Inferência via abordagens computacionalmente intensivas

Walmes Zeviani

Introdução

A lógica dos testes de hipótese frequentistas:

1. Definir a **hipótese nula** e hipótese alternativa.
2. Determinar uma **estatística de teste** calculada a partir dos dados.
3. Estabelecer a **região crítica** para tomar decisão.

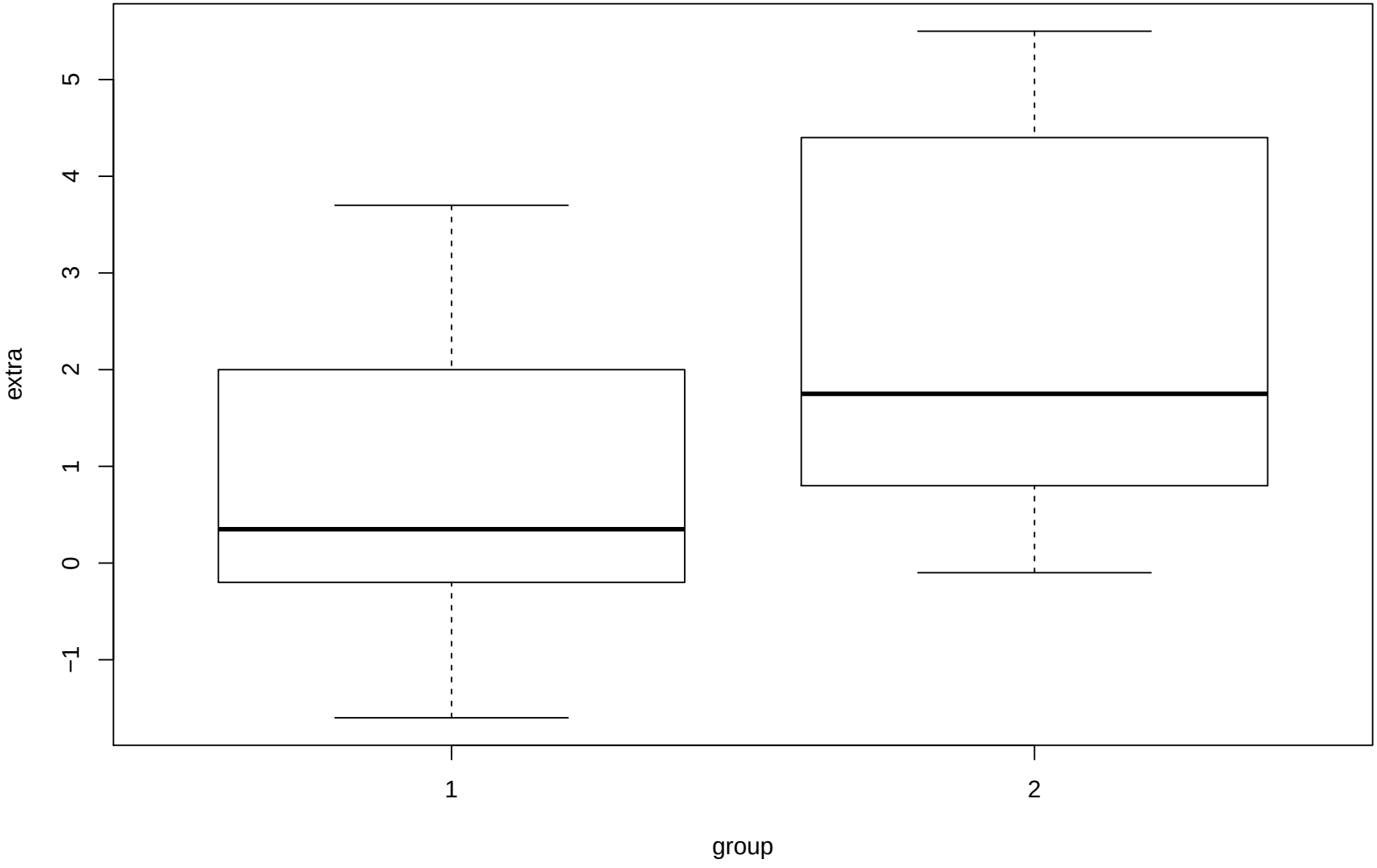
A região crítica é baseada na **distribuição amostral** da estatística de teste sob a hipótese nula.

