

# **Estatística Computacional — Módulo 1**

Notas de apoio (Cap. 2)

**Manuel Cabral Morais**

**Secção de Estatística e Aplicações**

**Instituto Superior Técnico**

Lisboa, Setembro–Outubro de 2003

# Contents

<b>2</b>	<b>Geração de números aleatórios. Método de Monte Carlo</b>	<b>49</b>
2.1	Geração de quantidades aleatórias discretas e contínuas por transformação de números aleatórios. Aplicação a distribuições específicas. . . . .	51
2.1.1	Geração de números pseudo-aleatórios da distribuição uniforme . . . . .	52
2.1.2	O método da transformada inversa . . . . .	54
2.1.3	Alguns algoritmos de geração de números pseudo-aleatórios de algumas distribuições . . . . .	58
2.1.4	Métodos de aceitação/rejeição . . . . .	71
2.2	Método de Monte Carlo . . . . .	76
2.2.1	Monte de Carlo hit-or-miss . . . . .	77
2.2.2	Monte de Carlo ordinário . . . . .	79
2.2.3	Técnicas de redução de variância de estimadores de Monte Carlo . . . . .	81
2.2.4	Técnicas de redução de variância na simulação do sistema M/M/1 . . . . .	86
2.2.5	Integrais múltiplos . . . . .	90
2.2.6	Método de Monte Carlo em inferência estatística	91
2.3	Referências . . . . .	93

## Chapter 2

# Geração de números aleatórios.

## Método de Monte Carlo

O papel da **simulação** na maior parte das ciências tem aumentado de importância de há umas décadas a esta parte. A crescente capacidade/velocidade de cálculo dos computadores a par da evolução notória de diversos métodos de simulação são responsáveis por se considerar que a simulação é uma **abordagem metodológica** possível para a **resolução aproximada** de problemas como o é a abordagem experimental.

Não surpreende pois que os **métodos de simulação** sejam usados em áreas como a física, a química, a criptografia, a biotecnologia, a indústria e, naturalmente, a **Estatística**.

Nesta última área é frequente ser preciso identificar a **distribuição** (exacta) ou os quantis de uma estatística, calcular probabilidades de eventos que dizem respeito a vectores aleatórios com densidades não necessariamente triviais (estamos a falar de integrais múltiplos sofisticados), ou a função potência de um teste de hipóteses.

Para o efeito será necessário obter concretizações de amostras aleatórias — ou seja, **simular/gerar amostras** — em **condições**

absolutamente

- **controladas e**
- **reprodutíveis,**

e de modo eficiente e passível de documentação cuidada e completa.

Este capítulo debruçar-se-á sobre diversos métodos de obtenção de números (pseudo-)aleatórios de diversas distribuições univariadas e multivariadas, sobre o método de Monte-Carlo e sobre métodos de redução de variância.

**Texto de apoio:** Gentle (1998, pp. vii–ix).

## 2.1 Geração de quantidades aleatórias discretas e contínuas por transformação de números aleatórios. Aplicação a distribuições específicas.

Os estatísticos necessitam muitas vezes de gerar amostras provenientes de populações ou distribuições específicas e como tal é fundamental disporem de uma fonte de **números aleatórios**.

É com vista a suprir esta necessidade que era usual muitos livros de texto incluírem em apêndice **tabela de números aleatórios**. Veja-se, por exemplo, as páginas 7 a 10 da seguinte obra em português: Murteira, B.J.F. (1980). *Probabilidades e Estatística (Vol. II)*. McGraw-Hill de Portugal, Lda.

Era com base em tabelas destas que se seleccionava amostras de uma população ou se decidia uma atribuição de tratamentos num planeamento experimental. Por exemplo, para obter uma amostra casual de uma população bastava atribuir um número a cada elemento da população, entrando depois numa página e coluna de qualquer tabela de números aleatórios, anotando os números obtidos e retirando da população os elementos correspondentes. As tabelas de números aleatórios são construídas de forma a garantir a natureza aleatória dos números que delas constam. Por sinal, os números das tabelas de Murteira (1980, pp. 7–10) foram obtidos com uma espécie de **roleta electrónica** e submetidos a numerosos **ensaios para testar a casualidade dos números gerados**.

Hoje em dia raramente se usa tabelas de números aleatórios. É de longe mais usual recorrer a computadores para gerar números *aleatórios*. A palavra *aleatórios* foi intencionalmente colocada em itálico pois, na

verdade, um computador é incapaz de gerar números aleatórios como se verá já a seguir.

**Texto de apoio:** Gentle (1998, pp. 1–6).

### 2.1.1 Geração de números pseudo-aleatórios da distribuição uniforme

Os números gerados por um computador são obtidos com recurso a **geradores determinísticos** com **carácter recursivo** — daí a **perda de aleatoriedade** e a designação de **números pseudo-aleatórios** (n.p.a.). Estes números constituem uma sequência com a particularidade de os últimos  $k$  números gerados serem responsáveis pela geração do seguinte. Esta sequência é finita e o seu tamanho é usualmente denominado de **período** ou **tamanho do ciclo**.

Os geradores determinísticos mais comuns são sem sombra de dúvida os **geradores congruenciais lineares**.

Estes geradores foram propostos por D.H. Lehmer em 1948. De acordo com este gerador cada número gerado  $x_{i-1}$  determina recursivamente o seguinte  $x_i$  por intermédio da fórmula:

$$x_i = (ax_{i-1} + c) \bmod m, \quad i = 1, 2, \dots, \quad 0 \leq x_i < m, \quad (2.1)$$

onde  $a$ ,  $c$  e  $m$  são inteiros e  $x_0$  é outro inteiro frequentemente denominado de **semente** já que em conjunto com  $a$ ,  $c$  e  $m$  definem univocamente uma sequência de números pseudo-aleatórios.

$x_i$  corresponde ao **resto da divisão inteira** de  $(ax_{i-1} + c)$  por  $m$ . Dividindo  $x_i$  por  $m$  tem-se, para cada  $i$  e para valores de  $a$  e  $m$  fixos convenientemente escolhidos, um **número**

$$u_i = \frac{x_i}{m} \quad (2.2)$$

**aproximadamente Uniforme no intervalo (0,1).**

Caso se considere  $c = 0$  o gerador é denominado de **gerador congruencial multiplicativo** e tem-se

$$x_i = (ax_{i-1}) \bmod m, \quad i = 1, 2, \dots, \quad 0 \leq x_i < m. \quad (2.3)$$

Para  $c \neq 0$  o gerador é também denominado de **gerador congruencial misto**.

A **escolha** de  $a$ ,  $c$  e  $m$  **requer algum cuidado** já que se pode correr o risco de a sequência de números pseudo-aleatórios rapidamente se repetir. Senão veja-se o caso em que  $a = 3$ ,  $c = 0$ ,  $m = 30$  e  $x_0 = 1$ , associado à sequência  $\{3, 9, 27, 21, 3, 9, 27, 21, \dots\}$ .

De facto, convém notar que o gerador produz quando muito  $m$  números pseudo-aleatórios distintos antes de voltar a repetir a sequência.

A **NAG** utiliza um gerador congruencial multiplicativo ( $c = 0$ ), com  $a = 13^{13}$  e  $m = 2^{59}$ , que possui um período igual a  $2^{57} \simeq 1.44 \times 10^{17}$ . A semente  $x_0$  é determinada pelo relógio interno do computador mas pode também ser pré-especificada por forma a permitir que se repita a mesma sequência de números pseudo-aleatórios, repetição absolutamente crucial em estudos comparativos.

O Capítulo 6 de Morgan (1984) descreve testes estatísticos que permitem averiguar se a sequência gerada é consistente com a distribuição Uniforme(0, 1) e não correlacionada.

Na secção 3.5 de Morgan (1984) são feitas considerações sobre resultados teóricos que norteiam a escolha de  $a$ ,  $c$  e  $m$ .

Para uma descrição mais detalhada sobre as propriedades dos geradores congruenciais lineares bem como a implementação dos mesmos, consulte-se Gentle (1998, pp. 6–22).

Por outro lado, Gentle (1998, pp.22–38) descreve outros geradores congruenciais, como são o caso dos

- **geradores congruenciais multiplicativos múltiplos**

$$x_i = (a_1x_{i-1} + \dots + a_kx_{i-k}) \bmod m,$$

- **geradores congruenciais não lineares**

$$x_i = (dx_{i-1}^2 + ax_{i-1} + c) \bmod m,$$

- **lagged Fibonacci**

$$x_i = (x_{i-j} + \dots + x_{i-k}) \bmod m,$$

ou ainda os **geradores congruenciais combinados** e os **geradores baseados em sistemas caóticos**.

**Textos de apoio:** Gentle (1998, pp.1–38) e Morgan (1984).

### 2.1.2 O método da transformada inversa

A geração de números pseudo-aleatórios provenientes de distribuições contínuas (e também discretas) distintas da Uniforme(0,1) é usualmente efectuada tirando partindo do seguinte resultado teórico.

**Teorema 2.1** Seja  $X$  uma v.a. contínua com f.d.p.  $f(x)$  e f.d.  $F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt, x \in \mathbb{R}$ . Então a v.a.  $Y = F(X)$  — que resulta da mera substituição de  $x$  por  $X$  na expressão geral da f.d. de  $X$  — verifica

$$Y = F(X) \sim \text{Uniforme}(0,1), \tag{2.4}$$



Este resultado, cuja demonstração é trivial caso se considere a restrição da f.d. de  $X$  para a qual  $F(x)$  é invertível, sugere o seguinte **algoritmo** de obtenção de um **número pseudo-aleatório** da distribuição de **v.a. contínua**  $X$  à custa de um (ou mais) número(s) pseudo-aleatório(s) da distribuição Uniforme(0,1)



**Método da transformada inversa** (v.a. contínua) — O algoritmo compreende os seguintes passos:

1. Gerar um número pseudo-aleatório  $u$  da distribuição Uniforme(0, 1).
2. Efectuar a atribuição

$$x = F^{-1}(u), \quad (2.5)$$

quantil de ordem  $u$  da distribuição da v.a. contínua  $X$ . ■

Este método só é vantajoso quando existe expressão fechada para a inversa da f.d. de  $X$ , i.e., para o quantil de ordem  $u$ ,  $F^{-1}(u)$ .

