

MONOGRAFIAS DE MATEMÁTICA Nº 49

**Simulação Estocástica:
Teoria e Algoritmos
Versão Completa**

Oscar H. Bustos

IMPA

Estrada Dona Castorina, 110

22460 Rio de Janeiro, RJ

Alejandro C. Frery

INPE — DPI

Avenida dos Astronautas, 1758

12201 São José dos Campos, SP

BRASIL

I S B N

85-244-0053-6

Copyright ©1992 Oscar H. Bustos / Alejandro C. Frery

Conselho Nacional de Desenvolvimento Científico e Tecnológico

INSTITUTO DE MATEMÁTICA PURA E APLICADA

Estrada Dona Castorina, 110

22.460 - Rio de Janeiro-RJ

MONOGRAFIAS DE MATEMÁTICA

(títulos já publicados)

- 1) Azevedo, Alberto / Piccinini, Renzo - INTRODUÇÃO À TEORIA DOS GRUPOS (1970) / reprodução (1984)
- 2) Santos, Nathan M. - VETORES E MATRIZES (1970) - esgotada
- 3) Carmo, Manfredo P. do - INTRODUÇÃO À GEOMETRIA DIFERENCIAL GLOBAL (1970) - esgotada
- 4) Palis Junior, Jacob - SEMINÁRIO DE SISTEMAS DINÂMICOS (1971) - esgotada
- 5) Carvalho, João Pitombeira de - INTRODUÇÃO À ÁLGEBRA LINEAR (1971) - esgotada
- 6) Fernandez, Pedro Jesus - INTRODUÇÃO À TEORIA DAS PROBABILIDADES (1971) - esgotada
- 7) Robinson, R. C. - LECTURES ON HAMILTONIAN SYSTEMS (1972) - esgotada
- 8) Carmo, Manfredo P. do - NOTAS DE GEOMETRIA RIEMANNIANA (1972) - esgotada
- 9) Hönig, Chaim S. - ANÁLISE FUNCIONAL E O PROBLEMA DE STURM-LIOUVILLE (1972) - esgotada
- 10) Melo, Wellington de - ESTABILIDADE ESTRUTURAL EM VARIEDADES DE DIMENSÃO 2 (1972) - esgotada
- 11) Lesmes, Jaime - TEORIA DAS DISTRIBUIÇÕES E EQUAÇÕES DIFERENCIAIS (1972) - esgotada
- 12) Vilanova, Clóvis - ELEMENTOS DA TEORIA DOS GRUPOS E DA TEORIA DOS ANÉIS (1972) - esgotada
- 13) Douai, Jean Claude - COHOMOLOGIE DES GROUPES (1973) - esgotada
- 14) Lawson Jr., H. Blaine - LECTURES ON MINIMAL SUBMANIFOLDS, Vol. I (1973) - esgotada
- 15) Lima, Elon Lages - VARIEDADES DIFERENCIÁVEIS (1973) - esgotada
- 16) Mendes, Pedro - TEOREMAS DE \mathcal{A} -ESTABILIDADE E ESTABILIDADES ESTRUTURAL EM VARIEDADES ABERTAS (1973) - esgot.
- 17) Amann, Herbert - LECTURES ON SOME FIXED POINT THEOREMS (1974) - esgotada
- 18) - EXERCÍCIOS DE MATEMÁTICA / IMPA (1974) - esgotada
- 19) Figueiredo, Djairo Guedes de - NÚMEROS IRRACIONAIS E TRANSCEDENTES (1975) - esgotada
- 20) Zeeman, C. E. - UMA INTRODUÇÃO INFORMAL À TOPOLOGIA DAS SUPERFÍCIES (1975) - esgotada
- 21) Carmo, Manfredo P. do - NOTAS DE UM CURSO DE GRUPOS DE LIE (1975) - esgotada
- 22) Prestel, Alexander - LECTURES ON FORMALLY REAL FIELDS (1975) - esgotada
- 23) Simis, Aron - INTRODUÇÃO À ÁLGEBRA (1976) - esgotada
- 24) Lesmes, Jaime - SEMINÁRIO DE ANÁLISE FUNCIONAL (1976) - esgotada
- 25) Brauer, Fred - SOME STABILITY AND PERTURBATION PROBLEM FOR DIFFERENTIAL AND INTEGRAL EQUATIONS (1976) - esg.
- 26) Rodriguez, Lúcio - GEOMETRIA DAS SUBVARIEDADES (1976) - esgotada
- 27) Miranda, Mário - FRONTIÈRE MINIME (1976)
- 28) Cardoso, Fernando - RESOLUBILIDADE LOCAL DE EQUAÇÕES DIFERENCIAIS PARCIAIS (1977) - esgotada
- 29) Becker, Eberhard - HEREDITARILY-PYTHAGOREAN FIELDS AND ORDERINGS OF HIGHER LEVEL (1978)
- 30) Bass, Hyman - PROJECTIVE MODULES AND SYMMETRIC ALGEBRAS (1978)
- 31) Neyman, J. - PROBABILIDADE FREQUENTISTA E ESTATÍSTICA FREQUENTISTA (1978)
- 32) Dumortier, Freddy - SINGULARITIES OF VECTOR FIELDS (1978)
- 33) Viswanathan, T. M. - INTRODUÇÃO À ÁLGEBRA E ARITMÉTICA (1979) - esgotada
- 34) Thayer, F. Javier - NOTES ON PARTIAL DIFFERENTIAL EQUATIONS (1980)
- 35) Bierstone, Edward - THE STRUCTURE OF ORBIT SPACES AND THE SINGULARITIES OF EQUIVARIANT MAPPINGS (1980)
- 36) Thayer, F. Javier - THÉORIE SPECTRALE (1982)
- 37) Carmo, Manfredo P. do - FORMAS DIFERENCIAIS E APLICAÇÕES (1983)
- 38) Prestel, Alexander / Roquette, Peter - LECTURES ON FORMALLY p-ADIC FIELDS (1983)
- 39) Lequain, Yves / Garcia, Arnaldo - ÁLGEBRA: UMA INTRODUÇÃO (1983) - esgotada
- 40) Barbosa, J. Lucas / Colares, A. Gervásio - MINIMAL SURFACES IN \mathbb{R}^3 (1986)
- 41) Bérard, Pierre H. - SPECTRAL GEOMETRY: DIRECT AND INVERSE PROBLEMS (1986)
- 42) Bérard, Pierre H. - ANALYSIS ON RIEMANNIAN MANIFOLDS AND GEOMETRIC APPLICATIONS: AN INTRODUCTION (1987)
- 43) Torres, Felipe Cano - DESINGULARIZATION STRATEGIES FOR THREE-DIMENSIONAL VECTOR FIELDS (1988)
- 44) Endler, Otto - TEORIA DOS CORPOS (1988)
- 45) Bruns, Winfried / Vetter, Udo - DETERMINANTAL RINGS (1988)
- 46) Nefez, Abramo - INTRODUÇÃO À GEOMETRIA PROJETIVA (1990)
- 47) Gouvêa, Fernando Quadros - FORMAS MODULARES: UMA INTRODUÇÃO (1990)
- 48) Jørgensen, Bent - EXPONENTIAL DISPERSION MODELS (1991)
- 49) Bustos, Oscar H. / Frery, Alejandro C. - SIMULAÇÃO ESTOCASTICA: TEORIA E ALGORITMOS (Versão Completa) (1992)
- 50) Lötac, Gérard - LECTURES ON NATURAL EXPONENTIAL FAMILIES AND THEIR VARIANCE FUNCTIONS (1992)

A las Conejas y a los viejis.
A Irma, Ofelia y Mirta.

Estas notas, elaboradas para o mini-curso do mesmo nome a ser ministrado no X SINAPE, pretendem servir a, basicamente, dois propósitos:

1. constituir-se num mini-texto de Simulação Estocástica a um nível de complexidade intermediário, isto é que um estudante de final de Mestrado em áreas afins (Matemática, Estatística, Pesquisa Operacional, Engenharia, etc.) possa acompanhar com relativamente pouca dificuldade; e
2. servir como referência para aqueles usuários da Simulação Estocástica interessados em fazer trabalhos relativamente sofisticados, com qualidade de um nível mínimo de aceitação (digamos, de publicação em revistas especializadas).

O leitor interessado meramente nas aplicações deveria ou estudar estas notas e depois prosseguir os estudos na direção desejada, ou deixá-las ao lado e (eventualmente) consultá-las durante a leitura de outros textos mais direcionados nesse sentido. O mais perto que aqui chegaremos delas será à confecção de alguns *tijolos* básicos (porém sólidos) para a posterior construção de casas, edifícios, ... o que as necessidades e capacidades do leitor venham a produzir.

Para tanto, os leitores deveriam ter alguma familiaridade com Probabilidade e Estatística, ao nível mencionado acima. Também será de ajuda dispor de ferramentas de Análise e de Álgebra Linear com a mesma profundidade.

Alguns dos exercícios propostos requerem o uso de pacotes para análise de dados, sendo suficiente uma versão para micro-computador de STATGRAPHICS, GLIM, ou ainda algum mais simples: basta poder calcular estatísticas básicas (médias, variâncias, correlações), gerar histogramas e tabelas, realizar testes χ^2 e de Kolmogorov-Smirnov e ajustar e testar modelos de regressão linear uni- e bi-variados. Outros precisam de familiaridade com alguma linguagem de programação, por exemplo BASIC, PASCAL, FORTRAN ou C. De preferência, algumas implementações de algoritmos deverão ser rodadas com o uso de co-processador matemático.

O conteúdo destas notas não pretende ser enciclopedista; antes bem o leitor deve estar ciente que os assuntos abordados refletem, por um lado, os pontos de vistas dos autores sobre o que é relevante na área de Simulação Estocástica; pelo outro, no que diz respeito aos exemplos e às aplicações, incluímos aqueles tópicos que a nossa experiência ditou como mais importantes. Portanto, o leitor pode sentir-se a vontade de achar que estas notas são excessivas e/ou incompletas.

Estas notas foram possíveis graças à confluência de vários fatores favoráveis. Muitas pessoas contribuíram positivamente para o resultado que o leitor tem entre mãos. Embora qualquer enumeração seja injusta pela necessária omissão de nomes, consideramos importante fixar alguns deles. (OHB): à Fundação Desenvolvimento da UNICAMP-FUNCAMP, ao Mestrado em Qualidade Processo 48.91, ao CNPq, e à possibilidade de trabalhar nestas notas como parte das atividades do verão de 1992; (ACF): aos professores e aos amigos Maria Eulália Vares (IMPA), Gerald Banon (INPE), Nelson Mascarenhas (INPE), Diego Rial, José Raúl Martínez, Mario Cardona e Enilson Costa.

O IMPA, como sempre, constituiu-se no ambiente perfeito para a redação (em \LaTeX) destas notas.

Índice

1	Introdução	1
1.1	O que é Simulação Estocástica?	1
1.2	Quais são os problemas que devem ser estudados usando Simulação Estocástica?	2
1.3	Quais são os problemas que podem ser estudados usando Simulação Estocástica?	4
2	Ferramentas e Noções de Probabilidade e Estatística	7
2.1	Notações	7
2.2	Algumas distribuições importantes	7
2.2.1	Casos Contínuos	7
2.2.1.1	Variáveis Uniformes	7
2.2.1.2	Variáveis Normais	8
2.2.1.3	Variáveis Cauchy	8
2.2.1.4	Variáveis Gamma	8
2.2.1.5	Variáveis von Mises	9
2.2.1.6	Variáveis Beta	9
2.2.2	Casos discretos	9
2.2.2.1	Variáveis Binomiais	9
2.2.2.2	Variáveis Poisson	9
3	Geradores de variáveis aleatórias uniformes	10
3.1	Propriedades desejadas	10
3.2	Breve Resenha Histórica	11
3.3	Algoritmos Congruenciais Lineares	13
3.3.1	Geradores Congruenciais Lineares Recomendados	18
3.3.1.1	Gerador 1	19
3.3.1.2	Geradores "Ótimos"	19
3.4	Algoritmos de Registro de deslocamento	19
3.5	Outros Geradores	21
3.5.1	Geradores de Fibonacci	21
3.6	Combinação e Embaralhado de Geradores	22
3.7	Considerações Computacionais	22
3.7.1	Problemas com o tamanho do registro dos números	22
3.7.2	Problemas com a Precisão	23

3.7.3	Problemas com a Velocidade dos Cálculos	23
3.8	Testes de Geradores	24
3.8.1	Teste χ^2	25
3.8.2	Teste de Kolmogorov-Smirnov	25
3.8.3	Teste Serial	26
3.8.4	Teste do Intervalo	26
3.8.5	Teste Póker	27
3.8.6	Testes das Corridas	28
3.8.7	Testes das Permutações	29
3.8.8	Teste do Máximo-de- T	29
3.8.9	Teste da Dependência Linear	30
3.9	Geradores Fornecidos por certos "Pacotes"	30
3.10	Exercícios	32
4	Métodos gerais para geração de variáveis aleatórias não uniformes	34
4.1	Introdução	34
4.1.1	Exemplos Simples	34
4.2	O Método de Inversão	35
4.3	O Método de Rejeição	36
4.4	O Método de Composição	41
4.5	Quociente de Uniformes	43
4.6	Exercícios	45
5	Geração de variáveis aleatórias discretas	47
5.1	Método de inversão	48
5.1.1	Algoritmo de Busca Sequencial	48
5.1.2	Variáveis Poisson por Busca Sequencial	49
5.1.3	Variáveis Aleatórias Discretas Uniformemente Distribuídas	49
5.2	Métodos baseados em Tabelas	49
5.2.1	O Princípio da Tabela	50
5.2.2	O Método do Histograma	50
5.3	Método da flecha (alias)	50
5.4	O Método de Rejeição	51
5.5	Exercícios	52
6	Casos especiais de variáveis aleatórias	55
6.1	Variáveis Aleatórias Normais	56
6.1.1	Método Polar ou de Box-Muller	56
6.1.1.1	Algoritmo Geral do Método Polar	57
6.1.1.2	Algoritmo N1	58
6.1.1.3	Algoritmo N2	58
6.1.1.4	Algoritmo N3	58
6.1.1.5	Algoritmo N4	58
6.1.2	Outros Métodos para Geração de $\mathcal{N}(0,1)$	59

6.1.2.1	Métodos I e II de Johnson-Ramberg	59
6.1.2.2	Algoritmo de von Neumann	59
6.1.2.3	Algoritmo por Rejeição de Cauchy	60
6.1.2.4	Algoritmo de Marsaglia e Bray	60
6.2	Variáveis Aleatórias Gamma	61
6.2.1	Casos Simples de Geração de Gammas	61
6.2.2	Geração de Gammas com $\alpha > 1$ e $\beta = 1$	61
6.2.2.1	Algoritmo XG	62
6.2.2.2	Algoritmo GB	62
6.2.3	Geração de Gammas com $\alpha < 1$ e $\beta = 1$	62
6.2.3.1	Mistura de Gamma com Exponencial	62
6.2.3.2	Mistura de Gamma com Beta	62
6.2.3.3	Algoritmo de Rejeição Geral com Weibull	63
6.3	Variáveis Aleatórias von Mises	63
6.4	Distribuição de Poisson	64
6.4.1	Algoritmo dos Tempos de Chegada	64
6.4.2	Algoritmo do Produto de Uniformes	65
6.5	Distribuição Binomial	65
6.5.1	Geração com Uma Uniforme	66
6.5.2	Busca Sequencial	66
6.5.3	Tempos de Espera	66
6.6	Exercícios	67
7	Tópicos e aplicações especiais de Simulação Estocástica	69
7.1	Uso da Simulação Estocástica em Inferência Estatística	69
7.1.1	Testes Monte Carlo	69
7.1.2	Estimação do desvio padrão de um estimador usando "bootstrap"	74
7.1.3	Intervalos de Confiança Monte Carlo	76
7.2	Otimização combinatória usando relaxação estocástica (simulated annealing)	78
7.3	Integração Estocástica ou Monte Carlo	80
7.4	Elementos de Simulação Estocástica Multivariada	82
7.4.1	Geração de Normal Bi-variada	83
7.4.2	Geração de Bernoulli bi-variada e de Tabelas de Contingência 2×2	85
7.5	Simulação de processos físicos	86
7.5.1	Um Modelo para o Ferromagnetismo	86
7.5.2	Barcos em Tempos de Guerra	87
7.5.3	Projéteis	88
7.5.4	Localização de Árvores	89
7.5.5	Modelos de Crescimento	90
7.5.5.1	Crescimento Balístico	90
7.5.5.2	Crescimento por Agregação	90
7.6	Exercícios	91

8	Análise dos Resultados obtidos em uma Simulação por Computador	95
8.1	Introdução	95
8.2	Variáveis Negativamente Correlacionadas	100
8.3	Variáveis de Controle	103
8.4	Variáveis de Controle com Regressão	106
8.5	Amostragem Condicional	112
8.6	Amostragem Seletiva	114
8.7	Amostragem Estratificada	117
8.8	Sugestões sobre Apresentação de um Relatório de uma Experiência (ou Estudo) de Simulação	119
8.9	Exercícios	125
A	Estimação dos Quantis	134
A.0.1	Algumas Propriedades da Distribuição do Estimador \hat{x}_α	135
B	Técnica de Particionamento para Cálculo de Variâncias	137
•	Bibliografia	141

Tabelas

8.1	Apresentação de Resultados	110
8.2	Apresentação de Resultados	111
8.3	Estimadores de Parâmetros Amostrais por Partição	123
8.4	Estimadores de Parâmetros Amostrais	123
8.5	Estimadores de Parâmetros Amostrais por Tamanho de Sub-amostra	124
8.6	Exercício 8.11	129
8.7	Exercício 8.12	130
8.8	Exercício 8.13	132
8.9	Exercício 8.14	133

Capítulo 1

Introdução

“Simulação”, como palavra empregada corriqueiramente, não possui o sentido pretendido nestas notas. Vejamos o que o Aurélio diz ao respeito:

simulação: [Do lat. *simulare*] S. f. 1. Ato ou efeito de simular. 2. Disfarce, fingimento; simulacro... 3. Hipocrisia, fingimento, impostura... 4. Experiência ou ensaio realizado com o auxílio de modelos... ♦ **simulação digital.** Experiência ou ensaio que consiste numa série de cálculos numéricos e decisões de escolha limitada, realizado segundo um conjunto de regras predeterminadas e apropriado ao emprego de computadores digitais.

O leitor deverá ter sentido no último aparte —que é, evidentemente, o sentido que mais nos interessará nestas notas— um forte cheiro a *algoritmo*.

Já, ao procurar **aleatório**, o mesmo dicionário fornece uma definição bastante clássica, e até referencia termos como *variáveis aleatórias*.

A plataforma do nosso trabalho será o computador digital. Devemos salientar que, historicamente, as primeiras simulações foram feitas em computadores analógicos, e tiveram como motivação a solução (gráfica e) automática de equações diferenciais. Os assuntos que nestas notas pretendemos abordar —embora muito parcialmente— tem por origem os trabalhos de von Neumann, dos que falaremos no Capítulo 3.

Decidimos empregar a associação dos termos **Simulação** e **Estocástica**, acompanhando a Ripley, para limitar um pouco o universo da Simulação. Este universo inclui a solução numérica de equações diferenciais, o estudo de *sistemas*, a avaliação de algoritmos complexos, a otimização e o teste de técnicas de comunicação e de processamento de sinais etc.

1.1 O que é Simulação Estocástica?

Suponhamos que temos entre mãos um problema bem formulado. Entenderemos por Simulação Estocástica à arte de gerar amostras de variáveis aleatórias num ambiente computacional e usar ditas amostras para a obtenção de um certo resultado.

- Por que dizemos *arte*? Porque precisamos lançar mão de uma classe enorme de resultados para atingir o objetivo com o máximo de qualidade e o mínimo de custo. Essa “caixa de

ferramentas” que precisamos conhecer e saber usar inclui, basicamente, noções de teoria de números, de probabilidade, de processos estocásticos, de estatística, de ciência da computação, de análise numérica, e das particularidades do problema sob consideração. Os Capítulos 2 a 6 destas notas pretendem fornecer elementos mínimos destas áreas.

- Por que *variáveis aleatórias*? Porque muitos problemas interessantes ou são de natureza não determinística ou porque é mais apropriado o tratamento estocástico (um belo exemplo disto é a solução de problemas de otimização combinatória usando relaxação estocástica: o *simulated annealing* que comentaremos no Capítulo 7).
- Por que *num ambiente computacional*? Porque suporemos que a complexidade e/ou o volume de dados do problema a ser resolvido são tais que está fora de cogitação a procura manual de uma solução. O leitor encontrará muitos exemplos nestas notas onde o ambiente computacional é central: alguns muito importantes no Capítulo 3 e outros espalhados pelo texto.
- Por que *de um certo resultado*? Porque o bom uso da Simulação Estocástica não leva a simular um resultado; pelo contrário, ele pode fornecer resultados aproximados aos resultados exatos (se eles existirem). Formas de avaliar esses resultados serão apresentadas no Capítulo 8.

1.2 Quais são os problemas que devem ser estudados usando Simulação Estocástica?

Em algumas aplicações, a simulação estocástica é o primeiro passo para se ter uma idéia intuitiva das generalidades do problema, antes de tentar resolvê-lo com outras ferramentas. Às vezes, quando dispõe-se de diferentes alternativas para achar uma solução, ela constitui-se na escolha que mais rápido e facilmente produz resultados aceitáveis. Porém, é bom salientar que *nenhuma solução aproximada é tão boa quanto a solução exata*.

A seguir enunciaremos a regra de ouro da Simulação Estocástica, que responde à pergunta formulada no título desta Secção:

Nenhum problema para o qual existam soluções teóricas alcançáveis no tempo requerido e a custos toleráveis deve ser resolvido usando Simulação Estocástica.

O leitor poderá fazer as seguintes perguntas: Como eu, que tenho que resolver o problema X no tempo Y a um custo inferior a Z vou saber se existe a tal solução exata (digamos S) satisfazendo as restrições impostas? Valerá a pena investir numa procura obsessiva de S , deixando de lado o desenho de uma experiência Monte Carlo que poderia dar-me uma idéia, embora não a exata, e correndo o risco de não achar nem S nem ter tempo de fornecer uma solução aproximada? Valerá a pena investir obsessivamente no desenho de um experimento Monte Carlo correndo o risco de que a concorrência chegue, também em tempo inferior a Y e com custos menores que Z , à solução exata?

Os autores não possuem, na data da redação destas notas, de respostas absolutamente gerais às perguntas acima (e existe forte evidência de que não as possuirão num tempo finito).

Apenas podemos dar algumas dicas: (i) enquanto lhe seja possível, estude para conhecer mais problemas e ferramentas para resolvê-los (dessa forma não correrá os riscos de tentar provar o Teorema de Pitágoras usando Simulação Estocástica nem de querer calcular a magnetização média de 10^{23} partículas que interagem ferromagneticamente em forma exata); (ii) procure resolver, usando Simulação Estocástica, tantos problemas simples como possível: não há melhor conselheiro do que a experiência; e (iii) sempre que possível trabalhe em equipes inter-disciplinares (a solução exata de alguns problemas pode ser evidente para alguns especialistas, e não para o leitor).

No decorrer destas notas empregaremos a seguinte notação: $\Pi_A(x)$ a função indicadora do conjunto A , isto é

$$\Pi_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{caso contrário.} \end{cases}$$

Vejamos um exemplo que ilustra este ponto de vista. Suponha o leitor que é convidado a estudar o seguinte problema de autômatas celulares: seja $\Lambda \subset \mathbb{Z}^2$ um subconjunto da forma $\Lambda = \{1, \dots, L\} \times \{1, \dots, L\}$. Diremos que os pontos $\ell \in \Lambda$ são vizinhos do ponto $j \in \Lambda$ se ℓ e j têm uma coordenada igual e a diferença das outras é, em valor absoluto, 1. Usaremos a notação ∂_j para denotar o conjunto dos vizinhos do ponto j . Seja $p_\Lambda \in [0, 1]$ fixa, $i = 0$ e $P = 0$ (diremos que P é o “indicador de ocupação completa”). Considere a seguinte dinâmica (suponha desconhecer resultados teóricos a respeito dela):

1. Observar u_1, \dots, u_{L^2} amostras independentes da variável aleatória U com distribuição uniforme no intervalo $(0, 1)$.
2. Para cada $j \in \Lambda$, se $u_j < p_\Lambda$ atribuir à variável $s(i)_j$ o valor 1; caso contrário fazer $s(i)_j = 0$. Isto é, as variáveis aleatórias $(S(i)_j)_{j \in \Lambda}$ são independentes e têm distribuição de Bernoulli com probabilidade p_Λ de sucesso.
3. Atribuir $S(i+1)_j = \Pi_{\{2,3,4\}}(\sum_{\partial_j} S(i)_j)$.
4. Atribuir $P = \Pi_{\{L^2\}}(\sum_{j \in \Lambda} (S(i+1)_j))$.
5. Se $P = 1$ parar; caso contrário retornar ao passo 1 com $i = i + 1$.