Exemplo

```r
# Tabela.
unstack(sleep, form = extra ~ group)
```

```
##       X1   X2
## 1    0.7  1.9
## 2   -1.6  0.8
## 3   -0.2  1.1
## 4   -1.2  0.1
## 5   -0.1 -0.1
## 6    3.4  4.4
## 7    3.7  5.5
## 8    0.8  1.6
## 9    0.0  4.6
## 10   2.0  3.4
```

```r
# Gráfico.
plot(extra ~ group, data = sleep)
```



```r
# Teste de hipótese.
t.test(extra ~ group, data = sleep, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  extra by group
## t = -1.8608, df = 18, p-value = 0.07919
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.363874  0.203874
## sample estimates:
## mean in group 1 mean in group 2
##            0.75            2.33
```

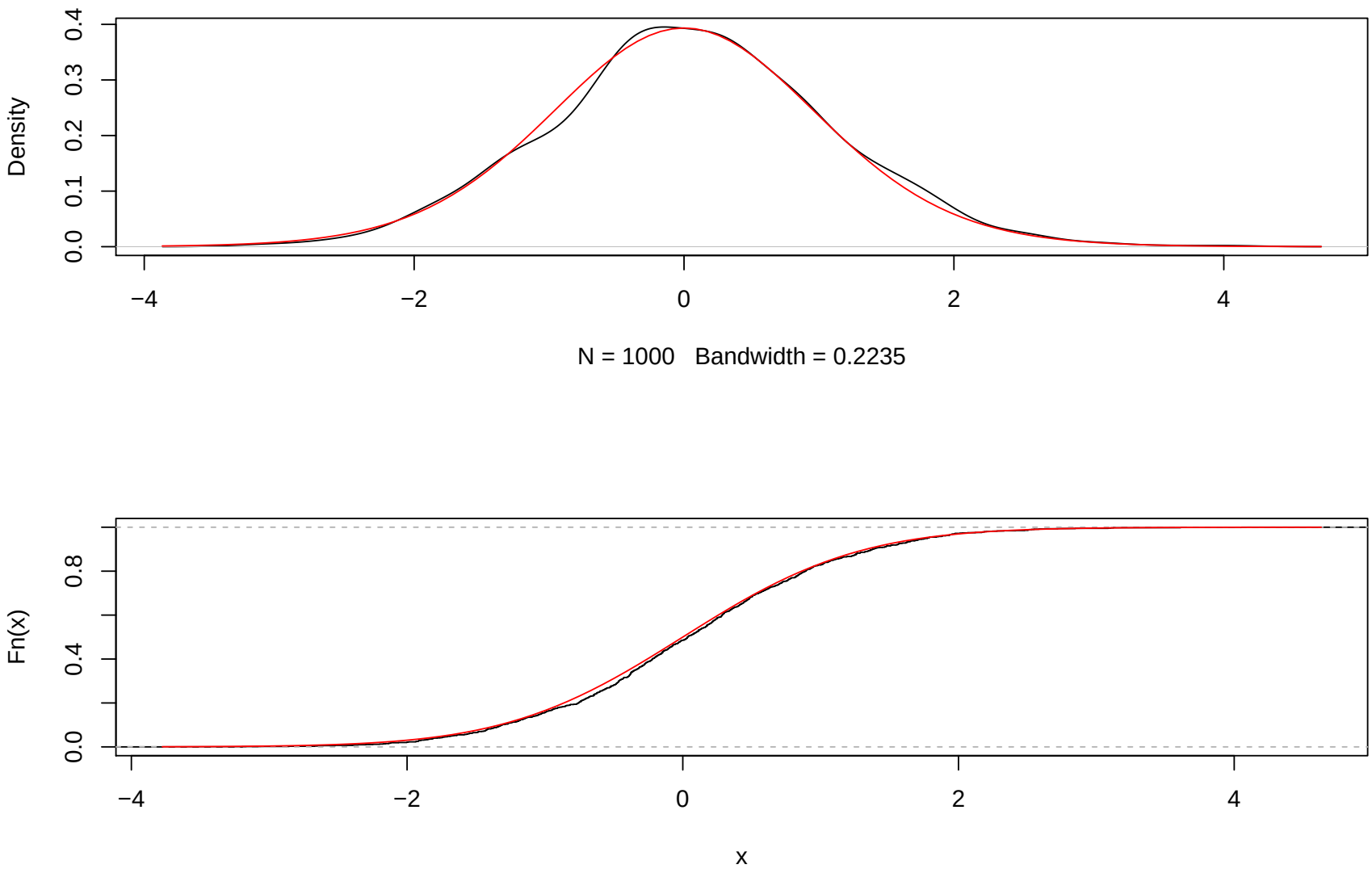
Sob a hipótese nula $H_0 : \delta = \mu_1 - \mu_2 = 0$, a estatística