De notar que, mesmos nos casos em que existe fórmula fechada para a inversa da f.d., a avaliação do número pseudo-aleatório  $x = F^{-1}(u)$  pode ser particularmente lenta e ser preferível recorrer a outro método alternativo e mais rápido de obtenção do número pseudo-aleatório  $x$ . Por outro lado, caso não exista expressão fechada para a inversa  $F^{-1}(u)$ , a obtenção deste quantil de ordem  $u$  passa pela resolução numérica da equação  $F(x) - u = 0$ .

Não surpreende pois que, para algumas distribuições, hajam outras escolhas possíveis de algoritmos para a geração de números pseudo-aleatórios. Como refere Gentle (1998, p. 41), estes algoritmos diferem em velocidade, precisão, necessidades de armazenamento na memória e complexidade do código.

**Exercício 2.2** — Considere que  $X \sim Exponencial(\lambda)$ , i.e.,

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0. \quad (2.6)$$

Defina o Passo 2 do algoritmo de geração de observações desta distribuição.

Repita o exercício para as seguintes distribuições:

(a)  $X \sim Weibull(\alpha, \beta)$

$$f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right], \quad x \geq 0; \quad (2.7)$$

(b)  $X \sim Pareto(\alpha, \beta)$ , i.e.,

$$F(x) = 1 - \left(\frac{\alpha}{x}\right)^\beta, \quad x \geq \alpha; \quad (2.8)$$

e ainda para

(c) a v.a.  $X$  com f.d.p. igual a  $f(x) = 2(x-1), 0 \leq x \leq 1$ . ■

**Exercício 2.3** — Escreva um programa na linguagem de programação que entender mais razoável para gerar 20 observações de cada uma das distribuições contínuas do exercício anterior, recorrendo para o efeito ao gerador congruencial multiplicativo utilizado pela *NAG*, i.e., com  $a = 13^{13}$ ,  $c = 0$ ,  $m = 2^{59}$  e  $u_0 = 1$ . ■

Coloca-se, muito naturalmente, a questão se não será possível um algoritmo similar ao apresentado para a **geração de números pseudo-aleatórios** respeitantes a **distribuições discretas**. A resposta à questão é afirmativa. Basta pensar na seguinte definição de quantil de ordem  $u$  de uma v.a. discreta com f.d.  $F(x)$ :

$$F^{-1}(u) = \inf\{m \in \mathbb{R} : u \leq F(m)\}, \quad 0 < u < 1. \quad (2.9)$$

**Método da transformada inversa** (v.a. discreta) — É natural que se considere os seguintes passos para gerar observações de uma v.a. discreta com f.d.  $F(x)$ :

1. Gerar um número pseudo-aleatório  $u$  da distribuição Uniforme(0, 1).

2. Efectuar a atribuição

$$x = F^{-1}(u) = \inf\{m \in \mathbb{R} : u \leq F(m)\}, \quad (2.10)$$

quantil de ordem  $u$  da distribuição da v.a. discreta  $X$ , onde  $F^{-1}$  representa a inversa generalizada de  $F$ . ■

Veja-se a representação esquemática deste algoritmo na Figura 2.2 de Gentle (1998, p. 44).<sup>1</sup>

Este algoritmo pressupõe que se construa uma tabela com todos os valores possíveis da v.a. discreta  $X$  e os respectivos valores da f.d. de  $X$ , e se efectue uma pesquisa nessa tabela (*table lookup*) para identificar o quantil de ordem  $u$  de  $X$ . Esta pesquisa não é necessariamente expedita para algumas distribuições discretas, sugerindo pois a utilização de algoritmos alternativos, tais como os descritos em Gentle (1998, pp. 45–47) que se devem a:

- Marsaglia, Norman e Cannon;
- Chen e Asau.

**Exercício 2.4** — Escreva um programa numa linguagem de programação que lhe seja familiar para gerar 20 observações de cada uma das seguintes distribuições discretas:

- (a) *Bernoulli*( $p$ );
- (b) *Binomial*(10, 0.2);
- (c) *Geométrica*( $p$ );
- (d) *Poisson*(1).

Recorra novamente ao gerador congruencial multiplicativo com  $a = 13^{13}$ ,  $c = 0$ ,  $m = 2^{59}$  e  $u_0 = 1$  e tente apurar o tempo de execução de cada algoritmo para diferentes valores de  $u_i$ . ■

---

<sup>1</sup>Gentle (1998, p. 43) define equivalentemente este quantil de ordem  $u$ :  $x = F^{-1}(u) : F(x^-) < u \leq F(x)$ .

Coloca-se ainda um problema em relação à utilização do método da transformada inversa: a geração de vectores pseudo-aleatórios contínuos — que digam respeito ao vector aleatório  $\underline{X} = (X_1, \dots, X_d)$ , i.e., à **distribuição multivariada contínua** — está à partida comprometida. Há, no entanto, uma excepção quando a f.d. conjunta  $\underline{X}$  é passível da seguinte decomposição à custa de f.d.'s condicionais:

$$\begin{aligned} F_{\underline{X}}(\underline{x}) &= P(X_1 \leq x_1, \dots, X_d \leq x_d) \\ &= F_{X_1}(x_1) \times F_{X_2|X_1=x_1}(x_2) \\ &\quad \times \dots \times F_{X_d|X_1=x_1, \dots, X_{d-1}=x_{d-1}}(x_d). \end{aligned} \quad (2.11)$$

Nesta situação é possível obter os seguintes quantis de ordem  $u_1, \dots, u_d$ :

$$\begin{aligned} x_1 &= F_{X_1}^{-1}(u_1) \\ x_2 &= F_{X_2|X_1=x_1}^{-1}(u_2) \\ &\dots \\ x_d &= F_{X_d|X_1=x_1, \dots, X_{d-1}=x_{d-1}}^{-1}(u_d). \end{aligned} \quad (2.12)$$

Este método pode ser devidamente adaptado para a geração de vectores pseudo-aleatórios discretos como se poderá ver na subsecção seguinte.

**Texto de apoio:** Gentle (1998, pp. 41–47).

### 2.1.3 Alguns algoritmos de geração de números pseudo-aleatórios de algumas distribuições

É tirando de um modo geral partido de resultados da **Teoria das Probabilidades** que é possível adiantar algoritmos que permitem obter de uma forma mais expedita números pseudo-aleatórios de distribuições discretas e contínuas específicas.

Ao falar-se de **eficiência de algoritmos** refere-se à velocidade com que o algoritmo gera os números pseudo-aleatórios. E ao falar-se em velocidade é preciso ter em conta que esta depende também do número de constantes que é preciso obter e armazenar antes de iniciar a geração propriamente dita (*setup*) e do número de operações e valores da  $Uniforme(0, 1)$  que é necessário gerar para obter o valor da distribuição específica.

Tenha-se também presente que a geração de números pseudo-aleatórios de uma distribuição pode fazer-se à custa de um algoritmo para um conjunto de valores dos parâmetros (e.g.  $\alpha \in \mathbb{N}$  onde  $\alpha$  é o parâmetro de forma da distribuição Gama) e determinado valor da ordem do quantil (e.g.  $u > 0.95$ ) e à custa de outro algoritmo distinto para outros valores dos parâmetros (e.g.  $\alpha \notin \mathbb{N}$ ) e da ordem desse mesmo quantil (e.g.  $u \leq 0.95$ ).

A Tabela 2.1 possui a f.p. (f.d.p.) bem como outras características diversas distribuições discretas (contínuas) importantes.

A ela seguem-se exemplos de algoritmos de geração de números pseudo-aleatórios de algumas dessas distribuições.

- **Exponencial** (Gentle (1998, p. 92)) — Ao considerar-se que  $\lambda$  representa o inverso do parâmetro de escala da v.a. exponencial, deve proceder-se do seguinte modo:

1. Gerar um número pseudo-aleatório  $u$  da distribuição  $Uniforme(0, 1)$ .
2. Efectuar a atribuição

$$x = -\frac{\ln(u)}{\lambda}. \quad (2.13)$$

De notar que a aplicação do método da transformada inversa conduziria à atribuição  $x = -\frac{\ln(1-u)}{\lambda}$ . No entanto,  $U \stackrel{st}{=} 1 - U \sim$

Table 2.1: Algumas distribuições discretas e contínuas importantes.