Problema: determinar, se existir, o valor p_c para o qual com $\Lambda = \mathbb{Z}^2$ vale

$$\text{Pr}(P = 1) = \begin{cases} 0 & \text{se } p_\Lambda < p_c \\ 1 & \text{se } p_\Lambda \geq p_c. \end{cases}$$

Alguns autores, usando técnicas de inferência para ajuste de escala, informaram não apenas esse valor p_c mas também intervalos de confiança associados... só que Schonmann (ver [114, e as referências aí citadas]) provou que $\text{Pr}(P = 1)$ para todo $p_c > 0$.

Advertência Importante: não procure responder com Simulação Estocástica perguntas que fogem ao seu escopo natural de aplicação; por exemplo, tratando de atingir o infinito.

Resumindo, o bom usuário da Simulação Estocástica tem o cuidado de deixá-la como última opção... conselho válido exceto para quem fará os Exercícios que complementam a teoria destas Notas.

1.3 Quais são os problemas que podem ser estudados usando Simulação Estocástica?

Embora as linhas acima possam sugerir que o escopo de utilização da Simulação Estocástica é limitado, existe um número muito grande de situações onde ela pode ser empregada com sucesso.

Os físicos interessados em Mecânica Estatística (veja, por exemplo, os textos de Binder [22, 21]) sabem bem disso, e vêm utilizando os Métodos Monte Carlo desde, praticamente, a aparição dos computadores digitais em forma proveitosa.

Alguns problemas de interesse prático que podem ser resolvidos usando Simulação Estocástica, são os que aparecem no desenho de estratégias ótimas para atendimento de usuários.

Exemplo 1: O problema a seguir é um exemplo típico onde o uso de Simulação Estocástica pode ser adequado. Considere a situação de (um modelo simples para) uma agência de banco que dispõe de oito caixas C_1, \dots, C_8 para atendimento ao público. Sabe-se que o tempo que cada caixa leva para atender um freguês é bem modelado por uma variável aleatória com distribuição dada pela densidade

$$\left[\frac{p}{\lambda_1} \exp\{-x/\lambda_1\} + \frac{1-p}{\lambda_2} \exp\{-x/\lambda_2\} \right] \Pi_{(0,+\infty)}(x),$$

com $0 < p < 1$ e $0 < \lambda_1 < \lambda_2 < +\infty$; isto é, com probabilidade p o cliente será “fácil” e, em média, sairá em um tempo λ_1 , já com probabilidade $1 - p$ o cliente será “difícil” e, em média, ocupará o caixa um tempo λ_2 . Pode-se supor que o tempo entre chegadas de clientes ao banco pode ser modelado como variáveis aleatórias exponenciais com média $\lambda_3(t)$. Esta última média $\lambda_3: [10, 16] \rightarrow (0, \lambda_{\max}]$ leva em conta as mudanças devidas aos horários pico, por exemplo uma função polinômica em t com mínimo nas 13 horas.

O gerente de atendimento está interessado em escolher uma das seguintes estratégias:

1. fazer uma fila única F_u , ou
2. fazer uma fila para cada caixa F_1, \dots, F_8 .

Da observação do funcionamento de outras agências foram estimados os valores de p e de $\{\lambda_i\}_{1 \leq i \leq 8}$ e sabe-se que com a estratégia 1 acima, se à chegada de um cliente a fila F_u contiver $\#F_u \leq 80$ ele entrará na fila; já se $\#F_u \geq 80$ ele entrará na fila com probabilidade $\exp\{80 - \#F_u\}$, caso contrário ele procurará uma outra agência. Também sabe-se que, nas agências similares onde foi implementada a estratégia 2 acima, os clientes ficavam se a menor fila era de 10 ou menos pessoas, já se $\inf_{1 \leq i \leq 8} \{\#F_i\} \geq 10$ ficavam com probabilidade $\exp\{10 - \inf_{1 \leq i \leq 8} \{\#F_i\}\}$, caso contrário iam para outra agência; e que o último freguês de cada fila espera um tempo exponencial de média 1 segundo ao cabo do qual “pula” para a menor fila vizinha (a primeira e última filas tem só uma fila vizinha cada).

O problema consiste em escolher a estratégia acima que minimize o número de fregueses perdidos no sentido da esperança matemática, isto é, durante um tempo infinito. O problema poderia ainda levar em conta a possibilidade de abrir mais caixas, mas neste caso os custos (com as obras, contratação de pessoal, etc.) também deveriam estar presentes no modelo.

Mesmo sem esta consideração de custos, os autores não conhecem de teoria alguma que permita resolver o problema acima em forma exata para condições realísticas: portanto esse é um caso onde a Simulação Estocástica (bem utilizada) pode fornecer respostas adequadas. Exemplos deste tipo são comuns em linhas de montagem, companhias aéreas, etc.

Salientemos que o leitor *não resolverá* um problema como apresentado acima em forma direta com as ferramentas contidas nestas notas. No entanto, aqui veremos como gerar ocorrências das variáveis aleatórias envolvidas e como aproveitar tais ocorrências para inferir a respeito do processo geral (tamanho das amostras, variâncias dos estimadores, etc.) Para resolver o problema e escolher uma das opções (F_u ou F_1, \dots, F_8), o leitor deverá dispor de um **ambiente de simulação** cujas componentes estocásticas são o alvo de estudo deste trabalho. Dito ambiente precisa das estruturas adequadas como, por exemplo, filas, listas encadeadas, etc. que são objeto de estudo de áreas da Ciência da Computação como Programação Orientada a Objetos, Estruturas de Dados, etc.

Porém, existem alguns ambientes de simulação disponíveis no mercado mas, como quase todo produto comercial, eles são pacotes fechados. O bom aproveitamento do material aqui exposto permitirá a construção de um ambiente dedicado e de boa qualidade para experiências Monte Carlo.

Exemplo 2: Vejamos um caso onde a Simulação Estocástica pode auxiliar no desenvolvimento de intuição a respeito da modelagem de uma classe de processos físicos. Suponha que o objeto de estudo seja a formação de moléculas bidimensionais e que uma teoria preveja que a formação dela obedece as seguintes hipóteses:

1. cada elemento base da molécula é um entre $b^{(1)}, \dots, b^{(5)}$, elementos unidimensionais de comprimento c_1, \dots, c_5 ;
2. cada molécula está formada por um número aleatório de elementos base com distribuição Poisson de parâmetro $\lambda > 0$. Seja $N(\omega) = n$ uma ocorrência dessa variável aleatória;
3. escolha $b_0 \in \{b^{(1)}, \dots, b^{(5)}\}$ uniformemente, digamos $b_0 = b^{(j)}$ e coloque-o com um extremo na origem de \mathbb{R}^2 e o outro em um ponto escolhido uniformemente em $C_0(c_j)$ (a circunferência de centro na origem e raio c_j).
4. a lei de formação da molécula é: para todo $1 \leq i \leq n$
 - (a) escolha $b_i \in \{b^{(1)}, \dots, b^{(5)}\}$ segundo a matriz de probabilidades $\{p_{k\ell}\}$, onde $1 \leq k \leq 5$ e $1 \leq \ell \leq 5$, e $p_{k\ell} = \text{IPr}(b_i = b^{(k)} | b_{i-1} = b^{(\ell)})$ descreve as preferências de vizinhança dos elementos base,
 - (b) coloque um extremo de b_i no extremo livre de b_{i-1} e o outro num ponto escolhido uniformemente em $C_{b_{i-1}}(c_i)$, isto é a circunferência de raio c_i e centro no extremo livre de b_{i-1} ,
 - (c) se a seqüência de elementos base b_0, \dots, b_i não apresentar auto-intersecções, prosseguir. Se não, a molécula é considerada inaceitável e destruída, e o processo de formação é interrompido.

Existem muitas perguntas a serem respondidas para este processo: diâmetro médio das moléculas formadas, o número médio de passos até a rejeição de uma molécula, etc. Mas, se

o leitor montar uma experiência Monte Carlo para simular este processo, além de (tal vez) responder essas perguntas em forma aproximada, poderá ter uma idéia visual de como são as moléculas previstas pela teoria; coisa não muito evidente da mera leitura das hipóteses. Já que para a implementação deste modelo são necessárias apenas ferramentas contidas nestas notas (além de alguns algoritmos gráficos bem simples), o leitor será convidado a resolver o problema no Exercício 7.11.

Capítulo 2

Ferramentas e Noções de Probabilidade e Estatística

Este breve Capítulo tem o intuito de relembrar algumas definições que serão úteis no decorrer do texto. Não aprofundaremos nestes assuntos, pois toda vez que seja necessário introduziremos as noções necessárias no ponto onde serão empregadas; no entanto, o leitor poderá voltar a estas páginas sempre que precise de alguma definição ou propriedade. Também aqui fixaremos as parametrizações que utilizaremos no resto das notas.

2.1 Notações

Para os temas básicos sobre Probabilidade, variáveis aleatórias, esperança de variáveis aleatórias, leis forte e fraca dos grandes números, teorema central do limite, etc. o leitor pode consultar os livros de James [65], de Hoel e outros [60] e o (mais avançado) de Chung [32].

Seja X uma variável aleatória. Se X é discreta, poremos $f_X(x) = \text{IPr}(X = x)$. Se X admite densidade (variável de tipo contínuo) f_X denotará essa densidade.

2.2 Algumas distribuições importantes

2.2.1 Casos Contínuos

2.2.1.1 Variáveis Uniformes

A variável aleatória U é dita ter distribuição uniforme no conjunto $A \subset \mathbb{R}^d$ (em símbolos $U \sim \mathcal{U}(A)$) se tem densidade dada por

$$f_U(u) = \frac{1}{\int_A du} \mathbb{I}_A(u) \text{ para todo } u \in \mathbb{R}^d.$$

Um caso que aparecerá com muita frequência é quando o conjunto A é um intervalo (aberto, fechado ou semiaberto) da reta. Sejam então $-\infty < a < b < +\infty$; U é dita ter

distribuição uniforme no intervalo (a, b) (ou, equivalentemente nos intervalos $(a, b]$, $[a, b)$ ou $[a, b]$) se

$$f_U(u) = \frac{1}{b-a} \mathbb{I}_{(a,b)}(u),$$

neste caso $\mathbb{E}\{U\} = (b-a)/2$, $\text{Var}(U) = (b-a)^2/12$ e $F_U(u) = \begin{cases} 0 & \text{se } u < a \\ \frac{u-a}{b-a} & \text{se } a \leq u \leq b \\ 1 & \text{se } u > b \end{cases}$.

2.2.1.2 Variáveis Normais

A variável aleatória X é dita ter distribuição Normal com média $\mu \in \mathbb{R}$ e variância $\sigma^2 > 0$ (em símbolos $X \sim \mathcal{N}(\mu, \sigma^2)$) se tem densidade dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \text{ para todo } x \in \mathbb{R}.$$

2.2.1.3 Variáveis Cauchy

A variável aleatória X é dita ter distribuição de Cauchy com parâmetros $\alpha \in \mathbb{R}$ e $\beta > 0$ se tem densidade dada por

$$f_X(x) = \frac{\beta}{\pi(\beta^2 + (x-\alpha)^2)} \text{ para todo } x \in \mathbb{R}.$$

Estas não têm esperança bem definida e têm variância infinita; também temos que $F_X(x) = \pi^{-1} \arctan((x-\alpha)/\beta) + \pi/2$.

2.2.1.4 Variáveis Gamma

A variável aleatória X é dita ter distribuição Gamma com parâmetros $\alpha > 0$ e $\beta > 0$ (em símbolos $X \sim \mathcal{G}(\alpha, \beta)$) se tem densidade dada por

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \exp\{-x/\beta\} x^{\alpha-1} \mathbb{I}_{(0,+\infty)}(x).$$

Neste caso vale que $\mathbb{E}\{X\} = \alpha\beta$ e $\text{Var}(X) = \alpha\beta^2$. Não se conhece forma fechada para F_X em geral. Alguns casos particulares da $X \sim \mathcal{G}(\alpha, \beta)$ são quando

- $\alpha = 1$, então $X \sim \mathcal{E}(\beta)$: a distribuição exponencial.
- $\alpha = k/2$, com $k \in \mathbb{N} \setminus \{0\}$, e $\beta = 2$, então $X \sim \chi_k^2$: a distribuição χ^2 com k graus de liberdade.

2.2.1.5 Variáveis von Mises

Diremos que a variável aleatória Θ tem distribuição de von Mises se tem densidade dada por

$$f_{\Theta}(\theta) = \frac{\exp\{\kappa \cos(\theta)\}}{2\pi I_0(\kappa)} \mathbb{I}_{[-\pi, +\pi]}(\theta),$$

onde $\kappa > 0$ é o parâmetro de forma e I_0 é a função de Bessel modificada de primeira classe e ordem zero, isto é

$$I_0(x) = \sum_{0 \leq j < \infty} \frac{1}{j!} (x/2)^{2j}.$$

2.2.1.6 Variáveis Beta

A variável aleatória X é dita ter distribuição Beta com parâmetros $\alpha > 0$ e $\beta > 0$ se tem densidade dada por

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbb{I}_{(0,1)}(x).$$

Neste caso vale que $\mathbb{E}\{X\} = \alpha/(\alpha + \beta)$ e $\text{Var}(X) = \alpha\beta/((\alpha + \beta)^2(\alpha + \beta + 1))$.

2.2.2 Casos discretos

2.2.2.1 Variáveis Binomiais

Diremos que a variável aleatória discreta X tem distribuição Binomial com parâmetros $p \in [0, 1]$ e $n \in \mathbb{N} \setminus 0$ (em símbolos $X \sim \text{Bi}(p, n)$) se

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \mathbb{I}_{\{0,1,\dots,n\}}(x).$$

Vale que $\mathbb{E}\{X\} = np$ e $\text{Var}(X) = np(1-p)$.

2.2.2.2 Variáveis Poisson

Diremos que a variável aleatória discreta X tem distribuição de Poisson com parâmetro $\lambda > 0$ (em símbolos $X \sim \mathcal{P}(\lambda)$) se

$$f_X(x) = \frac{\lambda^x}{x!} \exp\{-\lambda\} \mathbb{I}_{\{0,1,\dots\}}(x).$$

Vale que $\mathbb{E}\{X\} = \lambda$ e $\text{Var}(X) = \lambda$.

Capítulo 3

Geradores de variáveis aleatórias uniformes

3.1 Propriedades desejadas

Em simulação estocástica as variáveis aleatórias com distribuição uniforme no intervalo $[0, 1]$ são empregadas de muitas maneiras:

- Em forma direta.
- Para gerar distribuições discretas e contínuas.
- Para gerar conjuntos de variáveis aleatórias dependentes (processos estocásticos e/ou distribuições multivariadas).

Devemos ter bem em claro que na maioria das vezes, a performance de uma simulação estará fortemente correlacionada com a do gerador de uniformes usado. Daí a importância do assunto deste Capítulo.

Lembremos que uma sequência de variáveis aleatórias (U_1, \dots, U_N) é dita “uma amostra de tamanho N ” da distribuição uniforme em $(0, 1)$ —em símbolos: (U_1, \dots, U_N) são iid $\mathcal{U}(0, 1)$ — se

(3.1) para cada $k \geq 2$ e cada $1 \leq i_1 < \dots < i_k \leq N$ vale $\mathbb{P}\text{r}(U_{i_1} \leq u_1, \dots, U_{i_k} \leq u_k) = \mathbb{P}\text{r}(U_{i_1} \leq u_1) \dots \mathbb{P}\text{r}(U_{i_k} \leq u_k) = u_1 \cdot u_2 \dots u_k$ quaisquer que sejam u_1, \dots, u_k em $(0, 1)$, e

(3.2) para cada $i = 1, \dots, N$ e cada $u \in \mathbb{R}$ vale

$$\mathbb{P}\text{r}(U_i \leq u) = \begin{cases} 0 & \text{se } u < 0 \\ u & \text{se } 0 \leq u \leq 1 \\ 1 & \text{se } u > 1. \end{cases}$$

Um gerador “razoável” de uma sequência u_1, \dots, u_N que seja uma realização das variáveis aleatórias U_1, \dots, U_N , além de passar testes verificando se (3.1) e (3.2) são satisfeitas, deveria satisfazer (na maior medida possível) as duas seguintes propriedades:

(3.3) “repetibilidade” e “portabilidade”, e

(3.4) “velocidade computacional”.

Por “repetibilidade” entende-se que, dados os mesmos parâmetros que o definem, o gerador produza a mesma sequência sempre que assim o desejarmos. Mais adiante veremos a utilidade desta propriedade.

“Portabilidade” significa que, sob as mesmas condições definidoras, a sequência seja a “mesma”, independentemente da linguagem computacional usada para implementar o algoritmo de geração, e do computador usado. Na verdade, isto é bem difícil de ser atingido, particularmente com micro-computadores. Daí que neste trabalho não iremos abordar este assunto. O leitor interessado pode ver as referências [50,115].

A “velocidade computacional” está estreitamente ligada à precisão desejada nos resultados finais da simulação na qual é usado o gerador. Quanto mais rápido seja o gerador, mais resultados serão obtidos no mesmo tempo de uso de computador. Isto, em princípio, fará com que a precisão possa ser acrescentada. Em geral, para obter maior velocidade computacional será conveniente programar o algoritmo de geração usando uma linguagem de “baixo nível” ou de “nível intermediário” (como Assembler ou C, respectivamente), em vez de uma de “alto nível” (como FORTRAN), mais isto entrará em conflito com o requerimento de “portabilidade”.

De fato, toda vez que devamos fazer uma simulação, deveremos aceitar uma solução de compromisso entre os requerimentos expostos. Tal vez os prioritários devam ser sempre (3.1) e (3.2).

3.2 Breve Resenha Histórica

Os primeiros cientistas e pessoas interessadas em obter sequências de números aleatórios usaram procedimentos físicos. É provável que todos nós tenhamos jogado moedas ou dados como exercícios práticos nos primeiros cursos de Probabilidade. Experimentos de simulação baseados em tais procedimentos têm uma longa história, e ainda hoje são usados em jogos tais como bingos e em cassinos, lotéricas, etc. Tal vez os primeiros trabalhos sérios sejam o bem conhecido de Student e o de Tippett [121], ambos com já quase um século. O último produziu uma tabela de 40.000 dígitos aleatórios empregando dados censais. Infelizmente, tais dígitos não satisfaziam nem (3.1) nem (3.2).

Mais tarde, Kendall e Babbington-Smith [72] empregaram um dispositivo mecânico.

Já em 1955, a Rand Corporation [2] editou uma tabela de dígitos pseudo-aleatórios construídos em base a fenômenos de ruído eletrônico. Esses métodos têm recebido muita aceitação, possivelmente pela sua performance quando submetidos a certos testes, ou simplesmente porque *funcionavam* na prática. Mas eles não satisfazem o requerimento (3.3). Ainda são usados em situações onde a repetibilidade não é buscada, por exemplo em lotéricas; e sob condições que facilitem uma freqüente verificação da *aleatoriedade*, por exemplo em roletas nos cassinos. Todavia podem ser úteis em certas aplicações estatísticas, como em amostragem aleatória de populações não muito grandes. Também em simulações estocásticas por computador, quando dispõe-se de suficiente memória como para guardar uma tabela e quando somente são necessários poucos números.

Mais precisamente, um *bom* gerador de seqüências numéricas aleatórias (ou “pseudo-aleatórias”, para sermos exatos) deve satisfazer alguns princípios gerais, segundo Lewis e Orav [88]:

- A seqüência gerada deve ser intuitivamente aleatória.
- Essa aleatoriedade deve ser estabelecida teoricamente ou, no mínimo, deve passar certos testes de aleatoriedade. A aleatoriedade de uma seqüência jamais deve ser assumida sem essas verificações.
- Deve conhecer-se alguma coisa sobre as propriedades teóricas do gerador.

Esses princípios gerais indicam que os geradores *ad hoc* devem ser evitados. Também são aplicados à análise cuidadosa de geradores como os que veremos neste Capítulo.

Um dos primeiros trabalhos que sugeriu um método bem definido de geração de uma seqüência determinística tentando *imitar* uma seqüência aleatória, foi o de von Neumann [123]. Esse método é conhecido como “mid-square” e, sem muito rigor, pode ser escrito da seguinte maneira:

1. Seja X_0 um número inteiro de 5 dígitos decimais (pode ser que o dígito mais à esquerda seja 0). Faça $i = 0$.
2. Calcular X_i^2 . Eventualmente adicionar “0” à esquerda tantos como seja necessário para que X_i^2 resulte escrito como $X_i^2 = d_9d_8d_7d_6d_5d_4d_3d_2d_1d_0$, com d_j no conjunto $\{0, 1, \dots, 9\}$, $0 \leq j \leq 9$.
3. Definir $X_{i+1} = d_7d_6d_5d_4d_3$, isto é os cinco dígitos centrais ou “middle”.
4. Faça $i = i + 1$ e prossiga em 2.

Esse método não é usado, embora certas seqüências obtidas sejam bem aleatórias e se repita só depois de um número bem grande de termos. O que acontece é que ainda não se conhecem bem as suas propriedades, que dependem fortemente do valor inicial X_0 . Por exemplo: $X_0 = 2100 \Rightarrow X_1 = 4100, X_2 = 8100, X_3 = 6100, X_4 = 2100, \dots$; já para $X_0 = 3792$ temos $X_1 = X_2 = \dots = 3792$.

O leitor interessado em mais detalhes sobre este assunto pode consultar o livro de Donald Knuth [78] e os Exercícios 3.2 e 3.2.

Modernamente o método de geração mais usado é o chamado *congruencial linear*, que será o assunto da Secção 3.3 deste Capítulo.

Outro método em uso, embora não seja muito aconselhável, é o chamado de *registro de deslocamento*, que estudaremos brevemente na Secção 3.4.

Outros métodos serão vistos na Secção 3.5, e alguns sugeridos na década dos '80 podem ver-se em [97]. A Criptografia tem sugerido alguns geradores muito bons, mas que não são muito portáteis. O leitor interessado pode ver o algoritmo e o programa em FORTRAN dado em Press e outros [108].

Atualmente vários compiladores de linguagens de alto nível, como BASIC e PASCAL, oferecem *built-in* sub-rotinas de geração de seqüências numéricas pseudo-aleatórias. Infelizmente,

a maioria das vezes sem documentação suficiente. Isto é mais notável nos compiladores oferecidos pelos vendedores de “micro-” e de “mini-” computadores. Daí que é bom estar ciente do que disse Ripley [110, página 45]:

Microcomputer implementations work incorrectly surprinsignly often from faulty compilers or side effects of operating systems.

Já são bem melhores as subrotinas oferecidas por *pacotes* como IMSL [1,6] ou NAG [3] para computadores de grande porte. Também pacotes de grande uso em estatística como GLIM [14] tem sub-rotinas razoavelmente boas de geração.

Mas... sempre será bom olhar com certo cuidado as seqüências numéricas pseudo-aleatórias produzidas por um dado gerador. E todo trabalho Monte Carlo deveria usar, no mínimo, dois métodos diferentes: não apenas para checar a estabilidade dos resultados, mas também para evitar surpresas devidas à existência de estruturas insuspeitas.

3.3 Algoritmos Congruenciais Lineares

Nesta Secção e na seguinte para gerar uma seqüência u_1, \dots, u_N em $[0, 1)$, na verdade vamos gerar uma seqüência de inteiros y_1, \dots, y_N em $\{0, 1, \dots, M-1\}$, com M inteiro positivo (“grande”) e então tomar $u_k = y_k/M$ para todo $k = 1, 2, \dots, N$.

