

2 Simulação Estocástica

Em experimentos aleatórios não temos como saber qual será exatamente o resultado, entretanto, seu espaço de probabilidade pode ser estabelecido de forma conveniente de modo a avaliar a probabilidade de qualquer evento de interesse da σ -álgebra \mathcal{F} . Este fato dá origem ao que chamamos de *variável aleatória*.

Definição: Uma variável aleatória Y em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathcal{P})$ é uma função $Y : \Omega \rightarrow \mathbb{R}$ tal que

$$Y^{-1}(I) = \{\omega \in \Omega : Y(\omega) \in I\} \in \mathcal{F},$$

para todo intervalo $I \subset \mathbb{R}$, o que significa que Y é uma função mensurável em \mathcal{F} .

Dessa forma, a estrutura probabilística de quantidades associadas aos experimentos aleatórios podem ser estudadas, assim como fazer inferências sobre os parâmetros dos modelos relacionadas a essa estrutura. Porém, em situações, nas quais não conseguimos mostrar analiticamente as propriedades dos estimadores de máxima verossimilhança ou o comportamento dos resíduos a serem propostos, podemos realizar um estudo de simulação.

Nesse contexto, uma simulação estocástica consiste em reproduzir numericamente uma dada variável aleatória, cujo comportamento é representado matematicamente, de modo que cada simulação tenha um resultado diferente. Os métodos de simulação estocástica são procedimentos que envolvem a geração de números pseudo-aleatórios, com a seguinte estrutura básica:

- i. Identificar a variável aleatória de interesse, denotada por Y .
- ii. Gerar uma amostra aleatória Y_1, \dots, Y_n com a mesma distribuição de Y , ou seja, variáveis aleatórias independentes e identicamente distribuídas (iid).
- iii. Estimar a esperança da variável aleatória, denotada por $E(Y)$ (utilizando $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$) e avaliar a acurácia de estimativa (utilizando intervalos de confiança).

Em outras palavras, dada uma distribuição de probabilidade com função densidade ou função de probabilidade f , queremos determinar um algoritmo que possa reproduzir amostras aleatórias a partir dessa distribuição. Dessa forma, o “bloco de construção” de uma simulação estocástica é habilidade de gerar variáveis aleatórias.

Apesar da geração de variáveis aleatórias com distribuição de probabilidades conhecidas na literatura já estarem implementadas na linguagem R, a proposta desta Seção é mostrar como podemos, por esforço próprio, fazer isso para que tenhamos as ferramentas para simular variáveis aleatórias que ainda não estejam implementadas.

2.1 Geração de números aleatórios

Os métodos de simulação são embasados na geração de variáveis aleatórias independentes uniformemente distribuídas no intervalo $[0, 1]$. Porém, na prática, não é possível gerar de forma realmente aleatória. Ao invés disso, geramos *números pseudo-aleatórios* que tem uma aparência de números aleatórios, mas são de fato determinísticos. Uma vantagem de utilizar os números pseudo-aleatórios consiste em poder repetir exatamente a simulação posteriormente.

► **Gerador congruencial:** Dado um número inicial $Y_0 \in \{0, 1, \dots, m-1\}$ para $m \in \mathbb{N}$, definimos a sequência de números $Y_i \in \{0, 1, \dots, m-1\}$, $i = 1, 2, \dots$, por

$$Y_i = (aY_{i-1} + b) \bmod m,$$

em que a e b são números inteiros positivos e a expressão $d \bmod m$ corresponde ao resto da divisão de d por m . Como os computadores utilizam aritmética binária, podemos considerar $m = 2^k$ para algum k . De acordo com o teorema de Hull-Dobell, assumir $k = 32$, $a = 1103515245$ e $b = 12345$ garante que a sequência é de período M para cada valor inicial Y_0 . Por outro lado, Jones et al. (2014) sugere $k = 32$, $a = 1664525$ e $b = 1013904223$ para um bom método congruencial. Um bom gerador de números pseudo-aleatórios possui as seguintes características:

- i. Passar em testes estatísticos de aleatoriedade;
- ii. Possuir um longo período;

- iii. Eficiente;
- iv. Repetibilidade.

Agora se definimos

$$U_i = \frac{Y_i}{m},$$

obtemos valores entre 0 e 1 que se comportam como variáveis aleatórias uniformes, ou seja, $U \sim \mathcal{U}(0, 1)$.

► *Semente*: O número Y_0 é chamado de semente. Se conhecemos a semente (assim como m , a e b), então podemos reproduzir exatamente toda a sequência. A linguagem R utiliza o gerador de números pseudo-aleatórios denominado de Mersenne-Twister, que tem propriedades similares ao gerador congruencial e com um período muito mais longo. Mas, ainda precisa de uma semente para gerar os números pseudo-aleatórios.

