Técnicas de Bootstrap

- **Técnica introduzida por Efron como abordagem ao cálculo de intervalos de confiança de parâmetros, em circunstâncias em que outras técnicas não são aplicáveis, em particular no caso em que o número de amostras é reduzido**
- **esta técnica foi extrapolada para a resolução de muitos outros problemas de difícil resolução através de técnicas de análise estatística tradicionais (baseadas na hipótese de um elevado número de amostras)**

- **A técnica de bootstrap tenta realizar o que seria desejável realizar na prática, se tal fosse possível:**

  repetir a experiência

- **As observações são escolhidas de forma aleatória e as estimativas re-calculadas**

Técnicas de Bootstrap

- **Ideia básica:**
  - » **Uma vez que não se dispõe de toda a população de amostras (observações) faça-se o melhor com o que se dispõe que é o conjunto amostra**

$$X = \left(X_1, \cdots X_N\right)$$

- **A técnica de bootstrap trata a amostra observada como se esta representasse exactamente toda a população (conj. de experiências, realizações)**

- **Seja** $X = \left(X_1, \cdots X_N\right)$ **a amostra contendo N observações.**
- **Construir *B* amostras** $X^{*(1)}, \cdots, X^{*(B)}$ **iid de comprimento *N* cada**
- **Na terminologia de *bootstrapping* as *B* amostras iid construídas a partir da população finita** $\left(X_1, \cdots X_N\right)$ **corresponde a amostrar com substituição a partir do conjunto *X***

Exemplo

- **Seja a estimativa da média de uma v.a. *X***

$$\hat{\mu} = \frac{X_1 + \cdots + X_N}{N}$$

- **Intervalo de confiança:**

  » **Passo 0: Realização da experiência: conjunto amostra**

  $$X = (-2.41, 4.86, 6.06, 9.11, 10.2, 12.81, 13.17, 14.1, 15.77, 15.79)$$

  $$N = 10 \quad ; \quad \hat{\mu} = 9.946$$

  » **Passo 1: Usando um gerador de números aleatórios, selecionar aleatoriamente 10 amostras a partir de *X***

  $$\text{E}x : X^{*(1)} = (9.11, 9.11, 6.06, 13.17, 10.2, -2.41, 4.86, 12.81, -2.41, 4.86)$$

  » **Passo 2: Obter a estimativa de bootstrap**

  $$\hat{\mu}_1^* = 6.54$$

  » **Passo 3: Repetir os passos 1 e 2 um nº elevado de vezes para obter as estimativas**

  $$\hat{\mu}_1^*, \cdots, \hat{\mu}_B^* \qquad \text{E}x : B = 1000$$

  » **Passo 4: Aproximação da distribuição de $\hat{\mu}^*$**

  – **ordenar as estimativas por ordem crescente**

  $$\hat{\mu}_{(1)}^* \leq, \hat{\mu}_{(2)}^* \cdots, \hat{\mu}_{(B)}^*$$

  $\hat{\mu}_{(k)}^*$ é o *k*ésimo valor mais pequeno de $\hat{\mu}_1^*, \cdots, \hat{\mu}_B^*$