$X$	$P(X = x)$	$E[X]$	$V[X]$
Uniforme ( $\{1, 2, \dots, n\}$ )	$1/n, x = 1, 2, \dots, n$	$(n + 1)/2$	$(n^2 - 1)/12$
Binomial ( $n, p$ )	$C_x^n p^x (1 - p)^{n-x}, x=0,1,2,\dots,n$	$np$	$np(1 - p)$
Geométrica ( $p$ )	$(1 - p)^{x-1}p, x = 1, 2, \dots$	$1/p$	$(1 - p)/p^2$
Binomial Negativa ( $r, p$ )	$C_{r-1}^{x-1} p^r (1 - p)^{x-r}, x = r, r + 1, \dots$	$r/p$	$r(1 - p)/p^2$
Poisson ( $\lambda$ )	$e^{-\lambda} \lambda^x / x!, x = 0, 1, \dots$	$\lambda$	$\lambda$

---

$X$	$f(x)$	$E[X]$	$V[X]$
Uniforme ( $a, b$ )	$\frac{1}{b-a}, a \leq x \leq b$	$(a + b)/2$	$(b - a)^2/12$
Exponencial ( $\lambda$ )	$\lambda e^{-\lambda x}, x \geq 0$	$1/\lambda$	$1/\lambda^2$
Gama ( $\alpha, \lambda$ )	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, x \geq 0$	$\alpha/\lambda$	$\alpha/\lambda^2$
Erlang ( $n, \lambda$ )	$\frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}, x \geq 0$	$n/\lambda$	$n/\lambda^2$
$\chi^2_{(\nu)}$	$\frac{(1/2)^{\nu/2}}{\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, x \geq 0$	$\nu$	$2\nu$
Beta ( $\alpha, \beta$ )	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1 - x)^{\beta-1}, 0 \leq x \leq 1$	$\nu$	$2\nu$
Normal ( $\mu, \sigma^2$ )	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty$	$\mu$	$\sigma^2$

Uniforme(0, 1). Assim, poupa-se uma operação aritmética (uma subtracção) ao efectuar-se a atribuição (2.13).

- **Gama** (parâmetro de forma  $\alpha$  inteiro; Gentle (1998, p.95)) — Quando  $\alpha = n \in \mathbb{N}$ , a v.a.  $X \sim Gama(n, \lambda)$  diz-se com distribuição de *Erlang*( $n, \lambda$ ) e pode escrever-se à custa da seguinte soma de  $n$  v.a.s  $X =_{st} \sum_{i=1}^n X_i$  onde  $X_i \sim_{i.i.d.} exponencial(\lambda)$ . Não surpreende pois que se proceda da seguinte forma para gerar números pseudo-aleatórios da distribuição *Gama*( $n, \lambda$ ):

1. Gerar  $n$  números pseudo-aleatórios independentes  $u_1, \dots, u_n$  da distribuição *Uniforme*(0, 1).

2. Efectuar a atribuição

$$x = - \sum_{i=1}^n \frac{\ln(u_i)}{\lambda}. \quad (2.14)$$

- **Normal** (Gentle (1998, pp. 88–92)) — Ao pretender-se números pseudo-aleatórios normalmente distribuídos pode recorrer-se a diversos métodos de geração com precisão e velocidade distintas. Estes reportam-se à geração de números pseudo-aleatórios da v.a.  $Z \sim Normal(0, 1)$ . Note-se, no entanto, que ao efectuar-se a transformação  $x = \sigma z + \mu$  obtém-se um número pseudo-aleatório referente à v.a.  $X \sim Normal(\mu, \sigma^2)$ .

(a) **Teorema do Limite Central (TLC)** — Ao invocar-se o *TLC* para  $U_i \sim_{i.i.d.} Uniforme(0, 1)$ ,  $i = 1, \dots, n$ , conclui-se que

$$Z = \frac{\sum_{i=1}^n U_i - n/2}{\sqrt{n/12}} \sim_a Normal(0, 1). \quad (2.15)$$

Coloca-se muito naturalmente a questão: Quão grande deverá ser  $n$  para que a qualidade dos números gerados seja razoável? Uma escolha conveniente é  $n = 12$ .

Assim sendo a geração de números pseudo-aleatórios da v.a.  $Z \sim Normal(0, 1)$  compreende os dois seguintes passos:

1. Gerar 12 números pseudo-aleatórios independentes  $u_1, \dots, u_{12}$  da distribuição *Uniforme*(0, 1).
2. Efectuar a atribuição

$$z = \sum_{i=1}^{12} u_i - 6. \quad (2.16)$$

De notar que os valores de  $z$  pertencem em qualquer dos casos ao intervalo  $(-6, +6)$ : estão pois excluídos os restantes

valores reais que por sinal ocorrem com probabilidade igual a  $P(|Z| \geq 6) \simeq 10^{-10}$ .

De salientar também que este método não é propriamente rápido (pois afinal são necessários 12 números pseudo-aleatórios uniformes para gerar um normal) e que existem alternativas bem mais expeditas para gerar um número pseudo-aleatório da distribuição *Normal*(0, 1).

(b) **Método de Box-Müller** — A aplicação deste método encontra justificação num resultado probabilístico deixado para o Exercício 2.5 e compreende também dois passos muito simples:

1. Gerar 2 números pseudo-aleatórios independentes  $u_1, u_2$  da distribuição *Uniforme*(0, 1).

2. Efectuar as atribuições

$$z_1 = \sqrt{-2 \ln(u_1)} \cos(2\pi u_2) \quad (2.17)$$

$$z_2 = \sqrt{-2 \ln(u_1)} \sin(2\pi u_2). \quad (2.18)$$

Gentle (1998, pp. 88–89) chama a atenção para o facto de este método conduzir a resultados não necessariamente razoáveis quando se lida com geradores congruenciais *pobres*.

De referir também que apesar da independência entre as v.a.s associadas às concretizações definidas em (2.17) e (2.18) há autores que demonstraram que  $z_1$  e  $z_2$  não só assentam numa espiral como é de questionar a sua independência quanto mais não seja porque estão associados a dois números pseudo-aleatórios relacionados recursivamente. Por este motivo tais autores sugerem que se utilize somente um dos valores gerados para a normal padrão.



Há outras possibilidades de geração de números pseudo-aleatórios normais. Uma delas é um método do tipo aceitação/rejeição, o método de **Polar–Marsaglia** que será descrito na subsecção seguinte.

**Exercício 2.5** — Prove que, caso  $U_i \sim_{i.i.d.} \text{Uniforme}(0, 1)$ ,  $i = 1, 2$ , se tem

$$Z_1 = \sqrt{-2 \ln(U_1)} \cos(2\pi U_2) \sim \text{Normal}(0, 1) \quad (2.19)$$

$$Z_2 = \sqrt{-2 \ln(U_1)} \sin(2\pi U_2) \sim \text{Normal}(0, 1), \quad (2.20)$$

com a particularidade de  $Z_1$  e  $Z_2$  serem v.a.s independentes. ■

**Exercício 2.6** — Admita que  $Z \sim \text{Exponencial}(1)$  e que a v.a. discreta  $S$  toma valores  $+1$  e  $-1$  com probabilidade  $1/2$  em qualquer dos casos.

(a) Prove que a v.a.  $X = SZ/\lambda$  possui f.d.p. igual a

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|}, \quad x \in \mathbb{R}, \quad (2.21)$$

i.e., possui distribuição de *Laplace* com parâmetro  $\lambda$  ( $\lambda > 0$ ) (ou distribuição *Dupla Exponencial*).

(b) Após ter definido os passos do algoritmo de geração de números pseudo-aleatórios da distribuição de *Laplace*, gere 500 observações desta distribuição para  $\lambda = 0.5, 1, 2$  e desenhe os respectivos histogramas. ■

Em Gentle (1998, pp. 93–99) pode encontrar-se a descrição de algoritmos para a geração de números pseudo-aleatórios de diversas distribuições contínuas:

- **Gama** (Gentle (1998, pp. 93–96)) — Recomenda-se o recurso aos algoritmos de Cheng-Feast (1979), de Best-Ahrens-Dieter (1983) e de Philippe (1997), respectivamente, na geração de números pseudo-aleatórios das distribuições *Gama* com parâmetro de forma maior e menor que 1 ou truncadas à direita.
- **Beta** (Gentle (1998, p. 97)) — Caso ambos os parâmetros da distribuição *Beta* sejam menores que 1 deve recorrer-se ao algoritmo de Jöhnk
- **Qui-quadrado** (Gentle (1998, pp. 98–99)) — O recurso ao facto da v.a. do qui-quadrado com  $n$  graus de liberdade resultar da soma de  $n$  quadrados de v.a.s i.i.d. normais padrão implica a geração de  $n$  números pseudo-aleatórios desta distribuição, pelo que é mais vantajoso tirar partido do seguinte resultado

$$X \sim \chi_{(n)}^2 \Leftrightarrow X \sim Gama(n/2, 1/2) \quad (2.22)$$

e recorrer aos algoritmos de geração de números pseudo-aleatórios da distribuição *Gama*.

- **F-snedcor** (Gentle (1998, p. 99)) — Tira-se partido do facto de

$$Y = \frac{\nu_2}{\nu_1} \frac{X}{1 - X} \sim F_{(\nu_1, \nu_2)} \quad (2.23)$$

caso  $X \sim Beta(\nu_1/2, \nu_2/2)$  para gerar números pseudo-aleatórios da distribuição  $F_{(\nu_1, \nu_2)}$ . Há ainda a possibilidade de a geração se fazer à custa dois números pseudo-aleatórios de distribuições do qui-quadrado já que

$$X = \frac{\frac{Y_1}{\nu_1}}{\frac{Y_2}{\nu_2}} \sim F_{(\nu_1, \nu_2)} \quad (2.24)$$

caso  $Y_1$  e  $Y_2$  sejam v.a.s independentes tais que  $Y_i \sim \chi_{(\nu_i)}^2$ ,  $i = 1, 2$ .

- **T-student** (Gentle (1998, pp.97-98)) — O algoritmo de Bailey para a geração de números pseudo-aleatórios da distribuição *T – student* usa o facto de a v.a.  $X \sim t_{(\nu)}$  corresponder à raiz quadrada da v.a.  $Y \sim F_{(1,\nu)}$
- **Weibull** (Gentle (1998, p.99)) — Neste caso basta recorrer ao método da transformada inversa.

**Exercício 2.7** — Considere que  $Y$  e  $Z$  são v.a.s independentes com distribuições  $Gama(\alpha, \lambda)$  e  $Gama(\beta, \lambda)$ , respectivamente ( $\alpha, \beta, \lambda > 0$ ).