Definição 3.3.1 *Seja M inteiro positivo ≥ 2 . Uma seqüência (y_1, y_2, \dots) em $\{0, 1, \dots, M-1\}$ é dita “gerada pelo gerador congruencial linear com parâmetros a , c e M e semente y_0 ” se*

$$y_i = (a y_0 + c) \bmod M, \quad 1 \leq i,$$

onde a , c e y_0 são inteiros do conjunto $\{0, \dots, M-1\}$.

A terminologia usual é: a é dito o *multiplicador*, c é o *incremento* e M é o *módulo*; se $c \neq 0$ o gerador é dito *mixto*, se $c = 0$ é dito *multiplicativo*. Eles foram sugeridos pela primeira vez no trabalho de Lehmer [85].

Vejamos se o método da Definição 3.3.1 satisfaz os requerimentos já mencionados nas Secções anteriores. Um desses requerimentos era que a *seqüência gerada deve ser intuitivamente aleatória*. Precisando um pouco este conceito, essa seqüência deve ser praticamente imprevisível, isto é conhecendo y_0, y_1, \dots, y_n seja muito difícil conhecer y_{n+1} a menos que a regra formal que gera a seqüência toda seja conhecida. Por exemplo, a seqüência

$$0, 1, 6, 15, 12, 13, 2, 11, 8, 9, \dots \quad (3.5)$$

com $n = 9$ foi gerada por um gerador congruencial linear. O leitor poderia disser qual será o próximo número (entre 0 e 15)?

Já a seqüência

$$1, 12, 1, 12, 1, 12, 1, 12, 1, 12, \dots \quad (3.6)$$

com $n = 9$ tem sido gerada por um outro gerador congruencial, e facilmente deduz-se que o próximo número tem que ser 1.

Vemos assim que quanto mais números aparecem na sequência sem repetir-se, mais imprevisível será o próximo número.

A seguir vejamos algumas definições e resultados que nos possibilitarão saber quão imprevisível uma sequência gerada por um gerador congruencial é. Notemos primeiramente que uma sequência gerada por um gerador congruencial é da forma

$$y_{n+1} = f(y_n), \text{ para } n = 0, 1, 2, \dots$$

onde f é uma função definida sobre um conjunto da forma $\{0, 1, \dots, M-1\}$ e com valores nesse mesmo conjunto. A Proposição a seguir permitirá formalizar o conceito de *imprevisível* associado ao de números gerados sem repetição.

Proposição 3.3.1 *Sejam $M \geq 1$ inteiro, $f: \{0, 1, \dots, M-1\} \rightarrow \{0, 1, \dots, M-1\}$, $y_0 \in \{0, 1, \dots, M-1\}$, y_1, y_2, \dots inteiros tais que $y_{n+1} = f(y_n)$ para todo $n = 0, 1, 2, \dots$. Então existem λ e μ tais que*

$$y_0, y_1, \dots, y_{\mu+\lambda-1} \text{ são todos distintos;} \quad (3.7)$$

$$y_{n+\lambda} = y_n \text{ para todo } n \geq \mu. \quad (3.8)$$

Neste caso definimos

Definição 3.3.2 *O número $K = \min\{\lambda: \exists \mu \text{ satisfazendo (3.7) e (3.8)}\}$ é dito o período da sequência (y_0, y_1, \dots) .*

É fácil ver que $K \leq M$.

(3.9) Ora, como escolher a , c , M e y_0 para obter as sequências com maior período possível? Uma primeira observação indica uma limitante superior para M : o computador trabalha apenas com um número finito de inteiros.

Por exemplo, a maioria dos computadores reserva, no máximo, 32 bits para guardar valores inteiros: quer dizer que todo número inteiro (na base binária) é escrito como uma sequência de 32 "0" e "1". Logo, todo número inteiro z só poderá ser um número entre 0 e $2^{31} - 1$. Daí que M deva ser escolhido na base de como o computador guarda os inteiros. Normalmente o fabricante fornece dita informação.

A seguinte observação mostra que a resposta à pergunta (3.9) não é coisa fácil, se queremos sequências aleatórias *bem comportadas*. Seja M qualquer número, se tomarmos $a = c = 1$ e $y_0 = 0$ obtemos a sequência

$$0, 1, 2, 3, \dots, M-1, \quad (3.10)$$

que é de período completo (isto é $K = M$), mas... que ninguém diria que é aleatória.

Os resultados a seguir ajudam a encontrar sequências de período completo que, além disso têm outras propriedades desejáveis para a aleatoriedade.

Teorema 3.3.1 *Sejam a , c , M , y_0 , (y_1, y_2, \dots) como na Definição 3.3.1, com $c \neq 0$. Então o período da sequência (y_1, y_2, \dots) é M se e somente se*

1. $\text{mcd}(c, M) = 1$: máximo comum divisor entre c e M .

2. $a \equiv 1 \pmod p$, qualquer que seja p , fator primo de M .
3. Se 4 divide a M (em símbolos $4/M$), então $a \equiv 1 \pmod 4$.

O leitor pode ver uma demonstração (nada trivial) deste teorema em [110].

Corolário: Se $c \neq 0$ e M é primo, então período de $(y_1, y_2, \dots) = M \implies a \equiv 1$.

Mais adiante veremos que estudar os geradores com $c = 0$ é suficiente para ter bons geradores. Porém é impossível obter uma sequência de período completo com $c = 0$, pois não pode ser $y_n = 0$ para nenhum n ; se assim fosse teríamos $y_{n+1} = ay_n \pmod M = a0 \pmod M = 0$ (y_{n+1} deve estar entre 0 e $M - 1$), e daí $y_{n+2} = 0$ e assim por diante. Com tudo, o seguinte resultado diz como podemos conseguir seqüências de período $M - 1$ com $c = 0$ (geradores multiplicativos). Mas antes disso precisamos de uma definição técnica.

Definição 3.3.3 Seja M primo; $a \neq 0$ é dito "raiz primitiva" de M se $a^{(M-1)/p}$ não é múltiplo de M qualquer que seja p , fator primo de $M - 1$.

Teorema 3.3.2 Sejam $a, c, M, y_0, (y_1, y_2, \dots)$ como na Definição 3.3.1, K o período da seqüência (y_1, y_2, \dots) , e $c \neq 0$. Então

1. $K = M - 1 \implies M$ é primo.
2. $M - 1$ é múltiplo de K .
3. $K = M - 1 \iff a$ é raiz primitiva de M .

Uma demonstração deste Teorema pode encontrar-se em Ripley [110] ou em Knuth [78].

Na verdade, os resultados desse Teorema não são de muita utilidade do ponto de vista prático. Em primeiro lugar não é coisa fácil calcular as raízes primitivas de um número. Alguns resultados teóricos que facilitam esses cálculos podem estudar-se em Knuth [78], e um programa em FORTRAN para o cálculo efetivo de raízes primitivas aparece em Ripley [110]. Também Abramowitz e Stegun [9] têm uma tabela com as raízes primitivas de vários números.

O resultado a seguir, provado na referência [110], ajuda no cálculo de raízes primitivas quando uma delas já é conhecida.

Teorema 3.3.3 Sejam M primo e a uma raiz primitiva de M . Se k é um inteiro tal que $\text{mcd}(k, M - 1) = 1$, então $a^k \pmod M$ é raiz primitiva de M .

Baseados nos Teoremas 3.3.1 e 3.3.2 podem sugerir-se as duas seguintes regras para a escolha dos parâmetros de um gerador congruencial linear.

Regra 1: Escolher a, c , e M como no Teorema 3.3.1 e usar um gerador cujo período seja conhecido e igual a M .

Regra 2: Tomar $c = 0$, a e M como no Teorema 3.3.2 e usar $y_0 \neq 0$ como semente; logo trabalhar com um gerador com período conhecido e igual a $M - 1$.

Exemplo: Um bom gerador multiplicativo ($c = 0$) para computadores da linha IBM 360/370 (ver [57,87]). Seja $M = 2^{31} - 1$ (primo de Mersenne), pode ver-se que

$$M - 1 = 2^{31} - 2 = 2 \cdot 3^2 \cdot 7 \cdot 11 \cdot 31 \cdot 151 \cdot 331,$$

e que 7 é raiz primitiva de M . Então $7^5 = 16807$ também o é, pois $\text{mcd}(5, 2^{31-2}) = 1$ e aplicando o Teorema 3.3.3.

Assim o período de uma seqüência (y_1, y_2, \dots) gerada pelo gerador multiplicativo $a = 16807$, $M = 2^{31} - 1$, $c = 0$ e semente $y_0 \neq 0$ terá período $2^{31} - 2 = 2147483646$.

Para acrescentar a velocidade computacional de um gerador multiplicativo pode-se escolher M convenientemente, ainda que pagando o preço de não se ter um gerador com período $M - 1$. Por exemplo, $M = 2^k$ (sendo k um inteiro) é muito conveniente para computadores binários pois nesse caso a operação “cálculo do módulo” é igual à operação “deslocamento de bits”. O resultado a seguir, generalizado e demonstrado em [78], orienta-nos nas escolhas do multiplicador e da semente quando $c = 0$ e $M = 2^k$.

Teorema 3.3.4 *Seja $k > 2$ inteiro. Então*

1. *O gerador congruencial linear multiplicativo com módulo 2^k tem período máximo igual a 2^{k-2} se e somente se a é raiz primitiva de M .*
2. *Nesse caso, a seqüência (y_1, y_2, \dots) gerado pelo gerador descrito acima atinge o período máximo se e somente se y_0 (a semente) é ímpar.*

Ora, não é muito recomendável usar geradores congruenciais multiplicativos com módulo 2^k nos computadores binários devido ao que poderíamos chamar de *fenômeno de não-aleatoriedade dos bits menos significativos*. Para entender o que isso quer dizer, vejamos um resultado cuja demonstração pode-se omitir (embora seja bem fácil, como o leitor poderá verificar).

Proposição 3.3.2 *Seja (z_1, z_2, \dots) uma seqüência gerada pelo gerador congruencial linear com parâmetros a , c e M e semente z_0 . Seja d um divisor de M . Para cada $i = 0, 1, 2, \dots$ seja*

$$y_i = z_i \bmod d. \quad (3.11)$$

Então $y_i = (ay_i + c) \bmod d$, para todo $i = 0, 1, 2, \dots$

Demonstração: Como d é divisor de M existe r inteiro tal que $M = r \cdot d$. Dado que $z_{i+1} = (az_i + c) \bmod M$ para todo $i = 0, 1, 2, \dots$, temos que para cada $i = 0, 1, 2, \dots$ existe s_{i+1} tal que $z_{i+1} = az_i + c + s_{i+1}M$. Por 3.11, para cada $i = 0, 1, 2, \dots$ existe r_i inteiro tal que $y_i = z_i + r_i d$. Então, para cada $i = 0, 1, 2, \dots$ temos

$$\begin{aligned} y_{i+1} &= z_{i+1} + r_{i+1}d = az_i + c + s_{i+1}M + r_{i+1}d \\ &= a(y_i - r_i d) + c + s_{i+1}rd + r_{i+1}d \\ &= ay_i + c + (-ar_i + s_{i+1}r + r_{i+1})d, \end{aligned}$$

donde se segue a Proposição.

Agora sim, vejamos o que diz a Proposição 3.3.2: quando $M = 2^k$ com $k \geq 2$ inteiro. Nesse caso, se $d = 2$ temos que (y_0, y_1, \dots) é a seqüência formada com o bit menos significativo de (z_0, z_1, \dots) e o seu período é, no máximo, igual a 2; daí que (y_0, y_1, \dots) só possa ser uma das seguintes quatro seqüências: $(0, 1, 0, 1, \dots)$, $(0, 0, 0, 0, \dots)$, $(1, 1, 1, 1, \dots)$ ou $(1, 0, 1, 0, \dots)$.

Esse comportamento não aleatório das seqüências formadas com os bits menos significativos de um gerador congruencial com módulo não primo é uma propriedade indesejável, que pode agravar-se com o uso de certas técnicas utilizadas para acelerar o procedimento de geração (veja a referência [88]). O trabalho de Hutchinson [64] destaca que quando o módulo M do gerador congruencial multiplicativo é primo e a é raiz primitiva de M (isto é o gerador tem período máximo) então as seqüências formadas com os bits menos significativos têm a tendência de exibir um comportamento razoável, embora não exista teoria alguma para justificar essa afirmação.

Muita coisa poderia se dizer em torno das propriedades dos geradores congruenciais. Neste trabalho apenas apresentamos as que consideramos essenciais para a compreensão das definições básicas e das principais regras que orientam as escolhas dos parâmetros. Encerraremos esta Secção com um resultado que indica as limitações dos geradores congruenciais lineares, e com uma lista dos geradores recomendáveis dentro desta classe.

Teorema 3.3.5 *Seja y_1, y_2, \dots uma seqüência gerada pelo gerador congruencial linear com parâmetros a, c e M . Seja $k \geq 1$ e $\mathcal{A} = \{(y_i, \dots, y_{i+k-1}) : i = 0, 1, \dots, M-k\}$. Então \mathcal{A} está contido numa família de no máximo $(k!M)^{1/k}$ hiper-planos paralelos.*

A demonstração do Teorema 3.3.5, e de outros resultados de interesse para o estudo da *estrutura de rede* apresentada pelos geradores congruenciais lineares podem ver-se nas referências [35, 98, 110], entre outras.

Uma consequência prática do último Teorema é que se, por exemplo $M = 2^{32}$ e $k = 10$ então todas as 10-uplas do conjunto estarão contidas, no máximo, em 41 hiper-planos. O exemplo a seguir, famoso na literatura de geradores congruenciais, mostra que em certos casos esse número de hiper-planos pode ser bem menor que o limitante máximo dado pelo Teorema, e que essa situação pode passar despercebida mesmo para pessoas muito experientes.

Exemplo: Um gerador ruim e (infelizmente) muito usado: RANDU. Este é um gerador do tipo congruencial linear multiplicativo ($c = 0$) com $M = 2^{31}$ e $a = 2^{16} + 3$, fornecido pelo IBM Scientific Subroutine Package e ainda pelos micro-computadores da linha DEC-VAX. O seguinte resultado vale para este tipo particular de geradores

Proposição 3.3.3 *Seja (y_1, y_2, \dots) uma seqüência gerada pelo gerador congruencial multiplicativo com $M = 2^b$, $a = 2^n + 3$ e $2n \geq b$. Então*

1. $y_{j+1} - 6y_j + 9y_{j-1} = 0 \pmod{M}$.
2. Seja $u_j = y_j/M$. Então para todo $j = 1, 2, \dots$ temos que

$$(u_{j+1}, u_j, u_{j-1}) \in \left\{ 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1 \text{ e } z - 6y + 9x = r \right. \\ \left. \text{com } r = -5, -4, \dots, 8, 9 \right\}.$$

Demonstração: Para cada $j = 1, 2, \dots$ seja r_j inteiro tal que $x_j = (2^n + 3)x_{j-1} + r_j 2^b$. Então

$$x_{j+1} = (2^n + 3)x_j + r_{j+1}2^b = (2^n + 3)((2^n + 3)x_{j-1} + r_j 2^b) + r_{j+1}2^b$$

$$\begin{aligned}
&= (2^n + 3)(2^n + 3)x_{j-1} + ((2^n + 3)r_j + r_{j+1})2^b \\
&= 2^{2n}x_{j-1} + 6 \cdot 2^n x_{j-1} + 9x_{j-1} + ((2^n + 3)r_j + r_{j+1})2^b \\
&= 6((2^n + 3)x_{j-1}) - 9x_{j-1} + (2^{2n-b}x_{j-1} + (2^n + 3)r_j + r_{j+1})2^b \\
&= 6x_j - 9x_{j-1} + (-6r_j + 2^{2n-b}x_{j-1} + (2^n + 3)r_j + r_{j+1})2^b,
\end{aligned}$$

donde deduz-se a parte 1. A parte 2 segue de observar que (u_{i+1}, u_i, u_{i-1}) deve ser uma terna no cubo tri-dimensional unitário tal que (pela parte 1) $u_{i+1} - 6u_i + 9u_{i-1}$ é um inteiro.

A parte 2 da última Proposição diz que todas as ternas (u_{i+1}, u_i, u_{i-1}) estão cobertas por apenas 15 planos dentro do cubo tri-dimensional unitário, o que é muito ruim para várias aplicações da simulação estocástica.

Mesmo sabendo que toda seqüência gerada por um gerador congruencial linear tem uma estrutura de rede, é de se esperar que com uma escolha adequada dos parâmetros a , c e M essa estrutura não seja muito relevante desde o ponto de vista prático, pelo menos para dimensões da ordem menor ou igual a 6, digamos. Dessa maneira, os resultados de uma simulação não veriam-se afetados.

Fica assim válida a questão prática da escolha dos parâmetros quando temos resolvido usar um gerador congruencial linear. Para tanto, vários pesquisadores têm seguido o seguinte procedimento, que é simples e direto, embora um tanto demorado:

1. Escolher valores a , c e M dos quais saiba-se fornecem um gerador congruencial linear de período razoável.
2. Submeter as seqüências geradas pelo gerador com os parâmetros escolhidos em 1 a vários testes empíricos como os que serão estudados mais adiante. Documentar os resultados.
3. Submeter o gerador a testes teóricos. Em particular, o teste espectral de Coveyou e MacPherson [34] é usado atualmente devido a sua reconhecida capacidade para distinguir geradores bons de geradores ruins. Documentar os resultados.
4. Na medida que apareçam testes novos e mais severos, por exemplo os *stringent tests* sugeridos em [97], submeter os geradores em uso a esses testes e... documentar os resultados.

3.3.1 Geradores Congruenciais Lineares Recomendados

Não é possível recomendar gerador nenhum como sendo *o melhor* em sentido absoluto. Levando em consideração o dito no item 4 acima, geradores que são recomendáveis hoje poderiam deixar de sê-lo num futuro. Os geradores apresentados a seguir são recomendados por Lewis e Orav [88], e têm sido testados extensivamente mostrando propriedades razoavelmente boas. Porém, é muito importante levar em consideração certos aspectos computacionais referentes à velocidade e à precisão da aritmética do sistema computacional (computador, sistema operacional, linguagem, etc.) sob onde estiver implementado o gerador em questão. Alguma coisa sobre esse assunto pode ser vista numa próxima Secção neste Capítulo.

3.3.1.1 Gerador 1

Usado no pacote IMSL, foi sugerido por Lewis, Goodman e Miller [87] e os resultados dos testes realizados sobre ele estão em [82,83]. É o gerador congruencial linear multiplicativo ($c = 0$) com $a = 16.807$ e $M = 2^{31} - 1$. Entre as suas vantagens: é fácil de implementar em linguagens de alto nível, o seu uso está muito espalhado, o seu período é $2^{31} - 2$ ($= 2.147.483.646$) o que é suficiente para a maioria das aplicações. Devido a que é muito usado, se fizermos um trabalho de simulação nele baseados, nossos resultados terão maior chance de serem aceitos.

A sua principal desvantagem é que tem mostrado problemas com a sua estrutura de rede em dimensões altas

3.3.1.2 Geradores "Ótimos"

Foram recomendados por Fishman e Moore [44] depois de um estudo exhaustivo incluindo mais de 267 milhões de possíveis geradores com $M = 2^{31} - 1$ (entre eles o Gerador 1 acima) e os mais severos testes até a data de publicação do trabalho [97].

Todos os geradores recomendados são congruenciais lineares multiplicativos ($c = 0$) com $M = 2^{31} - 1$ e:

$$\begin{aligned} a &= 950.706.376 \\ &= 742.938.285 \\ &= 1.226.874.159 \\ &= 62.089.911 \\ &= 1.343.714.438. \end{aligned}$$

Todos eles têm período $2^{31} - 2$, e a principal desvantagem é que ay_i pode não ser computada corretamente mesmo em precisão dupla, em certos computadores.

3.4 Algoritmos de Registro de deslocamento

Foram propostos por Golomb em 1967 [51]. Operam sobre vetores binários (0 ou 1) pseudo-aleatórios da mesma maneira que os congruenciais o fazem sobre inteiros. Para obter-se um número em $[0, 1)$ deve-se converter o vetor binário em um inteiro, e então dividir este inteiro por 2^n se os vetores binários são de dimensão n .

Daremos apenas definições relativas a estes geradores e alguns exemplos. O leitor interessado em maiores detalhes pode consultar, por exemplo, as referências [27,90,88,97,108,110].

Na verdade, ainda está em discussão a performance destes geradores. Assim, [90] apresentam resultados de testes empíricos que avaliam uma razoavelmente boa performance. No entanto, [97] surpreende-se pelo fato de que ainda tenha pessoas que usam esses geradores *tão ruins*, e baseia as suas observações nos resultados de outros testes. Os autores deste trabalho têm usado um tipo de geradores desta classe, e têm em andamento um trabalho de testes e comparações com os congruenciais visando analisar as suas propriedades com maior detalhe.

Neste trabalho, incluímos referências a esses geradores para que o leitor saiba da sua existência.

Notação: Por *bit* entenderemos um qualquer dos inteiros "0" ou "1". Por *operação binária* entendemos qualquer operação aritmética realizada entre bits módulo 2. Assim, por exemplo, soma binária $a \oplus b = (a + b) \bmod 2$ com a e b bits; produto binário $a \otimes b = ab \bmod 2$ com a e b bits, etc. Finalmente, *vetor binário n -dimensional* será um vetor (b_0, \dots, b_{n-1}) onde b_i é um bit para todo $i = 0, \dots, n-1$. Se (b_0, \dots, b_{n-1}) é um vetor binário poremos $y = (b_0, \dots, b_{n-1})_2$ onde y é o inteiro definido por

$$y = 2^{n-1}b_0 + 2^{n-2}b_1 + \dots + 2^1b_{n-2} + 2^0b_{n-1}.$$

Logo, a operação de conversão de um vetor binário para inteiro e viceversa fica assim definida.

O exemplo a seguir ilustra o algoritmo típico destes geradores. Suponhamos querer gerar vetores binários 5-dimensionais (equivalentemente inteiros entre 1 e $2^5 - 1 = 31$), então começando com um vetor $(b_0, b_1, b_2, b_3, b_4)$ (binário) chamado "semente" geramos um novo vetor fazendo:

1. Escolha da semente: $(b_0, b_1, b_2, b_3, b_4)$.
2. Deslocar a semente dois posições para a esquerda: $(b_2, b_3, b_4, 0, 0)$.
3. Fazer uma soma binária dos vetores de 1 e 2: $(b_0 \oplus b_2, b_1 \oplus b_3, b_2 \oplus b_4, b_3, b_4)$.
4. Deslocar o vetor anterior três posições para a direita obtendo: $(0, 0, 0, b_0 \oplus b_2, b_1 \oplus b_3)$.
5. Somar os dois vetores de 3 e 4 obtendo o novo vetor $(b_0 \oplus b_2, b_1 \oplus b_3, b_2 \oplus b_4, b_0 \oplus b_2 \oplus b_3, b_1 \oplus b_3 \oplus b_4)$.

Por exemplo:

1. Semente: $(1, 1, 1, 1, 1)$.
2. Deslocar 2 à esquerda: $(1, 1, 1, 0, 0)$.
3. Somar: $0, 0, 0, 1, 1$.
4. Deslocar 3 à direita: $(0, 0, 0, 0, 0)$.
5. Novo vetor $(0, 0, 0, 1, 1)$.

Tomando o "novo vetor" como "semente" e aplicando as etapas 2, 3 e 4 obteríamos outro "novo":

1. $(0, 0, 0, 1, 1)$.
2. $(0, 1, 1, 0, 0)$.
3. $(0, 1, 1, 1, 1)$.
4. $(0, 0, 0, 0, 1)$.
5. $(0, 1, 1, 1, 0) \dots$

e assim por diante (veja o Exercício 3.10).

Vemos assim que temos três parâmetros a escolher: n (comprimento do vetor), número de posições a deslocar para a esquerda e número de posições a deslocar para a direita. A popularidade deste algoritmo talvez se deva à facilidade (aparente) de implementá-lo.

Em geral, o gerador pode caracterizar-se pelo uso de uma matriz binária T que é usada para multiplicar (operação binária) o vetor semente b_0 (binário) para produzir o vetor binário b_0T . Os vetores pseudo-aleatórios sucessivos são então gerados por novas multiplicações à direita pela matriz T (aqui, os “vetores” são pensados como “matrizes linha”, quer dizer vetores n -dimensionais como matrizes $1 \times n$): ficando então a sequência gerada na forma $b_0T, b_0T^2, b_0T^3, b_0T^4, \dots$

Todas as propriedades destes geradores podem ser obtidas como resultados da álgebra de matrizes binárias. Pelo menos as propriedades teóricas, já que as *práticas* tem que ser descobertas usando testes.