» **Passo 5: Intervalo de confiança**

– **O intervalo de confiança a (1-α)100% é dado por**

$$\left(\hat{\mu}^*_{(q_1)}, \hat{\mu}^*_{(q_2)}\right)$$

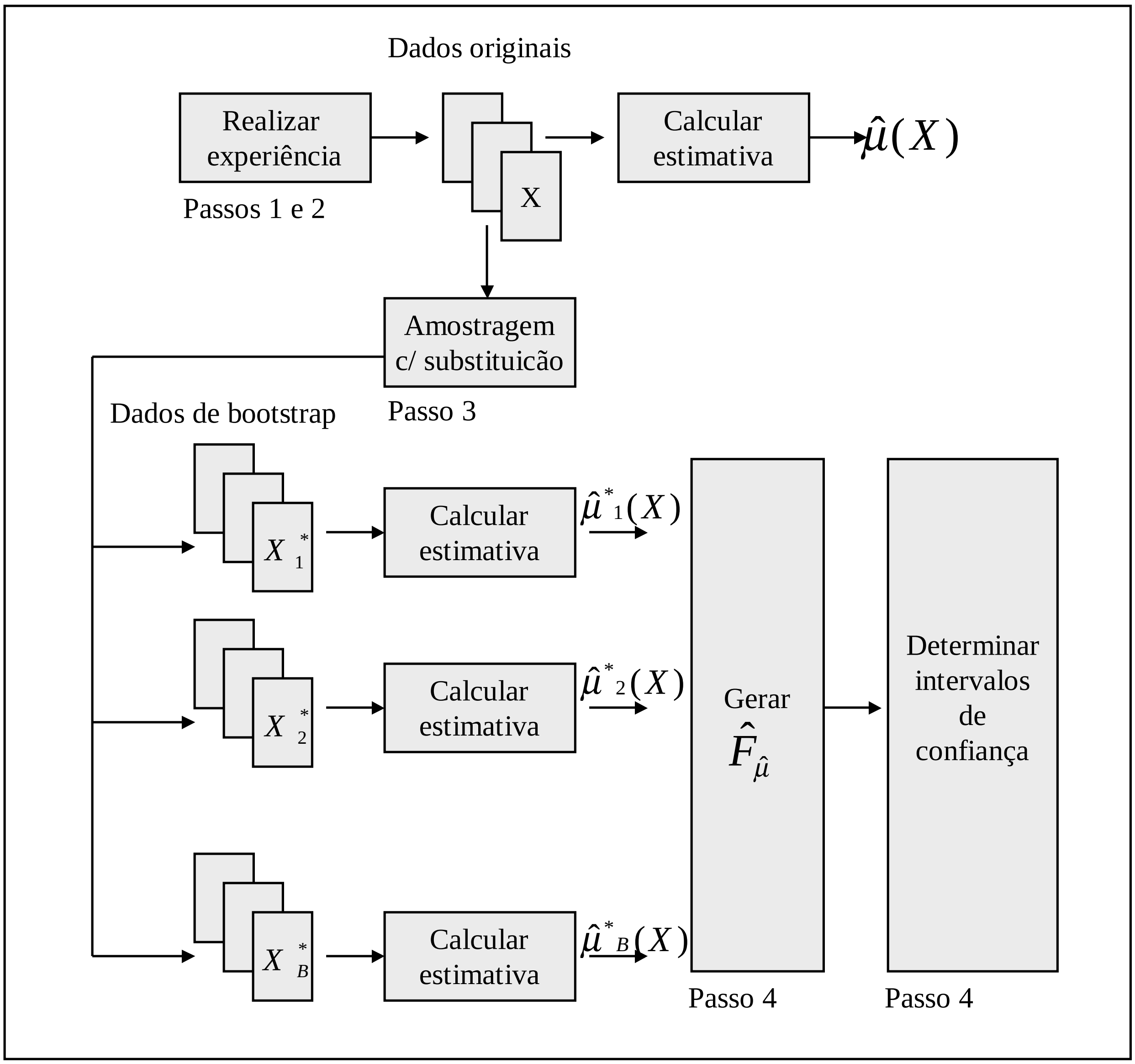
em que

$$q_1 = \text{parte\_inteira}(B\alpha/2)$$

$$q_2 = B - q_1 + 1$$

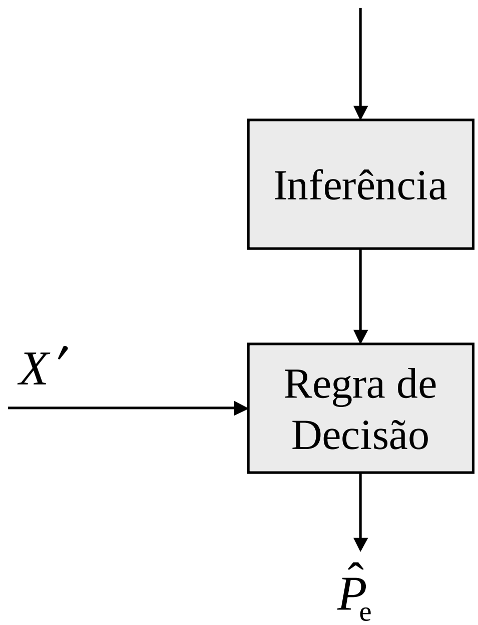
Para α=0.005 e *B*=1000 vem $q_1 = 25$ $q_2 = 976$

e o intervalo é (6.27,13.19)



