

Bootstrap

Uma leitura de Efron e Tibshirani (1986)

Lucas Machado Moschen

Escola de Matemática Aplicada
Fundação Getulio Vargas

8 de dezembro de 2021

- ▶ Técnica desenvolvida com objetivo de medir a acurácia de estimadores;
- ▶ É mais utilizado na abordagem frequentista, mas pode ser interpretado na Bayesiana ([Rubin, 1981](#));
- ▶ A ideia geral é gerar novos pontos a partir dos iniciais e proceder a estimação em cada conjunto de dados, gerando uma distribuição para o estimador.

A construção de [Efron and Tibshirani \(1986\)](#) supõe que observamos $\mathbf{x} = (x_1, x_2, \dots, x_n)$ de

$$\mathbf{X} = (X_1, X_2, \dots, X_n), X_i \stackrel{iid}{\sim} F,$$

em que F é distribuição desconhecida. Podemos relaxar para dados dependentes. ([Lahiri, 1999](#)).

A construção de [Efron and Tibshirani \(1986\)](#) supõe que observamos $\mathbf{x} = (x_1, x_2, \dots, x_n)$ de

$$\mathbf{X} = (X_1, X_2, \dots, X_n), X_i \stackrel{iid}{\sim} F,$$

em que F é distribuição desconhecida. Podemos relaxar para dados dependentes. ([Lahiri, 1999](#)).

Sejam $R(\mathbf{X}, F)$ a variável de interesse como a estatística $\hat{\theta}(\mathbf{X})$, e \hat{F} a distribuição empírica de \mathbf{x} .

A construção de [Efron and Tibshirani \(1986\)](#) supõe que observamos $\mathbf{x} = (x_1, x_2, \dots, x_n)$ de

$$\mathbf{X} = (X_1, X_2, \dots, X_n), X_i \stackrel{iid}{\sim} F,$$

em que F é distribuição desconhecida. Podemos relaxar para dados dependentes. ([Lahiri, 1999](#)).

Sejam $R(\mathbf{X}, F)$ a variável de interesse como a estatística $\hat{\theta}(\mathbf{X})$, e \hat{F} a distribuição empírica de \mathbf{x} .

Vamos utilizar estimadores de Monte Carlo para aproximar aspectos de da distribuição de R , como $\mathbb{E}_{\hat{F}}[R]$.

Um exemplo usual é o erro padrão

$$\sigma_F = \left(\text{Var}_F(\hat{\theta}(\mathbf{X})) \right)^{1/2}$$

de uma estatística $\hat{\theta}(\mathbf{X})$ que pode ser aproximada por

$$\hat{\sigma}_{\hat{F}} = \left(\text{Var}(\hat{\theta}_{\hat{F}}(\mathbf{x})) \right)^{1/2}$$

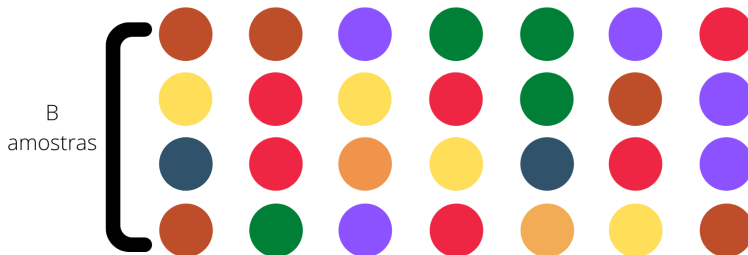
que é o desvio padrão com $F = \hat{F}$.

Re-amostragem

Amostra original



Amostras bootstrap



Estimador Monte Carlo

Para cada amostra, avaliamos $R^{(i)} = R(\mathbf{x}^{(i)}, \hat{F})$ e obtemos

$$\{R^{(1)}, \dots, R^{(B)}\}.$$

Assim, teremos que

$$\frac{1}{B} \sum_{i=1}^B R^{(i)} \xrightarrow{P} \mathbb{E}_{\hat{F}}[R],$$

pela Lei Forte dos Grandes Números.

Estimador Monte Carlo

Para cada amostra, avaliamos $R^{(i)} = R(\mathbf{x}^{(i)}, \hat{F})$ e obtemos

$$\{R^{(1)}, \dots, R^{(B)}\}.$$

Assim, teremos que

$$\frac{1}{B} \sum_{i=1}^B R^{(i)} \xrightarrow{P} \mathbb{E}_{\hat{F}}[R],$$

pela Lei Forte dos Grandes Números.

No caso do erro padrão,

$$\hat{\sigma}_B = \left(\frac{1}{B-1} \sum_{i=1}^B (\hat{\theta}(\mathbf{x}^{(i)}) - \hat{\theta}^B)^2 \right)^{1/2} \rightarrow \hat{\sigma}_{\hat{F}}.$$

Por que tamanho n ?

Temos que $R(\mathbf{X}, F)$ depende de n através de \mathbf{X} . A distribuição do estimador muda com n . Isso é fácil de ver quando $\hat{\theta}(\mathbf{x})$ é consistente. A demonstração clássica de Monte Carlo supõe que as variáveis são identicamente distribuídas.

Por que tamanho n ?

Temos que $R(\mathbf{X}, F)$ depende de n através de \mathbf{X} . A distribuição do estimador muda com n . Isso é fácil de ver quando $\hat{\theta}(\mathbf{x})$ é consistente. A demonstração clássica de Monte Carlo supõe que as variáveis são identicamente distribuídas.

([Bickel and Freedman, 1981](#)) apresenta o seguinte teorema:

THEOREM 2.1. *Suppose X_1, X_2, \dots are independent, identically distributed, and have finite positive variance σ^2 . Along almost all sample sequences X_1, X_2, \dots , given (X_1, \dots, X_n) , as n and m tend to ∞ :*

- (a) *The conditional distribution of $\sqrt{m}(\mu_m^* - \mu_n)$ converges weakly to $N(0, \sigma^2)$.*
- (b) *$s_m^* \rightarrow \sigma$ in conditional probability: that is, for ϵ positive,*

$$P\{|s_m^* - \sigma| > \epsilon | X_1, \dots, X_n\} \rightarrow 0 \text{ a.s.}$$

Se **acreditamos** que F vem de uma família, podemos usar seu Estimador de Máxima Verossimilhança para realizar o processo de re-amostragem.

- ▶ Método anterior ao bootstrap para estimar viés e variância de estimadores;
- ▶ A ideia básica é deixar uma amostra de fora para cada re-amostragem e fazer o processo de estimar em cada amostra:

$$\mathbf{x}_{(i)} = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n).$$

- ▶ O estimador de Jackknife para σ_F é

$$\hat{\sigma}_J = \sqrt{\frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}(\mathbf{x}_{(i)}) - \hat{\theta}^J)^2}.$$

Referências I

- Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The annals of statistics*, pages 1196–1217.
- Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, pages 54–75.
- Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Annals of Statistics*, pages 386–404.
- Rubin, D. B. (1981). The bayesian bootstrap. *The annals of statistics*, pages 130–134.