3.5 Outros Geradores

3.5.1 Geradores de Fibonacci

O leitor é convidado a lembrar a definição da Série de Fibonacci, para entender o por que do nome dos geradores definidos a seguir.

Definição 3.5.1 *Seja \mathcal{X} um conjunto e $\odot: \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{X}$ uma operação (em vez de $\odot(x, y)$ poremos $x \odot y$). Diz-se que uma sequência (ξ_1, ξ_2, \dots) em \mathcal{X} é gerada por um gerador de Fibonacci com parâmetros $r \geq 1, s \geq 1, r \neq s$ e operador \odot se*

$$\xi_i = \xi_{i-r} \odot \xi_{i-s} \text{ para todo } i \geq \max(r, s) + 1.$$

Exemplos:

1. $\mathcal{X} = \{0, 1, \dots, M-1\}$, $M \geq 2$ inteiro e \odot é a soma, a multiplicação ou a subtração módulo M .
2. $\mathcal{X} = \{0, 1\}^n$, com $n \geq 2$ inteiro e \odot é a soma, a multiplicação, a subtração, ou o “or” (inclusivo ou exclusivo) entre vetores n -dimensionais de bits.
3. $\mathcal{X} = [0, 1]$ e \odot é a soma ou a subtração módulo 1.

Os primeiros trabalhos sobre estes geradores, segundo o conhecimento dos autores, foram os citados nas referências [55,92]. Eles foram esquecidos devido a que os casos particulares tratados não conduziram a geradores com boas propriedades. Porém, na última década, foram resuscitados em [97], onde são mostradas características destes geradores que abrem um campo muito promissor para as pesquisas. A principal dificuldade está na escolha adequada dos elementos envolvidos: r, s, \odot e as sementes $\xi_1, \dots, \xi_{\max(r,s)}$.

3.6 Combinação e Embaralhado de Geradores

Certas modificações das saídas de geradores simples podem fornecer *melhores* seqüências. Alguns resultados teóricos podem ver-se, por exemplo nas referências [25,112,97]. Neste último trabalho prova-se um resultado donde deduz-se que os períodos dos geradores individuais podem ser aumentados com o uso das combinações.

Um gerador do tipo *combinação de geradores congruenciais* que apresenta propriedades notáveis e fácil de implementar, embora ainda não exhaustivamente testado, é o apresentado em [127,126]. A desvantagem mais importante é que o tempo de geração de um gerador “combinado” é maior que o correspondente aos geradores simples que o definem. Além disso, pouco se conhece em torno das estruturas resultantes das combinações. Daí que certos autores [110] sejam relutantes ao seu uso.

Outra modificação da saída de um gerador consiste em *embaralhar* essa saída antes da escolha do número que, finalmente, será usado. Em termos gerais, os algoritmos que implementam dita técnica consistem em, essencialmente, as seguintes etapas: sejam G_1 um gerador que produz uma seqüência em $[0,1]$, G_2 um gerador que produz uma seqüência em $\{1, \dots, N\}$ (N inteiro), ambas as duas uniformemente aleatórias. Então

1. Gerar uma tabela, digamos x_1, \dots, x_N com G_1 . Seja $i = 1$.
2. Gerar v com G_2 .
3. Definir $u_i = x_v$.
4. Gerar x com G_1 e definir $x_v := x$.
5. Se mais u_i 's são necessários, voltar ao passo 2 com $i = i + 1$.

O método descrito acima foi sugerido em [94], e uma implementação usando o gerador congruencial linear multiplicativo com $a = 16807$ e $m = 2^{31} - 1$ aparece em [91]. Um outro esquema de embaralhado pode ver-se em [11]. Finalmente, os pacotes [1,6] fornecem sub-rotinas que implementam algumas dessas técnicas.

3.7 Considerações Computacionais

No início deste Capítulo dissemos que não estudariamos neste trabalho questões referentes à portabilidade. Porém, um aspecto da portabilidade deve ser destacado:

O computador que estejamos usando deve ser capaz de computar corretamente o algoritmo de geração que escolhemos.

Essa afirmação parece óbvia, mas não é se levamos em consideração que existem alguns problemas ocultos. Por exemplo:

3.7.1 Problemas com o tamanho do registro dos números

Consideremos o Gerador 1 da Secção 3.3:

$$y_{i+1} = 16807y_i \bmod (2^{31} - 1).$$

Dado que a sequência assim construída está contida no conjunto $\{1, \dots, 2^{31} - 2\}$ parece que poderíamos codificar o algoritmo de geração correspondente usando apenas inteiros de 32 bits. Porém, se $y_i = 2^{31} - 2$ teríamos que $16807y_i$ não cabe num registro desse tamanho (32 bits); assim, quando tentemos fazer esse produto teremos um erro de "overflow". Se o compilador em uso não estiver desenhado para avisar a ocorrência de tais erros, obteríamos inadvertidamente uma sequência que não se corresponde às geradas verdadeiramente pelo gerador em questão. Assim, nada poderíamos dizer sobre as propriedades da sequência obtida.

Uma versão correta do algoritmo deveria basear-se sobre operações em aritmética real em precisão dupla; veja, por exemplo, [46].

O mesmo tipo de problemas pode apresentar-se quando tenta-se a implementação de outros geradores, mesmo não sendo eles do tipo congruencial (veja o Exercício 3).

3.7.2 Problemas com a Precisão

Mesmo não ocorrendo erros de overflow, os compiladores —particularmente os disponíveis para micro-computadores— apresentam sérias deficiências no tocante à precisão de diversas operações aritméticas.

Por exemplo, se usarmos o gerador multiplicativo congruencial com $a = 950.706.376$ —um dos geradores ótimos (final da Secção 3.3)— e se $y_i = 2^{31} - 2 = 2.147.483.646$ então $a \cdot y_i$ em aritmética de precisão dupla não dá overflow trabalhando com reais de 64 bits (por exemplo num PC/AT ou XT com co-processador 80287 ou 8087, ou num micro da linha 386 com co-processador correspondente).

No entanto, usando um MICROTEC 386 com sistema operacional DOS 4.01, linguagem FORTRAN 77 e compilador MICROSOFT MS Versão 4 esse produto é

$$2.041.626.394.607.927.000,$$

sendo que o verdadeiro valor é (calculado a mão)

$$2.041.626.394.607.926.896$$

... e sabemos que outros números diferentes são obtidos para esse mesmo produto usando diferentes equipamentos, sistemas, linguagens e/ou compiladores.

Às vezes, embora muito infreqüentemente, é possível ter acesso à documentação sobre as limitações das aritméticas implementadas, que deveriam formar parte importante do conjunto dos manuais fornecidos pelo vendedor do equipo computacional em uso. Se não, seria bom tentar averiguar alguma coisa sobre a precisão do seu computador.

Finalmente:

3.7.3 Problemas com a Velocidade dos Cálculos

Este item é muito importante quando deseja-se realizar simulações de tamanho considerável, e resulta de muito valor quando é precisa alta eficiência.

O ótimo seria codificar o gerador numa linguagem *assembler*. Mas se isso não é possível, bem podem usar-se linguagens de alto nível em forma eficiente. As características e especificações dependerão do equipamento, da linguagem, etc., mas algumas dicas podem ser de ajuda.

1. Usar constantes definidas; por exemplo $A=2147483647.000$ e não valores calculados: $A=2.000**31-1.000$.
2. Usar operações mais rápidas; por exemplo checar se o equipamento em uso faz mais rapidamente $A=2147483647.000$ e $B=X/A$, do que ter associado previamente à uma constante, digamos C o valor $1./2147483647.000$, e então fazer $B=C*X$.
3. Se o trabalho for feito num micro-computador, então *sempre* usar co-processador aritmético nas operações de ponto flutuante.

3.8 Testes de Geradores

Esta área constitui um campo de pesquisa muito ativo nos tempos atuais, e é extremamente importante para toda pessoa envolvida em trabalhos de simulação. Porém, não é fácil e leva muito tempo realizar corretamente tais testes. Por isto não é aconselhável fazer testes das seqüências de uniformes toda vez que se faça um trabalho de simulação: mais razoável é saber de cada gerador em uso a sua performance respeito a vários desses testes, e saber em que medida essa performance pode afetar os resultados finais do nosso trabalho. A partir daí, se for necessário apurar certos detalhes, pode procurar-se na literatura especializada a informação relevante.

Com essa idéia em mente é que neste Secção estudaremos alguns dos testes mais comuns (e outros nem tanto) aos que costuma-se someter as seqüências numéricas geradas por geradores. Estes são os chamados *testes empíricos*. Aplicam-se a amostras da seqüência completa que pode obter-se com um dado gerador (aquele que deseja-se avaliar) e começando com diferentes sementes. Obviamente, o tamanho das seqüências a serem testadas será também um fator de peso na avaliação dos resultados dos testes. O que procura-se verificar é se ditas seqüências apresentam estruturas indesejáveis para o uso que delas faremos no restante do trabalho. De fato, o número potencial de tais testes é infinito, pois infinito é o escopo de diferentes tipos de desvios da teórica *uniformemente independentemente e identicamente distribuída*. Do ponto de vista prático, é suficiente considerar alguns testes clássicos como χ^2 , Kolmogorov-Smirnov, etc. aplicáveis não apenas à distribuição uniforme, outros específicos para dita distribuição (*gaps*, *serial*, etc.) e... talvez alguns mais severos (ver [97]).

Existe uma outra classe de testes, chamada de “testes teóricos”. Eles têm a ver com a estrutura no espaço n -dimensional das seqüências $(u_i, u_{i+1}, \dots, u_n)$, $(u_{i+1}, \dots, u_{n+1})$,... formadas a partir da seqüência completa (u_1, u_2, \dots) gerada por um dado gerador. Essa estrutura, nos geradores do tipo congruencial e os de registro de deslocamento, depende dos parâmetros que definem o gerador e daí que não seja necessário conhecer toda a seqüência (u_1, u_2, \dots) —embora subsistam problemas relacionados às sementes.

Mesmo reconhecendo a importância dos testes teóricos, não estudaremos nenhum deles neste trabalho pois hoje em dia ainda é uma área muito pouco desenvolvida e a sua aplicação é restrita apenas a geradores de certo tipo (por exemplo, pouco se conhece sobre este tipo de testes para geradores “mixtos” ou para os de “embaralhado”). Deixamos, assim, ao leitor interessado uma brevíssima lista de referências sobre o tema: [12,34,44,62,78].

Voltando aos testes empíricos, veremos com algum detalhe os seguintes: χ^2 , Kolmogorov-Smirnov, Serial, do Intervalo (*gap*), Póker, das Corridas (*runs*), das Permutações, do Máximo-

de- T , e da Dependência Linear.

Até o final desta Secção, suporemos que U_1, \dots, U_N são valid com distribuição $\mathcal{U}(0,1)$, u_1, \dots, u_N são os N números obtidos como saída do gerador que quer-se testar, H_0 (ou hipótese nula) é a hipótese de que u_1, \dots, u_N é uma amostra de U_1, \dots, U_N . Seja $t: \mathbb{R}^N \rightarrow \mathbb{R}$ uma função, $p_t(u_1, \dots, u_N)$ (ou simplesmente p_t) será o p -valor da estatística $t(U_1, \dots, U_N)$ sob H_0 , isto é

$$p_t(u_1, \dots, u_N) = \text{Pr}(t(U_1, \dots, U_N) \geq t(u_1, \dots, u_N)).$$

Lembremos que se $p_t(u_1, \dots, u_N) = \alpha$, então diz-se que o teste de H_0 definido por t rejeita H_0 ao nível α ou, equivalentemente, “a probabilidade de errar ao rejeitar H_0 é menor ou igual que $p_t(u_1, \dots, u_N)$ ”. Daí que rejeitaremos H_0 se p_t é “pequeno”, e não rejeitaremos se p_t é “grande” (um valor razoável para considerar p_t pequeno é $p_t < 0.01$).

Notemos que para calcular p_t é necessário conhecer a distribuição de $t(U_1, \dots, U_N)$, coisa que nem sempre é possível; todavia, pode-se substituir a distribuição exata de $t(U_1, \dots, U_N)$ pela sua distribuição assintótica (se ela existir). Do ponto de vista prático, não devemos preocuparnos demasiadamente, pois a maioria das vezes o valor de p_t pode obter-se de tabelas adequadas e, freqüentemente, a mesma sub-rotina que calcula $t(u_1, \dots, u_N)$ fornece o valor p_t . Assim acontece com as implementações em pacotes como IMSL e NAG.

3.8.1 Teste χ^2

Seja $2 \leq k < N$. Para cada $i = 1, \dots, k$ sejam

$$f_i = \frac{1}{N} \# \left\{ 1 \leq j \leq N : x_j \in \left[\frac{i-1}{k}, \frac{i}{k} \right) \right\}$$

$$t = \sum_{1 \leq i \leq k} \frac{(f_i - \frac{N}{k})^2}{\frac{N}{k}}.$$

Temos que $p_t \approx \text{Pr}(Y \geq t)$, com $Y \sim \chi_{k-1}^2$ (denotamos “aproximadamente igual” com “ \approx ”). O k deve guardar uma certa relação com N ; por exemplo, se $N = 1000$ então $k = 59$ (ver a referência [6] para mais detalhes).

3.8.2 Teste de Kolmogorov-Smirnov

Seja $(u_{(1)}, \dots, u_{(N)})$ o vetor (u_1, \dots, u_N) com os elementos permutados em forma tal que

$$u_{(1)} \leq \dots \leq u_{(N)}.$$

Por definição sejam $u_{(0)} = 0$ e $u_{(N+1)} = 1$. Seja $\hat{F}: \mathbb{R} \rightarrow [0, 1]$ dada por

$$\hat{F}(x) = \frac{i}{N} \text{ se } u_{(i)} \leq x < u_{(i+1)}, \quad i = 0, 1, \dots, N$$

e seja $t = \max\{|\hat{F}(x) - x| : x \in [0, 1]\}$.

Pode ver-se, por exemplo em [52] ou na documentação da sub-rotina correspondente (G08CAF no NAG [3]) que, para N "grande" e sob H_0 , $p_t \approx 1 - L(\sqrt{N}t)$, onde

$$L(z) = \begin{cases} 0 & \text{se } z \leq 0.27 \\ \frac{\sqrt{2\pi}}{z} \sum_{1 \leq k \leq 3} \exp\left\{-\frac{(2k-1)^2 \pi^2}{8z^2}\right\} & \text{se } 0.27 < z < 1 \\ 1 - 2 \sum_{1 \leq k \leq 4} (-1)^{k-1} \exp\{-2k^2 z^2\} & \text{se } 1 \leq z < 3.1 \\ 1 & \text{se } z \geq 3.1. \end{cases}$$

3.8.3 Teste Serial

Suponhamos que N é par (mais precisamente, $N = 2n$). Seja $2 \leq k < n$. Para cada $1 \leq i, j \leq k$ definimos

$$\begin{aligned} C(i, j) &= \left[\frac{i-1}{k}, \frac{i}{k}\right) \times \left[\frac{j-1}{k}, \frac{j}{k}\right) \\ f_{(i,j)} &= \#\{1 \leq \ell \leq n: (u_{2\ell-1}, u_{2\ell}) \in C(i, j)\} \\ t &= \sum_{1 \leq i \leq k} \sum_{1 \leq j \leq k} \frac{(f_{(i,j)} - \frac{n}{k^2})^2}{\frac{n}{k^2}}. \end{aligned}$$

Pode ver-se em [35] que para N grande e sob H_0 tem-se que $p_t \approx \text{IPr}(Y \geq t)$, com $Y \sim \chi_{k^2-1}^2$.

IMSL [6] fornece sub-rotinas (GTPR e GTPST) que implementam este teste para as seqüências $(u_1, u_2), (u_2, u_3), (u_3, u_4), \dots$ no lugar de $(u_1, u_2), (u_3, u_4), \dots$ segundo Good [54, 53].

3.8.4 Teste do Intervalo

Sejam $0 \leq \alpha < \beta \leq 1$. O teste considera o comprimento das seqüências consecutivas da forma (u_j, \dots, u_{j+r}) tais que

$$\begin{aligned} u_{j+i} &\notin (\alpha, \beta) \text{ para } i = 0, \dots, r-1 \\ \text{e } u_{j+r} &\in (\alpha, \beta). \end{aligned}$$

O algoritmo a seguir implementa uma das variantes deste teste, também implementada em [3]. Para mais detalhes sobre este e outros testes, uma boa referência é [78]. Seja T inteiro positivo fixo (usualmente $T = 9$).

1. Sejam $j = 0, z_0 = 0, \dots, z_T = 0$ e $n = 0$.
2. $r = 0$.
3. $j = j + 1$.
4. Se $j \leq N$ seguir em 5. Se não terminar.
5. Se $u_j \in (\alpha, \beta)$ fazer $n = n + 1$ e seguir em 6. Se não fazer $r = r + 1$ e seguir em 3.
6. Se $r \geq T$ fazer $z_T = z_T + 1$ e seguir em 2. Se não fazer $z_r = z_r + 1$ e seguir em 2.

Seja $t = \sum_{r=0}^T \frac{(z_r - np_r)^2}{np_r}$, onde

$$\begin{aligned} p_0 &= p = \beta - \alpha \\ p_r &= p(1-p)^r \text{ se } 1 \leq r \leq T-1 \\ p_T &= (1-p)^T. \end{aligned}$$

Para N grande e sob H_0 vale que $p_t(u_1, \dots, u_N) \approx \text{IPr}(Y \geq t)$, com $Y \sim \chi_T^2$.

Como acontece toda vez que aplicamos testes baseados na estatística χ^2 , devemos verificar que o número de pontos em cada categoria seja suficientemente grande; por exemplo neste caso $np_r \geq 5$ para $r = 0, \dots, T$.

3.8.5 Teste Póker

A descrição a seguir corresponde ao algoritmo implementado pela sub-rotina GTPOK do IMSL [6]. Para cada $j = 1, \dots, N$ seja $(b_{j,k})_{k \geq 1}$ uma sequência em $\{0, 1\}$ (bits) tal que

$$u_j = \sum_{1 \leq k < \infty} b_{j,k} 2^{-k},$$

(Nota: se existir $k_0 \geq 2$ tal que $b_{j,k_0-1} = 0$ e $b_{j,k} = 1 \forall k \geq k_0$ redefinimos os b 's da forma $b_{j,k_0-1} = 1$ e $b_{j,k} = 0 \forall k \geq k_0$). Em outras palavras, $b_{j,k}$ é o k -ésimo bit da parte fracionária do registro em ponto flutuante de u_j .

Sejam $1 \leq i_1 < i_2$ e $r = i_2 - i_1 + 1$,

$$C = \{(b_{j,i_1}, \dots, b_{j,i_2}) \in \{0, 1\}^r : j = 1, \dots, N\}.$$

Para cada $j = 1, \dots, N$ e cada $\mathbf{b}_j = (b_{j,i_1}, \dots, b_{j,i_2})$ sejam $I_{0,j} = \#\{i_1 \leq i \leq i_2 : b_{j,i} = 0\}$ e $I_{1,j} = \#\{i_1 \leq i \leq i_2 : b_{j,i} = 1\}$. Consideremos os seguintes $k_1 = \lceil r/2 \rceil + 1$ sub-conjuntos de C :

$$\begin{aligned} C_1 &= \{\mathbf{b}_j : I_{0,j} = r \text{ ou } I_{1,j} = r\}, \\ C_2 &= \{\mathbf{b}_j : I_{0,j} = r-1 \text{ ou } I_{1,j} = r-1\}, \\ &\dots \\ C_{k_1} &= \{\mathbf{b}_j : I_{0,j} = r - \lceil r/2 \rceil \text{ ou } I_{1,j} = r - \lceil r/2 \rceil\}. \end{aligned}$$

Para cada $i = 1, \dots, k_1$ seja $Z(i) = \#\{j \leq N : \mathbf{b}_j \in C_i\}$. Pode ver-se que vale o resultado seguinte: se $(U_j)_{j \geq 1}$ são valid com distribuição $\mathcal{U}(0, 1)$ e se para cada $k = 1, 2, \dots$ o $b_{j,k}$ é o k -ésimo bit da parte fracionária de U_j , então

$$\text{IPr}(b_{j,k} = 1) = \text{IPr}(b_{j,k} = 0) = \frac{1}{2}.$$

Daí que seja de interesse comparar os valores $Z(i)$ com

$$\text{IPr}(i) = \binom{r}{i} \frac{1}{2^{r-1}},$$

para $i = 1, \dots, k_1$. Para fazer esta comparação emprega-se a estatística

$$t = \sum_{1 \leq i \leq k_1} \frac{(Z(i) - N\text{IPr}(i))^2}{N\text{IPr}(i)}$$

que, sob H_0 fornece um p -valor aproximado $p_t \approx \text{IPr}(Y \geq t)$, com $Y \sim \chi_{k_1-1}^2$.

Para a correta aplicação deste teste N , i_1 e i_2 devem ser tais que $N\text{IPr}(i) \geq 5$ para todo $i = 1, \dots, k_1$.

3.8.6 Testes das Corridas

Analisa o comprimento das sub-sequências ascendentes (ou descendentes) da sequência a ser testada. Suponhamos que a sequência começa assim: 0.415, 0.919, 0.800, 0.677, 0.594, 0.059, 0.444, 0.618, 0.429, 0.779, 0.788, 0.878, 0.914, 0.952 ...; então

1. 0.415, 0.919 é uma corrida ascendente de comprimento 2 (que se interrompe a causa de que o seguinte, 0.800, é menor que o anterior, 0.919).
2. 0.800 é uma corrida ascendente de comprimento 1.
3. 0.677 é uma corrida ascendente de comprimento 1.
4. 0.594 é uma corrida ascendente de comprimento 1.
5. 0.059, 0.444, 0.618 é uma corrida ascendente de comprimento 3.
6. 0.429, 0.779, 0.788, 0.878, 0.914, 0.952 é uma corrida ascendente de comprimento 6, etc.

Para cada $p = 1, 2, \dots, T-1$ (T usualmente é 6) sejam R_p o número de corridas ascendentes de comprimento p presentes na sequência sob teste (u_1, \dots, u_N) , e R_T^l o número de corridas ascendentes de comprimento igual ou maior a T .

Um dos *testes das corridas* mais populares compara as frequências observadas com as esperadas. Essas últimas estão calculadas, por exemplo, em [86]. Note-se que não é possível aplicar o teste χ^2 dado que as corridas consecutivas não são independentes.

Um teste bastante complexo é dado em Knuth [78, págs. 65 e seguintes]. Esse teste tem sido implementado em IMSL [1,6] nas sub-rotinas GTRN e RUNS, respectivamente, e em NAG [3].

Veremos a seguir um teste das corridas mais simplificado, sugerido em [78] e analisado com detalhe em [35]: consiste em, simplesmente, obter corridas independentes pulando o último elemento de cada sub-sequência. Assim, no exemplo anterior, somente são consideradas as corridas

1. 0.415, 0.919 (corrida de comprimento 2).
2. 0.677 (corrida de comprimento 1).
3. 0.059, 0.444, 0.618 (corrida de comprimento 3).
4. 0.779, 0.788, 0.878, 0.914, 0.952 (corrida de comprimento 5), etc.

Pode ver-se que sob H_0 a probabilidade de uma corrida de comprimento k é $k/(k+1)!$.

Vejamos agora a estatística e a probabilidade de distribuição sob H_0 correspondente a este teste simplificado. Sejam R_p para $p = 1, \dots, T-1$ e R'_T como antes, ν o número total de corridas ascendentes e

$$\begin{aligned} f_p &= R_p/\nu \text{ para } p = 1, \dots, T-1 \\ f_T &= R'_T/\nu \\ t &= \sum_{1 \leq p \leq T-1} \frac{(f_p - \frac{\nu\nu}{(p+1)!})^2}{\frac{\nu\nu}{(p+1)!}} + \frac{(f_T - \frac{\nu}{T!})^2}{\frac{\nu}{T!}}. \end{aligned}$$

O já citado [35] fornece a referência sobre a prova de que sob H_0 , $t \sim \chi^2_{T-1}$ e daí é fácil deduzir um teste.

