.20	1.13	1.22
Ave.	1.09	1.01	1.15	1.15	1.15	1.15	1.20
S.D.	.26	.30	.23	.23	.23	.23	.22

*Ten Monte Carlo trials of $X_i \sim_{\text{ind}} \mathcal{N}(0, 1)$, $i = 1, 2, \ldots, 13$ were used to compare different bootstrap methods for estimating the expected value of random variable (3.12). The true expectation is 0.95. The quantities tabled are $E_* R^*$, the bootstrap expectation for that trial. The values in the first two columns are for the bootstrap as described originally, and for the symmetrized version (3.8)–(3.10). The smoothed bootstrap expectations were approximated using a secondary Monte Carlo simulation for each trial, $N = 50$, as described in "Method 2," Section 2. Each of these entries estimates the actual value of $E_* R^*$ unbiasedly with a standard error of about .15. The column "$d = 0$" would exactly equal column "(3.6)" if $N \to \infty$.

Error rate estimation in discriminant analysis

Consider a standard linear discriminant analysis problem. The data consists of independent random samples from two unknown distributions F and G.

$$X_i = x_i \sim F, i = 1, 2, \cdots, m$$
$$Y_j = y_j \sim G, j = 1, 2, \cdots, n.$$

Error rate estimation in discriminant analysis

Consider a standard linear discriminant analysis problem. The data consists of independent random samples from two unknown distributions F and G.

$$X_i = x_i \sim F, i = 1, 2, \cdots, m$$
$$Y_j = y_j \sim G, j = 1, 2, \cdots, n.$$

- Define the region B by

$$B = \{z : (\bar{y} - \bar{x})' S^{-1} (z - \frac{\bar{x} + \bar{y}}{2}) > \log m/n\},$$

where $S = (\sum (x_i - \bar{x})'(x_i - \bar{x}) + (y_i - \bar{y})'(y_i - \bar{y}))/(m + n)$.

Error rate estimation in discriminant analysis

Consider a standard linear discriminant analysis problem. The data consists of independent random samples from two unknown distributions F and G.

$$X_i = x_i \sim F, i = 1, 2, \cdots, m$$
$$Y_j = y_j \sim G, j = 1, 2, \cdots, n.$$

- Define the region B by

$$B = \{z : (\bar{y} - \bar{x})' S^{-1} (z - \frac{\bar{x} + \bar{y}}{2}) > \log m/n\},$$

  where $S = (\sum (x_i - \bar{x})'(x_i - \bar{x}) + (y_i - \bar{y})'(y_i - \bar{y}))/(m + n)$.
- Define $\widehat{\mathrm{error}}_F = \#\{i : x_i \in B\}/m$, $\mathrm{error}_F = \mathrm{Prob}_F(X \in B)$.

Error rate estimation in discriminant analysis

Consider a standard linear discriminant analysis problem. The data consists of independent random samples from two unknown distributions F and G.

$$X_i = x_i \sim F, i = 1, 2, \cdots, m$$
$$Y_j = y_j \sim G, j = 1, 2, \cdots, n.$$

- Define the region B by

$$B = \{z : (\bar{y} - \bar{x})' S^{-1} (z - \frac{\bar{x} + \bar{y}}{2}) > \log m/n\},$$

  where $S = (\sum (x_i - \bar{x})'(x_i - \bar{x}) + (y_i - \bar{y})'(y_i - \bar{y}))/(m + n)$.
- Define $\widehat{\text{error}}_F = \#\{i : x_i \in B\}/m$, $\text{error}_F = \text{Prob}_F(X \in B)$.
- We will be interested in $R((\mathbf{X}, \mathbf{Y}), (F, G)) = \text{error}_F - \widehat{\text{error}}_F$.

Error rate estimation in discriminant analysis

The brute force application of the bootstrap via simulation: bootstrap random samples

$$X_i^* = x_i^* \sim \hat{F}, Y_j^* = y_j^* \sim \hat{G}.$$

Error rate estimation in discriminant analysis

The brute force application of the bootstrap via simulation: bootstrap random samples

$$X_i^* = x_i^* \sim \hat{F}, Y_j^* = y_j^* \sim \hat{G}.$$

The bootstrap random variable is

$$R^* = R((\mathbf{X}^*, \mathbf{Y}^*), (\hat{F}, \hat{G})).$$

Error rate estimation in discriminant analysis

The brute force application of the bootstrap via simulation: bootstrap random samples

$$X_i^* = x_i^* \sim \hat{F}, Y_j^* = y_j^* \sim \hat{G}.$$

The bootstrap random variable is

$$R^* = R((\mathbf{X}^*, \mathbf{Y}^*), (\hat{F}, \hat{G})).$$

Repeated independent generation yields a sequence of independent realizations of R^*, say $R^{*(1)}, R^{*(2)}, \cdots, R^{*(N)}$. Approximate the expectation/variance of R by sample mean/variance of $R^{*(t)}$.