- (a) Mostre que a v.a.  $X = Y/(Y + Z)$  possui distribuição  $Beta(\alpha, \beta)$ , i.e., a sua f.d.p. é igual a

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1}, 0 \leq x \leq 1. \quad (2.25)$$

- (b) Use este resultado para descrever um método de geração de números pseudo-aleatórios da distribuição  $Beta(\alpha, \beta)$  no caso em que ambos os parâmetros são inteiros maiores que 1.
- (c) Gere 1000 observações da distribuição  $Beta(4, 5)$  e desenhe o respectivo histograma. ■

**Exercício 2.8** — A v.a.  $X$  diz-se com distribuição Logística( $\mu, \sigma$ ) se possuir f.d.p. dada por:

$$f(x) = \frac{e^{-\frac{x-\mu}{\sigma}}}{\sigma \left(1 + e^{-\frac{x-\mu}{\sigma}}\right)^2}, -\infty < x < +\infty. \quad (2.26)$$

- (a) Defina um método para gerar números pseudo-aleatórios da distribuição Logística( $\mu, \sigma$ ).
- (b) Gere 1000 observações da distribuição *Logística* padrão ( $\mu = 0, \sigma = 1$ ) e desenhe o respectivo histograma. ■

Abaixo podem encontrar-se os algoritmos de geração de números pseudo-aleatórios de algumas distribuições discretas:

- **Bernoulli** (Gentle (1998, p. 47)).<sup>2</sup> — Para gerar números pseudo-aleatórios com distribuição de *Bernoulli*( $p$ ) basta recorrer ao método da transformada inversa:

1. Gerar um número pseudo-aleatório da distribuição *Uniforme*(0, 1).
2. Efectuar a atribuição

$$x = \begin{cases} 0, & \text{se } u \geq p \\ 1, & \text{se } u < p. \end{cases} \quad (2.27)$$

À semelhança do que se verificou na aplicação do método da transformada inversa para a distribuição exponencial opta-se por poupar uma operação aritmética ao efectuar-se a atribuição em (2.27) ao invés de

$$x = \begin{cases} 0, & \text{se } 1 - u \leq p \\ 1, & \text{se } 1 - u > p. \end{cases} \quad (2.28)$$

- **Binomial** (Gentle (1998, pp. 99–100)) — Escusado dizer que, para gerar um número pseudo-aleatório da distribuição *Binomial*( $n, p$ ), é necessário gerar  $n$  números pseudo-aleatórios da distribuição de *Bernoulli*( $p$ ) e de seguida contabilizar o número de sucessos. Assim a geração compreende os seguintes passos:

1. Gerar  $n$  números pseudo-aleatórios independentes  $u_1, \dots, u_n$  da distribuição *Uniforme*(0, 1).

---

<sup>2</sup>Ao que tudo indica o algoritmo em Gentle (1998, p. 47) não está correcto.

2. Efectuar as atribuições

$$y_i = \begin{cases} 0, & \text{se } u_i \geq p \\ 1, & \text{se } u_i < p, \end{cases} \quad (2.29)$$

para  $i = 1, \dots, n$ , e

$$x = \sum_{i=1}^n y_i. \quad (2.30)$$

Para mais detalhes acerca de outros procedimentos de geração consulte-se Gentle (1998, p. 100).

- **Poisson** (Gentle (1998, p. 100)) — Caso o valor esperado da v.a. de  $Poisson(\lambda)$  seja pequeno pode recorrer-se ao método da transformada inversa para gerar números pseudo-aleatórios.

Pode também recorrer-se ao seguinte algoritmo:

1. Gerar números pseudo-aleatórios  $u_1, u_2, \dots$  da distribuição Uniforme(0, 1).
2. Atribuir o valor

$$x = \inf\{m \in \mathbb{N}_0 : \sum_{i=1}^m [-\ln(u_i)] \leq \lambda\}. \quad (2.31)$$

- **Geométrica** (Gentle (1998, p. 101)) — Uma vez que existe expressão para o quantil de ordem  $u$  da distribuição Geométrica( $p$ ) o procedimento de geração de números pseudo-aleatórios compreende os seguintes passos:

1. Gerar um número pseudo-aleatório  $u$  da distribuição Uniforme(0, 1).
2. Efectuar a atribuição

$$\begin{aligned}
x &= \inf\{m \in \mathbb{R} : u \leq F(m)\} \\
&= \left\lfloor \frac{\ln(1-u)}{\ln(1-p)} \right\rfloor.
\end{aligned} \tag{2.32}$$

onde  $\lfloor y \rfloor$  representa a parte inteira do real  $y$ .

Escusado será dizer que pode poupar-se uma operação se se considerar  $u$  ao invés de  $(1-u)$  na Equação (2.32).

**Exercício 2.9** — Justifique o procedimento descrito para a geração de números pseudo-aleatórios da distribuição de *Poisson*, recorrendo, se entender necessário, a um *Processo de Poisson* com taxa unitária. Gere 1500 observações da distribuição de *Poisson*(10) e desenhe o respectivo histograma. ■

Segue-se a apresentação dos algoritmos de geração de números pseudo-aleatórios de duas v.a.s multivariadas:

- **Normal multivariada** (Gentle (1998, pp. 105–106)) — O vector aleatório  $\underline{X} = (X_1, \dots, X_d)$  diz-se com distribuição normal  $d$ -variada com valor esperado

$$\underline{\mu} = (\mu_1, \dots, \mu_d) \tag{2.33}$$

e matriz de covariância não singular

$$\underline{\Sigma} = [\sigma_{ij}]_{i,j=1,\dots,d} = [cov(X_i, X_j)]_{i,j=1,\dots,d} \tag{2.34}$$

— escrevendo-se neste caso

$$\underline{X} \sim N_d(\underline{\mu}, \underline{\Sigma}) \tag{2.35}$$

— caso  $\underline{X}$  possua f.d. conjunta

$$f(\underline{x}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\underline{\Sigma}|} \exp \left[ -\frac{1}{2} (\underline{x} - \underline{\mu})^\top \underline{\Sigma}^{-1} (\underline{x} - \underline{\mu}) \right], \tag{2.36}$$

para  $\underline{x} \in \mathbb{R}^d$ .

Tendo em conta que, caso  $Z_i \sim_{i.i.d.} Normal(0, 1)$ ,  $i = 1, \dots, d$ , e  $\underline{X} = \mathbf{T}^\top \underline{Z} + \underline{\mu}$  — onde  $\underline{Z} = (Z_1, \dots, Z_d)$  e  $\mathbf{T}$  é uma matriz  $d \times d$  que verifica  $\mathbf{T}^\top \mathbf{T} = \underline{\Sigma}$  —, se tem

$$\underline{X} \sim N_d(\underline{\mu}, \underline{\Sigma}), \quad (2.37)$$

não surpreende que um algoritmo para gerar um vector pseudo-aleatório da distribuição  $N_d(\underline{\mu}, \underline{\Sigma})$  compreenda os seguintes passos:

1. Gerar  $d$  números pseudo-aleatórios independentes  $z_1, \dots, z_d$  da distribuição  $Normal(0, 1)$ .
2. Efectuar a atribuição  $\underline{x} = \mathbf{T}^\top \underline{z} + \underline{\mu}$ .

Gentle (1998, p. 106) refere outros procedimentos de gerar números pseudo-aleatórios desta distribuição multivariada.

- **Multinomial** (Gentle (1998, p. 106)) — O vector aleatório  $\underline{X} = (X_1, \dots, X_d)$  diz-se com distribuição multinomial  $(d-1)$ -variada com parâmetros  $n$  e  $\underline{p} = (p_1, \dots, p_d)$ , onde  $\sum_{i=1}^d X_i = n$  e  $\underline{p} \in \{(p_1, \dots, p_d) \in [0, 1]^d : \sum_{i=1}^d p_i = 1\}$  — escrevendo-se neste caso

$$\underline{X} \sim Multinomial_{d-1}(n, \underline{p}) \quad (2.38)$$

— caso  $\underline{X}$  possua f.p. conjunta igual a

$$P(\underline{X} = \underline{x}) = \frac{n!}{\prod_{i=1}^d x_i!} \prod_{i=1}^d p_i^{x_i} \quad (2.39)$$

para  $\underline{x} \in \{(n_1, \dots, n_d) \in \{0, 1, \dots, n\}^d : \sum_{i=1}^d n_i = n\}$ .

Admita-se que  $\underline{X}$  possui distribuição multinomial com parâmetros  $n$  e vector de probabilidades  $\underline{p} = (p_1, \dots, p_d)$ .

A leitura de Gentle (1998, p.106) sugere que a geração de vectores pseudo-aleatórios desta distribuição passa por tirar-se partido do facto de as distribuições marginais serem binomiais dependentes ( $cov(X_i, X_j) = -p_i p_j$ ) e das distribuições condicionais serem iguais a:

$$X_1 \sim \text{Binomial}(n, p_1) \quad (2.40)$$

$$(X_2|X_1 = x_1) \sim \text{Binomial}\left(n - x_1, \frac{p_2}{1 - p_1}\right) \quad (2.41)$$

$$(X_3|X_1 = x_1, X_2 = x_2) \\ \sim \text{Binomial}\left(n - x_1 - x_2, \frac{p_3}{1 - p_1 - p_2}\right) \quad (2.42)$$

...

$$(X_{d-1}|X_1 = x_1, \dots, X_{d-2} = x_{d-2}) \\ \sim \text{Binomial}\left(n - \sum_{i=1}^{d-2} x_i, \frac{p_{d-1}}{1 - \sum_{i=1}^{d-2} p_i}\right) \quad (2.43)$$

$$(X_d|X_1 = x_1, \dots, X_{d-1} = x_{d-1}) = n - \sum_{i=1}^{d-1} x_i. \quad (2.44)$$

A geração faz-se sequencialmente à custa destas distribuições condicionais, tal como descrito no final da Subsecção 2.1.2. E, por forma a acelerar a geração destes números pseudo-aleatórios, recomenda-se que a geração seja feita sequencialmente por ordem decrescente das probabilidades de sucesso  $p_1, \dots, p_d$ , i.e., começar pela componente com maior probabilidade de sucesso e terminar com a componente com a menor delas.