3.8.7 Testes das Permutações

Divide-se a sequência (u_1, u_2, \dots) em n grupos de T elementos cada (deve tomar-se $N = nT$). Os elementos em cada grupo podem estar ordenados de $T!$ maneiras. A frequência com a qual cada ordenamento aparece é então comparada com a sua probabilidade teórica sob H_0 (isto é com $1/T!$).

Para fixar idéias, seja $T = 3$ e suponhamos $N = 3n$. Sejam

$$\begin{aligned} f_1 &= \#\{0 \leq j \leq n-1: u_{3j+1} < u_{3j+2} < u_{3j+3}\} \\ f_2 &= \#\{0 \leq j \leq n-1: u_{3j+1} < u_{3j+3} < u_{3j+2}\} \\ f_3 &= \#\{0 \leq j \leq n-1: u_{3j+2} < u_{3j+1} < u_{3j+3}\} \\ f_4 &= \#\{0 \leq j \leq n-1: u_{3j+2} < u_{3j+3} < u_{3j+1}\} \\ f_5 &= \#\{0 \leq j \leq n-1: u_{3j+3} < u_{3j+1} < u_{3j+2}\} \\ f_6 &= \#\{0 \leq j \leq n-1: u_{3j+3} < u_{3j+2} < u_{3j+1}\}. \end{aligned}$$

Consideremos a estatística

$$t = \sum_{1 \leq k \leq 6} \frac{(f_k - \frac{n}{3!})^2}{\frac{n}{3!}}.$$

É fácil verificar que sob H_0 $p_i \approx \text{IPr}(Y \geq t)$, com $Y \sim \chi^2_{T!-1}$. No caso do exemplo $T! - 1 = 5$.

3.8.8 Teste do Máximo-de- T

Como no teste anterior, a sequência a ser testada (u_1, \dots, u_N) é dividida em n grupos de T elementos cada um. Dentro de cada grupo escolhe-se o maior elemento e compara-se a distribuição empírica da sub-amostra com a distribuição teórica sob H_0 , isto é com

$$F(y) = y^T \Pi_{(0,1)}(y).$$

Na comparação recomenda-se usar o teste de Kolmogorov-Smirnov para amostras de uma população só.

Para fixar idéias, sejam $T = 3$ e suponhamos $N = 3n$. Para cada $j = 0, 1, \dots, n-1$ seja

$$v_j = \max(u_{3j+1}, u_{3j+2}, u_{3j+3}).$$

Consideremos o vetor v_0, \dots, v_{n-1} ordenado, isto é

$$v_{(0)} \leq v_{(1)} \leq \dots \leq v_{(n-1)}$$

e sejam

$$\begin{aligned} K^+ &= \sqrt{n} \max \left\{ \left(\frac{j+1}{n} - v_{(j)}^3 \right) : j = 0, \dots, n-1 \right\} \\ K^- &= \sqrt{n} \max \left\{ \left(v_{(j)}^3 - \frac{j}{n} \right) : j = 0, \dots, n-1 \right\}. \end{aligned}$$

O teste rejeita H_0 se $\max\{K^+, K^-\} \geq 1.5174$ ($\approx \sqrt{\frac{1}{2} \ln \frac{1}{0.01}}$) ou se $\min\{K^+, K^-\} \leq 0.07089$ ($\approx \sqrt{\frac{1}{2} \ln \frac{1}{0.99}}$).

É aconselhável fazer um gráfico simultâneo da distribuição empírica das v 's e da função $y \mapsto y^T \Pi_{(0,1)}(y)$. O leitor interessado poderá ver os detalhes deste e do teste anterior, por exemplo, em [78].

3.8.9 Teste da Dependência Linear

Uma maneira de testar a dependência entre os elementos da sequência (u_1, \dots, u_N) é analisar a dependência linear entre elementos separados; quer dizer, analisar sequências da forma $(u_1, u_{1+k}), (u_2, u_{2+k}), \dots$ para diferentes valores de k .

Para tanto, seja $k \geq 1$ fixo (usualmente escolhe-se $k = 1, 2$ ou 3). Definimos

$$C_k = \frac{1}{N-k} \sum_{1 \leq i \leq N-k} \left(u_i - \frac{1}{2} \right) \left(u_{i+k} - \frac{1}{2} \right).$$

Pode provar-se que, sob H_0 vale que $\mathbb{E}(C_k) = 0$ e $\text{Var}(C_k) = (144(N-k))^{-1}$. Para $N \gg k$ Fishman [43] recomenda comparar com os percentis da $\mathcal{N}(0,1)$ os percentis amostrais da estatística $t = 12C_k\sqrt{N-k}$.

3.9 Geradores Fornecidos por certos "Pacotes"

Às vezes pode não ser conveniente (ou diretamente impossível) escolher um gerador. Por exemplo, certos pacotes orientados à análise de dados tem incorporado um gerador e não dão a opção de usar geradores alternativos. Mas, todo usuário de um pacote deveria estar ciente do tipo de gerador nele implementado e, aplicando o estudado nas Seções anteriores, conhecer as limitações do mesmo.

Quando se trabalha diretamente com programas escritos por nós mesmos, ou usando pacotes de sub-rotinas como IMSL, NAG em linguagens computacionais como C, BASIC, FORTRAN ou PASCAL existe a possibilidade de escolher entre o gerador incorporado como

parte da linguagem (em FORTRAN tal coisa não existe) ou a implementação de um gerador tal como os recomendados no final da Secção 3.3, ou ainda testar o gerador que será usado.

Para ajudar nesta tarefa, concluiremos este Capítulo com uma síntese do que dizem Lewis e Orav [88] sobre vários pacotes e sistemas computacionais (software e hardware) em uso.

SAS: ver [4]. Usa os seguintes geradores

1. multiplicativo congruencial com $a = 397.204.094$ e $M = 2^{31} - 1$.
2. um embaralhado do gerador congruencial multiplicativo com $a = 16.807$ e $M = 2^{31} - 1$.

Como já vimos na Secção 3.6, ambos os dois possuem boas propriedades.

SPSS: ver [5]. Oferece o gerador congruencial multiplicativo com $a = 16.807$ e $M = 2^{31} - 1$.

APL: ver [71]; em várias das suas implementações usa o gerador do SPSS.

IMSL(a): ver [1]. Oferece duas sub-rotinas que implementam os geradores do SAS e também uma sub-rotina que implementa um embaralhado destas.

IMSL(b): ver [6]. Além dos geradores da versão anterior, oferece um outro gerador congruencial multiplicativo com $a = 950.706.376$

NAG: este pacote de sub-rotinas usa um gerador congruencial multiplicativo com $M = 2^{59}$ e $a = 13^{13}$. As origens deste gerador são pouco conhecidas, embora seja mencionado em [109,120]. Segundo [88] não é um gerador aconselhável devido a que a sua implementação requer aritmética inteira em quádrupla precisão na maioria dos computadores. Além disso, nada está documentado sobre as suas propriedades estatísticas. Em [110] algumas boas qualidades deste gerador são mencionadas, e num apêndice fornecem-se os fontes de programas em FORTRAN 77 que possibilitam a implementação correta de geradores deste tipo. Sem dúvida, se for possível usar esse pacote, o uso dessa sub-rotina (G05CAF) é recomendável.

Micros: os geradores que acompanham o software fornecido por vendedores de micro-computadores devem ser considerados como suspeitos. Pode-se ver alguma coisa em torno desse problema sobre geradores para micros da linha APPLE em [117]. Segundo [88] o gerador do IBM/PC BASIC é totalmente defeituoso.

DEC: Lewis e Orav [88] avisam que a DEC tem oferecido junto com o sistema VMS para os seus equipamentos VAX uma implementação da sub-rotina RANDU, da qual já vimos a sua ruim estrutura tri-dimensional (Proposição 3.3.3). Porém, Ripley [110] diz que a DEC para os VAX oferece o gerador congruencial mixto com $M = 2^{32}$, $a = 69.069$ e $c = 1$, que apresenta propriedades razoáveis (ver [99]). Em fim... cuidado ao usar esse equipamento!

GLIM3: ver [14]; oferece o gerador congruencial mixto com $M = 2^{35}$, $a = 8.404.997$ e $c = 1$. Algumas das suas propriedades podem ver-se em [110, pág. 39].

3.10 Exercícios

Nos exercícios onde pede-se realizar um programa de computador, pode-se utilizar qualquer linguagem que lhe seja familiar.

Exercício 3.1: Defina o método de von Neumann para números de 4 dígitos decimais e prove que se $X_n = (d_3 d_2 d_1 d_0)_{10}$ (isto é $d_j \in \{0, \dots, 9\}, 0 \leq j \leq 3$) com $d_3 = d_2 = 0$ então

1. $X_n > X_{n+1} > X_{n+2} \dots$ e
2. $\exists k \geq 1: X_{n+k} = 0$, portanto $X_{n+p} = 0 \forall p \geq k$.

Exercício 3.2: Usando o método do Exercício anterior veja o que se passa se existir $X_n = (d_3 d_2 d_1 d_0)_{10}$ com $d_0 = d_1 = 0$.

Exercício 3.3: (Precisa de computador) Usando o método de von Neumann para números de 5 dígitos decimais começando com $X_0 = 12345$ obtenha X_1, \dots, X_N (digamos $N \geq 50$). Faça um gráfico (X, Y) com os pontos $(X_1, X_2), (X_3, X_4), \dots$. Existe algum padrão evidente nesse gráfico que invalide a aleatoriedade da sequência obtida?

Exercício 3.4: O seguinte resultado estabelece uma condição necessária e suficiente para que um gerador multiplicativo com módulo $M = 2^k$ tenha período máximo 2^{k-1} (lembre o Teorema 3.3.4): a sequência gerada pelo gerador congruencial multiplicativo com módulo $M = 2^k$ atinge o período máximo 2^{k-1} se e somente se $a = 3 \bmod 8$ ou $a = 5 \bmod 8$, e y_0 é ímpar. Implemente o gerador da (u_0, u_1, \dots) em $(0, 1)$ dado por $u_i = 1013u_{i-1} \bmod 1$ em aritmética de ponto flutuante e (v_0, v_1, \dots) dado por $v_i = u_i/M, M = 2^{16}$ e (y_0, y_1, \dots) gerada por $y_i = 1013y_{i-1} \bmod M$. Quais são os períodos obtidos em ambas as sequências?

Exercício 3.5: Computar as sequências obtidas (y_1, y_2, \dots) a partir de $y_i = (ay_{i-1} + c) \bmod 64$ com y_0 à sua escolha e $a = 29$ e $c = 17$, $a = 9$ e $c = 1$, $a = 13$ e $c = 0$ e, finalmente, $a = 11$ e $c = 0$.

Exercício 3.6: Encontre os períodos dos geradores congruenciais multiplicativos com $M = 67$ e $a = 10$, $a = 12$, $a = 16$ e $a = 18$.

Exercício 3.7: Faça gráficos (u_i, u_{i+1}) e (u_{2i}, u_{2i+1}) com $u_i = y_i/M$ para as sequências (y_1, y_2, \dots) geradas nos Exercícios 3.5 e 3.6.

Exercício 3.8: Considere o seguinte gerador de Fibonacci: $\mathcal{X} = (0, 1)$, \odot é a soma módulo 1; isto é, seja (u_1, u_2, \dots) uma sequência em $[0, 1]$ tal que

$$u_i = (u_{i-1} + u_{i-2}) \bmod 1, \quad i \geq 2.$$

Prove que jamais pode acontecer $u_{i-2} < u_i < u_{i-1}$. Porém, prove que se (U_1, U_2, \dots) são variáveis com distribuição $\mathcal{U}(0, 1)$, então $\mathbb{P}(U_{i-2} < U_i < U_{i-1}) = \frac{1}{6} \forall i$.

Exercício 3.9: Considere o seguinte gerador do tipo “combinação linear de geradores congruenciais” [127]: sejam

1. (x_1, x_2, \dots) gerada por $x_{i+1} = (171x_i) \bmod 30269, i \geq 1, 2, \dots$
2. (y_1, y_2, \dots) gerada por $y_{i+1} = (172y_i) \bmod 30307, i \geq 1, 2, \dots$

3. (z_1, z_2, \dots) gerada por $z_{i+1} = (170z_i) \bmod 30323$, $i \geq 1, 2, \dots$

com sementes x_0 , y_0 e z_0 respectivamente.

Para cada $i \geq 1, 2, \dots$ seja

$$u_i = \left(\frac{x_i}{30269} + \frac{y_i}{30307} + \frac{z_i}{30323} \right) \bmod 1.$$

1. Faça um programa que implemente este gerador das seqüências (u_1, u_2, \dots) .
2. Prove que se U e V são variáveis $\mathcal{U}(0, 1)$ então $W = (U + V) \bmod 1 \sim \mathcal{U}(0, 1)$. Daí deduz-se facilmente que as seqüências das u_i acima são razoavelmente seqüências de variáveis com distribuição $\mathcal{U}(0, 1)$.
3. Submeta as seqüências geradas pelo programa desenvolvido acima a vários dos testes estudados neste Capítulo.

Exercício 3.10: Considere o exemplo da Secção 3.4. Sejam

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad L = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad R = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(a) Prove que o algoritmo dado na Secção 3.4 é

$$\text{"vetor novo"} = \text{"semente"}(I + L^2)(I + R^3).$$

(b) Obtenha a seqüência $b_0T, b_0T^2, b_0T^3, b_0T^4, \dots, b_0T^{32}, \dots$. Observe a partir de que potência de T a seqüência começa a repetir-se.

Capítulo 4

Métodos gerais para geração de variáveis aleatórias não uniformes

4.1 Introdução

Seja F uma função de distribuição definida sobre \mathbb{R} e (X_1, X_2, \dots) variável com distribuição comum F . Queremos gerar uma realização (x_1, x_2, \dots) da (X_1, X_2, \dots) . Para tanto suponhamos ter gerado uma realização (u_1, u_2, \dots) de uma sequência (U_1, U_2, \dots) de variável com distribuição comum $\mathcal{U}(0, 1)$. Neste capítulo estudaremos algoritmos para transformar (u_1, u_2, \dots) numa (x_1, x_2, \dots) .

Em geral, dada F disporemos de vários algoritmos razoáveis para fazer o que desejamos, mas, geralmente, não existirá “o melhor” algoritmo. A escolha dependerá de um compromisso entre vários critérios.

Entre esses critérios podemos considerar:

1. O método deverá ser fácil de se entender e de programar.
2. A execução do programa que implementa o método terá que ser razoavelmente rápida. Embora não é aconselhável dar demasiada ênfase neste assunto na maioria dos trabalhos de simulação de tamanho médio. Ainda em aqueles de certo tamanho considerável, o custo computacional não é o maior entre outrem. Veja, por exemplo, [10].
3. Os algoritmos não deverão acentuar as deficiências do algoritmo usado na geração da (u_1, u_2, \dots) .

4.1.1 Exemplos Simples

Exemplo 4.1.1: $F = \mathcal{N}(0, 1)$. A seguir um método simples para gerar variáveis normais padrão, isto é $\mathcal{N}(0, 1)$, quando não é necessário muita qualidade.

Suponhamos querer gerar (x_1, \dots, x_N) .

1. Seja $i = 1$.

2. Gerar $u_{i,1}, \dots, u_{i,12}$
3. Definir $x_i = \sum_{k=1}^{12} u_{i,k} - 6$.
4. Fazer $i = i + 1$.
 - Se $i \leq N$, então prosseguir em 2.
 - Se $i > N$, então terminar.

Esse algoritmo baseia-se no fato de que se U_1, \dots, U_{12} são variáveis $\mathcal{U}(0,1)$, então $X = \sum_{i=1}^{12} U_i - 6$ tem média nula e variância unitária. Por outro lado, o Teorema Central do Limite diz que X terá uma distribuição *aproximadamente* $\mathcal{N}(0,1)$.

Exemplo 4.1.2: $F = \mathcal{E}(\beta)$. Lembremos que uma variável aleatória X tem distribuição exponencial com média β ($\beta > 0$), se X tem a sua densidade dada por

$$f_X(x) = \beta^{-1} \exp\{-x/\beta\} \mathbb{I}_{(0,\infty)}(x).$$

A seguinte proposição sugere um método bem simples para gerar (x_1, \dots, x_N) a partir de (u_1, \dots, u_N) .

Proposição 4.1.1 *Se U é uma variável aleatória com distribuição $\mathcal{U}(0,1)$, então a variável aleatória $X = -\beta \ln(U)$ tem distribuição $\mathcal{E}(\beta)$.*

Exemplo 4.2.3: $F = \chi_n^2$.

•Caso 1: Quando n é par, isto é $n = 2m$ com $m \geq 1$, a seguinte Proposição sugere um método bem simples para gerar (x_1, \dots, x_N) com distribuição χ_{2m}^2 sabendo gerar exponenciais

Proposição 4.1.2 *Se Y_1, \dots, Y_m são variáveis com distribuição $\mathcal{E}(1/2)$, então $X = \sum_{i=1}^m Y_i$ tem distribuição χ_{2m}^2 .*

•Caso 2: Quando n é ímpar, isto é $n = 2m + 1$ com $m \geq 0$, e sabendo gerar variáveis χ_{2m}^2 é fácil gerar variáveis χ_{2m+1}^2 usando a Proposição a seguir

Proposição 4.1.3 *Sejam Y e Z variáveis aleatórias independentes, $Y \sim \chi_{2m}^2$ e $Z \sim \mathcal{N}(0,1)$, então $X = Y + Z^2$ tem distribuição χ_{2m+1}^2 .*

4.2 O Método de Inversão

Este é um método que funciona bem para geração de variáveis de tipo contínuo, embora possa servir, em princípio, para gerar variáveis quaisquer. Baseia-se no Teorema a seguir, cuja demonstração requer certa maturidade matemática.

Teorema 4.2.1 *Seja $F: \mathbb{R} \rightarrow [0,1]$ uma função de distribuição acumulada; $F^-(t): [0,1] \rightarrow \mathbb{R}$ definida por $F^-(t) = \inf\{x: t \leq F(x)\}$. Se $U \sim \mathcal{U}(0,1)$ e $X = F^-(U)$, então F é a função de distribuição da variável aleatória X .*

Veremos a seguir três exemplos de geração usando este Teorema.

Exemplo 4.2.1: $F = Bi(1, p)$. Seja $F: \mathbb{R} \rightarrow [0, 1]$ definida por

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - p & \text{se } 0 \leq x \leq 1 \\ 1 & \text{se } 1 \leq x. \end{cases}$$

Afirmção: $F^-(t) = \mathbb{I}_{[1-p, 1]}(t)$ para todo t em $[0, 1]$. Daí se quisermos gerar (x_1, \dots, x_N) , “amostra” de tamanho N da $Bi(1, p)$ faremos

1. Seja $i = 1$.
2. Gerar u_i com distribuição $\mathcal{U}(0, 1)$.
3. Definir $x_i = F^-(u_i)$, isto é

$$x_i = \begin{cases} 1 & \text{se } 1 - p \leq u_i \leq 1 \\ 0 & \text{se } 0 \leq u_i < 1 - p. \end{cases}$$

4. Fazer $i = i + 1$
 - Se $i \leq N$, então seguir em 2.
 - Se $i > N$, então terminar.

Exemplo 4.2.2: $F = Weibull(\beta)$. Seja $F: \mathbb{R} \rightarrow [0, 1]$ definida pela função de distribuição acumulada $F(x) = (1 - \exp\{-x^\beta\})\mathbb{I}_{(0, +\infty)}(x)$.

Afirmção: $F^-(t) = (-\ln(1 - t))^{1/\beta}$ para todo $0 < t < 1$. Pelo Teorema 4.2.1, se quisermos gerar (x_1, \dots, x_n) , “amostra” de tamanho N da $\mathcal{W}(\beta)$ faremos como no Exemplo 4.2.1 com a F^- da Afirmção acima.

Exemplo 4.2.3: F Cauchy padrão. Seja $F: \mathbb{R} \rightarrow [0, 1]$ definida por $F(x) = 1/2 + 1/\pi \tan^{-1}(x)$, $x \in \mathbb{R}$.

Afirmção: $F^-(t) = \tan(\pi(t - 1/2))$, $t \in (0, 1)$. Daí se quisermos gerar variáveis aleatórias com distribuição Cauchy padrão poderemos proceder como nos Exemplos anteriores.

É possível usar métodos numéricos para calcular F^- quando não dispomos de uma fórmula explícita para calculá-la. Por exemplo, podemos gerar $X \sim \mathcal{N}(\mu, \sigma^2)$ definindo $X = \mu + \sigma \Phi^{-1}(U)$, onde $U \sim \mathcal{U}(0, 1)$ e Φ é a função de distribuição acumulada da $\mathcal{N}(0, 1)$. Veja o Exercício 4.9.

Embora o Método de Inversão seja universalmente válido, pode resultar muito lento se não está disponível uma boa rotina para calcular F^- . Daí que será útil estudar outros métodos de geração.

4.3 O Método de Rejeição

Historicamente esta classe de métodos começou a aparecer a partir do trabalho de von Neumann [123]. Baseiam-se na seguinte

Proposição 4.3.1 *Sejam $(\Omega, \mathcal{F}, \mathbb{P}r)$ um espaço probabilístico; $Y: \Omega \rightarrow \mathbb{R}$ uma variável aleatória com densidade g ; $Z: \Omega \rightarrow \mathbb{R}$ uma variável aleatória; $A \subset \mathbb{R}$ (Borel) mensurável tal que $\mathbb{P}r(Z \in A) > 0$; $\mathcal{F}' = \mathcal{F} \cap Z^{-1}(A) = \{B \cap Z^{-1}(A) : B \in \mathcal{F}\}$; $\mathbb{P}r': \mathcal{F}' \rightarrow [0, 1]$ dada por $\mathbb{P}r'(B \cap Z^{-1}(A)) = \frac{\mathbb{P}r(B \cap Z^{-1}(A))}{\mathbb{P}r(Z^{-1}(A))}$; $X: Z^{-1}(A) \rightarrow \mathbb{R}$ dada por $X(\omega) = Y(\omega)$ para todo $\omega \in \Omega$. Então*

1. $(Z^{-1}(A), \mathcal{F}', \mathbb{P}r')$ é um espaço probabilístico.
2. X como variável aleatória sobre $(Z^{-1}(A), \mathcal{F}', \mathbb{P}r')$ admite uma densidade dada por $f(x) = \frac{h(x)g(x)}{\int h g}$, onde $h(Y) = \mathbb{P}r(Z^{-1}(A)|Y)$.

Demonstração:¹ O primeiro item é fácil. Vejamos o segundo.

Seja $x \in \mathbb{R}$ arbitrário. Definamos $B = X^{-1}((-\infty, x]) = Y^{-1}((-\infty, x] \cap Z^{-1}(A))$. Então $\mathbb{P}r'(X \leq x) = \mathbb{P}r'(B) = \frac{\mathbb{P}r(Y^{-1}((-\infty, x] \cap Z^{-1}(A)))}{\mathbb{P}r(Z^{-1}(A))}$ (4.3.1). Mas $\mathbb{P}r(Y^{-1}((-\infty, x]) \cap Z^{-1}(A)) = \int_{Y^{-1}((-\infty, x])} h(Y) d\mathbb{P}r = \int_{(-\infty, x]} h(y) d\mathbb{P}r_Y(y) = \int_{-\infty}^x h(y) g(y) dy$; onde $\mathbb{P}r_Y$ é a probabilidade sobre $(\mathbb{R}, \mathcal{B}_1)$ induzida por Y e $\mathbb{P}r$.

Também $\mathbb{P}r(Z^{-1}(A)) = \int h(Y) d\mathbb{P}r = \int_{-\infty}^{\infty} h(y) g(y) dy$. Daí substituindo em (4.3.1) temos

$$\mathbb{P}r'(X \leq x) = \frac{1}{\int_{-\infty}^{\infty} h(y) g(y) dy} \int_{-\infty}^x h(y) g(y) dy,$$

onde resulta a Proposição.

A Proposição anterior aplica-se para gerar um valor de X da forma a seguir:

1. Gerar um valor z de Z e um valor y de Y .
2. Se $z \in A$ (o que acontece com probabilidade $h(y)$) definir $X(\omega) = y$. Caso contrário, voltar a 1.

Devido à estrutura de este procedimento, o conjunto A é dito *região de aceitação* de Y .

O resultado a seguir sugere um algoritmo que implementa o chamado **Método Geral de Rejeição**.