Error rate estimation in discriminant analysis

TABLE 2*

Random Variable	$m = n = 10$ Mean	(S.E.)	S.D.	$m = n = 20$ Mean	(S.E.)	S.D.	Remarks
Error Rate Diff. (4.4) R	**.062**	(.003)	**.143**	**.028**	(.002)	**.103**	Based on 1000 trials
Bootstrap Expectation $E_* R^*$	**.057**	(.002)	**.026** [.023]	**.029**	(.001)	**.015** [.011]	Based on 100 trials; $N = 100$ Bootstrap Replications per trial. (Figure in brackets is S.D. if $N = \infty$.)
Bootstrap Standard Deviation $SD_*(R^*)$	**.131**	(.0013)	**.016**	**.097**	(.002)	**.010**	
Cross–Validation Diff. $\tilde{R}$	**.054**	(.009)	**.078**	**.032**	(.002)	**.043**	Based on 40 trials

*The error rate difference (4.4) for linear discriminant analysis, investigated for bivariate normal samples (4.8). Sample sizes are $m = n = 10$ and $m = n = 20$. The values for the bootstrap method were obtained by Method 2, $N = 100$ bootstrap replications per trial. The bootstrap method gives useful estimates of both the mean and standard deviation of R. The cross-validation method was nearly unbiased for the expectation of R, but had about three times as large a standard deviation. All of the quantities in this table were estimated by repeated Monte Carlo trials; standard errors are given for the means.

Relationship with the jackknife

The Jackknife is shown to be a linear approximation method for the bootstrap. To be precise, approximating the bootstrap distribution by Taylor series expansion is the same as Jaeckel's infinitesimal jackknife.

Relationship with the jackknife

The Jackknife is shown to be a linear approximation method for the bootstrap. To be precise, approximating the bootstrap distribution by Taylor series expansion is the same as Jaeckel's infinitesimal jackknife.

- Define $P_i^* = \#\{i : X_i^* = x_i\}$, and
$$\mathbf{P}^* = (P_1^*, \cdots, P_n^*).$$
- Write $R(\mathbf{P}^*) = R(\mathbf{X}^*, \hat{F})$.
- We have
$$R(\mathbf{P}^*) \approx R(\mathbf{e}/n) + (\mathbf{P}^* - \mathbf{e}/n)\mathbf{U} + \frac{1}{2}(\mathbf{P}^* - \mathbf{e}/n)\mathbf{V}(\mathbf{P}^* - \mathbf{e}/n)'.$$

Relationship with the jackknife

Here, we used homogeneous extension $R(\mathbf{P}^*) = R(\mathbf{P}^*/\sum_{i=1}^n P_i^*)$.

Relationship with the jackknife

Here, we used homogeneous extension $R(\mathbf{P}^*) = R(\mathbf{P}^*/\sum_{i=1}^n P_i^*)$. We have

- $\mathbb{E}_*\mathbf{P}^* = \mathbf{e}/n$, $\text{Cov}_*\mathbf{P}^* = \mathbf{I}/n^2 - \mathbf{e}'\mathbf{e}/n^3$.
- $\mathbf{eU} = 0, \mathbf{eV} = -n\mathbf{U}', \mathbf{eVe}' = 0$.

Relationship with the jackknife

Here, we used homogeneous extension $R(\mathbf{P}^*) = R(\mathbf{P}^*/\sum_{i=1}^n P_i^*)$. We have

- $\mathbb{E}_*\mathbf{P}^* = \mathbf{e}/n$, $\text{Cov}_*\mathbf{P}^* = \mathbf{I}/n^2 - \mathbf{e}'\mathbf{e}/n^3$.
- $\mathbf{eU} = 0, \mathbf{eV} = -n\mathbf{U}', \mathbf{eVe}' = 0$.