Em Gentle (1998, p.106) são referidas outras formas de obter vectores pseudo-aleatórios desta mesma distribuição.



**Exercício 2.10** — Faça um levantamento das rotinas/métodos de geração de números pseudo-aleatórios de distribuições discretas e contínuas dos *packages* *BMDP*, *Maple*, *Matlab*, *Mathematica*, *NAG*, *R*, *SAS*, *SPSS*, *Statistica* e outros *packages* com que esteja familiarizado. ■

O capítulo 7 de Gentle (1998) faz um apanhado geral sobre os cuidados a ter na geração de números pseudo-aleatórios (como é o caso do controlo da semente e do método de geração).

Este mesmo capítulo acrescenta ainda algumas notas sobre a geração de números pseudo-aleatórios com o *software* *S-Plus*.

Por seu lado, Morgan(1984, pp. 236–249) faz o mesmo com o *software* *NAG*.

**Textos de apoio:** Gentle (1998, p.47 e pp. 85–106).

#### 2.1.4 Métodos de aceitação/rejeição

Os métodos de aceitação/rejeição são caracterizados por gerarem números pseudo-aleatórios de uma v.a.  $X$  com f.d.p.  $f(x)$  (directa ou indirectamente) à custa de números pseudo-aleatórios de uma outra v.a.  $Y$  com f.d.p.  $g(x)$  *similar* à de  $X$ , números estes de muito mais fácil geração que os respeitantes à distribuição de  $X$ .

Descrever-se-ão nesta subsecção dois métodos de aceitação/ rejeição.

**Método de aceitação/rejeição simples** — Sejam:

- $X$  uma v.a. que possui f.d.p.  $f(x)$  e contradomínio limitado e igual a  $[a, b]$ ;
- $c$  uma constante que verifique  $f(x) \leq c, x \in [a, b]$ .

A geração de números pseudo-aleatórios da distribuição de  $X$  passa por:

1. Gerar um ponto  $(u, v)$  no rectângulo de vértices  $(a, 0)$ ,  $(a, c)$ ,  $(b, 0)$  e  $(b, c)$ , gerando para o efeito dois números pseudo-aleatórios  $u_1$  e  $u_2$  da distribuição  $Uniforme(0, 1)$  e efectuando as atribuições  $u = a + (b - a)u_1$  e  $v = cu_2$ .
2. Caso  $v \leq f(u)$ , efectuar a atribuição  $x = u$ ; caso contrário, voltar ao Passo 1. (Esquema...) ■

A **probabilidade de aceitação** (do número pseudo-aleatório  $v$ ) no Passo 2 do **método de aceitação/rejeição simples** é igual ao quociente entre a área sob  $f(x)$  e sob o rectângulo acima referido, i.e.,

$$P(\text{aceitação método aceit./rej. simples}) = \frac{1}{c(b - a)}. \quad (2.45)$$

É pois fundamental que se escolha o valor de  $c$  o mais pequeno possível por forma a maximizar a probabilidade de aceitação neste método — a escolha mais razoável para  $c$  é sem sombra de dúvida

$$c = \max_{a \leq x \leq b} f(x). \quad (2.46)$$

**Exercício 2.11** — Gere 1000 números pseudo-aleatórios da distribuição  $Beta(4, 5)$  recorrendo ao método de aceitação/rejeição simples.

- (a) Obtenha a probabilidade de aceitação deste algoritmo.
- (b) Compare a velocidade de execução deste algoritmo com a daquele que utilizou para resolver a alínea (c) do Exercício 2.7. ■

Importa salientar que o método de aceitação/rejeição simples não pode ser utilizado quando  $X$  possui contradomínio não limitado e que a probabilidade de rejeição é considerável caso a f.d.p. de  $X$  seja *point-aguda*. (Esquema...)

As desvantagens acima referidas não se colocam em relação ao método que se descreve já de seguida.

**Método de aceitação/rejeição geral** — Sejam:

- $X$  uma v.a. que possui f.d.p.  $f(x)$ ;
- $Y$  uma v.a. com f.d.p.  $g(x)$  e cujos números pseudo-aleatórios são mais fáceis de gerar;
- $k$  uma constante maior que 1 e que verifique  $f(x) \leq k g(x)$ ,  $-\infty < x < +\infty$ , i.e.,  $k g(x)$  majora  $f(x)$ .<sup>3</sup>

A geração de números pseudo-aleatórios da distribuição de  $X$  passa por:

1. Gerar um número pseudo-aleatório  $y$  da distribuição de  $Y$ .
2. Gerar um número pseudo-aleatório  $u$  da distribuição de  $Uniforme(0, 1)$ .
3. Caso  $u \leq \frac{f(y)}{k g(y)}$ , efectuar a atribuição  $x = y$ ; caso contrário, voltar ao Passo 1. ■

A **probabilidade de aceitação** (do número pseudo-aleatório  $u$ ) no Passo 3 do **método de aceitação/rejeição geral** é igual a

$$P(\text{aceitação método aceit./rej. geral}) = \frac{1}{k}. \quad (2.47)$$

e deve ser maximizada a todo o custo. Assim, uma vez fixa a função  $g(x)$ , deve escolher-se

$$k = \max_{-\infty < x < +\infty} \frac{f(x)}{g(x)}. \quad (2.48)$$

**Exercício 2.12** — Prove que os números pseudo-aleatórios gerados pelos métodos de aceitação/rejeição simples e geral possuem de facto f.d.p.  $f(x)$ . ■

---

<sup>3</sup> $k g(x)$  é o que na literatura anglófona se designa por *envelope function*.

**Exercício 2.13** — O método de aceitação/rejeição simples é um caso particular do método de aceitação/rejeição geral. Identifique-o e reescreva os passos daquele método. ■

**Exercício 2.14** — Considere a geração de números pseudo-aleatórios da distribuição de  $Normal(0, 1)$  usando um método de aceitação/rejeição baseado na f.d.p. logística

$$g(x) = \frac{\rho e^{-\rho x}}{(1 + e^{-\rho x})^2}, \quad -\infty < x < +\infty, \quad (2.49)$$

onde  $\rho > 0$ .

Suponha que  $0 < \rho \leq \sqrt{2}$  e considere que  $\phi(x)$  representa a f.d.p. da  $Normal(0, 1)$ .

- (a) Mostre que  $\phi(x)/g(x)$  possui máximo em  $x = 0$ .
- (b) Descreva um método de geração de números pseudo-aleatórios da distribuição de  $Normal(0, 1)$  que minimize a probabilidade de rejeição. ■

**Exercício 2.15** — Admita que se pretende gerar números pseudo-aleatórios da distribuição de  $Semi - Normal(1)$ , i.e., da v.a.  $X$  com f.d.p.

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = 2\phi(x), \quad x > 0. \quad (2.50)$$

Considere a v.a.  $Y \sim Exponencial(1)$  cuja f.d.p. é igual a  $g(x) = e^{-x}, x > 0$ .

- (a) Mostre que  $f(x)/g(x)$  possui máximo em  $x = 1$ , identifique tal valor máximo e determine a probabilidade de aceitação no Passo 3 do método de aceitação/rejeição geral.

- (b) Haveria vantagem em definir este método à custa de uma distribuição *Exponencial* com outros valores do parâmetro de escala?
- (c) Use o método de aceitação/rejeição geral baseado na distribuição de  $Y$  para gerar 1000 números pseudo-aleatórios da distribuição de  $X$ .
- (d) Obtenha 1000 números pseudo-aleatórios da distribuição  $Normal(0, 1)$  à custa dos obtidos na alínea anterior. ■

Para a descrição de variantes do método de aceitação/rejeição geral bem como da aplicação do mesmo à geração de números pseudo-aleatórios referentes a distribuições discretas é favor referir-se a Gentle (1998, pp. 49–55).

**Texto de apoio:** Gentle (1998, pp. 47–55).

## 2.2 Método de Monte Carlo

O método de solução numérica de problemas denominado de método de Monte Carlo baseia-se essencialmente na **simulação de variáveis aleatórias**. Entre tais problemas contam-se a estimação do integral definido simples

$$\theta = \int_a^b h(x)dx \quad (2.51)$$

ou do integral definido múltiplo

$$\theta_p = \int_{x_p} h(\underline{x})d\underline{x} = \int_{x_p} h(x_1, \dots, x_p)dx_1 \dots dx_p, \quad (2.52)$$

onde  $A \subset \mathbb{R}^p$  ( $p > 1$ ).

A gênese do método é marcada pela publicação, em 1949, de um trabalho da autoria de Metropolis e Ulam. A existência deste método muito se deve aos esforços de von Neumann. O seu nome deve-se à capital do principado do Mónaco, célebre pelo seu casino onde se pode encontrar um dos dispositivos mecânicos mais simples de geração de números pseudo-aleatórios, a roleta.

Só com o advento dos computadores (electrónicos) se difundiu amplamente o método de Monte Carlo.

Descrever-se-ão de seguida dois métodos de Monte Carlo, cujo objectivo é a estimação do integral simples definido em (2.51), facilmente generalizáveis para o cálculo do integral definido múltiplo em (2.52).