2.2 Geração de variáveis aleatórias

2.2.1 Método da transformação inversa

Definição: A função de distribuição acumulada (ou simplesmente função de distribuição) de uma variável aleatória Y , denotada por F , é definida por

$$F(y) = P(Y \leq y), \quad y \in \mathbb{R}. \quad (1)$$

Além disso, a função de distribuição acumulada definida em (1) de uma variável aleatória Y , obedece as seguintes propriedades:

- F1) $\lim_{y \rightarrow -\infty} F(y) = 0$ e $\lim_{y \rightarrow \infty} F(y) = 1$.
- F2) F é contínua à direita, isto é, $\lim_{\epsilon \rightarrow 0} F(y + \epsilon) = F(y)$ para $\epsilon > 0$.
- F3) F é não decrescente, isto é, $F(y) \leq F(x)$ sempre que $y \leq x$, $\forall x, y \in \mathbb{R}$.

Proposição: Sejam Y uma variável aleatória com função de distribuição acumulada F definida em (1), F^{-1} denotando a função inversa de F e a variável aleatória $U \sim \mathcal{U}(0, 1)$. Então, $Y = F^{-1}(U)$ tem função de distribuição acumulada F .

Temos que $F^{-1}(u) \leq y$ é equivalente a $u \leq F(y)$. Logo, para $U \sim \mathcal{U}(0, 1)$, observamos que

$$P(Y \leq y) = P[F^{-1}(U) \leq y] = P[U \leq F(y)] = F(y), \quad (2)$$

isto é, F é a função de distribuição acumulada de $F^{-1}(U)$.

Portanto, para simular uma amostra aleatória de uma variável aleatória Y com função de distribuição acumulada F pelo método da transformação inversa, temos os seguintes algoritmos:

► *Variável aleatória discreta*: uma variável aleatória Y em $(\Omega, \mathcal{F}, \mathcal{P})$ é classificada como discreta, se assume somente um número enumerável de valores (finito ou infinito). A função de probabilidade, denotada por $f(y_i) = P(Y = y_i)$, $i = 1, 2, \dots$ é uma função que satisfaz as seguintes condições:

- a) $0 \leq f(y_i) \leq 1, \forall i = 1, 2, \dots$;
- b) $\sum_i f(y_i) = 1$.

Assim, temos:

- P1. Definir k, pk, Fy ;
- P2. Gerar $U \sim \mathcal{U}(0, 1)$;
- P3. Enquanto $U > Fy$ então calcular $py, Fy = Fy + py$ (atualizar Fy), $k = k+1$ (incrementar k) e volta ao P2;
- P4. Declarar $y = k$.

- Variável aleatória contínua: uma variável aleatória Y em $(\Omega, \mathcal{F}, \mathcal{P})$ com função de distribuição acumulada F é classificada como contínua, se existir uma função não negativa f , tal que

$$F(y) = \int_{-\infty}^y f(\omega) d\omega, \text{ para todo } y \in \mathbb{R}.$$

Então, a função f é denominada função densidade de probabilidade e satisfaz as seguintes condições:

- a) $f(y) \geq 0, \forall y \in \mathbb{R}$;
- b) $\int_{-\infty}^{\infty} f(\omega) d\omega = 1$.

Logo, temos:

- P1. Gerar $U \sim \mathcal{U}(0, 1)$;
- P2. Retornar $y = F^{-1}(U)$.

★ Observação:

- Função quantílica: é a inversa da função de distribuição acumulada;
- Fórmula recursiva: é uma fórmula que define os termos de uma sequência utilizando termos anteriores, cujas partes são:
- i. o valor de um termo inicial;
 - ii. uma expressão em função de termos anteriores.

Dessa forma, certas probabilidade podem ser expressas em uma forma recursiva a partir de

$$\frac{P(Y = k + 1)}{P(Y = k)}, \quad k = 0, 1, \dots,$$

em que $P(\cdot)$ é uma função de probabilidade de uma variável aleatória Y .

2.2.2 Método da aceitação e rejeição

Sejam Y uma variável aleatória com função densidade ou função de probabilidade f e X uma variável aleatória com função densidade ou função de probabilidade g , em que g é parecida com f . Suponhamos que gerar variáveis aleatórias de Y seja uma tarefa difícil, mas obter amostras aleatórias de X seja mais fácil. Então, $g(x)$ é um envelope para $f(x)$ se existe uma constante $M > 0$ tal que

$$f(x) \leq M g(x) \quad \text{para todo } x,$$

em que $M = \max_x \frac{f(x)}{g(x)}$.

