Efron (1979) – Bootstrap Methods

Spencer Woody

UT-Austin, SDS 190

spencer.woody@utexas.edu

November 4, 2016

Overview

Problem Statement

We want to estimate the sampling distribution of a random variable $R(X, F)$ given a sample size n

$$X = \{X_1, X_2, \dots X_n\}$$
$$X_i \overset{\text{iid}}{\sim} F, \; i = 1, 2, \dots, n$$
$$X_i = x_i \text{ observed data}$$

Jackknife

$R(X, F) = t(X) - \theta(F)$, $t(X)$ is some estimator of $\theta(F)$. some parameter of F which is of interest.
We obtain $\hat{\text{Bias}}(t)$, $\hat{\text{Var}}(t)$ by recomputing $t(\cdot)$ n times, *each time leaving one observation from* X *out*.

$$R'(X, F) = \frac{t(X) - \hat{\text{Bias}}(t) - \theta(F)}{\left(\hat{\text{Var}}(t)\right)^{1/2}} \sim t_{n-1}$$

Bootstrap

(1) Construct $\hat{F}$, discrete uniform with mass $\frac{1}{n}$ on each x_i

(2) Draw the *bootstrap sample* of size n from $\hat{F}$. This is a *resampling* scheme.

$$X^* = \{X_1^*, X_2^*, \dots X_n^*\}$$
$$X_i^* \overset{\text{iid}}{\sim} \hat{F},\ i = 1, 2, \dots, n$$
$$X_i^* = x_i^* \text{ realization of bootstrap sample}$$

(3) Approximate $R(X, F)$ with $R^* = R(X^*, \hat{F})$.

Implementation of Bootstrap

I. Direct theoretical calculation

Use theoretical framework without creating simulations

II. Monte Carlo approximation with N simulations

1. Generate $X^{*1}, \dots, X^{*N}$
2. Calculate $R^{*1} = R(X^{*1}, \hat{F}), \dots, R^{*N} = R(X^{*N}, \hat{F})$
3. Create histogram of $R^{*1}, \dots, R^{*N}$ to make inferences on $R(X, F)$

Trivial theoretical calculation: Bernoulli

- $F \sim \text{Bernoulli}(\theta)$
- $X = \{X_1, X_2, \ldots X_n\}$
- $X_i \overset{\text{iid}}{\sim} F$
- $R(X, F) = \bar{X} - \theta$ (sample average – pop. parameter)

Generate X^* and construct $\hat{F}$ as described previously

- $X_i^* \overset{\text{iid}}{\sim} \text{Bernoulli}(\bar{x})$
- $R(X, F) \approx R^*(X^*, \hat{F}) = \bar{X}^* - \bar{x}$

$$\text{E}(R^*) = 0 \qquad \text{Var}(R^*) = \text{Var}(\bar{X}^*) = \frac{\bar{x}(1-\bar{x})}{n}$$

$\bar{X}$ is unbiased in θ with variance $\bar{x}(1-\bar{x})/n$

MSE of sample median

Again, $X = \{X_1, X_2, \dots X_n\}$, $X_i \overset{\text{iid}}{\sim} F$. Suppose n is odd, $n = 2m - 1$

- $\theta(F)$ is the population median
- $t(X) = X_{(m)}$ is the sample median
- $R(X, F) = t(X) - \theta(F)$

Generate bootstrap X^*

$$\begin{aligned} R^* = R(X^*, \hat{F}) &= t(X^*) - \theta(\hat{F}) \\ &= X^*_{(m)} - x_{(m)} \end{aligned}$$

MSE of sample median

Problem: Find expected MSE of $t(X)$. Note that $R^* = X^*_{(m)} - x_{(m)}$

$$\begin{aligned}
E_F((t(X) - \theta(F))^2) &= E((R(X,F))^2) \\
&\approx E_*((R_*)^2) \\
&= E_*(X_{(m)} - x_{(m)})^2 \\
&= \sum_{\ell=1}^{n} (x_{(\ell)} - x_{(m)})^2 P(R^* = x_{(\ell)} - x_{(m)}) \\
&= \sum_{\ell=1}^{n} (x_{(\ell)} - x_{(m)})^2 P(X^*_{(m)} = x_{(\ell)})
\end{aligned}$$

Find the distribution of random variable $X^*_{(m)}$

MSE of sample median

$$
\begin{aligned}
P(X^*_{(m)} > x_{(\ell)}) &= P(x_{(\ell)} < X^*_{(m)}) \\
&= P(\text{Binomial}(n, \ell/n) \leq m - 1) \\
&= \sum_{j=0}^{m-1} \binom{n}{j} \left(\frac{\ell}{n}\right)^j \left(\frac{n-\ell}{n}\right)^{n-j}
\end{aligned}
$$