Proposição 4.3.2 *Sejam $(\Omega, \mathcal{F}, \mathbb{P}r)$ um espaço probabilístico; $X: \Omega \rightarrow \mathbb{R}$ uma variável aleatória com densidade f e $Y: \Omega \rightarrow \mathbb{R}$ uma variável aleatória com densidade g . Suponhamos que existe $M < \infty$ tal que $\frac{f(x)}{g(x)} \leq M$ para todo x em \mathbb{R} . Sejam $U: \Omega \rightarrow \mathbb{R}$ uma variável aleatória independente de Y com distribuição $U(0, 1)$ e a variável aleatória $Z: \Omega \rightarrow \mathbb{R}$ definida por $Z(\omega) = \frac{f(Y(\omega))}{g(Y(\omega))U(\omega)}$. Sejam ainda $A = [M, \infty)$, \mathcal{F}' e $\mathbb{P}r'$ como na Proposição 4.3.1 e, finalmente, $X': Z^{-1}(A) \rightarrow \mathbb{R}$ definida por $X'(\omega) = Y(\omega)$ para todo $\omega \in Z^{-1}(A)$. Então X' como variável aleatória sobre o espaço probabilístico $(Z^{-1}(A), \mathcal{F}', \mathbb{P}r')$ admite a f como densidade.*

¹Numa primeira leitura pode-se pular as provas das Proposições desta Secção.

Demonstração: Aplicando a Proposição 4.3.1 temos que $h(y) = \frac{f(y)}{g(y)M}$; com efeito,

$$\mathbb{P}\text{r}(Z \in A|Y) = \mathbb{P}\text{r}\left(\frac{f(Y)}{g(Y)U} \geq M|Y\right) = \mathbb{P}\text{r}\left(U \leq \frac{f(Y)}{g(Y)M}|Y\right) = \mathbb{P}\text{r}\left(U \leq \frac{f(Y)}{g(Y)M}\right),$$

a última desigualdade sendo válida pela independência entre U e Y . Daí

$$h(Y) = \frac{f(Y)}{g(Y)M} = \mathbb{P}\text{r}(Z \in A|Y).$$

Assim, aplicando a Proposição 4.3.1 a densidade de X' resulta ser

$$\frac{h(x)g(x)}{\int h g} = \frac{\frac{f(x)}{g(x)M}g(x)}{\int \frac{f}{gM}g} = \frac{\frac{f(x)}{M}}{M^{-1} \int f} = f(x).$$

Temos então o seguinte **Algoritmo de Método de Rejeição Geral** para obter uma amostra da variável aleatória X com densidade f , dispondo de um gerador da variável Y com densidade g , de um gerador da variável $U \sim \mathcal{U}(0,1)$ quando $\frac{f(y)}{g(y)} \leq M$ para todo y em \mathbb{R} :

1. Gerar um valor y da Y .
2. Gerar um valor u da U .
3. Se $Mu \leq f(y)/g(y)$, definir $x = y$; caso contrário voltar ao 1.

É bom saber se teremos que esperar muito tempo até gerar um valor da variável aleatória X . A Proposição a seguir fornece uma resposta a esta questão.

Proposição 4.3.3 *Sejam $(\Omega, \mathcal{F}, \mathbb{P}\text{r})$, Y , X , g , f e M como na Proposição 4.3.2. Para cada $i = 1, 2, \dots$ sejam $U_i: \Omega \rightarrow \mathbb{R}$ tais que $(U_i)_{i \geq 1}$ são iid com distribuição $\mathcal{U}(0,1)$ e Y, U_1, U_2, \dots são independentes. Para cada $i = 1, 2, \dots$ seja $Z_i: \Omega \rightarrow \mathbb{R}$ dada por*

$$Z_i(\omega) = \frac{f(Y(\omega))}{g(Y(\omega)) \cdot U_i(\omega)}.$$

Seja $N: \Omega \rightarrow \{0, 1, 2, \dots\}$ definida por

$$N(\omega) = \begin{cases} 0 & \text{se } Z_1(\omega) \geq M, \\ 1 & \text{se } Z_1(\omega) < M \text{ e } Z_2(\omega) \geq M, \end{cases}$$

em geral, $N(\omega) = n$ se $Z_1(\omega) < M, \dots, Z_n(\omega) < M$ e $Z_{n+1}(\omega) \geq M$. Então:

1. $\mathbb{P}\text{r}(N = n|Y = y) = \left(1 - \frac{f(y)}{g(y)M}\right)^n \frac{f(y)}{g(y)M}.$
2. $\mathbb{E}\{N\} = M \int_{-\infty}^{+\infty} \frac{g(t)^2}{f(t)} dt.$

Demonstração: Suponhamos provada a primeira parte. Então a distribuição da variável aleatória N , condicional à Y , é uma geométrica de razão $\theta = \frac{f(Y)}{g(Y)M}$. Daí

$$\mathbb{E}\{N|Y\} = (\theta(Y))^{-1} = M \frac{g(Y)}{f(Y)}.$$

Logo,

$$\mathbb{E}\{N\} = \mathbb{E}\{\mathbb{E}\{N|Y\}\} = M \int_{-\infty}^{\infty} \frac{g(t)}{f(t)} d\mathbb{P}_Y(t) = M \int_{-\infty}^{\infty} \frac{g(t)}{f(t)} g(t) dt,$$

e a segunda parte fica provada.

Vejamos a primeira parte:

$$\begin{aligned} \mathbb{P}_Y(N = 0|Y) &= \mathbb{P}_Y(Z_1 \geq M|Y) = \mathbb{P}_Y\left(\frac{f(Y)}{g(Y)U_1} \geq M \mid Y\right) \\ &= \mathbb{P}_Y\left(U_1 \leq \frac{f(Y)}{g(Y)M} \mid Y\right) = \mathbb{P}_Y\left(U_1 \leq \frac{f(Y)}{g(Y)M}\right) = \frac{f(Y)}{g(Y)M} \end{aligned}$$

pela independência entre U_1 e Y e $U_1 \sim \mathcal{U}(0, 1)$. Seja agora $n \geq 1$, então

$$\begin{aligned} \mathbb{P}_Y(N = n|Y) &= \mathbb{P}_Y(Z_1 < M, \dots, Z_n < M, Z_{n+1} \geq M|Y) \\ &= \mathbb{P}_Y\left(U_1 > \frac{f(Y)}{g(Y)M}, \dots, U_n > \frac{f(Y)}{g(Y)M}, U_{n+1} \leq \frac{f(Y)}{g(Y)M} \mid Y\right) \\ &= \mathbb{P}_Y\left(U_1 > \frac{f(Y)}{g(Y)M}, \dots, U_n > \frac{f(Y)}{g(Y)M}, U_{n+1} \leq \frac{f(Y)}{g(Y)M}\right) \\ &= \prod_{1 \leq i \leq n} \mathbb{P}_Y\left(U_i > \frac{f(Y)}{g(Y)M}\right) \cdot \mathbb{P}_Y\left(U_{n+1} \leq \frac{f(Y)}{g(Y)M}\right) \\ &= \left(1 - \frac{f(Y)}{g(Y)M}\right)^n \frac{f(Y)}{g(Y)M} \end{aligned}$$

Exemplo 4.3.1: Geração de variáveis $\text{Beta}(\alpha, \beta)$ pelo método de Rejeição Geral. Seja $f_1(x) = x^{\alpha-1}(1-x)^{\beta-1}\mathbb{I}_{(0,1)}(x)$ com $\alpha > 0, \beta > 0$.

Afirmção:

1. f_1 é limitada se e somente se $\alpha \geq 1$ e $\beta \geq 1$.
2. Se $\alpha > 1$ e $\beta > 1$ então $f_1(x) \leq \frac{(\alpha-1)^{\alpha-1}(\beta-1)^{\beta-1}}{(\alpha+\beta-2)^{\alpha+\beta-2}} \forall x \in (0, 1)$.

Daí resulta a seguinte

Regra 1: Para gerar amostras da $\text{Beta}(\alpha, \beta)$ onde $\alpha \geq 1$ e $\beta \geq 1$, pode-se aplicar o algoritmo do Método de Rejeição Geral com

$$\begin{aligned} f(x) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \mathbb{I}_{(0,1)}(x) \\ g(x) &= \mathbb{I}_{(0,1)}(x) \text{ isto é } Y \sim \mathcal{U}(0, 1) \\ M &= \frac{\Gamma(\alpha + \beta)(\alpha - 1)^{\alpha-1}(\beta - 1)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)(\alpha + \beta - 2)^{\alpha+\beta-2}}. \end{aligned}$$

Nota: Se $\alpha > 0$ e $\beta \geq 1$ então para $x \in (0, 1)$ vale que

$$0 \leq \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\alpha x^{\alpha-1}} \leq (1-x)^{\beta-1} \leq 1.$$

Além disso

$$g(x) = \alpha x^{\alpha-1} \mathbb{I}_{(0,1)}(x)$$

é a densidade de $Y = V^{1/\alpha}$, com $V \sim \mathcal{U}(0, 1)$. Daí resulta a seguinte

Regra 2: Para gerar amostras da $\text{Beta}(\alpha, \beta)$ onde $\alpha > 0$ e $\beta \geq 1$ pode-se aplicar o algoritmo do Método de Rejeição Geral com

$$\begin{aligned} f(x) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \mathbb{I}_{(0,1)}(x) \\ g(x) &= \alpha x^{\alpha-1} \mathbb{I}_{(0,1)}(x) \\ M &= \frac{\alpha \Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}. \end{aligned}$$

A arte no uso dos métodos de rejeição consiste em encontrar a g conveniente. Esse método pode-se aplicar também à geração de variáveis aleatórias discretas, mas nesses casos é bem difícil achar funções g adequadas.

A seguir veremos uma aplicação da Proposição 4.3.1 que funciona muito bem para gerar variáveis com distribuição do tipo $f(x) \propto \exp(-b(x))$, com “ \propto ” significando “proporcional a”. Baseia-se na seguinte

Proposição 4.3.4 *Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço probabilístico, $f: \mathbb{R} \rightarrow \mathbb{R}$ uma densidade e $Y: \Omega \rightarrow \mathbb{R}$ uma variável aleatória com densidade g . Suponhamos que existe uma constante M tal que $0 \leq b(x) + \ln(M) < 1$, onde $b(x) = -\ln(f(x)/g(x))$, para todo $x \in \mathbb{R}$. Seja $(U_i)_{i \geq 1}$ uma sequência de variáveis com distribuição $\mathcal{U}(0, 1)$ tal que (Y, U_1, U_2, \dots) são independentes, e seja $U_0 = b(Y) + \ln(M)$. Definimos $Z: \Omega \rightarrow \{1, 2, \dots\}$ por $Z(\omega) = \min\{i \leq 1: U_i(\omega) \geq U_{i-1}(\omega)\}$ isto é, Z é o primeiro índice tal que $U_1 \geq U_{i-1}$. Seja A o conjunto dos números ímpares, ou seja $A = \{2k+1: k = 0, 1, 2, \dots\}$. Então $\mathbb{P}(Z \in A | Y = y) = \exp\{-(b(y) + \ln(M))\}$.*

Demonstração: Seja $y \in \mathbb{R}$ e escrevamos $t = b(y) + \ln(M)$. Provaremos primeiro que

Afirmção 1: $\mathbb{P}(Z > n | Y = y) = t^n/n!$ para todo $n = 1, 2, \dots$

Demonstração de Af. 1:

$$\begin{aligned} \mathbb{P}(Z > n | Y = y) &= \mathbb{P}(U_0 > U_1 > \dots > U_n | U_0 = t) \\ &= \mathbb{P}(t > U_1 > \dots > U_n) \text{ pela independência de } Y, U_1, U_2, \dots \\ &= \int_0^t \left[\int_0^{u_1} \dots \int_0^{u_{n-1}} du_n du_{n-1} \dots du_2 \right] du_1 \\ &= \frac{t^n}{n!} \end{aligned}$$

Também $\text{IPr}(Z > 0|Y = y) = 1$. Logo, para todo $n = 1, 2, \dots$ temos

$$\text{IPr}(Z = n|Y = y) = \text{IPr}(Z > n-1|Y = y) - \text{IPr}(Z > n|Y = y) = \frac{t^{n-1}}{(n-1)!} - \frac{t^n}{n!},$$

e daí

$$\text{IPr}(Z = n|Y = y) = \sum_{0 \leq k < \infty} \text{IPr}(Z = 2k+1|Y = y) = \exp\{-t\}.$$

Seja agora $X: Z^{-1}(A) \rightarrow \mathbb{R}$ é dada por $X(\omega) = Y(\omega)$ para todo $\omega \in Z^{-1}(A)$, então X como variável aleatória sobre $(Z^{-1}(A), \mathcal{F} \cap Z^{-1}(A), \text{IPr}')$ admite uma densidade dada por

$$\frac{\exp\{-b(x) - \ln(M)\} g(x)}{\int_{-\infty}^{+\infty} \exp\{-b(u) - \ln(M)\} g(u) du} = f(x) \quad \forall x,$$

onde IPr' é a probabilidade sobre $\mathcal{F} \cap Z^{-1}(A)$ dada por

$$\text{IPr}'(B \cap Z^{-1}(A)) = \frac{\text{IPr}(B \cap Z^{-1}(A))}{\text{IPr}(Z^{-1}(A))}.$$

O algoritmo a seguir implementa o método de geração de variáveis com distribuição f sugerido pela Proposição 4.3.4.

Algoritmo 4.3.1: Sejam f, g, b e M como na Proposição 4.3.4. Para gerar uma amostra da variável aleatória X com densidade f :

1. Gerar uma amostra y da variável aleatória Y com densidade g .
2. Sejam $t = b(y) + \ln(M)$, $k = 0$ e $u = t$.
3. Gerar u^* da U_{2k+1} .
4. Se $u < u^*$ definir $X = y$ e parar. Se não seguir em 5.
5. Gerar um valor u' da U_{2k+2} e redefinir $u = u'$.
6. Se $u < u^*$ fazer $k = k + 1$ e seguir em 3. Se não (ou seja se $u \geq u^*$, o primeiro índice i tal que $U_i \geq U_{i-1}$ foi par), voltar ao 1.

4.4 O Método de Composição

Suponhamos que estejamos interessados em obter amostras de variáveis com densidade f , que por sua vez pode ser escrita na forma

$$f = \sum_{i \in I} p_i f_i,$$

onde $I \subset \mathbb{N}$ é finito ou não, $p_i > 0$ para todo $i \in I$, $\sum_{i \in I} p_i = 1$ e f_i é uma densidade para todo $i \in I$. Então podemos primeiramente escolher $i \in I$ segundo as probabilidades $\{p_j; j \in I\}$ e depois obter uma amostra de f_i .

O uso mais comum dessa técnica começa por particionar a imagem da X com distribuição f em intervalos. Para fixar idéias, vejamos a seguinte aplicação para gerar amostras da Beta(α, β) sugerida por Atkinson e Whittaker [13]. Mas antes disso temos a seguinte

Proposição 4.4.1 *Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço probabilístico; $(U_1, U_2, U_3): \Omega \rightarrow [0, 1]^3$ variáveis com distribuição $\mathcal{U}(0, 1)$; $Y: \Omega \rightarrow \mathbb{R}$, $Z: \Omega \rightarrow \mathbb{R}^3$ dadas por $Y = \frac{1}{2}U_1^{1/\alpha}$ e $Z = (U_1, U_2, U_3)$. Sejam também $A_1 = \{(u_1, u_2, u_3): u_2 \leq [2(1 - \frac{1}{2}u_1^{1/\alpha})]^{\alpha-1}, u_3 \leq \frac{1}{2}\}$, $A_2 = \{(u_1, u_2, u_3): u_2 \leq [2(1 - \frac{1}{2}u_1^{1/\alpha})]^{\alpha-1}, u_3 > \frac{1}{2}\}$, $A = A_1 \cup A_2$; $X: Z^{-1}(A) \rightarrow \mathbb{R}$ definida por*

$$X(\omega) = \begin{cases} Y(\omega) & \text{se } Z(\omega) \in A_1 \\ 1 - Y(\omega) & \text{se } Z(\omega) \in A_2 \end{cases}.$$

Então X como variável aleatória sobre $(Z^{-1}(A), \mathcal{F} \cap Z^{-1}(A), \mathbb{P}')$ tem distribuição Beta(α, α), onde \mathbb{P}' é definida como na Proposição 4.3.1.

Demonstração: Seja $x \in \mathbb{R}$ qualquer,

$$\mathbb{P}(X \leq x) = \frac{\mathbb{P}(X \leq x, Z \in A)}{\mathbb{P}(Z \in A)}. \quad (4.1)$$

Ora, $\mathbb{P}(X \leq x, Z \in A) = \mathbb{P}(X \leq x, Z \in A_1) + \mathbb{P}(X \leq x, Z \in A_2)$, e

$$\begin{aligned} \mathbb{P}(X \leq x, Z \in A_1) &= \int_{Y^{-1}((-\infty, x])} \mathbb{P}(Z \in A_1 | Y = y) d\mathbb{P}_Y(y) \\ &= \alpha 2^\alpha \int_0^x \mathbb{P}(Z \in A_1 | Y = y) y^{\alpha-1} \mathbb{I}_{(0,1/2)}(y) dy \\ &= \alpha 2^\alpha \int_0^x \left[\frac{1}{2} 2^{\alpha-1} (1-y)^{\alpha-1} \right] y^{\alpha-1} \mathbb{I}_{(0,1/2)}(y) dy. \end{aligned}$$

Também

$$\mathbb{P}(X \leq x, Z \in A_2) = \alpha 2^\alpha \int_{1-x}^{1/2} \left[\frac{1}{2} 2^{\alpha-1} (1-y)^{\alpha-1} \right] y^{\alpha-1} \mathbb{I}_{(0,1/2)}(y) dy.$$

Finalmente,

$$\begin{aligned} \mathbb{P}(Z \in A) &= \int [\mathbb{P}(Z \in A_1 | Y = y) + \mathbb{P}(Z \in A_2 | Y = y)] d\mathbb{P}_Y(y) \\ &= \alpha 2^\alpha \int_0^{1/2} (1-y)^{\alpha-1} y^{\alpha-1} dy \end{aligned}$$

Substituindo em 4.1 temos

$$\mathbb{P}'(X \leq x) = \frac{\frac{1}{2} \int_0^x (1-y)^{\alpha-1} y^{\alpha-1} \mathbb{I}_{(0,1/2)}(y) dy + \int_{1-x}^{1/2} (1-y)^{\alpha-1} y^{\alpha-1} \mathbb{I}_{(0,1/2)}(y) dy}{\int_0^{1/2} (1-y)^{\alpha-1} y^{\alpha-1} dy}.$$

onde,

$$\begin{aligned} f_X(x) &= \left(2 \int_0^{1/2} (1-y)^{\alpha-1} y^{\alpha-1} dy \right)^{-1} \frac{d}{dx} \left(\int_0^x (1-y)^{\alpha-1} y^{\alpha-1} \mathbb{I}_{(0,1)}(y) dy \right) \\ &= \left(2 \int_0^{1/2} (1-y)^{\alpha-1} y^{\alpha-1} dy \right)^{-1} (1-x)^{\alpha-1} x^{\alpha-1} \mathbb{I}_{(0,1)}(x) \\ &= \frac{\Gamma(2\alpha)}{\Gamma(\alpha)\Gamma(\alpha)} (1-x)^{\alpha-1} x^{\alpha-1} \mathbb{I}_{(0,1)}(x) \end{aligned}$$

já que $2 \int_0^{1/2} (1-y)^{\alpha-1} y^{\alpha-1} = \int_0^1 (1-y)^{\alpha-1} y^{\alpha-2}$

A Proposição anterior sugere a seguinte regra para gerar amostras da $\text{Beta}(\alpha, \alpha)$, com $\alpha < 1$.

Algoritmo 4.4.1 para gerar variáveis com distribuição $\text{Beta}(\alpha, \alpha)$, com $\alpha < 1$: Sejam U_1, U_2, U_3 valid com distribuição $\mathcal{U}(0, 1)$

1. Gerar u_1 e u_2 de U_1 e U_2 respectivamente, e definir $y = \frac{1}{2} u_1^{1/\alpha}$.
2. Se $u_2 \leq [2(1-y)]^{\alpha-1}$ então seguir em 3. Se não voltar ao 1.
3. Gerar u_3 de U_3 . Se $u_3 > \frac{1}{2}$ redefinir $y = 1 - y$.
4. Definir $X = y$

4.5 Quociente de Uniformes

O resultado a seguir é a base do método dado em Kinderman e Monahan [73]:

Teorema 4.5.1 *Seja $h: \mathbb{R} \rightarrow (0, +\infty)$ uma função tal que $\int h(x) dx < \infty$. Seja $C_h = \{(u, v) \in \mathbb{R}^2: 0 \leq u \leq \sqrt{h(v/u)}\}$. Então*

1. $\text{Area}(C_h) := \int_{C_h} du dv = \frac{1}{2} \int h(x) dx$, e portanto é finita.
2. *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade; $(U, V): \Omega \rightarrow \mathbb{R}^2$ dada pela densidade*

$$f_{(U,V)}(u, v) = \frac{1}{\text{Area}(C_h)} \mathbb{I}_{C_h}(u, v).$$

Então $X := V/U$ tem densidade dada por $f(x) = (\int h)^{-1} h(x)$.

Demonstração: Vejamos 1:

$$\int_{C_h} du dv = \int_{-\infty}^{+\infty} \int_0^{\sqrt{h(v/u)}} du dv = \left| \begin{array}{l} u = y \\ v = xy \end{array} \right| = \int_{-\infty}^{+\infty} \int_0^{\sqrt{h(x)}} y dy dx = \int_{-\infty}^{+\infty} \frac{1}{2} h(x) dx.$$

Vejamos 2: $\mathbb{P}(X \leq x) = \frac{1}{\text{Area}(C_h)} \int_{A \cap C_h} du dv$, onde $A = \{(u, v): v \leq ux\}$ e $C_h = \{(u, v): 0 \leq u \leq \sqrt{h(v/u)}\}$. Considerando a mudança de variáveis $t_1 = v/u$ e $t_2 = u$ temos que

$$\mathbb{P}(X \leq x) = \frac{1}{\text{Area}(C_h)} \int_{-\infty}^x \int_0^{\sqrt{h(t_1)}} t_2 dt_2 dt_1.$$

Daí segue que X tem uma distribuição dada por

$$f(x) = \frac{1}{\text{Area}(C_h)} \int_0^{\sqrt{h(x)}} u du = \frac{1}{\text{Area}(C_h)} h(x) = \frac{1}{\int h} h(x).$$

O problema de gerar uma amostra da distribuição uniforme sobre C_h , requerida no Teorema anterior, é simples quando C_h está limitada por um retângulo. Com efeito:

Algoritmo 4.5.1: Geração de uma amostra da distribuição uniforme sobre C_h quando C_h está limitada pelo retângulo $\{(u, v): 0 \leq u \leq a, b_- \leq v \leq b_+\}$. Sejam (U_1, U_2) valid com distribuição $\mathcal{U}(0, 1)$

1. Gerar u_1 de U_1 e u_2 de U_2 .
2. Definir $u = au_1$, $v = b_- + (b_+ - b_-)u_2$. Se $(u, v) \in C_h$ definir $(U, V) = (u, v)$. Se não, voltar ao 1.

Algoritmo 4.5.2: Geração de variáveis com distribuição de Cauchy padrão pelo método de Quocientes de Uniformes. Para tanto aplicamos o Teorema 4.5.1 na seguinte situação:

$$h(x) = \frac{1}{1+x^2} \text{ logo, } f(x) = \frac{1}{\pi(1+x^2)}, \quad a = 1, \quad b_- = -1, \quad b_+ = 1.$$

Dáí, sendo U_1, U_2 variad com distribuição $\mathcal{U}(0, 1)$,

1. Gerar u_1 de U_1 e u_2 de U_2 .
2. Definir $u = u_1$ e $v = 2u_2 - 1$, isto é $u = au_1$, $v = b_- + (b_+ - b_-)u_2$.
3. Se $u \leq \sqrt{h(v/u)}$ (ou mesma coisa: $u^2 + v^2 \leq 1$) definir $X = v/u$ e sair. Se $u > \sqrt{h(v/u)}$ voltar ao 1.