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \delta}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t_{\text{Student}}(\nu = n_1 + n_2 - 2)$$

```r
# Simulação.
N <- 1000
n <- 10

t_val <- replicate(N, {
    # Amostras independentes da mesma população (H_0 verdadeira).
    x_1 <- rnorm(n, mean = 0, sd = 1)
    x_2 <- rnorm(n, mean = 0, sd = 1)
    # Diferença entre médias (H_0: delta == 0).
    d <- mean(x_1) - mean(x_2)
    # Variância combinada.
    s2 <- ((n - 1) * var(x_1) + (n - 1) * var(x_2))/(2 * n - 2)
    # Estatística do teste.
    t <- d/sqrt(s2 * (2/n))
    return(t)
})


# Distribuição empírica vs distribuição teórica.
par(mfrow = c(2, 1))
plot(density(t_val), main = NA)
curve(dt(x, df = 2 * n - 2), add = TRUE, col = 2)
plot(ecdf(t_val), main = NULL)
curve(pt(x, df = 2 * n - 2), add = TRUE, col = 2)
layout(1)
```


Density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
N = 1000 Bandwidth = 0.2235
Fn(x)
0.0
0.4
0.8
−4
−2
0
2
4
x

- Distribuição amostral é a distribuição de uma estatística (qualquer função da amostra) ao longo de todas as amostras de mesmo tamanho de uma população.
- Algumas estatísticas de teste tiveram a distribuição amostral determinada, e.g., t de Student, F de Snedecor, etc.
- Com a distribuição amostral pode-se fazer:
  - Testes de hipótese;
  - Intervalos de confiança;
  - Determinação de tamanho de amostra;
- A distribuição de uma estatística de teste pode ser exata ou aproximada.
- Com isso o teste pode ser exato ou aproximado.

Algumas situações

- Não possuem um teste de hipótese apropriado.
- As suposições para os testes não são atendidas.
- O teste tem aproximação ruim com a amostra pequena.

Abordagens consideradas

- Teste de aleatorização (permutação).
- Métodos de Jackknife.
- Métodos de Bootstrap.
- Métodos de Monte Carlo.

Testes de Aleatorização

- Abordagem baseada em permutação das observações.
- São considerados testes livre de distribuição.
- Faz suposições sobre o processo gerador dos dados.
- Cálculo da estatística de teste:
  - No conjunto de todos os arranjos possíveis (exaustivo): distribuição amostral exata.
  - Amostra do conjunto completo de arranjos (reamostragem sem reposição).
- Sob a hipótese nula os dados são **permutáveis**.

Uma senhora toma chá

- Aconteceu com Fisher e Muriel Bristol.
- Fisher descreve em seu livro em 1935.
- A senhora declarou saber discriminar bebida conforme a ordem em que chá e leite eram adicionados à xícara.
- H_0: a senhora não sabe distinguir (classifica aleatoriamente).
- Experimento: 8 xícaras, 4 de cada tipo servidas aleatoriamente.
- Resposta: a classificação de 4 xícaras de um tipo.

Perguntas

- Quantos arranjos possíveis?
- Qual a chance da senhora acertar todas por mero acaso?
- Qual a chance de acertar 3 em 4?
- Qual a região crítica?

Respostas

- $\binom{8}{4} = \frac{8!}{4!(8-4)!} = 70$.
- É $1/70$ pois só existe uma forma correta no universo das 70.
- "Arranjos de 3 corretos em 4 selecionados" $\times$ "arranjos de 1 errado em 4 selecionados": $\binom{4}{3} \cdot \binom{4}{1} = 16$, então $16/70 \approx 0.23$.
- Ao nível de 5%, a hipótese nula será rejeitada apenas se a senhora acertar as 4 xícaras pois $1/70 \approx 0.14 < 0.05$.

Jackknife

- Jackknife é uma espécie de canivete suiço.
- Equipado com várias ferramentas, fácil transporte.
- Mas ferramentas especializadas são melhores que as desse canivete.
- Proposto por Tukey.

A inspiração para a abordagem

A ideia é fundamentada no estimador da média

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

A média com a j-ésima observação removida, $\bar{X}_{-j}$, é

$$\bar{X}_{-j} = \frac{1}{n-1} \left[\left(\sum_{i=1}^{n} X_i \right) - X_j \right].$$

Combinando as expressões anteriores, pode-se determinar o valor de X_j por

$$X_j = n\bar{X} - (n-1)\bar{X}_{-j}.$$

Essa expressão não tem valor para o caso da média, que serviu apenas de inspiração. Mas tem utilidade para outras estatísticas.

