

Simulação Monte Carlo

Tópicos de Inferência Estatística

José Passos

ISEG-ULisboa

10 de Outubro de 2019



- A simulação Monte Carlo remonta aos anos 40 do século XX, associada aos trabalhos de pesquisa em torno do desenvolvimento da primeira bomba atómica.
- O objectivo é resolver problemas complexos utilizando a força bruta computacional.
- O método de Monte Carlo assenta em técnicas de amostragem e probabilísticas.
- Tem a virtude de poder simular um fluxo "infindável" de variáveis aleatórias, para uma distribuição conhecida, possibilitando-nos de uma forma simples, a utilização dos resultados frequencistas e assintóticos (Lei dos grandes números e TLC).
- Esta simulação baseia-se na distribuição $\mathcal{U}(0, 1)$, que constitui a representação probabilística básica de aleatoriedade num computador.

- No R os valores $\mathcal{U}(0, 1)$ são gerados a partir da instrução `runif()`.
- Em rigor, os valores supostamente aleatórios, gerados num computador, são efectivamente pseudo aleatórios pois são gerados com recurso a um algoritmo, de forma determinística.
- De modo simplificado fala-se simplesmente em números aleatórios ou gerador de aleatórios.
- Uma sequência de números pseudo aleatórios deverá ter as mesmas propriedades estatísticas de uma sequência *i.i.d.*. Este é o critério que se utiliza para avaliar a qualidade de um gerador de (pseudo) aleatórios.

- Em geral pretendem-se valores aleatórios provenientes de outras distribuições, para além da uniforme, $\mathcal{U}(0, 1)$.
- Existe uma técnica (transformação inversa) que nos permite transformar qualquer v.a. numa uniforme e *vice versa*.

Definição

Seja X uma v.a. contínua com f.d.p. f e f.d. F . A v.a. $U = F(X)$ tem distribuição $\mathcal{U}(0, 1)$

- **Exemplo:** se $X \sim \mathcal{E}(1)$ então $F(x) = 1 - e^{-x}$. Resolvendo $u = 1 - e^{-x}$ em ordem a x obtém-se $x = -\log(1 - u)$. Portanto, se $U \sim \mathcal{U}(0, 1)$ então $X = -\log(U) \sim \mathcal{E}(1)$. Na prática gera-se computacionalmente u e resolve-se a equação anterior em ordem a x para obter valores com a distribuição desejada.

- Para gerar v.a.s discretas usa-se também a técnica de transformação inversa.
- Para gerar $X \sim P_\theta$, onde P_θ tem suporte inteiro, calculam-se as probabilidades,

$$p_0 = P_\theta(X \leq 0)$$

$$p_1 = P_\theta(X \leq 1)$$

$$p_2 = P_\theta(X \leq 2)$$

...

- De seguida gera-se $U \sim \mathcal{U}(0, 1)$ e faz-se

$$X = k \text{ se } p_{k-1} < U < p_k$$

- Existem distribuições onde a técnica da transformação inversa não funciona. Neste caso existem métodos alternativos designados métodos indirectos (método de aceitação-rejeição).
- O R tem a vantagem de gerar directamente a generalidade das distribuições com que trabalhamos, não sendo necessário a aplicação dos métodos referidos.

Tabela: Geradores de aleatórios no R

Distribuição	Comando R
$N(\mu, \sigma^2)$	<code>rnorm(n,mu,sigma)</code>
$B(\alpha, \beta)$	<code>rbeta(n,alpha,beta)</code>
$Bi(N, p)$	<code>rbinom(n,N,p)</code>
$Po(\theta)$	<code>rpoi(n,theta)</code>
$\chi^2_{(\nu)}$	<code>rchisq(n,nu)</code>
$t^2_{(\nu)}$	<code>rt(n,nu)</code>
...	...

Teorema: Lei Forte dos Grandes Números

Seja X_1, X_2, \dots, X_n uma amostra casual proveniente de uma população X com média μ finita. Então quando $n \rightarrow \infty$,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{q.c.} \mu$$

Aplicação frequente:

- aproximar $E(X)$ a partir de \bar{x} , para uma amostra observada.

Integração de Monte Carlo:

- pretende-se calcular o integral:

$$E[h(X)] = \int_{\mathcal{X}} h(x)f(x)dx$$

- a utilização do método de Monte Carlo para calcular $E[h(X)]$ consiste em gerar uma amostra de dimensão n (com n grande) a partir de f , propondo como aproximação,

$$\bar{h}_n = \frac{1}{n} \sum_{i=1}^n h(x_i)$$

- pela lei forte dos grandes números \bar{h}_n converge quase certamente para $E[h(X)]$

- adicionalmente, se $E[h^2(X)]$ existe finito tem-se pelo TLC,

$$\frac{\bar{h}_n - E[h(X)]}{\sqrt{v_n}} \rightarrow^d N(0, 1)$$

onde v_n é uma estimativa de $var(\bar{h}_n)$,

$$v_n = \frac{1}{n} \sum_{i=1}^n [h(x_i) - \bar{h}_n]^2$$

Outras utilizações: o método de Monte Carlo pode ser utilizado para aproximar a distribuição por amostragem de uma estatística $T(X)$:

- vamos supor a população $X \sim F_\theta$, com F_θ conhecida;
- vamos fixar a dimensão da amostra em n ;
- vamos gerar $g = 1, 2, \dots, G$ amostras da população F_θ cada uma com dimensão n :

amostra 1: $x_1^1, x_2^1, \dots, x_n^1$

amostra 2: $x_1^2, x_2^2, \dots, x_n^2$

...

amostra G: $x_1^G, x_2^G, \dots, x_n^G$

- para cada amostra g , calcula-se o valor da estatística, t^g :

$$t^1 = T(x_1^1, x_2^1, \dots, x_n^1)$$

$$t^2 = T(x_1^2, x_2^2, \dots, x_n^2)$$

...

$$t^G = T(x_1^G, x_2^G, \dots, x_n^G)$$

- no final tem-se uma amostra de dimensão G , proveniente da distribuição por amostragem da estatística $T(X_1, \dots, X_n)$, que pode ser utilizada para estudar as suas propriedades, calcular o histograma, probabilidades, momentos, etc.