Portanto, o algoritmo para gerar $Y \sim f$ pelo método da aceitação e rejeição é dado por

- P1. Gerar $X \sim g$;
- P2. Gerar $U \sim \mathcal{U}(0, 1)$;
- P3. Se $U \leq \frac{f(X)}{Mg(X)}$ então $Y = X$ (“aceita”). Caso contrário, retorne ao passo 1 (“rejeita”).

2.2.3 Geração de variáveis aleatórias com distribuição normal

Dizemos que uma variável aleatória Y é normalmente distribuída com média $\mu \in \mathbb{R}$ e variância $\sigma^2 > 0$ se a sua função densidade de probabilidade é dada por

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} (y - \mu)^2 \right], \quad -\infty < y < \infty$$

Logo, se $Y \sim N(\mu, \sigma^2)$, temos que

$$Z = \frac{Y - \mu}{\sqrt{\sigma^2}} \sim N(0, 1), \quad (3)$$

ou seja, a função densidade de probabilidade da variável aleatória Z é dada por

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right), \quad -\infty < z < \infty, \quad (4)$$

e conhecida como distribuição normal padrão.

A partir do resultado (3) temos que

$$Y = \mu + \sigma Z,$$

em que $\sigma = \sqrt{\sigma^2}$ e $Z \sim N(0, 1)$.

Portanto, para simular uma variável aleatória normal com parâmetros μ e σ^2 é suficiente simular variáveis aleatórias a partir da distribuição normal padrão (4). Então, para simular variáveis aleatórias normais, podemos utilizar a transformação Box-Muller. Sejam $X \sim N(0, 1)$ e $Y \sim N(0, 1)$ variáveis aleatórias independentes e a função densidade de probabilidade conjunta é dada por

$$f_{X,Y}(x, y) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}(x^2 + y^2)\right], \quad -\infty < x < \infty, -\infty < y < \infty.$$

Dessa forma, para gerar números aleatórios da distribuição normal padrão, vamos considerar a mudança de variáveis por coordenadas polares, ou seja,

$$x = r \cos(\theta) \quad \text{e} \quad y = r \sin(\theta),$$

em que r presente o raio e θ o ângulo.

Logo, pelo método do Jacobiano

$$f_{R,\Theta}(r, \theta) = f_{X,Y}(x, y) \left| \begin{array}{cc} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{array} \right|,$$

a função densidade conjunto de R e Θ é dada por

$$f_{R,\Theta}(r, \theta) = \frac{1}{2\pi} r \exp\left(-\frac{1}{2}r^2\right), \quad r > 0, \quad 0 < \theta < 2\pi. \quad (5)$$

Agora, obtendo as funções densidades marginais da função conjunta (5), podemos utilizar o método da transformação inversa para gerar os números aleatórios de r e θ . Então, a função densidade marginal de Θ é dada por

$$f(\theta) = \int_0^\infty \frac{1}{2\pi} r \exp\left(-\frac{1}{2}r^2\right) dr = \frac{1}{2\pi}, \quad \text{se} \quad 0 < \theta < 2\pi,$$

ou seja, $\Theta \sim \mathcal{U}(0, 2\pi)$.

Observação: Se $U \sim \mathcal{U}(0, 1)$ e $X \sim \mathcal{U}(a, b)$, então podemos gerar os números aleatórios de X utilizando U da seguinte forma

$$x = a + (b - a)u.$$

Por outro lado, a função densidade marginal de R é dada por

$$f(r) = \int_0^{2\pi} \frac{1}{2\pi} r \exp\left(-\frac{1}{2}r^2\right) d\theta = r \exp\left(-\frac{1}{2}r^2\right) \quad \text{se} \quad r > 0. \quad (6)$$

Logo, podemos observar que a função densidade (6) é distribuição Weibull com parâmetro $\lambda = \sqrt{2}$ e $\alpha = 2$. Dessa forma, os números aleatórios de r podem ser gerados a partir da seguinte expressão

$$r = \lambda[-\log(1 - u)]^{1/\alpha} = \sqrt{-2 \ln(1 - u)},$$

em que $U \sim \mathcal{U}(0, 1)$.

Portanto, o algoritmo para gerar números aleatórios de uma distribuição normal é dado por

P1. Gerar $U \sim \mathcal{U}(0, 1)$;

P2. Gerar $\theta \sim \mathcal{U}(0, 2\pi)$;

P3. Fazer $r = \sqrt{-2 \ln(1 - U)}$;

P4. Fazer $x = r \cos(\theta)$;

P5. Fazer $y = r \sin(\theta)$;

P6. Fazer $z = [x, y]$.