O resultado a seguir pode ser de ajuda para determinar se C_h está dentro de um retângulo.

Teorema 4.5.2 *Seja $h: \mathbb{R} \rightarrow (0, +\infty)$ uma função tal que h e a função $x \mapsto x^2 h(x)$ são limitadas. Sejam*

$$\begin{aligned} C_h &= \{(u, v): 0 \leq u \leq \sqrt{h\left(\frac{u}{v}\right)}\}, \\ a &= \sqrt{\sup\{h(x): x \in \mathbb{R}\}}, \\ b_+ &= \sqrt{\sup\{x^2 h(x): x \geq 0\}}, \\ b_- &= -\sqrt{\sup\{x^2 h(x): x \leq 0\}}; \end{aligned}$$

então $C_h \subset [0, a] \times [b_-, b_+]$.

Demonstração: Seja $(u, v) \in C_h$; é imediato ver que deve ser $0 \leq u \leq a$. Ora, se $v \geq 0$ temos que ou $0 < u^2 < h(v/u)$ ou $v^2 \leq u^2 h(u/v)$ com $t = v/u$, portanto ou $v^2 \leq b_+^2$ ou $v \leq b_+$. Analogamente prova-se que se $v < 0$ deve ser que $v \geq b_-$.

Exemplo 4.5.1: Geração de variáveis com distribuição exponencial pelo Método do Quociente de Uniformes. Aplicamos o Teorema 4.5.1 e o Algoritmo 4.5.1 na situação

$$h(x) = \exp\{-x\} \mathbb{I}_{(0, +\infty)}(x), \text{ logo } f(x) = h(x), \quad a = 1, \quad b_- = 0, \quad b_+ = 2/e.$$

Neste caso $(u, v) \in C_h \iff 0 < u \text{ e } u^2 \leq \exp\{-v/u\} \iff 0 < u \text{ e } v \leq -2u \ln(u)$.

Temos então o seguinte algoritmo onde (U_1, U_2) são variad com distribuição $\mathcal{U}(0, 1)$:

1. Gerar u_1 de U_1 e u_2 de U_2 .
2. Definir $u = u_1$ e $v = 2u_2/e$.

3. Se $v \leq -2u \ln(u)$ definir $X = v/u$ e sair. Se não voltar ao 1.

Exemplo 4.5.2: Geração de variáveis com distribuição $\mathcal{N}(0,1)$ pelo Método do Quociente de Uniformes. Neste caso temos

$$h(x) = \exp\left\{-\frac{1}{2}x^2\right\}, \quad a = 1, \quad b_+ = \sqrt{\frac{2}{e}}, \quad b_- = -\sqrt{\frac{2}{e}}.$$

Assim, é fácil ver que

$$(u, v) \in C_h \iff v^2 \leq -4u^2 \ln(u).$$

Se (U_1, U_2) são variáveis com distribuição $\mathcal{U}(0,1)$ temos o seguinte algoritmo:

1. Gerar u_1 de U_1 e u_2 de U_2 .
2. Definir $u = u_1$ e $v = \sqrt{2e^{-1}}(2u_2 - 1)$.
3. Se $v^2 \leq -4u^2 \ln(u)$ definir $X = v/u$ e sair. Se não voltar ao 1.

Também é possível “mergulhar” C_h dentro de um polígono que não seja um retângulo, para assim conseguir um processo de geração mais eficiente. O leitor interessado pode consultar o trabalho de Hsuan [63].

4.6 Exercícios

Nos exercícios onde pede-se realizar um programa de computador, pode-se utilizar qualquer linguagem que lhe seja familiar.

Exercício 4.1: a) Faça um programa que implemente o algoritmo do Exemplo 4.2.1. b) Faça um gráfico (X, Y) com os pontos $(x_1, x_2), (x_3, x_4), \dots, (x_{N-1}, x_N)$ (N deve ser par) gerados ao rodar o programa feito na parte anterior. Existe algum padrão evidente nesse gráfico que invalide a normalidade da sequência x_1, \dots, x_N ? c) Faça um “Normal-Plot” dos pontos x_1, \dots, x_N . Interprete-o.

Exercício 4.2: a) Prove a Proposição 4.2.1. b) Faça um programa que implemente o método de exponenciais sugerido pela Proposição 4.2.1. c) Submeta sequências x_1, \dots, x_N geradas pelo programa anterior a um teste de bondade do ajuste (χ^2 , Kolmogorov-Smirnov ou outro).

Exercício 4.3: a) Prove a Proposição 4.2.2. b) Faça um programa que implemente o método de geração de variáveis χ^2_{2m} usando a Proposição 4.2.2 e o programa da parte “b)” do Exercício 4.2. c) Submeta sequências x_1, \dots, x_N geradas pelo programa anterior a um teste de bondade do ajuste.

Exercício 4.4: Prove a Proposição 4.2.3. b) Faça um programa que implemente o método de geração de variáveis χ^2_{2m+1} usando a Proposição 4.2.3 e o programa da parte “b)” do Exercício 4.3. c) Submeta sequências x_1, \dots, x_N geradas pelo programa anterior a um teste de bondade do ajuste.

Exercício 4.5: Prove o Teorema 4.2.1. (veja, por exemplo, o livro de Ripley [110]).

Exercício 4.6: a) Prove a Afirmação do Exemplo 4.2.1. b) Faça um programa que implemente o algoritmo do Exemplo 4.2.1. c) Submeta seqüências geradas rodando o programa realizado em “b)” a testes de bondade do ajuste.

Exercício 4.7: a) Prove a Afirmação do Exemplo 4.2.2. b) Idem à parte “b)” do Exercício 4.6 para a F do Exemplo 4.2.2 ($F = \text{Weibull}(\beta)$). c) Idem à parte “c)” do Exercício 4.6 para o caso deste Exercício.

Exercício 4.8: a) Prove a Afirmação do Exemplo 4.2.3. b) e c) como as partes “b)” e “c)” do Exercício 4.6 para F Cauchy e pelo método sugerido no Exemplo 4.2.3 (inversão).

Exercício 4.9: a) Implementar algoritmo apresentado em [17] para calcular Φ^{-1} . b) Faça um programa que implemente o Método de Inversão de variáveis normais padrão, isto é $\mathcal{N}(0,1)$ usando o programa realizado em “a)”. c) Idem a “b)” do Exercício 4.1 para o gerador da parte “b)”. d) idem ao “c)” do Exercício 4.1 para o gerador da parte “b)”.

Exercício 4.10: Prove a Afirmação do Exemplo 4.3.1.

Exercício 4.11: a) Faça um programa que implemente a Regra 1 do Exemplo 4.3.1 para gerar variáveis com distribuição $\text{Beta}(\alpha, \beta)$ onde $\alpha \geq 1$ e $\beta \geq 1$. b) Submeta seqüências geradas rodando o programa realizado em “a)” a testes de bondade do ajuste.

Exercício 4.12: a) Faça um programa que implemente a Regra 2 do Exemplo 4.3.1 para gerar variáveis com distribuição $\text{Beta}(\alpha, \beta)$ onde $\alpha > 0$ e $\beta \geq 1$. b) Submeta seqüências geradas rodando o programa realizado em “a)” a testes de bondade do ajuste.

Exercício 4.13: a) Faça um programa que implemente o Algoritmo 4.3.1 para o seguinte caso:

$$\begin{aligned} f(x) &= (1 - e^{-1})^{-1} e^{-x} \mathbb{I}_{(0,1)}(x) \\ g(x) &= \mathbb{I}_{(0,1)}(x) \\ M &= (1 - e^{-1})^{-1}. \end{aligned}$$

b) Submeta seqüências geradas rodando o programa realizado em “a)” a testes de bondade do ajuste.

Exercício 4.14: Faça um programa que implemente o Algoritmo 4.4.1 para gerar amostras da distribuição $\text{Beta}(\alpha, \alpha)$ com $\alpha < 1$. b) Submeta seqüências geradas rodando o programa realizado em “a)” a testes de bondade do ajuste.

Exercício 4.15: Faça um programa que implemente o Algoritmo 4.5.2 para gerar amostras da distribuição Cauchy padrão pelo Método de Quociente de Uniformes. b) Submeta seqüências geradas rodando o programa realizado em “a)” a testes de bondade do ajuste.

Exercício 4.16: Idem ao Exercício 4.15 para o Exemplo 4.5.1 de geração de variáveis com distribuição exponencial pelo Método de Quocientes de Uniformes.

Exercício 4.17: Idem ao Exercício 4.15 para o Exemplo 4.5.2 de geração de variáveis com distribuição $\mathcal{N}(0,1)$ pelo Método de Quocientes de Uniformes.

Capítulo 5

Geração de variáveis aleatórias discretas

A geração de variáveis aleatórias discretas possui algumas características especiais que fazem com que seja interessante dedicar um Capítulo ao assunto. Algumas destas particularidades estão diretamente associadas ao fato de estarmos lidando com os objetos que mais naturalmente podem ser tratados em um computador digital: os números inteiros. Usaremos a notação $\mathbb{N} = \{0, 1, \dots\}$ o conjunto dos inteiros não negativos.

Diremos que uma variável aleatória é discreta quando tem por contra-domínio (isto é, toma valores em) um conjunto \mathcal{D} enumerável de valores. Exemplos destas variáveis aleatórias são as possuidoras de distribuições Binomial ($X \in \mathcal{D} = \{0, 1, \dots, n\}$, $n \geq 1$), Binomial negativa ($X \in \mathcal{D} = \mathbb{N}$), Poisson ($X \in \mathcal{D} = \mathbb{N}$), geométrica ($X \in \mathcal{D} = \mathbb{N} \setminus \{0\}$), etc. Seja qual for $\mathcal{D} = \{\xi_0, \xi_1, \dots\}$ sempre é possível identificar todo elemento $\xi_j \in \mathcal{D}$ com um elemento $j \in \mathbb{N}$.

A distribuição de uma tal variável aleatória discreta fica unicamente determinada pelo vetor $\mathbf{p} = (p_0, \dots, p_n)$, onde $n \in \mathbb{N}$ ou $n = \infty$, tal que $\mathbb{P}\text{r}(X = \xi_j) = p_j$. É claro que $p_j \geq 0$ para todo índice j e que $\sum_{j=0}^n p_j = 1$. Neste Capítulo utilizaremos extensivamente esta caracterização da variável aleatória discreta X , usando o vetor $\mathbf{p} = (p_0, \dots, p_n)$, toda vez que $n < +\infty$. Também será de utilidade a função de distribuição acumulada $F: \mathbb{N} \rightarrow [0, 1]$ definida por $F(j) = \sum_{i \leq j} p_i$; e a relação $F(j) - F(j-1) = \mathbb{P}\text{r}(X = j)$. O leitor deverá lembrar que não é a única caracterização possível: também poderíamos usar a função característica da variável aleatória X .

Quando o contra-domínio \mathcal{D} é infinito enumerável, como nos casos Poisson e Binomial negativa acima, é computacionalmente impossível fornecer o vetor \mathbf{p} em forma explícita. Em vez disso, pode lançar-se mão da expressão analítica das probabilidades (se for possível dispor de uma). Lembremos, por exemplo, que a variável aleatória X tem distribuição de Poisson(λ), $\lambda > 0$, se e somente se $\mathbb{P}\text{r}(X = j) = p_j = (e^{-\lambda} \lambda^j) / j!$ para todo $j = 0, 1, \dots$: desta forma, fixando o valor do parâmetro λ estamos fixando p_j para todo $j = 0, 1, \dots$.

5.1 Método de inversão

Este método é o análogo ao visto para variáveis aleatórias contínuas na Secção 4.2. Para gerar uma amostra de tamanho 1 da variável aleatória X com distribuição caracterizada pelo vetor $\mathbf{p} = (p_0, \dots, p_n)$ (conhecido em forma explícita ou em forma implícita), e contradomínio $\{0, \dots, n\}$, $n \geq 1$, geramos uma amostra u da variável aleatória U com distribuição $\mathcal{U}(0, 1)$ e comparamos o valor de u com os valores de \mathbf{p} . Desta forma podemos formular o seguinte **Algoritmo de inversão geral** para variáveis discretas:

1. Gerar u , amostra de $U \sim \mathcal{U}(0, 1)$.
2. Definir x , amostra de $X \sim \mathbf{p}$ como $x = j$ tal que $\sum_{i < j} p_i < u \leq \sum_{i \leq j} p_i$.

Assim, obtemos $x = X(\omega)$ usando uma transformação monótona de $u = U(\omega)$. É claro que a validade deste Algoritmo decorre de $\text{IPr}(X = j) = F(j) - F(j - 1) = p_j$.

O leitor é convidado a pensar por que a solução do método (geral) anterior pode ser obtida *sempre* em tempo finito, para a classe de variáveis aleatórias consideradas neste Capítulo; e também por que, para variáveis aleatórias contínuas, essa afirmação é válida apenas para casos especiais.

Podemos obter a solução para a desigualdade do item 2 acima usando várias técnicas, algumas mais eficientes do que outras (do ponto de vista computacional) dependendo da forma do vetor \mathbf{p} . Vejamos alguns algoritmos particulares, e o Exercício 5.1 para uma melhoria do Algoritmo geral.

5.1.1 Algoritmo de Busca Sequencial

Este algoritmo é útil seja qual for a forma em que definimos a distribuição da variável aleatória X . Definido o vetor \mathbf{p} (seja explícita ou implicitamente), gerar $u = U(\omega)$ com $U \sim \mathcal{U}(0, 1)$ e fazer

1. Atribuir $x = 0$ e $s = p_0$.
2. Enquanto $u > s$ fazer:
 - (a) $x = x + 1$.
 - (b) $s = s + p_x$.
3. Retornar x .

A grande generalidade deste algoritmo acarreta alguns problemas que podem ser muito sérios. Os dois mais importantes são (a) a sucessiva avaliação de $s = s + p_x$ pode acumular erros numéricos, principalmente se os valores p_j estão dados em forma implícita, tais que as caudas da distribuição de X estejam totalmente deturpadas no nosso cálculo; e (b) o tempo que o algoritmo leva para gerar $X(\omega)$ é uma variável aleatória, cuja distribuição depende fortemente das caudas da distribuição de X . Devroye [36] sugere deixar este algoritmo como último recurso a ser usado.

5.1.2 Variáveis Poisson por Busca Sequencial

Já dizemos que $X \sim \text{Poisson}(\lambda)$ se $p_j = \text{Pr}(X = j) = \frac{e^{-\lambda} \lambda^j}{j!}$. Assim é fácil ver que $p_0 = e^{-\lambda}$ e $p_{j+1} = \frac{\lambda}{j+1} p_j$, e aproveitando esta forma recursiva das probabilidades p_j podemos escrever o seguinte Algoritmo para obter x , amostra de tamanho 1 de $X \sim \mathcal{P}(\lambda)$:

1. Gerar u , amostra da $U \sim \mathcal{U}(0, 1)$.
2. Fazer $x = 0$, $p = p_0$ e $s = p$
3. Enquanto $u > s$ fazer
 - (a) $x = x + 1$.
 - (b) $p = \frac{\lambda}{x}$.
 - (c) $s = s + p$.
4. Retornar x .

No livro de Devroye [36] pode ver-se uma melhora deste método; não a incluímos nestas notas pois mais adiante consideraremos outros algoritmos mais interessantes para a geração de amostras da $\mathcal{P}(\lambda)$.

Nessa mesma referência podem ver-se outros casos particulares: inversão por truncamento de variáveis aleatórias contínuas, inversão baseada em comparações, etc. O primeiro destes requer uma coincidência entre o vetor \mathbf{p} da X discreta a ser gerada com a densidade f_Y de uma variável aleatória contínua Y nos inteiros (isto é, precisa achar Y tal que $f_Y(j) = p_j \forall 0 \leq j \leq n$), e tal que a geração de amostras de Y seja rápida —por inversão, por exemplo—. O segundo (inversão baseada em comparações) precisa de alocação dinâmica de memória, o que é restritivo em termos de linguagens de programação que dispõem desse recurso. Portanto, ambos os casos são particulares o suficiente para não exigir tratamento nestas notas. Só cabe uma consideração: a do caso a seguir, que aparece com certa frequência em problemas de simulação.

5.1.3 Variáveis Aleatórias Discretas Uniformemente Distribuídas

Seja $X \in \{0, \dots, n-1\}$, $n > 1$ com $\mathbf{p} = (p_0, \dots, p_{n-1})$ e $p_j = \frac{1}{n}$ para todo $0 \leq j \leq n-1$. Usando a notação $[\nu] = \max\{i \in \mathbb{N} : i \leq \nu\}$, podemos obter x usando o resultado (Exercício 5.2) $[nU] \sim \mathcal{U}\{0, \dots, n-1\}$, onde $U \sim \mathcal{U}(0, 1)$.

5.2 Métodos baseados em Tabelas

Os métodos vistos anteriormente, de uma forma ou de outra, utilizam uma tabela de probabilidades e comparações sucessivas para gerar uma amostra de $X \sim \mathbf{p}$. Essa tabela, implícita em princípio, pode ser explicitada em alguns casos e assim acelerar os algoritmos já vistos.

Alguns dos algoritmos mais eficientes aproveitam o princípio de *hashing*, amplamente utilizado nas aplicações computacionais. O leitor interessado pode ver a definição deste princípio na referência [78], e para a sua aplicação aos algoritmos de geração de variáveis discretas o livro de Devroye [36, pp. 96–102].

Veremos alguns casos particulares a seguir.

5.2.1 O Princípio da Tabela

Imaginemos a seguinte situação: queremos gerar amostras de $X \sim \mathbf{p} = (p_0, \dots, p_n)$, onde $p_j = \frac{k_j}{M}$ com k_j e M inteiros para todo $0 \leq j \leq n < +\infty$; isto é, todas as probabilidades p_j são números fracionários com denominador comum M . Para gerar $x = X(\omega)$ podemos utilizar o algoritmo a seguir: (a) Construa-se um vetor $A = (a_1, \dots, a_M)$ com $a_\ell = 0$ para $1 \leq \ell \leq k_0$, $a_\ell = 1$ para $k_0 < \ell \leq k_0 + k_1, \dots$, $a_\ell = n$ para $k_0 + \dots + k_{n-1} < \ell \leq M$; isto é, o vetor A tem k_0 elementos "0", k_1 elementos "1", ..., k_n elementos "n". (b) Tome-se um elemento desse vetor escolhido ao acaso uniformemente em A .

A fraqueza deste método é a dificuldade de achar problemas práticos onde as p_j têm a forma requerida e, caso defrontemo-nos com um tal caso, tal vez M seja proibitivamente grande.

5.2.2 O Método do Histograma

O leitor deve estar familiarizado com a construção de histogramas baseados numa amostra (x_1, \dots, x_N) de tamanho N . Estes desenhos podem ser obtidos como saídas da grande maioria dos pacotes mais conhecidos (STATGRAPHICS, GLIM, etc.), e são de muita utilidade para uma análise prévia de dados.

Outra aplicação de grande interesse é, dado $H_C(x_1, \dots, x_N)$, o histograma da amostra (x_1, \dots, x_N) e a partição C , gerar uma amostra da distribuição empírica IPr_{H_C} induzida por H_C . Esta geração é muito usada associada a técnicas Bayesianas.

Precisando um pouco mais os termos, sejam a amostra de tamanho N dada $(x_1, \dots, x_N) \in \mathbb{R}^N$; $C = \{C_i\}$, $1 \leq i \leq k$; $C_i = (c_i^-, c_i^+]$, com $c_i^- < c_i^+$ reais para $i = 2, \dots, k-1$; $C_1 = (-\infty, c_1^+]$, $C_k = (c_k^-, +\infty)$ tais que $c_i^- = c_{i-1}^+$ para todo $1 < i \leq k$ (isto é, como $\bigcup_{i=1}^k C_i = \mathbb{R}$ e $\bigcap_{i=1}^k C_i = \emptyset$, C é uma partição da reta), então definimos $H_C: \mathbb{R} \rightarrow [0, 1]$, o histograma (de proporções, para alguns autores) como

$$H_C(t) = \begin{cases} \frac{1}{N} \# \{x_j: x_j < c_1^+\} & \text{para } t = c_1^+ - \frac{c_2^+ - c_2^-}{2} \\ \frac{1}{N} \# \{x_j: x_j \in C_i\} & \text{para } t = \frac{c_i^+ + c_i^-}{2} \text{ e } 1 < i < k \\ \frac{1}{N} \# \{x_j: x_j > c_k^-\} & \text{para } t = c_k^- + \frac{c_{k-1}^+ - c_{k-1}^-}{2} \\ 0 & \text{caso contrário.} \end{cases}$$

Assim H_C é equivalente a um vetor de probabilidades \mathbf{p} , só que com os índices em $\{c_1^+ - \frac{c_2^+ + c_2^-}{2}, \{\frac{c_i^+ + c_i^-}{2}\}_{1 < i < k}, c_k^- + \frac{c_{k-1}^+ - c_{k-1}^-}{2}\}$. Desta forma, podemos definir uma variável aleatória discreta $Y \sim \text{IPr}_{H_C}$ que assume valores no conjunto anterior de índices, e gerar amostras de Y usando algum dos métodos já apresentados. Diremos que a distribuição de Y foi induzida por H_C .

5.3 Método da flecha (alias)

O método a seguir, proposto por Walker [125,124] permite gerar amostras da variável aleatória $X \sim \mathbf{p} = (p_0, \dots, p_{k-1})$ com uma tabela cuja ordem de tamanho é $\mathcal{O}(k)$ e tempo máximo de busca independente de \mathbf{p} . Baseia-se no seguinte

Teorema 5.3.1 *Toda variável aleatória com distribuição caracterizada pelo vetor de probabilidades $\mathbf{p} = (p_0, \dots, p_{k-1})$ pode ser re-escrita como uma mistura equiprovável de k vetores aleatórios bi-dimensionais. Isto é, existem k pares de inteiros $(i_0, j_0), \dots, (i_{k-1}, j_{k-1})$ e um vetor de probabilidades $\mathbf{q} = (q_0, \dots, q_{k-1})$ tais que*

$$p_i = \frac{1}{k} \sum_{0 \leq \ell \leq k-1} (q_\ell \mathbb{I}_{\{i\}}(i_\ell) + (1 - q_\ell) \mathbb{I}_{\{i\}}(j_\ell)), \quad \forall 0 \leq i < k.$$

O leitor interessado na prova deste Teorema pode consultar as referências indicadas ou [36].

Para produzir um algoritmo utilizando este resultado precisamos (a) gerar efetivamente a mistura equiprovável e (b) gerar o valor. Este último item não oferece dificuldade nenhuma, pois a equiprobabilidade da mistura quer dizer que precisamos apenas *lançar uma flecha* no quadrado $[0, 1] \times [0, 1]$ e ler o índice da região flechada. Temos assim o seguinte **algoritmo da flecha**:

1. Geração da tabela para o método da flecha
 - (a) Atribuir aos conjuntos de inteiros G e S os valores iniciais $G = \emptyset$ e $S = \emptyset$.
 - (b) Para todo $0 \leq \ell \leq k-1$ fazer
 - i. $q_\ell = kp_\ell$.
 - ii. Se $q_\ell < 1$ então fazer $S = S \cup \{\ell\}$. Se não fazer $G = G \cup \{\ell\}$.
 - (c) Enquanto $S \neq \emptyset$ fazer
 - i. Pegar $k \in G$ e $\ell \in S$.
 - ii. Atribuir $j_\ell = k$.
 - iii. Atribuir $q_k = q_k - 1 + q_\ell$.
 - iv. Se $q_k < 1$ então atribuir $G = G \setminus \{k\}$ e $S = S \cup \{k\}$.
 - v. Fazer $S = S \setminus \{k\}$.
2. Obtenção da amostra x da X
 - (a) Gerar u da $\mathcal{U}(0, 1)$. Atribuir $x = [ku]$
 - (b) Gerar v da $\mathcal{U}(0, 1)$.
 - (c) Se $v < q_x$ atribuir $x = i_x$ e finalizar. Caso contrário atribuir $x = j_x$ e retornar.

Caso seja necessária uma amostra de tamanho n , deverá repetir-se n vezes apenas o passo 2 do Algoritmo. Na referência [36] o leitor achará métodos para sofisticar (e acelerar) o Algoritmo descrito acima.