This gives

$$\mathbb{E}_*R(\mathbf{P}^*) \approx R(\mathbf{e}/n) + \frac{1}{2n^2}\text{tr}(\mathbf{V}),$$

$$\text{Var}_*R(\mathbf{P}^*) = \sum_{i=1}^n U_i^2/n^2.$$

Relationship with the jackknife

Here, we used homogeneous extension $R(\mathbf{P}^*) = R(\mathbf{P}^*/\sum_{i=1}^n P_i^*)$. We have

- $\mathbb{E}_*\mathbf{P}^* = \mathbf{e}/n$, $\text{Cov}_*\mathbf{P}^* = \mathbf{I}/n^2 - \mathbf{e}'\mathbf{e}/n^3$.
- $\mathbf{eU} = 0, \mathbf{eV} = -n\mathbf{U}', \mathbf{eVe}' = 0$.

This gives

$$\mathbb{E}_*R(\mathbf{P}^*) \approx R(\mathbf{e}/n) + \frac{1}{2n^2}\mathrm{tr}(\mathbf{V}),$$
$$\mathrm{Var}_*R(\mathbf{P}^*) = \sum_{i=1}^n U_i^2/n^2.$$

In this case, consider $R(\mathbf{X}^*, \hat{F}) = \theta(\hat{F}^*) - \theta(\hat{F})$. $(R(\mathbf{e}/n) = 0)$
$\mathbb{E}_*[\theta(\hat{F}^*) - \theta(\hat{F})] \approx \frac{1}{2n^2}\mathrm{tr}(\mathbf{V})$ suggests $\mathbb{E}_F(\theta(\hat{F}) - \theta(F)) \approx \frac{1}{2n^2}\mathrm{tr}(\mathbf{V})$.
Similarly, $\mathrm{Var}_F(\theta(\hat{F})) \approx \sum_{i=1}^n U_i^2/n^2$.

Relationship with the jackknife

The approximations

$$\mathbb{E}_F(\theta(\hat{F}) - \theta(F)) \approx \frac{1}{2n^2}\mathrm{tr}(\mathbf{V})$$
$$\mathrm{Var}_F(\theta(\hat{F})) \approx \sum_{i=1}^{n} U_i^2/n^2$$

exactly agree with those given by Jaeckel's infinitesimal jackknife.

Relationship with the jackknife

The approximations

$$\mathbb{E}_F(\theta(\hat{F}) - \theta(F)) \approx \frac{1}{2n^2}\mathrm{tr}(\mathbf{V})$$
$$\mathrm{Var}_F(\theta(\hat{F})) \approx \sum_{i=1}^{n} U_i^2/n^2$$

exactly agree with those given by Jaeckel's infinitesimal jackknife.

The ordinary jackknife replaces the derivatives with finite differences.

Jackknife in 'unbalanced' situation

Question: In the two-sample problem, should we leave out one x_i at a time, then one y_j at a time, or should we leave out all mn pairs (x_i, y_j) at a time?

Jackknife in ‘unbalanced’ situation

Question: In the two-sample problem, should we leave out one x_i at a time, then one y_j at a time, or should we leave out all mn pairs (x_i, y_j) at a time?

In this paper, Brad answered this question by using Taylor series expansion on the bootstrap distribution. He showed that in two-sample situation, we should leave out one x_i at a time, then one y_j at a time.

Residual Bootstrap

Consider a general regression model $X_i = g_i(\beta) + \epsilon_i$.

Residual Bootstrap

Consider a general regression model $X_i = g_i(\beta) + \epsilon_i$.

- In the linear model case, where $g_i(\beta) = c_i'\beta$ (with c_i being a vector of covariates). Estimation $\hat{\beta}$ can be obtained by least squares.

Residual Bootstrap

Consider a general regression model $X_i = g_i(\beta) + \epsilon_i$.

- In the linear model case, where $g_i(\beta) = c_i'\beta$ (with c_i being a vector of covariates). Estimation $\hat{\beta}$ can be obtained by least squares.
- By resampling residuals ϵ_i^* from F, new bootstrap samples $X_i^* = g_i(\hat{\beta}) + \epsilon_i^*$ can be generated, leading to bootstrap estimates $\hat{\beta}^*$.

Residual Bootstrap

Consider a general regression model $X_i = g_i(\beta) + \epsilon_i$.

- In the linear model case, where $g_i(\beta) = c_i'\beta$ (with c_i being a vector of covariates). Estimation $\hat{\beta}$ can be obtained by least squares.
- By resampling residuals ϵ_i^* from F, new bootstrap samples $X_i^* = g_i(\hat{\beta}) + \epsilon_i^*$ can be generated, leading to bootstrap estimates $\hat{\beta}^*$.
- Bootstrap replications $\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_N^*$ allow estimation of the bootstrap distribution of the estimator, providing insight into its variability and bias.