**Texto de apoio:** Sobol (1983, p. 7).

### 2.2.1 Monte de Carlo hit-or-miss

Recorrer-se-á a um exemplo bastante simples para **motivar** o método de **Monte Carlo hit-or-miss** para o cálculo de uma estimativa de  $\theta$ . Suponha-se que se pretende determinar a área de uma **figura plana**  $S$ , com um aspecto arbitrário e não necessariamente conexa. Sem perda de generalidade, admita-se que  $S$  está contida num quadrado de área unitária.

Admita-se também que são gerados  $n$  pontos pseudo-aleatórios nesse quadrado dos quais somente  $r$  ( $r = 0, 1, \dots, n$ ) pertencem a  $S$ . Uma **estimativa** (intuitivamente) **razoável** para a **área de  $S$**  é o quociente  $r/n$ , concretização do estimador  $R/n$ . A precisão desta estimativa aumenta, naturalmente, com o valor de  $n$ .

**Exercício 2.16** — Qual a distribuição, valor esperado e variância do estimador  $R/n$ ? ■

O algoritmo acima descrito pode ser imediatamente generalizado para o cálculo do integral definido  $\theta = \int_a^b h(x)dx$ , onde o intervalo  $[a, b]$  é limitado.

**Método de Monte Carlo hit-or-miss** — Comece-se por considerar uma constante  $c$  que verifique  $h(x) \leq c, x \in [a, b]$ . A obtenção de uma estimativa de  $\theta$  pressupõe:

1. Gerar  $n$  pontos pseudo-aleatórios  $(u_1, v_1), \dots, (u_n, v_n)$  no rectângulo de vértices  $(a, 0), (a, c), (b, 0)$  e  $(b, c)$ .<sup>4</sup>
2. Contabilizar o número de pontos  $r$  abaixo da curva de  $h(x)$ .

---

<sup>4</sup>Relembre-se que a geração do ponto  $(u, v)$  no referido rectângulo passa pela geração de dois números pseudo-aleatórios  $z_1$  e  $z_2$  da distribuição *Uniforme*(0, 1) e pela efectuação das atribuições  $u = a + (b - a)z_1$  e  $v = cz_2$ .

3. Obter a estimativa de  $\theta$

$$\hat{\theta}_{hit} = \frac{rc(b-a)}{n}. \quad (2.53)$$

■

Escusado será referir que o valor  $r$  nos passos 2 e 3 do algoritmo não passa do valor observado da v.a.  $R \sim Binomial(n, p = \theta/[c(b-a)])$ .

Este algoritmo é particularmente simples já que pressupõe a geração de pontos pseudo-aleatórios, seguida de averiguação do conjunto a que eles pertencem e do cálculo de um quociente.

Este método faz também lembrar o método de aceitação/rejeição simples.

**Exercício 2.17** — Considere a estimação de  $\theta = \int_a^b h(x)dx$ .

(a) Identifique a distribuição, o valor esperado e a variância do estimador da probabilidade de um ponto pseudo-aleatório estar sob a curva de  $h(x)$ . Tratar-se-á de um estimador centrado de  $\theta$ ?

(b) Justifique a estimativa para  $\theta$ ,  $\hat{\theta} = rc(b-a)/n$ . ■

Embora se saiba o resultado do integral mencionado já a seguir, o exercício seguinte servirá para ilustrar o método de Monte Carlo *hit-and-miss* assim como para calcular o verdadeiro valor da variância do estimador de  $\theta$ .

**Exercício 2.18** — Considere

$$\theta = \int_0^1 \sqrt{1-x^2} dx. \quad (2.54)$$

(a) Elabore o gráfico da função  $\sqrt{1-x^2}$  no intervalo  $[0, 1]$ , identifique a área a que corresponde  $\theta$  e o verdadeiro valor desta constante. (Morgan (1984, p. 163).)



- (b) Identifique o estimador de  $\theta$ ,  $\hat{\Theta}$ , e determine o verdadeiro valor esperado e a variância deste estimador.
- (c) Estime  $\theta$  recorrendo ao método de Monte de Carlo *hit-and-miss* gerando para o efeito 500 pontos pseudo-aleatórios. ■

Importa ainda referir que o método de Monte Carlo *hit-and-miss* não pode ser aplicado quando o conjunto  $[a, b]$  não é limitado, ao contrário do método de Monte Carlo descrito seguidamente.

**Textos de apoio:** Morgan (1984, p. 163) e Sobol (1983, pp. 7–10).

### 2.2.2 Monte de Carlo ordinário

Quando é possível factorizar a função  $h(x)$  do seguinte modo

$$h(x) = g(x)f(x), \quad (2.55)$$

onde  $f(x)$  é a f.d.p. da v.a.  $X$  definida no intervalo  $[a, b]$ , passa a ter-se

$$\begin{aligned} \theta &= \int_a^b g(x)f(x)dx \\ &= E[g(X)]. \end{aligned} \quad (2.56)$$

Deste modo o problema da **avaliação do integral definido simples** transforma-se num problema de **estimação** bastante familiar: a estimação do **valor esperado** da v.a.  $g(X)$ .

Caso se considere que  $X \sim \text{Uniforme}(a, b)$ , tem-se:

$$f(x) = \frac{1}{b-a}, a \leq x \leq b; \quad (2.57)$$

$$g(x) = (b-a)h(x), a \leq x \leq b; \quad (2.58)$$

$$\theta = (b-a)E[h(X)]. \quad (2.59)$$

Uma estimativa razoável e centrada para  $\theta$  é sem sombra de dúvida

$$\hat{\theta} = (b-a) \frac{\sum_{i=1}^n h(x_i)}{n} \quad (2.60)$$

onde  $\underline{x} = (x_1, \dots, x_n)$  é uma amostra de números pseudo-aleatórios da distribuição  $Uniforme(a, b)$ .

Há a assinalar uma vantagem óbvia deste método de Monte Carlo neste caso particular, quando comparado com o método de Monte Carlo *hit-and-miss*: naquele método são necessários somente  $n$  números pseudo-aleatórios da distribuição  $Uniforme(a, b)$  contra os  $2n$  números pseudo-aleatórios imprescindíveis na geração dos  $n$  pontos pseudo-aleatórios do método de Monte Carlo *hit-and-miss*.

**Exercício 2.19** — Considere a estimação de  $\theta = \int_a^b h(x)dx$  pelo método de Monte Carlo ordinário da distribuição  $X \sim Uniforme(a, b)$ . Obtenha o valor esperado e a variância deste estimador de  $\theta$ . Será um estimador centrado de  $\theta$ ? ■

Uma vez feita a motivação do método de Monte Carlo ordinário (*crude Monte Carlo method*) é a vez de apresentar o algoritmo.

**Método de Monte Carlo ordinário** — Considere-se que

$$\theta = \int_a^b h(x)dx \quad (2.61)$$

e que se verifica

$$h(x) = g(x)f(x), x \in [a, b], \quad (2.62)$$

onde  $f(x)$  é a f.d.p. da v.a.  $X$  definida no intervalo  $[a, b]$ .

Então a determinação de uma estimativa de  $\theta$  compreende os seguintes passos:

1. Gerar  $n$  números pseudo-aleatórios  $x_1, \dots, x_n$  da distribuição de  $X$ .
2. Efectuar a atribuição

$$\hat{\theta}_{ord} = \frac{\sum_{i=1}^n g(x_i)}{n}. \quad (2.63)$$

■

De notar que no passo 1 do algoritmo  $x_1, \dots, x_n$  são as concretizações das v.a.s  $X_i \sim_{i.i.d.} X, i = 1, \dots, n$ .

**Exercício 2.20** — Retome o Exercício 2.18.

- (a) Estime  $\theta = \int_0^1 \sqrt{1-x^2} dx$  recorrendo ao método de Monte Carlo ordinário fazendo uso de  $2 \times 500$  números pseudo-aleatórios da distribuição  $X \sim Uniforme(0, 1)$ .
- (b) Determine o verdadeiro valor esperado e a variância do estimador de  $\theta = \int_0^1 \sqrt{1-x^2} dx$  associado ao método de Monte Carlo ordinário descrito em (a) e compare-os com os correspondentes valores obtidos na alínea (b) do Exercício 2.18. ■

**Exercício 2.21** — Considere a estimação do integral definido

$$\theta = \int_1^{+\infty} \frac{e^{-x}}{x} dx. \quad (2.64)$$

- (a) Após ter efectuado a transformação de variável conveniente recorra ao método de Monte Carlo ordinário usando 1000 números pseudo-aleatórios da distribuição  $X \sim Exponencial(1)$ .
- (b) Compare o valor obtido em (a) com o valor numérico obtido pelo *Mathematica*. ■

**Texto de apoio:** Gentle (1998, pp. 131–133) e Morgan (1984, pp. 163–165).

### 2.2.3 Técnicas de redução de variância de estimadores de Monte Carlo

Na secção anterior já foram feitas algumas comparações e considerações sobre a variância dos estimadores de Monte Carlo à laia de exercício.

Pode adiantar-se, a título de exemplo, que os dois estimadores de  $\theta = \int_a^b h(x)dx$  são centrados e possuem variâncias iguais a:

- **método de Monte Carlo hit-and-miss**

$$\begin{aligned} V(\hat{\Theta}_{hit}) &= V\left[\frac{1}{n}Rc(b-a)\right] \\ &= \frac{\theta[c(b-a) - \theta]}{n} \end{aligned} \quad (2.65)$$

onde  $n$  representa o número de pontos pseudo-aleatórios (i.e., é necessária a geração de  $2n$  números pseudo-aleatórios);

- **método de Monte Carlo ordinário**

$$\begin{aligned} V(\hat{\Theta}_{ord}) &= V\left[\frac{1}{n}\sum_{i=1}^n g(X_i)\right] \\ &= \frac{V[g(X)]}{n} \\ &= \frac{E[g^2(X)] - E^2[g(X)]}{n} \\ &= \frac{\int_a^b g^2(x)f(x)dx - \theta^2}{n} \end{aligned} \quad (2.66)$$

onde  $n$  corresponde ao total de números pseudo-aleatórios gerados.