Ex: estimação do erro de um classificador

$$X = \{X_1, \cdots X_N\} \quad , \text{ conjunto de amostras}$$



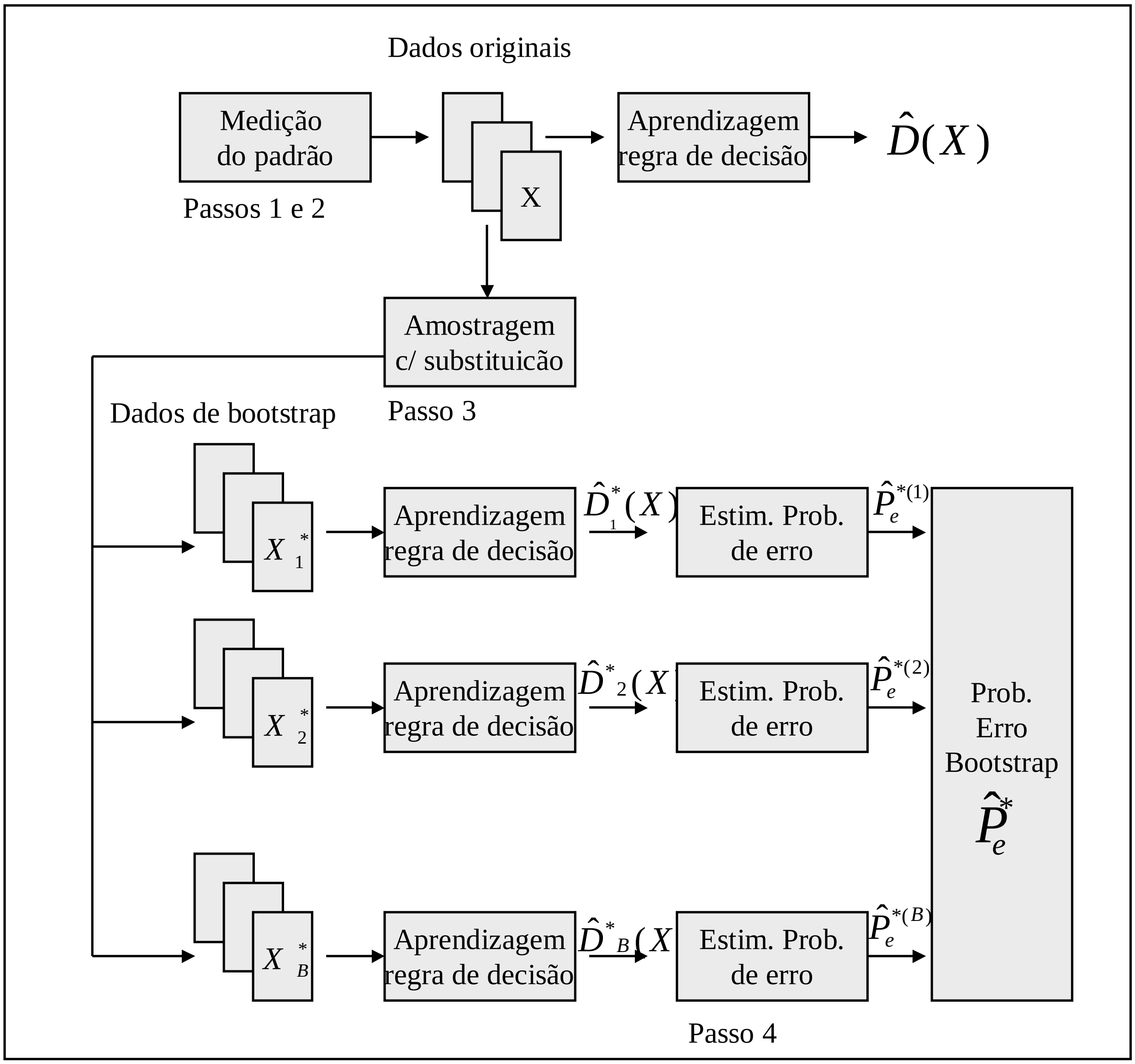
- **Se *X* é numeroso divide-se *X* em dois conjuntos**

$$X^{(1)}, X^{(2)} \qquad X^{(1)} \cup X^{(2)} = \mathrm{X}$$

$$X^{(1)} \cap X^{(2)} = \Phi$$

e usa-se $X^{(1)}$ **em treino e** $X^{(2)}$ **em teste**

- **Se *X* possui um número reduzido de amostras: aplicar bootstrapping**


Dados originais
Medição
do padrão
Passos 1 e 2
X
Aprendizagem
regra de decisão
$\hat{D}(X)$
Amostragem
c/ substituicão
Passo 3
Dados de bootstrap
X_1^*
X_2^*
X_B^*
Aprendizagem
regra de decisão
$\hat{D}_1^*(X)$
$\hat{D}^*_2(X$
$\hat{D}^*_B(X$
Estim. Prob.
de erro
$\hat{P}_e^{*(1)}$
$\hat{P}_e^{*(2)}$
$\hat{P}_e^{*(B)}$
Prob.
Erro
Bootstrap
$\hat{P}_e^*$
Passo 4

Bibliografia

- **B. Efron, R. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, 1993.**