Residual Bootstrap

For the linear model,

Residual Bootstrap

For the linear model,

- if C is the design matrix and $G = C'C$ (assumed nonsingular), the least squares estimator $\hat{\beta} = G^{-1}C'X$ has mean β and covariance $\sigma_\epsilon^2 G^{-1}$

Residual Bootstrap

For the linear model,

- if C is the design matrix and $G = C'C$ (assumed nonsingular), the least squares estimator $\hat{\beta} = G^{-1}C'X$ has mean β and covariance $\sigma_\epsilon^2 G^{-1}$
- The bootstrap values ϵ_i^* are independent with mean zero and variance $\hat{\sigma}^2 = \sum_{i=1}^n (x_i - g_i(\hat{\beta}))^2/n$. This implies that $\hat{\beta}^* = G^{-1}C'XX^*$ has bootstrap mean and variance

$$\mathbb{E}_*\hat{\beta}^* = \hat{\beta}, \quad \mathrm{Cov}_*\hat{\beta}^* = \hat{\sigma}^2 G^{-1}.$$

Residual Bootstrap

For the linear model,

- if C is the design matrix and $G = C'C$ (assumed nonsingular), the least squares estimator $\hat{\beta} = G^{-1}C'X$ has mean β and covariance $\sigma_\epsilon^2 G^{-1}$
- The bootstrap values ϵ_i^* are independent with mean zero and variance $\hat{\sigma}^2 = \sum_{i=1}^n (x_i - g_i(\hat{\beta}))^2/n$. This implies that $\hat{\beta}^* = G^{-1}C'XX^*$ has bootstrap mean and variance

$$\mathbb{E}_*\hat{\beta}^* = \hat{\beta}, \quad \text{Cov}_*\hat{\beta}^* = \hat{\sigma}^2 G^{-1}.$$

- This approach aligns with traditional theory, offering a practical tool for assessing estimator performance in finite samples.

Residual Bootstrap

For the linear model,

- if C is the design matrix and $G = C'C$ (assumed nonsingular), the least squares estimator $\hat{\beta} = G^{-1}C'X$ has mean β and covariance $\sigma_\epsilon^2 G^{-1}$
- The bootstrap values ϵ_i^* are independent with mean zero and variance $\hat{\sigma}^2 = \sum_{i=1}^n (x_i - g_i(\hat{\beta}))^2/n$. This implies that $\hat{\beta}^* = G^{-1}C'XX^*$ has bootstrap mean and variance

$$\mathbb{E}_* \hat{\beta}^* = \hat{\beta}, \quad \text{Cov}_* \hat{\beta}^* = \hat{\sigma}^2 G^{-1}.$$

- This approach aligns with traditional theory, offering a practical tool for assessing estimator performance in finite samples.
- On the contrary, covariance obtained by the jackknife methods looks very different. $\text{Cov}\hat{\beta} \approx G^{-1}(\sum_{i=1}^n c_i' c_i \hat{\epsilon}_i^2) G^{-1}$.

Theory behind Bootstrap

Assume the sample sapce is a finite set. Any distribution on F can be represented as a vector $\mathbf{P}$. Then the empirical distribution and the bootstrap distribution satisfies

$$\hat{\mathbf{P}}|\mathbf{P} \sim \text{Multinomial}(n, \mathbf{P}), \quad \hat{\mathbf{P}}^*|\hat{\mathbf{P}} \sim \text{Multinomial}(n, \hat{\mathbf{P}}).$$

Theory behind Bootstrap

Assume the sample sapce is a finite set. Any distribution on F can be represented as a vector $\mathbf{P}$. Then the empirical distribution and the bootstrap distribution satisfies

$$\hat{\mathbf{P}}|\mathbf{P} \sim \text{Multinomial}(n, \mathbf{P}), \quad \hat{\mathbf{P}}^*|\hat{\mathbf{P}} \sim \text{Multinomial}(n, \hat{\mathbf{P}}).$$

By asymptotics,

$$\sqrt{n}(\hat{\mathbf{P}} - \mathbf{P})|\mathbf{P} \approx \mathcal{N}(0, \Sigma_{\mathbf{P}}), \quad \sqrt{n}(\hat{\mathbf{P}}^* - \hat{\mathbf{P}})|\hat{\mathbf{P}} \approx \mathcal{N}(0, \Sigma_{\hat{\mathbf{P}}}), \mathbf{P} \approx \hat{\mathbf{P}}.$$