MSE of sample median

$$
\begin{aligned}
P(X^*_{(m)} = x_{(\ell)}) &= P(X^*_{(m)} > x_{(\ell-1)}) - P(X^*_{(m)} > x_{(\ell)}) \\
&= \sum_{k=0}^{m-1} \binom{n}{k} \left(\frac{\ell-1}{n}\right)^k \left(\frac{n-\ell-1}{n}\right)^{n-k} \\
&- \sum_{j=0}^{m-1} \binom{n}{j} \left(\frac{\ell}{n}\right)^j \left(\frac{n-\ell}{n}\right)^{n-j}
\end{aligned}
$$

Now we can use this to estimate

$$
\begin{aligned}
E((R(X,F))^2) &= E_F((t(X) - \theta(F))^2) \\
&\approx \sum_{\ell=1}^{n} (x_{(\ell)} - x_{(m)})^2 P(X^*_{(m)} = x_{(\ell)})
\end{aligned}
$$

No Monte Carlo sampling needed!

Error Rate in Linear Discriminant Analysis

We have two unknown continuous PDFs F and G in $\mathbb{R}^k$ space

$$X_i = x_i, \quad X_i \overset{\text{iid}}{\sim} F, \quad i = 1, 2, \ldots, m$$
$$Y_j = y_j, \quad Y_j \overset{\text{iid}}{\sim} G, \quad j = 1, 2, \ldots, n.$$

Given our data X and Y we partition $\mathbb{R}^k$ into two complementary regions A and B so as to ascribe a future observation z to F distribution if $z \in A$, or to G if $z \in B$.

$$\text{error}_F = P_F(X \in B)$$
$$\widehat{\text{error}}_F = \frac{\#\{x_i \in B\}}{m}$$
$$R((X, Y), (F, G)) = \text{error}_F - \widehat{\text{error}}_F$$

Error Rate in Linear Discriminant Analysis

Given observed data x and y, the MLE of B is

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

where $S = \left[\sum_i (x_i - \bar{x})(x_i - \bar{x})^T + \sum_j (y_j - \bar{y})(y_j - \bar{y})^T \right] / (m + n)$

Error Rate in Linear Discriminant Analysis

Estimating error using bootstrap

Construct bootstrap random samples

$$
\begin{aligned}
X_i^* &= x_i^*, \quad X_i^* \overset{\text{iid}}{\sim} \hat{F}, \quad i = 1, 2, \ldots, m \\
Y_j^* &= y_j^*, \quad Y_j^* \overset{\text{iid}}{\sim} \hat{G}, \quad j = 1, 2, \ldots, n.
\end{aligned}
$$

This yields a region B^* which is defined the same as before, except $\bar{x}^*$, $\bar{y}^*$, and S^* replace $\bar{x}$, $\bar{y}$, and S. One Monte Carlo simulation will produce

$$
R^* = R((X^*, Y^*), (F^*, G^*)) = \frac{\#\{x_i \in B^*\}}{m} - \frac{\#\{x_i^* \in B^*\}}{m}.
$$

Estimating error using cross validation

Leave one observation of x out. Construct $\tilde{B}$ similar to before

$$\tilde{B} = \left\{ z : (\bar{y} - \bar{x}) S^{-1} \left(z - \frac{\bar{x} + \bar{y}}{2} \right) > \log \frac{m}{n} \right\}.$$

Assign removed value according to this boundary. Repeat this m times, $\widetilde{\text{error}}_F$ is the proportion of times we wrongly assign the removed value. Do we decrease m to $m-1$ in this equation after leaving one out? *NO.*

$$\tilde{R} = \widetilde{\text{error}}_F - \widehat{\text{error}}_F$$

Error Rate in Linear Discriminant Analysis

Comparison of two methods

$$F : X \sim N\left(\begin{pmatrix} -\frac{1}{2} \\ 0 \end{pmatrix}, I\right), \quad G : Y \sim N\left(\begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}, I\right)$$

	$m = n = 10$			$m = n = 20$			
Random Variable	Mean	(S.E.)	S.D.	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

Figure: Error rate differences

Relationship between Jackknife and Bootstrap

Shotgun method?

"I also wish to thank the many friends who suggested names more colorful than *Bootstrap*, including *Swiss Army Knife*, *Meat Axe*, *Swan-Dive*, *Jack-Rabbit*, and my personal favorite, the *Shotgun*, which, to paraphrase Tukey, 'can blow the head off any problem if the statistician can stand the resulting mess.'" – Bradley Efron, 1979.

Thank you!

Questions

- Email: `spencer.woody@utexas.edu`
- Presentation: `github.com/spencerwoody`

Paper

Efron, B. "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics*. 1979. Vol. 7, No. 1, 1-26.