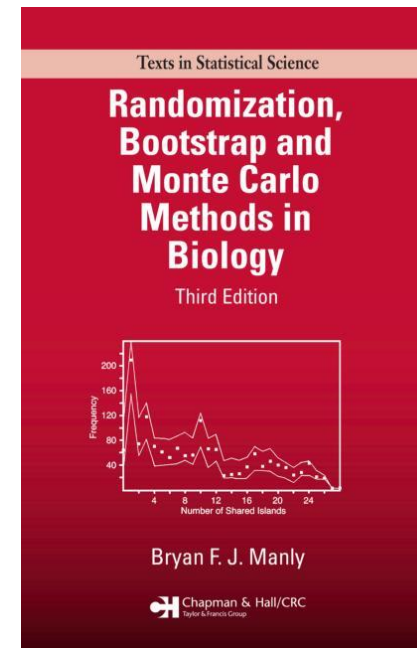


2

Aleatorização



- Manly, B. F. J. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 3ª ed., Chapman & Hall/CR Press, Boca Raton, 2007.

ESQUEMA DO CAPÍTULO

2.1 MÉTODOS DE ALEATORIZAÇÃO

2.2 MÉTODO JACKKNIFE

2.2 Método Jackknife

- Um **jackknife** é um faca de bolso com diversas ferramentas, como um canivete suíço;
- O valor da ferramenta é sua **portabilidade** e a possibilidade de utilização em diversas situações, embora raramente seja a ferramenta ideal;
- Em estatística, o nome jackknife foi usado por Tukey (1958) para descrever uma abordagem geral para teste de hipóteses e intervalos de confiança, em situações em que métodos melhores **não** poderiam ser facilmente utilizados;
- Para os padrões **atuais**, este método não é realmente computacionalmente intensivo.



2.2 Método Jackknife (cont.)

- Suponha que uma amostra aleatória de n valores $X_1, X_2, \dots, X_n$ são obtidos e a **média amostral**

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

é usada para estimar a média populacional.

- Suponha agora que a média amostral seja **recalculada** retirando-se a j -ésima observação

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

- Com o auxílio das duas equações acima, pode-se **obter**:

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

- Portanto, é possível determinar o valor amostral X_j a partir da média total e a média com o X_j removido.

2.2 Método Jackknife (cont.)

- O método é particularmente útil na estimativas de **outros** parâmetros que não a média populacional;
- Suponha que o **parâmetro θ** seja estimado por alguma função de uma amostra de tamanho n ,

$$\hat{\theta}(X_1, X_2, \dots, X_n)$$

com a remoção de X_j , tem-se a estimativa **parcial**,

$$\hat{\theta}_{-j}(X_1, X_2, \dots, X_{j-1}, X_{j+1}, \dots, X_n)$$

- Por analogia com o caso do cálculo da média, calculam-se os seguintes **pseudo-valores**

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

que fazem o papel dos valores X_j , no cálculo da média;

- A **estimativa pontual jackknife** para θ é a **média** destes pseudo-valores:

$$\hat{\theta}^* = \sum_{j=1}^n \hat{\theta}_j^* / n$$

2.2 Método Jackknife (cont.)

- A consideração dos pseudo-valores como amostras aleatórias de estimativas independentes sugere que **variância do estimador jackknife** é dado por

$$s^2/n,$$

em que s^2 é a variância dos pseudo-valores;

- Um **intervalo de confiança** aproximado de $(1-\alpha)100\%$ para θ é dado por

$$\hat{\theta}^* \pm t_{\alpha/2, n-1} s / \sqrt{n}$$

em que $t_{\alpha/2, n-1}$ é percentil $(1-\alpha/2)$ da distribuição t , com $n-1$ graus de liberdade;

- Assim, aparentemente, o jackknife transforma o problema de estimação do parâmetro θ , em um problema de estimação de **médias aritméticas**;
- Entretanto, como os pseudo-valores podem estar correlacionados, as estimativas podem ser **viciadas**.

2.2 Método Jackknife (cont.)

Exemplo 2.2.1:

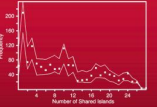
- Para a estimação do desvio padrão populacional, suponha que seja extraída uma amostra aleatória de tamanho 20 de uma certa população:

3,56; 0,69; 0,10; 1,84; 3,93; 1,25; 0,18; 1,13; 0,27; 0,50; 0,67;
0,01; 0,61; 0,82; 1,70; 0,39; 0,11; 1,20; 1,21; 0,72

- Suponha que você decide utilizar o estimador (viciado):

$$\hat{\sigma} = \sqrt{\left[\sum_{i=1}^n (x_i - \bar{x})^2 / 20 \right]}$$

- Questões relativas ao uso do jackknife:
 - O vício do estimador pode ser removido?
 - É possível estimar o erro padrão do estimador do desvio padrão?
 - Um intervalo de confiança de $(1-\alpha)100\%$ pode ser construído?