Theory behind Bootstrap

Assume the sample sapce is a finite set. Any distribution on F can be represented as a vector $\mathbf{P}$. Then the empirical distribution and the bootstrap distribution satisfies

$$\hat{\mathbf{P}}|\mathbf{P} \sim \text{Multinomial}(n, \mathbf{P}), \quad \hat{\mathbf{P}}^*|\hat{\mathbf{P}} \sim \text{Multinomial}(n, \hat{\mathbf{P}}).$$

By asymptotics,

$$\sqrt{n}(\hat{\mathbf{P}} - \mathbf{P})|\mathbf{P} \approx \mathcal{N}(0, \Sigma_{\mathbf{P}}), \quad \sqrt{n}(\hat{\mathbf{P}}^* - \hat{\mathbf{P}})|\hat{\mathbf{P}} \approx \mathcal{N}(0, \Sigma_{\hat{\mathbf{P}}}), \mathbf{P} \approx \hat{\mathbf{P}}.$$

So

$$\sqrt{n}(\hat{\mathbf{P}}^* - \hat{\mathbf{P}})|\hat{\mathbf{P}} \approx \sqrt{n}(\hat{\mathbf{P}} - \mathbf{P})|\mathbf{P}.$$

Conclusions

- Bootstrap is a simple but powerful tool in estimating bias and variance of the statistics of interest.
- The Jackknife methods can be viewed as first order approximation of Bootstrap.
- Statisticians need to feel comfortable with simulations.
- Bootstrap has earned its place as one of the most influential developments in statistical methodology.
- Its versatility and robustness have made it indispensable in a wide range of fields, including but not limited to finance, biology, engineering, and social sciences. Over 70k citations (20+50). In 2018, Brad was awarded the "International Prize in Statistics" in recognition of the bootstrap.

Prehistory of bootstrap

John Hubback, an Indian Civil Servant, introduced a version of the block bootstrap for spatial data. (1927)

P.C. Mahalanobism, the eminent Indian statistician, was inspired by Hubback's work and used Hubback's spatial sampling schemes explicitly for variance estimation. (1930s)

Prehistory of bootstrap

Julian Simon had published a number of resampling examples, including a bootstrap example, in his 1969 book Basic Research Methods in Social Science.

Prehistory of bootstrap

Julian Simon had published a number of resampling examples, including a bootstrap example, in his 1969 book Basic Research Methods in Social Science.

Prehistory of bootstrap

Julian Simon had published a number of resampling examples, including a bootstrap example, in his 1969 book Basic Research Methods in Social Science.

"Recently I have concluded that a bootstrap-type test has better theoretical justification than a permutation test in this case, although the two reach almost identical results with a sample this large" (Simon 1993)

Contribution of Brad Efron

- Efron's contributions were of course far-reaching. They vaulted forward from earlier ideas, of people such as Hubback, Mahalanobis, Hartigan and Simon, creating a fully fledged methodology that is now applied to analyse data on virtually all human beings.

Contribution of Brad Efron

- Efron's contributions were of course far-reaching. They vaulted forward from earlier ideas, of people such as Hubback, Mahalanobis, Hartigan and Simon, creating a fully fledged methodology that is now applied to analyse data on virtually all human beings.
- Efron combined the power of Monte Carlo approximation with an exceptionally broad view of the sort problem that bootstrap methods might solve.

Limitations

- Bootstrap method can be computationally intensive because we need to calculate statistics of interest for all bootstrap random samples.
- It is an approximate method. For small size of samples, the result may not be reliable.
- When the original dataset is small or contains outliers. The accuracy of the bootstrap estimates depends on the adequacy of the original dataset for representing the population.
- The bootstrap method assumes that observations in the original dataset are independent and identically distributed (IID). If this assumption is violated, such as in the case of time-series data or spatial data with autocorrelation, the bootstrap estimates may be biased or unreliable.

Further Developments of the Bootstrap Method

- Bayesian bootstrap
- The parametric bootstrap
- Bootstrap confidence intervals
- Bootstrap for time series
- ......

Origin of the Name

Jackknife: ‘it could work on anything.’

Origin of the Name

Jackknife: ‘it could work on anything.’

Bootstrap: ‘Pull yourself up by your bootstraps.’

Other proposed names

Swiss Army Knife, Meat Axe, Swan-Dive, Jack-Rabbi, and **Shotgun**

Other proposed names

Swiss Army Knife, Meat Axe, Swan-Dive, Jack-Rabbi, and **Shotgun**

Other proposed names

Swiss Army Knife, Meat Axe, Swan-Dive, Jack-Rabbi, and **Shotgun**

"It can blow the head off any problem if the statistician can stand the resulting mess."