Como em qualquer problema de estimação deve procurar-se reduzir a variância do estimador (que pode passar pelo aumento do valor de  $n$ ) e, simultaneamente, preservar outras propriedades do estimador como ser centrado.

A técnica de redução de variância motivada e descrita de seguida denomina-se de **técnica das variáveis antitéticas**.

Por forma a motivar este método comece-se por considerar um par de v.a.s identicamente distribuídas,  $Y_1$  e  $Y_2$ , com a particularidade

de serem dependentes e terem valor esperado comum e igual a  $\theta = \int_a^b g(x)f(x)dx$ . Neste caso pode concluir-se que  $\frac{Y_1+Y_2}{2}$  é estimador de  $\theta$  verificando:

$$E\left(\frac{Y_1 + Y_2}{2}\right) = \theta; \quad (2.67)$$

$$\begin{aligned} V\left(\frac{Y_1 + Y_2}{2}\right) &= \frac{1}{4} \times [V(Y_1) + V(Y_2) + 2cov(Y_1, Y_2)] \\ &= \frac{1}{2} V(Y_1)[1 + corr(Y_1, Y_2)]. \end{aligned} \quad (2.68)$$

Ora, pode reduzir-se a variância deste estimador, caso  $Y_1$  e  $Y_2$  sejam escolhidas de tal modo que se correlacionem negativamente. E, com efeito, tal ocorre caso se considere:

- $X$  uma v.a. com f.d.p.  $f(x)$  e f.d.  $F(x)$ ;
- $\theta = \int_a^b g(x)f(x)dx$ , onde  $g(x)$  é uma função monótona;
- $U \sim Uniforme(0, 1)$ ;
- $Y_1 = g[F^{-1}(U)]$  e  $Y_2 = g[F^{-1}(1 - U)]$ .

Importa notar que  $U$  e  $1 - U$  dizem-se variáveis antitéticas.

**Método de Monte Carlo ordinário** (variáveis antitéticas) — Estes resultados sugerem uma nova versão do método de Monte Carlo ordinário para a estimação de  $\theta$  compreendendo os seguintes passos:

1. Gerar  $n/2$  números pseudo-aleatórios  $u_1, \dots, u_{n/2}$  da distribuição  $U \sim Uniforme(0, 1)$ .
2. Efectuar as atribuições

$$y_{i1} = g[F^{-1}(u_i)] \quad (2.69)$$

$$y_{i2} = g[F^{-1}(1 - u_i)], \quad (2.70)$$

para  $i = 1, \dots, n/2$ .

### 3. Efectuar a atribuição

$$\begin{aligned}\hat{\theta}_{anti} &= \frac{\sum_{i=1}^{n/2} \frac{y_{i1} + y_{i2}}{2}}{n/2} \\ &= \frac{1}{n} \sum_{i=1}^{n/2} (y_{i1} + y_{i2}).\end{aligned}\tag{2.71}$$

Esta versão do método possui a vantagem de reduzir para metade a quantidade de números pseudo-aleatórios que é necessário gerar. ■

**Exercício 2.22** — Retome a estimação de  $\theta = \int_0^1 \sqrt{1-x^2} dx$ , onde  $g(x) = \sqrt{1-x^2}$  e  $f(x) = 1, 0 \leq x \leq 1$  (i.e.,  $X \sim Uniforme(0, 1)$ ).

- (a) Obtenha uma estimativa pontual de  $\theta$  recorrendo ao método de Monte de Carlo ordinário fazendo uso de 500 números pseudo-aleatórios da distribuição  $U \sim Uniforme(0, 1)$  (i.e.,  $n = 1000$ ) e de variáveis antitéticas.
- (b) Determine o verdadeiro valor da variância deste novo estimador de  $\theta$  e compare-o com o verdadeiro valor da variância do estimador de *theta* baseado no método de Monte Carlo ordinário com recurso a 1000 números pseudo-aleatórios da distribuição  $X \sim Uniforme(0, 1)$ . ■

**Exercício 2.23** — Repita o Exercício 2.21 considerando desta feita o método de Monte de Carlo ordinário que faz uso de 500 números pseudo-aleatórios da distribuição  $U \sim Uniforme(0, 1)$  e de variáveis antitéticas.

Compare estes resultados com os obtidos no Exercício 2.21. ■

**Exercício 2.24** — Considere que

$$\theta = \int_0^1 e^{-x^2} dx.\tag{2.72}$$

- (a) Estime  $\theta$  pelo método de Monte Carlo *hit-and-miss*. Recorra para o efeito a 1000 pontos pseudo-aleatórios da distribuição  $Uniforme(0, 1)$ .
- (b) Estime  $\theta$  pelo método de Monte Carlo ordinário. Use 2000 números pseudo-aleatórios da distribuição  $Uniforme(0, 1)$ .
- (c) Obtenha uma estimativa de  $\theta$  usando agora o método de Monte Carlo ordinário com recurso a variáveis antitéticas e aos 1000 primeiros números pseudo-aleatórios obtidos em (b).
- (d) Obtenha intervalos de confiança aproximados a 95% para  $\theta$  baseados nos métodos utilizados em (a)–(c). Compare as diferentes estimativas pontuais e intervalares de  $\theta$ . ■

Há outras possibilidades de redução de variância: é caso da **redução da variância por condicionamento**, que por sinal não é descrita em Gentle (1998).

Considere-se que  $Y = g(X)$  e que é possível identificar uma v.a.  $Z$  que permita a definição da v.a.  $E(Y|Z)$  e retome-se mais uma vez a estimação de  $\theta = E[g(X)] = E(Y)$ . Então note-se que

$$V(Y) = V[E(Y|Z)] + E[V(Y|Z)] \quad (2.73)$$

donde se conclui que

$$V[E(Y|Z)] \leq V(Y) \quad (2.74)$$

o que sugere que ao invés de se usar a média do valores gerados  $y_i = g(x_i)$  (i.e. a média  $\frac{1}{n} \sum_{i=1}^n g(x_i) = \frac{1}{n} \sum_{i=1}^n y_i$ ), se recorra à média dos valores gerados da v.a.  $E(Y|Z)$ .

Deste modo a estimativa de  $\theta$  é igual a

$$\hat{\theta}_{cond} = \frac{1}{n} \sum_{i=1}^n E(Y|Z = z_i). \quad (2.75)$$

**Exercício 2.25** — Estime  $\theta = E(X)$  — onde a v.a.  $X$  é tal que

- $X|Z = z \sim \text{Beta}(z, z^2 + 1)$
- $Z \sim \text{Poisson}(\lambda)$

(Morgan (1984, p. 173)) —, com base em 1000 números pseudo-aleatórios e considerando  $\lambda = 2.5$  ao método de Monte Carlo ordinário usando a técnica de redução da variância por condicionamento. ■

Podem encontrar-se outras técnicas de redução de variância em Gentle (1998, pp. 135–139) e Morgan (1984, p. 166–180).

**Texto de apoio:** Gentle (1998, pp. 135–139) e Morgan (1984, pp. 160–180).

#### 2.2.4 Técnicas de redução de variância na simulação do sistema M/M/1

Um sistema  $M/M/1$  é caracterizado por:

- tempos entre chegadas consecutivas i.i.d. com distribuição *Exponencial*( $\lambda$ );
- tempos de serviço, também eles i.i.d. com distribuição comum *Exponencial*( $\mu$ ) e independentes dos tempos entre chegadas;
- um único servidor que atende os clientes por ordem de chegada (política *First Come First Served*).

De notar que o **tempo de permanência no sistema** (*waiting time in system*) corresponde ao tempo que decorre desde a chegada do cliente ao sistema até que ele o abandone, i.e., corresponde ao tempo que o cliente dispende na fila de espera, aguardando pelo início do respectivo



serviço, ao qual é adicionado o tempo que o servidor dispense a servir esse mesmo cliente.

Apesar de a **distribuição** do tempo de permanência no sistema ser **conhecida** e *Exponencial*( $\mu - \lambda$ ),  $0 > \mu > \lambda$ , recorreremos à simulação para estimar o valor esperado do tempo de permanência no sistema — doravante representado por  $W^{(s)}$  — essencialmente para ilustrar duas técnicas de redução da variância.

Um estimador mais que óbvio para  $E[W^{(s)}]$  é  $\bar{W}^{(s)} = \frac{1}{n} \sum_{i=1}^n W_i^{(s)}$ , a média dos tempos de permanência no sistema de  $n$  clientes.

### **Simulação de tempos de permanência no sistema M/M/1 —**

De modo a efectuar a simulação desta média é essencial considerar as seguintes v.a.s, onde  $j = 1, \dots, n$ :

- $I_j$  representa o tempo decorrido entre a chegada do  $(j - 1)$ –ésimo e do  $j$ –ésimo ( $j > 2$ ) clientes e  $I_1$  (*interarrival time*) o instante de chegada do primeiro cliente;
- $S_j$  representa o tempo de serviço do  $j$ –ésimo cliente (*service time*);
- $W_j$  representa o tempo de permanência no sistema (*waiting time in system*).

Para além disso é necessário:

1. Gerar  $n$  números pseudo-aleatórios independentes  $i_1, \dots, i_n$ , referentes ao tempo entre chegadas consecutivas (logo com distribuição *Exponencial*( $\lambda$ )).<sup>5</sup>
2. Gerar  $n$  números pseudo-aleatórios independentes  $s_1, \dots, s_n$ , dizendo respeito aos tempos de serviço com distribuição *Exponencial*

---

<sup>5</sup>Esta geração far-se-á, é claro, à custa de  $n$  números pseudo-aleatórios independentes  $u_1, \dots, u_n$  da distribuição *Uniforme*(0, 1).