O caso geral

Suponha que θ seja um parâmetro a ser estimado a partir de uma função dos dados (amostra de tamanho n)

$$\hat{\theta} = f(X_1, X_2, \ldots, X_n).$$

A quantidade

$$\theta_j^* = n\hat{\theta} - (n-1)\hat{\theta}_{-j}$$

é denominada de *pseudo-valor* e se baseia nas diferenças entre a estimativa com todas as observações ($\hat{\theta}$) e a *estimativa parcial*, ou seja, aquela sem a j-ésima observação ($\hat{\theta}_{-j}$).

O estimador pontual de Jackknife é definido por

$$\hat{\theta}^* = \frac{1}{n}\sum_{j=1}^{n} \theta_j^*,$$

ou seja, **é a média dos pseudo-valores**.

Os valores $\hat{\theta}$ e $\hat{\theta}^*$ não são iguais para o caso da média amostral mas não necessariamente iguais nos casos gerais.

Se for assumido que os valores θ_j^*, $j = 1, \ldots, n$, são independentes, a variância do estimador de Jackknife (inspirado pelo caso da média) é dados por

$$\text{Var}(\hat{\theta}^*) = \frac{S^2_{\theta^*}}{n}, \quad S^2_{\theta^*} = \frac{1}{n-1} \sum_{j=1}^{n} (\theta_j^* - \hat{\theta}^* * *^2.$$

Informação adicional

- Os pseudo valores são correlacionados em algum grau, com isso, a variância do estimador Jackknife pode ser viciada.
- É possível usar leave-two-outs, leave-three-outs, mas isso aumenta o custo.
- Validação cruzada tem relação com Jackknife.

Bootstrap

Principal objetivo*

Determinar as propriedades da distribuição do estimador de certo parâmetro, mas sem fazer suposições sobre a forma da distribuição dos dados.

A ideia

O conjunto de valores observados $(x_1, \ldots, x_n)$ é considerado uma realização de uma amostra aleatória $(X_1, \ldots, X_n)$ de uma distribuição desconhecida F.

Considere que existe interesse no parâmetro θ que pode ser estimado pela estatística $T(X_1, \ldots, X_n)$, ou seja, $\hat{\theta} = T(X_1, \ldots, X_n)$.

- Qual o vício do estimador $\hat{\theta}$?
- Qual a variância do estimador $\hat{\theta}$?
- Como obter um intervalo de confiança para θ?
- Como testar hipóteses sobre θ a partir da conhecida amostra?

Distribuição empírica

Distribuição empírica é a distribuição discreta em que cada ponto amostral tem o mesmo peso, ou seja, cada $x_i, i = 1, \ldots, n$, tem peso $1/n$. Essa distribuição de probabilidades é representada por $\hat{F}$ e é uma estimativa de F baseada na amostra observada.

Principio plug-in: substituir a F desconhecida por sua estimativa conhecida $\hat{F}$. N bootstrap, $\hat{F}$ é considerada como se fosse F.

Tomadas B observações independentes e identicamente distribuídas de $\hat{F}$ é o mesmo que reamostrar com reposição a amostra original.

