

# Testes de aleatorização

Fundamentos e aplicações

Prof. Walmes Zeviani  
walmes@ufpr.br

Laboratório de Estatística e Geoinformação  
Departamento de Estatística  
Universidade Federal do Paraná

Atualizado em 2018-09-17

## Justificativas

- ▶ Métodos computacionalmente intensivos para inferência estatística são usados quando as abordagens tradicionais não são adequadas.
  - ▶ Resultados assintóticos em pequenas amostras.
  - ▶ Violação de pressupostos.
  - ▶ Não existência de mecanismos de inferência específicos.
- ▶ Tais métodos se baseiam em reamostragem e/ou simulação.
- ▶ Podem ser aplicados em muitos contextos.

## Objetivos

- ▶ TODO

# Bootstrap: visão geral

- ▶ Bootstrap foi apresentado de forma sistematizada por Efron 1979.
- ▶ O termo bootstrap foi usado por Efron 1979 com o mesmo espírito que Tukey 1958 usou Jackknife.
- ▶ O método já havia sido usado em circunstâncias anteriores.
- ▶ Bootstrap é um método de reamostragem que pode usado para avaliar propriedades de estimadores e fazer inferência.
- ▶ Bootstrap é um método de Monte Carlo pois usa a distribuição empírica dos dados como se fosse a verdadeira distribuição.
- ▶ Principais aplicações de bootstrap:
  - ▶ Avaliar propriedades da distribuição de estimadores para seleção, ajuste de vício, etc.
  - ▶ Substituir ou aprimorar a adequação de abordagens assintóticas em amostras pequenas: intervalos de confiança, testes de hipótese.

# Funcionamento

- ▶ Considere uma amostra de observações iid  $x_i, i = 1, \dots, n$ .
- ▶ Usando a distribuição empírica, cada valor  $x_i$  tem igual probabilidade de  $1/n$  de ocorrer.
- ▶ Considere que  $\theta$  seja um parâmetro de interesse que dispõe de um estimador  $\hat{\theta} = f(X_1, \dots, X_n)$ .
- ▶ Uma **amostra bootstrap** é um conjunto de valores extraídos ao acaso **com reposição** da amostra original.
- ▶ A estimativa de  $\theta$  na  $b$ -ésima reamostra bootstrap é  $\hat{\theta}_b^*$ .
- ▶ A estimativa pontual bootstrap é o valor médio

$$\hat{\theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* \quad (1)$$

# Intervalos de confiança

- ▶ O intervalo de confiança padrão bootstrap é calculado por

$$\text{estimativa bootstrap} \pm \text{quantil}_{\alpha/2} \cdot \text{erro padrão bootstrap} \quad (2)$$

$$\hat{\theta}^* \pm z_{\alpha/2} \left( \sum_{b=1}^B \frac{(\hat{\theta}_b^* - \hat{\theta}^*)^2}{B-1} \right)^{1/2}. \quad (3)$$

- ▶ Assume-se que

1.  $\hat{\theta}$  tem distribuição aproximadamente normal;
2.  $\hat{\theta}$  é um estimador não viciado.

- ▶ Este tipo de intervalo não requer um valor alto para  $B$ .
- ▶ O vício do estimador pode ser determinado pelo próprio procedimento.

# Intervalos de confiança

- ▶ O intervalo de confiança padrão bootstrap (IC-padrão) é válido e pode ser usado em situações em que inferência assintótica é difícil de aplicar.
- ▶ O IC-padrão é assintoticamente acurado tal como são os intervalos baseados na distribuição normal.
- ▶ Intervalos feitos usando quantis da distribuição  $t$  são mais acurados para estimadores em amostras pequenas.
- ▶ Muitas variações do IC-padrão foram desenvolvidas para produzir inferência de melhor qualidade.

# Intervalos de confiança

- ▶ O intervalo de confiança percentil bootstrap é determinado por

$$(\hat{\theta}_{\alpha/2}^*, \hat{\theta}_{1-\alpha/2}^*), \quad (4)$$

que correspondem aos percentis  $\alpha/2$  e  $1 - \alpha/2$ .

- ▶ Este intervalo
  1. não faz suposições sobre a distribuição de  $\hat{\theta}$ ;
  2. requer maior valor para  $B$  que o intervalo de confiança padrão.

# Ajuste de vício

- ▶ O bootstrap fornece uma abordagem intuitiva

$$E(\hat{\theta} - \theta) \approx E_B(\hat{\theta}_j^* - \hat{\theta}) = \hat{\theta}^* - \hat{\theta}. \quad (5)$$

- ▶ Ou seja, considere a distribuição empírica como sendo a distribuição verdadeira e determine o viés médio usando a média das amostras bootstrap.
- ▶ Atenção: ao estimar o vício deveria-se adicionar o erro amostral.

Conteúdo baseado em

[https://ocw.mit.edu/courses/sloan-school-of-management/15-450-analytics-of-finance-fall-2010/lecture-notes/MIT15\\_450F10\\_lec09.pdf](https://ocw.mit.edu/courses/sloan-school-of-management/15-450-analytics-of-finance-fall-2010/lecture-notes/MIT15_450F10_lec09.pdf).





## Próxima aula

- ▶ Intervalos de confiança com correção para o vício.
- ▶ Aplicações de bootstrap.
- ▶ Bootstrap paramétrico.

## Avisos

- ▶ Sabatina estará disponível a partir de Qua.

# Referências bibliográficas

- Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife". Em: *The Annals of Statistics* 7.1, pp. 1–26.
- Tukey, John W. (1958). "Bias and confidence in not quite large samples (abstract)". Em: *The Annals of Mathematical Statistics* 2.29, p. 614.