$(\mu)$ .<sup>6</sup>

3. Efectuar a seguinte atribuição, para  $j = 1, \dots, n$ ,

$$w_j = \begin{cases} s_i, & \text{se } w_j \leq i_j \text{ (cliente imediatamente atendido)} \\ w_{j-1} - i_j + s_j, & \text{se } w_{j-1} > i_j \text{ (caso contrário).} \end{cases} \quad (2.76)$$

Por forma a estimar  $V(\bar{W}^{(s)})$  deve repetir-se as simulações acima  $r$  vezes e a estimativa de tal variância mais não será que a variância da amostra de médias de conjuntos de  $n$  tempos de permanência no sistema. ■

**Simulação de tempos de permanência no sistema M/M/1 (variáveis antitéticas)** — A estimativa de  $V(\bar{W}^{(s)})$  pode ser reduzida recorrendo à técnica das variáveis antitéticas que compreenderá o seguintes passos:

1. Gerar  $w_1, \dots, w_n$  tal como se descreve no algoritmo anterior e muito em particular recorrendo à Equação (2.76).
2. Gerar  $w_1^*, \dots, w_n^*$  tal como no Passo 1 mas usando os números pseudo-aleatórios antitéticos de  $u_1, \dots, u_n$  e  $v_1, \dots, v_n$ , i.e.,  $1 - u_1, \dots, 1 - u_n$  e  $1 - v_1, \dots, 1 - v_n$ . Assim se obtém a média de  $2n$  tempos de permanência no sistema.
3. Repetir os Passos 1 e 2 até registar  $r/2$  amostras de  $2n$  tempos de permanência no sistema. ■

**Simulação de tempos de permanência no sistema M/M/1 (terceiro algoritmo)** — Há ainda outra possibilidade de redução da estimativa de  $V(\bar{W}^{(s)})$ . Ela passa por:

---

<sup>6</sup>Analogamente, usam-se  $n$  números pseudo-aleatórios independentes  $v_1, \dots, v_n$  da distribuição *Uniforme*(0, 1).

1. Gerar dois números pseudo-aleatórios independentes  $u_1$  e  $u_2$  da distribuição  $Uniforme(0, 1)$ .
2. Gerar dois números pseudo-aleatórios independentes  $i$  e  $s$  das distribuições  $Exponencial(\lambda)$  e  $Exponencial(\mu)$  (respectivamente), efectuando para isso as atribuições:

$$\begin{aligned} i &= -\frac{\ln(u_1)}{\lambda}; \\ s &= -\frac{\ln(u_2)}{\mu}. \end{aligned} \tag{2.77}$$

3. Gerar outros dois números pseudo-aleatórios  $i^*$  e  $s^*$  das distribuições  $Exponencial(\lambda)$  e  $Exponencial(\mu)$ , considerando para tal as atribuições:

$$\begin{aligned} i^* &= -\frac{\ln(u_2)}{\lambda}; \\ s^* &= -\frac{\ln(u_1)}{\mu}. \end{aligned} \tag{2.78}$$

**Exercício 2.26** — Considere um sistema  $M/M/1$  com taxa de chegadas  $\lambda = 0.6$  (chegadas por minuto) e taxa de serviço  $\mu = 1$  (serviços por minuto).

- (a) Simule tempos de permanência neste sistema por forma a estimar pontual e intervalarmente o tempo esperado de permanência no sistema  $E(W^{(s)})$ , considerando para o efeito  $n = 200$  e  $r = 100$  e uma estimativa da variância do estimador  $\bar{W}^{(s)}$ .

Confronte as estimativas pontual e intervalar como o verdadeiro valor de  $E(W^{(s)})$ .

(b) Compare os resultados da alínea (a) com os que obteria se tivesse recorrido a variáveis antitéticas.

Por forma a que esta comparação seja mais rigorosa/razoável, procure controlar a semente de modo a usar os mesmos números pseudo-aleatórios uniformes na geração dos tempos de permanência no sistema que terá usado na alínea (a). ■

**Texto de apoio:** Morgan (1984, pp. 173–180).

### 2.2.5 Integrais múltiplos

O método de Monte Carlo é de longe muito mais relevante no cálculo de um **valor aproximado do integral definido múltiplo**

$$\theta_p = \int_{\mathcal{X}_p} h(\underline{x}) d\underline{x} \quad (2.79)$$

(onde  $\mathcal{X}_p \subset \mathbb{R}^p$ ) que no caso de um integral definido unidimensional

$$\theta = \int h(x) dx. \quad (2.80)$$

A razão prende-se com a existência de variadíssimos métodos de integração numérica para a obtenção de valores aproximados de  $\theta$  por vezes superiores ao método de Monte Carlo.

A adaptação do método de Monte Carlo ordinário com ou sem o uso de variáveis antitéticas faz-se sem grande dificuldade. Senão veja-se:

- **Método de Monte Carlo ordinário** — Comece-se por efectuar a seguinte factorização:

$$h(x_1, \dots, x_p) = g(x_1, \dots, x_p) f(x_1, \dots, x_p), \quad (2.81)$$

onde  $f(x_1, \dots, x_p)$  é a f.d.p. do vector aleatório  $\underline{X} = (X_1, \dots, X_p)$  que toma valores exclusivamente em  $\mathcal{X}_p$ . Neste contexto

$$\theta_p = E[g(X_1, \dots, X_p)] \quad (2.82)$$

pelo que deve gerar-se  $n$  vectores pseudo-aleatórios  $\underline{x}_1, \dots, \underline{x}_n$  com a densidade conjunta comum  $f(x_1, \dots, x_p)$  e de seguida calcular a média  $\hat{\theta}_p = \frac{1}{n} \sum_{i=1}^n g(\underline{x}_i)$ . ■

- **Método de Monte Carlo ordinário** (variáveis antitéticas) — Este método só deve ser usado quando o vector aleatório é constituído por  $p$  v.a.s independentes  $X_i$  com f.d.  $F_i(x_i)$ .

Nesta situação deve considerar-se o par de v.a.s dependentes

$$Y_1 = g[F_1^{-1}(U_1), \dots, F_p^{-1}(U_p)] \quad (2.83)$$

$$Y_2 = g[F_1^{-1}(1 - U_1), \dots, F_p^{-1}(1 - U_p)] \quad (2.84)$$

onde  $U_i \sim_{iid} \text{Uniforme}(0, 1), i = 1, \dots, p$ . ■

### 2.2.6 Método de Monte Carlo em inferência estatística

Em Inferência Estatística, é frequente assumir que se conhece a **distribuição de uma estatística de teste** ou de um **estimador**. Esta situação é de um modo geral irrealista pois a **validade desse resultado distribuicional** pressupõe por sua vez a **validade de hipóteses de trabalho** demasiado **rigorosas** e geralmente **inadequadas aos dados** de que se dispõe.

Basta pensar no que acontece à **distribuição exacta** e muito especialmente na **potência do teste** que faz uso da estatística de teste

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \quad (2.85)$$

quando se pretende confrontar as hipóteses  $H_0 : \mu = \mu_0$  e  $H_1 : \mu \neq \mu_0$  — e não se está a lidar com uma população  $X$  normalmente distribuída mas sim uma normal contaminada por observações discordantes (ou se está na presença de uma estrutura de dependência, etc.).

A resposta analítica para a questão anterior e outras quantas similares não é trivial.

A **resposta numérica** para a questão encontra-se na **simulação** dessa estatística, i.e., em **estudos de Monte Carlo**.

Estes **estudos de simulação** devem ser encarados com autênticas **experiências** que requerem um **planeamento cuidado** e, no caso da estimação da potência da estatística de teste (2.85) passam por decidir:

- as distribuições da população  $X$  (e.g.  $Normal(0, 1)$ ,  $Normal(0, 1)$  com  $c\%$  de observações discordantes,  $Cauchy(0, 1)$ );
- as dimensões das amostras  $n$  (e.g.  $n = 20, 200, 2000$ );
- os níveis de significância do teste (e.g.  $\alpha = 0.01, 0.05$ );
- os números de replicações (e.g.  $r = 1000, 10000$ );
- os valores de  $\mu$  (e.g.  $\mu = \mu_0 \pm \{0, 2(0.1)\}$ ); etc.

Posto isto deve calcular-se o valor da potência do teste (e.g. a percentagem de rejeições da hipótese nula) para cada combinação dos factores acima e organizar-se os resultados de preferência graficamente por forma a permitir uma análise mais cuidada dos mesmos.

Para outro exemplo de um estudo de simulação Monte Carlo no contexto da regressão linear simples é favor referir-se a Gentle (1998, pp. 180–190).

O problema da estimação de quantis e outras medidas de importância crucial na caracterização do comportamento probabilístico de estatísticas de teste e estimadores será aprofundado posteriormente aquando do estudo dos **Métodos de Bootstrap** no **Módulo 2** desta disciplina.

**Texto de apoio:** Gentle (1998, pp. 177–190).

## 2.3 Referências

Gentle, J.E. (1998). *Random Number Generation and Monte Carlo Methods*. Springer-Verlag, New York, Inc. (QA298.GEN.50103)

Morgan, B.J.T. (1984). *Elements of simulation*. Chapman and Hall. (QA276/1.MOR.32020)

Sobol, I. (1983). *O Método de Monte Carlo*. Editora Mir Moscou.

## Agradecimento

O autor destas notas de apoio salienta que a preparação das mesmas muito deve ao seu caderno do curso *C96 – Computational Statistics* leccionado por R.J. Brooks, no ano lectivo de 1989/ 1990, no *Department of Statistical Science* da *University College London*.