Bootstrap não-paramétrico

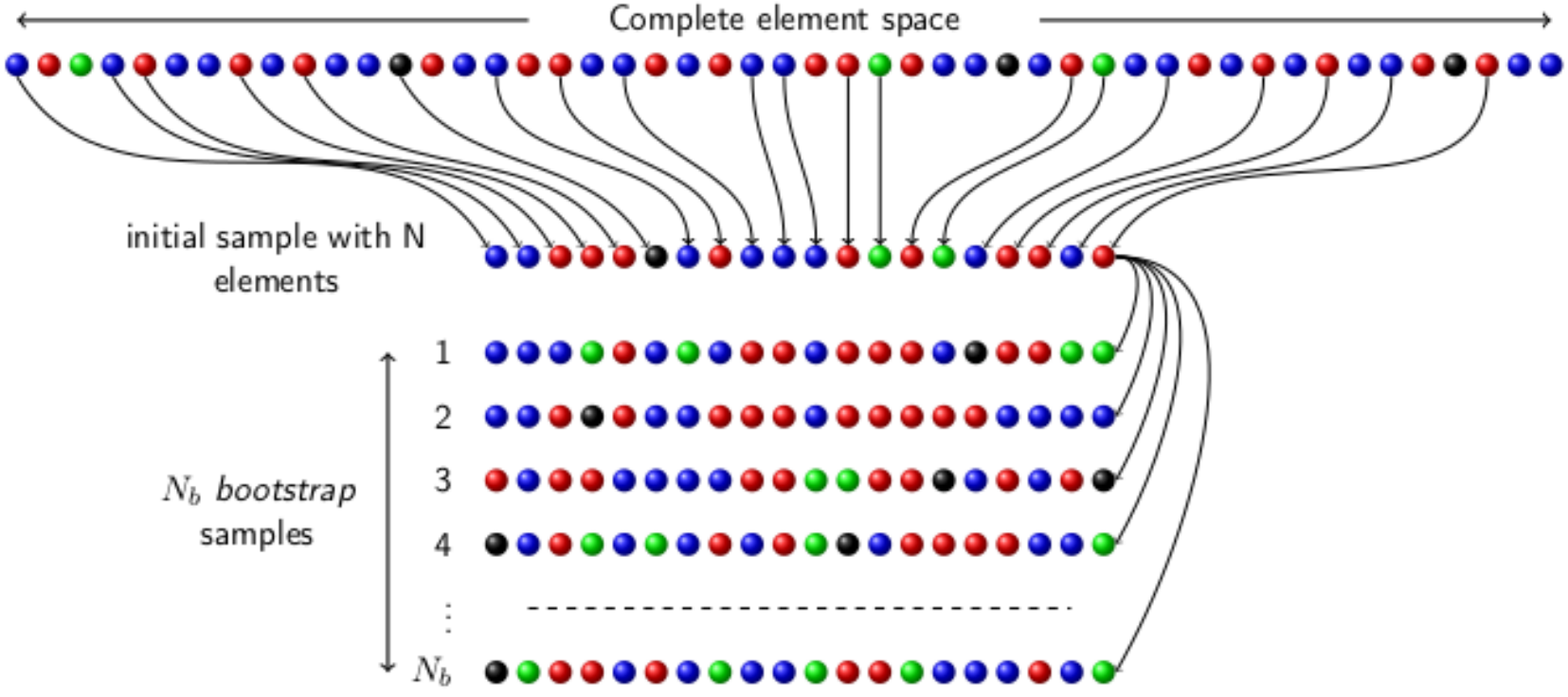
A amostra bootstrap é obtida através de reamostragem aleatória com reposição da amostra original.

Bootstrap paramétrico

A amostra bootstrap é obtida através de geração de números aleatórios da distribuição assumida para os dados. Os parâmetros da distribuição são estimados através da amostra original.

O algorítmo do bootstrap não paramétrico

1. Gere uma amostra com reposição da distribuição empírica dos dados (reamostragem com reposição).
2. Calcule $\hat{\theta} = T(x_1, \ldots, x_n)$ que é a estimativa bootstrap de θ.
3. Repita os passos 1 e 2 B vezes, onde B é suficientemente grande.
4. Resuma ou represente a distribuição formada pelos valores $\hat{\theta}_i, i = 1, \ldots, B$.


Complete element space
initial sample with N elements
N_b bootstrap samples
1
2
3
4
N_b

Mais detalhes

Existem muitos aspectos relacionados ao bootstrap que não serão abordados:

- Métodos para obtenção de intervalos de confiança.
- Correções para vício de estimadores.
- Inferência boostrap em amostras correlacionadas (séries temporais, dados espaciais).

Para mais detalhes visite
`http://conteudo.icmc.usp.br/CMS/Arquivos/arquivos_enviados/SECAO-POSGRAD_87_bootstrap-slides.pdf`.

Monte Carlo

A inferência por métodos Monte Carlos é baseada na geração de números aleatórios do modelo assumido para os dados. Esses métodos são utilizados para:

- Avaliar propriedades de um estimador pontual e/ou intervalar.
- Avaliar propriedades de um teste de hipóteses.
- Determinar tamanhos de amostra.
- Solucionar problemas otimização, integração, etc.

Os testes de aleatorização e boostrap são casos particulares de métodos Monte Carlo.

Nos métodos MC, deve-se **assumir uma distribuição de probabilidades** para algum componente aleatório do modelo, o que por vezes é considerada uma desvantagem da abordagem.

A partir da geração de amostras aleatórias do modelo **sob hipótese nula** são calculadas as estatísticas de interesse. A **distribuição amostral** das estatísticas é o ponto de partida para a inferência.