2.2 Método Jackknife (cont.)

Exemplo 2.2.1 (cont.):

- Cálculos para a estimativa jackknife:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	6	17	18	19	20	desvio padrão	pseudo-valor
amostra	3,56	0,69	0,10	1,84	3,93	1,25	0,18	1,13	0,27	0,50	0,67	0,01	0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	1,033	
1		0,69	0,10	1,84	3,93	1,25	0,18	1,13	0,27	0,50	0,67	0,01	0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	0,879	3,959
2	3,56		0,10	1,84	3,93	1,25	0,18	1,13	0,27	0,50	0,67	0,01	0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	1,056	0,586
3	3,56	0,69		1,84	3,93	1,25	0,18	1,13	0,27	0,50	0,67	0,01	0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	1,036	0,971
4	3,56	0,69	0,10		3,93	1,25	0,18	1,13	0,27	0,50	0,67	0,01	0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	1,043	0,840
5	3,56	0,69	0,10	1,84		1,25	0,18	1,13	0,27	0,50	0,67	0,01	0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	0,813	5,202
6	3,56	0,69	0,10	1,84	3,93		0,18	1,13	0,27	0,50	0,67	0,01	0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	1,059	0,544
7	3,56	0,69	0,10	1,84	3,93	1,25		1,13	0,27	0,50	0,67	0,01	0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	1,040	0,898
8	3,56	0,69	0,10	1,84	3,93	1,25	0,18		0,27	0,50	0,67	0,01	0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	1,059	0,527
9	3,56	0,69	0,10	1,84	3,93	1,25	0,18	1,13		0,50	0,67	0,01	0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	1,044	0,823
10	3,56	0,69	0,10	1,84	3,93	1,25	0,18	1,13	0,27		0,67	0,01	0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	1,052	0,671
11	3,56	0,69	0,10	1,84	3,93	1,25	0,18	1,13	0,27	0,50		0,01	0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	1,056	0,593
12	3,56	0,69	0,10	1,84	3,93	1,25	0,18	1,13	0,27	0,50	0,67		0,61	0,82	1,70	0,39	0,11	1,20	1,21	0,72	1,031	1,062
13	3,56	0,69	0,10	1,84	3,93	1,25	0,18	1,13	0,27	0,50	0,67	0,01		0,82	1,70	0,39	0,11	1,20	1,21	0,72	1,055	0,617
14	3,56	0,69	0,10	1,84	3,93	1,25	0,18	1,13	0,27	0,50	0,67	0,01	0,61		1,70	0,39	0,11	1,20	1,21	0,72	1,058	0,548
15	3,56	0,69	0,10	1,84	3,93	1,25	0,18	1,13	0,27	0,50	0,67	0,01	0,61	0,82		0,39	0,11	1,20	1,21	0,72	1,048	0,738
16	3,56	0,69	0,10	1,84	3,93	1,25	0,18	1,13	0,27	0,50	0,67	0,01	0,61	0,82	1,70		0,11	1,20	1,21	0,72	1,048	0,737
17	3,56	0,69	0,10	1,84	3,93	1,25	0,18	1,13	0,27	0,50	0,67	0,01	0,61	0,82	1,70	0,39		1,20	1,21	0,72	1,037	0,962
18	3,56	0,69	0,10	1,84	3,93	1,25	0,18	1,13	0,27	0,50	0,67	0,01	0,61	0,82	1,70	0,39	0,11		1,21	0,72	1,059	0,535</

2.2 Método Jackknife (cont.)

Exemplo 2.2.1 (cont.):

- A **média** dos pseudo-valores é a estimativa jackknife de θ :

$$\hat{\sigma}^* = \sum_{j=1}^n \hat{\sigma}_j^* / n = 1,096$$

- O **erro padrão** estimado é $s/\sqrt{20}$, em que s é o desvio padrão dos 20 pseudo-valores:

$$s / \sqrt{20} = \sqrt{\left[\sum_{j=1}^n (\hat{\sigma}_j^* - \hat{\sigma}^*)^2 / 19 \right]} / \sqrt{20} = 1,220 / \sqrt{20} = 0,273$$

- O **intervalo de confiança** de 95% é dado por:

$$1,096 \pm 2,09 \times 0,273 = [0,527; 1,665]$$

em que $t_{0,025;19} = 2,09$ corresponde ao percentil 1-0,025=0,975 da distribuição t , com $n-1=19$ graus de liberdade;

- O jackknife pode ser considerado **bem sucedido**, pois os dados (simulados de uma exponencial) têm média e desvio padrão iguais a 1.