5.4 O Método de Rejeição

O método visto na Secção 4.3 é válido também para o caso de variáveis aleatórias discretas. Neste caso a variável aleatória X tem a sua distribuição caracterizada pelo vetor $\mathbf{p} = (p_0, \dots)$, e poderemos usar o Método de Rejeição se existem um vetor de probabilidades $\mathbf{q} = (q_0, \dots)$ e uma constante $c \geq 1$ tais que $p_i \leq cq_i$ para todo $i \geq 0$; onde c é a constante de rejeição e a variável aleatória Y , cuja distribuição é caracterizada pelo vetor \mathbf{q} , é de fácil geração. Assim sendo, podemos utilizar o seguinte **Algoritmo de Rejeição**:

1. Repetir
 - (a) Gerar u da $U \sim \mathcal{U}(0, 1)$.
 - (b) Gerar j da $Y \sim q$.
2. Enquanto $ucqj \leq p_j$.
3. Retornar j .

A generalidade deste método faz com que ele seja mais demorado do que algumas técnicas especializadas. Outro problema é a necessidade de achar os q e c adequados (como toda vez que se emprega o Método de Rejeição). Porém, ele é bastante útil quando as variáveis aleatórias a serem geradas não são identicamente distribuídas, isto é quando precisamos obter amostras de distribuições que mudam frequentemente; nestos casos, a construção de tabelas para cada distribuição pode consumir muito tempo, e pode ser vantajoso aplicar o método acima descrito.

5.5 Exercícios

Nos exercícios onde pede-se realizar um programa de computador, pode-se utilizar qualquer linguagem que lhe seja familiar.

Exercício 5.1: Considere a variável aleatória discreta $X \in \{0, \dots, 100\}$ com distribuição dada pelo vetor p definido por

$$p_j = \begin{cases} \frac{1}{200} & \text{se } 0 \leq j \leq 99 \\ \frac{1}{2} & \text{se } j = 100 \end{cases}.$$

(a) Implemente o Algoritmo de inversão direto definido na Secção 5.1 para a geração de X .
 (b) Obtenha uma amostra (x_1, \dots, x_{1000}) de tamanho 1000 usando o algoritmo implementado em (a), tomando o cuidado de medir t_k , os tempos de CPU necessários para obter cada x_k , $1 \leq k \leq 1000$. Calcule $\bar{k} = \frac{1}{1000} \sum_{k=1}^{1000} t_k$. (c) Considere agora uma nova variável aleatória $\tilde{X} \sim \tilde{p}$ onde $\tilde{p}_0 = \frac{1}{2}$, $\tilde{p}_j = p_j$ se $1 \leq j \leq 99$ e $\tilde{p}_{100} = \frac{1}{200}$. Repita os passos “(a)” e “(b)” acima para a geração de \tilde{X} e compare \bar{k} com \tilde{k} . (d) Tire as suas conclusões e formalize uma melhora do Algoritmo de Inversão geral para variáveis aleatórias discretas usando permutações adequadas (ordenamento?) do vetor p ; isto é, encontre a permutação π tal que $\mathbb{E}_{\pi(p)}\{T\}$ seja mínima, onde T é a variável aleatória que representa o número de passos que o Algoritmo de Inversão geral tem que realizar até obter uma amostra de $\tilde{X} \sim \pi(p)$.

Exercício 5.2: Prove que $[nU] \sim \mathcal{U}\{0, \dots, n-1\}$, onde $U \sim \mathcal{U}(0, 1)$, para todo $n \geq 1$. Escreva o programa que implementa este algoritmo, para n geral; isto é, faça-o tal que n seja um argumento da sua função.

Exercício 5.3: (Verificação de uma possível) Acumulação de desvios das hipóteses de uniformidade do seu gerador. Aplicaremos sequencialmente o método do histograma 5.2.2 para ver se o gerador de uniformes que o leitor está utilizando não acumula desvios significativos das hipóteses desejadas. Para isto podemos considerar uma partição do intervalo $[0, 1]$ da seguinte forma: $C = \{(\frac{k}{10}, \frac{k+1}{10}]\}_{0 \leq k < 10}$.

1. Gere uma amostra de tamanho 1000 da $\mathcal{U}(0,1)$ fornecida pelo seu gerador de uniformes.
2. Forme o histograma $H_C^{(0)}$ da amostra gerada acima, com C a partição indicada previamente.
3. Para cada $1 \leq \ell \leq 100$ faça:
 - (a) Gere uma amostra de tamanho 1000 da distribuição induzida por $H_C^{(\ell-1)}$.
 - (b) Forme o histograma $H_C^{(\ell)}$, com C acima e os valores gerados na parte anterior.
4. Compare $H_C^{(100)}$ com $H_C^{(0)}$ e tire conclusões. O que acontece com as caudas?

Exercício 5.4: O pacote de análise de dados que o leitor já deve estar usando tem, quase com certeza, algum conjunto de dados univariados (caso esta conjectura seja falsa, arrume dados que tenham sido de interesse para alguém). Digamos que essa amostra é de tamanho N . (a) Faça o histograma H_C destes dados para alguma partição C apropriada aos mesmos. (b) Gere uma amostra de tamanho N da IPr_{H_C} induzida pelo histograma H_C . (c) Faça o teste de Kolmogorov-Smirnov (para duas amostras) e verifique se a hipótese H_0 (as amostras provêm da mesma distribuição) é rejeitada. Caso isto ocorra, explique por que.

Exercício 5.5: Dizemos que a variável aleatória X tem distribuição normal discreta com parâmetro $\sigma > 0$ se o p que a caracteriza é dado por

$$p_j = K \frac{\left(|j| + \frac{1}{2}\right)^2}{2\sigma^2} \text{ para todo } j \text{ inteiro,}$$

onde K é uma constante de normalização dada por $K = \sum_i \frac{(|i| + \frac{1}{2})^2}{2\sigma^2}$.

(a) Prove que $K = \sigma^{-1}((2\pi)^{-1/2} + o(1))$ quando $\sigma \rightarrow \infty$. (b) Escreva um programa que implementa algum algoritmo para a geração de amostras da X , tomando a providência de incluir uma medida do tempo de CPU necessário para gerar cada amostra (c) Prove que o algoritmo a seguir gera amostras da X :

1. Repetir:
 - (a) Gerar uma amostra y da $Y \sim \mathcal{N}(0,1)$.
 - (b) Definir $j = [y + 1/2]$.
 - (c) Definir $z = |j| + 1/2$.
 - (d) Gerar uma amostra u da $U \sim \mathcal{U}(0,1)$.
2. Enquanto $-2\sigma^2 \ln(u) \geq z^2 - j^2$.
3. Retornar j .

(d) Escreva o programa que implementa o algoritmo acima, com as mesmas providências de (b). (e) Gere amostras de tamanho 1000 para cada valor de $\sigma = 0.1, 1$ e 10 usando ambos os algoritmos. Compare os tempos.

Exercício 5.6: (Exemplo interessante para alguns problemas de Processamento de Imagens: as imagens de radar de abertura sintética (SAR) apresentam um ruído deste tipo) Dizemos que a variável aleatória X tem distribuição de Raighley discreta com parâmetro $\xi > 0$ se o \mathbf{p} que a caracteriza é dado por

$$p_j = \int_{(j,j+1]} \frac{y}{\xi^2} \exp\left\{-\frac{1}{2}\left(\frac{y}{\xi}\right)^2\right\} dy \text{ para todo } j \in \mathbb{N}.$$

(a) Escreva as probabilidades p_i em forma analítica usando a função de distribuição acumulada F_Y da variável aleatória Y que tem por densidade $f_Y(y)$ o argumento da integral acima. (b) Prove que $\mathbf{p} = (p_0, p_1, \dots)$ é um vetor de probabilidades. (c) Ache a posição dos máximos de \mathbf{p} em função de $\xi > 0$, isto é encontre o conjunto $j_\xi^* = \max_{\{0 \leq j\}} p_j$. (d) Escreva o programa que implementa algum algoritmo para a geração de amostras x_1, \dots, x_N da X acima, com $\xi > 0$ como parâmetro. (e) Usando a função de distribuição acumulada $F_Y(y)$, calculada acima, implemente a geração de amostras da variável aleatória contínua $Y \sim \mathcal{R}(\xi)$ pelo método de inversão. (f) Aproveite a implementação de item (e) para escrever outro algoritmo de geração de amostras da X , isto é sabendo gerar y_1, \dots, y_N , simplesmente retorne x_1, \dots, x_N onde $x_i = \lfloor y_i \rfloor$ para todo $1 \leq i \leq N$. (g) Compare os tempos de CPU necessários para ambos os algoritmos.

Capítulo 6

Casos especiais de variáveis aleatórias

Neste Capítulo veremos alguns algoritmos especializados para obter amostras das distribuições discretas e contínuas que aparecem mais freqüentemente na prática de Simulação Estocástica.

Quando o usuário destas técnicas (de Simulação Estocástica) resolve construir a sua própria biblioteca de funções —no ambiente computacional (máquina, linguagem, compilador, etc.) que lhe resulta mais familiar ou adequado aos seus propósitos—, pode parecer uma tarefa impossível a de construir programas que ou (i) sejam gerais o suficiente para satisfazer uma classe muito larga de distribuições desejadas e, ao mesmo tempo, que sejam aceitavelmente rápidos ou (ii) implementar algoritmos sofisticados, e rápidos, especializados para cada uma das distribuições que virão aparecer nas experiências.

Uma tendência mais ou menos generalizada é a de construir bons algoritmos (especializados e rápidos) para algumas distribuições importantes. Estes serão os *tijolos* com que o resto do **ambiente de simulação estocástica** será montado. Então, a partir destes algoritmos —devidamente testados e documentados— obter amostras de outras distribuições não consideradas como básicas.

A obtenção destas amostras pode ser feita aproveitando as propriedades probabilísticas das variáveis aleatórias envolvidas, por exemplo, *misturando* ocorrências das básicas. Desta maneira, com uma biblioteca especializada, relativamente reduzida e muito bem documentada, o usuário poderá ter acesso a um leque bastante abrangente de distribuições.

Para aqueles casos onde não é claro como *misturar* ocorrências de variáveis aleatórias básicas para obter o resultado desejado, o leitor sempre poderá lançar mão dos métodos gerais apresentados nos dois Capítulos anteriores.

Prosseguindo na ordem das notas, neste Capítulo veremos algoritmos especializados para algumas variáveis aleatórias contínuas e a seguir, para algumas discretas. Os casos considerados são: variáveis Normais, Gamma e von Mises, entre as primeiras, e Poisson e Binomiais para as segundas. Outros casos (Exponenciais, χ^2 , Cauchy, Beta, Raighley, Weibull, etc.) ou já foram estudados em outras partes destas notas ou decorrem trivialmente do material aqui exposto.

Consideraremos alguns casos de interesse particular, seja pela sua ampla utilidade ou seja

pelas características especiais da sua simulação.

6.1 Variáveis Aleatórias Normais

Sabemos que a variável aleatória $X: \Omega \rightarrow \mathbb{R}$ é dita ter distribuição Normal Padrão se a sua densidade é dada por

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\{-y^2/2\} \quad \forall y \in \mathbb{R};$$

então escreveremos $X \sim \mathcal{N}(0,1)$. Se $Y = \mu + \sigma X$ com $\mu \in \mathbb{R}$ e $\sigma > 0$, então escreveremos $Y \sim \mathcal{N}(\mu, \sigma^2)$.

Existem vários algoritmos para a geração de amostras de variáveis aleatórias Normais Padrão. Alguns já foram apresentados nos exemplos 4.1.1 (usando o Teorema Central do Limite) e 4.5.2 (pelo Método do Quociente de Uniformes) e nos exercícios 4.6, 4.9 (pelo Método de Inversão) e 4.17. Veremos a seguir alguns métodos especializados para este caso.

6.1.1 Método Polar ou de Box-Muller

Esta classe de algoritmos serve para gerar *pacotes* de amostras independentes de tamanho $d \geq 2$ da $X \sim \mathcal{N}(0,1)$. Veremos com detalhe implementações do caso $d = 2$, mas para tanto precisaremos de algumas definições e teoremas para sustentar a validade dos algoritmos. Para generalidade, apresentaremos também o algoritmo geral.

Definição 6.1.1 Diremos que vetor aleatório $X: \Omega \rightarrow \mathbb{R}^d$ tem simetria radial se para toda matriz A ortonormal de dimensão $d \times d$ a distribuição do vetor aleatório $Y = AX$ é a mesma que a do vetor X ; escreveremos “ X sr”. Se, além de satisfazer a condição anterior, vale que $\mathbb{P}(X = 0) = 0$, onde 0 é a origem de \mathbb{R}^d , então diremos que o vetor aleatório X tem simetria radial estrita; escreveremos “ X sre”.

Note-se que a definição acima diz respeito à distribuição do vetor d -dimensional Y , que é uma rotação do vetor X . Escreveremos $C_d = \{x = (x_1, \dots, x_d) \in \mathbb{R}^d: \sum_{i=1}^d x_i^2 = 1\}$, isto é a casca da esfera d -dimensional de raio unitário e centro na origem do \mathbb{R}^d .

Definição 6.1.2 Seja $X: \Omega \rightarrow \mathbb{R}^d$ com sre e densidade $f_X: \mathbb{R}^d \rightarrow \mathbb{R}^+$. Pode ver-se (por exemplo em [36]) que existe $g: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ tal que $\mathbb{P}(f_X(x) \neq g(\|x\|)) = 0$. Uma tal função g será chamada função definidora de f_X .

Esta função será de utilidade mais adiante, e é tal que caracteriza a distribuição de X (daí o seu nome). Pode ver-se que g é tal que

$$dV_d \int_0^{+\infty} r^{d-1} g(r) dr = 1 \text{ onde } V_d = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2} + 1)},$$

isto é V_d é o volume de C_d .

Teorema 6.1.1 Propriedades das distribuições na casca da esfera unitária:

1. Se \mathbf{X} sse então $\frac{\mathbf{X}}{\|\mathbf{X}\|} \sim \mathcal{U}(C_d)$.
2. Se $\mathbf{X} \sim \mathcal{U}(C_d)$ então o vetor X_1^2, \dots, X_d^2 tem a mesma distribuição do vetor $S^{-1}(Y_1, \dots, Y_d)$, onde Y_1, \dots, Y_d são vâiid com distribuição $\text{Gamma}(1/2, 1)$ (ou seja $f_{Y_i}(y) = (\Gamma(1/2)\sqrt{y}e^y)^{-1}$) e $S = \sum_{i=1}^d Y_i$.
3. Se $\mathbf{X} \sim \mathcal{U}(C_d)$ então $X_1^2 \sim \text{Beta}(1/2, (d-1)/2)$.

Teorema 6.1.2 A distribuição Normal: se N_1, \dots, N_d são vâiid com distribuição $\mathcal{N}(0, 1)$ então o vetor $\mathbf{X} = (N_1, \dots, N_d)$ tem sse com densidade $f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-d/2} \exp\{-\|\mathbf{x}\|^2/2\}$. Ainda mais, se (X_1, \dots, X_d) tem sse e as suas componentes X_i são independentes então as X_i são vâiid $\mathcal{N}(0, \sigma_i^2)$ com $\sigma_i > 0 \forall 1 \leq i \leq d$.

Teorema 6.1.3 A transformação radial:

1. Se \mathbf{X} tem sse em \mathbb{R}^d com função definidora g , então $R = \|\mathbf{X}\|$ tem densidade $dV_d r^{d-1} g(r) \forall r \geq 0$.
2. Se $\mathbf{X} \sim \mathcal{U}(C_d)$ e R é independente de \mathbf{X} com a densidade dada acima, então $R\mathbf{X}$ tem sse em \mathbb{R}^d com função definidora g .
3. Se \mathbf{X} tem sse em \mathbb{R}^d com função definidora g , e se R é uma variável aleatória independente de \mathbf{X} não negativa (ou seja, se $R: \Omega \rightarrow \mathbb{R}^+$) com densidade h então $R\mathbf{X}$ tem sr com função definidora

$$\tilde{g}(r) = \int_0^{+\infty} \frac{h(u)}{u^d} g\left(\frac{r}{u}\right) du.$$

Teorema 6.1.4 Se $(X_1, X_2) \sim \mathcal{U}(C_2)$ e $S = \sqrt{X_1^2 + X_2^2}$ então

1. S e $S^{-1}(X_1, X_2)$ são independentes.
2. $S^2 \sim \mathcal{U}(0, 1)$ e $(\cos(2\pi S^2), \sin(2\pi S^2)) \sim \mathcal{U}(C_2)$.
3. X_2/X_1 tem distribuição Cauchy padrão.
4. $S^{-1}(X_1, X_2) \sim \mathcal{U}(C_2)$.
5. $S^{-2}(X_1^2 - X_2^2, 2X_1X_2) \sim \mathcal{U}(C_2)$.

Para uma prova dos Teoremas acima o leitor pode consultar o livro de Devroye [36] ou textos avançados de Distribuições Multivariadas. Vejamos agora as técnicas que, aproveitando os resultados vistos, permitem gerar amostras da $\mathbf{X} \sim \mathcal{N}(0, 1)$.

Algoritmo Geral do Método Polar

Seja $d \geq 2$, o algoritmo a seguir permite gerar uma amostra de tamanho d de vâiid com distribuição $\mathcal{N}(0, 1)$ usando um vetor d -dimensional de Uniformes e uma Gamma:

1. Gerar x amostra da $\mathbf{X} \sim \mathcal{U}(C_d)$.

2. Gerar r amostra da R com densidade $f_R(r) = dV_d r^{d-1} \exp\{-\frac{r^2}{2}\} \mathbb{I}_{(0,+\infty)}(r)$, isto é $R = \sqrt{2G}$ com $G \sim \text{Gamma}(d/2, 1)$.
3. Retornar rx .

Para a grande maioria das aplicações podemos obter a sequência x_1, \dots, x_n , onde $n = 2k$ e $k = 1, 2, \dots$ inteiro, com k chamadas ao algoritmo desenhado com $d = 2$. Com esta consideração podemos usar qualquer dos seguintes algoritmos:

Algoritmo N1

Geração de uma amostra de tamanho 2 da $X \sim \mathcal{N}(0, 1)$ usando uma Exponencial e um vetor de Unifomes.

1. Gerar y da $Y \sim \mathcal{E}(1)$.
2. Gerar (v, w) da $(V, W) \sim \mathcal{U}(C_2)$.
3. Atribuir $r = \sqrt{v^2 + w^2}$.
4. Retornar $(x_1, x_2) = \frac{\sqrt{2y}}{r}(v, w)$.

Algoritmo N2

Geração de uma amostra de tamanho 2 da $X \sim \mathcal{N}(0, 1)$ usando uma Exponencial e uma Uniforme.

1. Gerar y da $Y \sim \mathcal{E}(1)$.
2. Gerar u da $U \sim \mathcal{U}(0, 1)$.
3. Retornar $(x_1, x_2) = \sqrt{2y}(\cos(2\pi u), \sin(2\pi u))$.

Algoritmo N3

Geração de uma amostra de tamanho 2 da $X \sim \mathcal{N}(0, 1)$ usando uma Exponencial e um vetor de Uniformes.

1. Gerar y da $Y \sim \mathcal{E}(1)$.
2. Gerar (v, w) da $(V, W) \sim \mathcal{U}(C_2)$.
3. Atribuir $s = v^2 + w^2$.
4. Retornar $(x_1, x_2) = \frac{\sqrt{2y}}{s}(v^2 - w^2, 2vw)$.

Algoritmo N4

Geração de uma amostra de tamanho 2 da $X \sim \mathcal{N}(0, 1)$ usando um vetor de Uniformes.

1. Gerar (v, w) da $(V, W) \sim \mathcal{U}(C_2)$.
2. Atribuir $r = \sqrt{v^2 + w^2}$.

3. Retornar $(x_1, x_2) = \frac{\sqrt{-4 \ln(r)}}{r}(v, w)$.

A validade dos algoritmos considerados decorre dos Teoremas 6.1.1, 6.1.2, 6.1.3 e 6.1.4. Em particular, o leitor pode ver o Exercício 56 do Capítulo II do livro de James [65, pág. 95] para uma forma direta de provar a validade do **Algoritmo N2**.

Já foram vistas técnicas para gerar $U \sim \mathcal{U}(A)$, com A um conjunto arbitrário. Dado que para os algoritmos acima o conjunto A tem uma forma bem comportada, o método de rejeição é apropriado para estes casos.

6.1.2 Outros Métodos para Geração de $\mathcal{N}(0, 1)$

Os métodos a seguir podem ser considerados variações de algumas das técnicas vistas no decorrer destas notas. Por isto, os apresentaremos sem provas, e o leitor interessado nelas poderá consultar o livro de Devroye [36]

Métodos I e II de Johnson-Ramberg

São métodos para gerar pacotes de amostras, e podem ser competitivos com as técnicas Polares já vistas pois não precisam obter a raiz quadrada.

Algoritmo JR I

1. Gerar x amostra da $X \sim \mathcal{U}(C_2)$.
2. Gerar r amostra da $R = \sqrt{2G}$ com $G \sim \text{Gamma}(3/2, 1)$.
3. Retornar rx

Algoritmo JR II

1. Gerar x amostra da $X \sim \mathcal{U}(C_3)$.
2. Gerar r amostra da $R = \sqrt{2G}$ com $G \sim \text{Gamma}(2, 1)$.
3. Retornar rx .

Algoritmo de von Neumann

É o Método de Rejeição para a variável aleatória Y com distribuição de Laplace, isto é $f_Y(y) = \frac{1}{2} \exp\{-|y|\} \forall y \in \mathbb{R}$, e constante de rejeição ótima:

1. Atribuir $u = 0$ e $v = +\infty$.
2. Repetir enquanto $2v \leq (u - 1)^2$
 - (a) Gerar u e v valid com distribuição $\mathcal{E}(1)$.
3. Gerar u da $U \sim \mathcal{U}(0, 1)$ e atribuir $s = -1$ se $u \leq 1/2$, $s = +1$ se não.
4. Retornar su .

6.1.2.3 Algoritmo por Rejeição de Cauchys

Para utilizar a distribuição de Cauchy em forma adequada, foi calculado o valor ótimo da constante. Daí resulta o seguinte Algoritmo:

1. Atribuir $a = \sqrt{e}/2$, $u_1 = +\infty$ e $s = 0$.
2. Repetir enquanto $u \leq a(1+s) \exp\{-s/2\}$:
 - (a) Gerar u_1 e u_2 amostras independentes da $U \sim \mathcal{U}(0,1)$.
 - (b) Atribuir $x = \tan(\pi u_2)$ e $s = x^2$ (agora x é uma amostra da X com distribuição Cauchy padrão).
3. Retornar x .

6.1.2.4 Algoritmo de Marsaglia e Bray

A validade do algoritmo a seguir depende de uma técnica que omitimos nestas notas: o Método de Composição-Rejeição. O leitor interessado pode consultar as referências [36, págs. 382–389] ou [74]. Porém, o algoritmo a seguir é de bastante interesse:

1. Gerar u amostra da $U \sim \mathcal{U}(0,1)$
2. Se $0 \leq u \leq 16/\sqrt{2\pi e}$ então faça
 - (a) Gerar u_1 e u_2 amostras independentes da $U \sim \mathcal{U}(0,1)$.
 - (b) Atribuir $x = 2.3153508u - 1 + u_1 + u_2$
 - (c) Retornar x .
3. Se $16/\sqrt{12\pi e} < u \leq 0.9745$ então faça
 - (a) Gerar u_1 amostra da $U \sim \mathcal{U}(0,1)$.
 - (b) Atribuir $x = \frac{3}{2}(u_1 - 1 + 9.0334237(u - 0.8638))$.
 - (c) Retornar x .
4. Se $0.9972002 < u \leq 1$ então atribua $x = u_1 = +\infty$
 - (a) Enquanto $xu_1 \leq \frac{9}{2}$ faça
 - i. Gerar u_1 e u_2 amostras independentes da $U \sim \mathcal{U}(0,1)$.
 - ii. Atribuir $x = \frac{9}{2} - \ln(u_2)$.
 - (b) Atribuir $s = \frac{u - 0.9986501}{|u - 0.9986501|}$ e $x = s\sqrt{2x}$.
 - (c) Retorne x .
5. Se $0.9745 < u \leq 0.9973002$ então atribua $s = w = -\infty$ e $v = 0$.
 - (a) Enquanto $u \leq 49.0024445 \exp\{-v^2/2\} - (s + w)$ faça
 - i. Gerar v amostra da $V \sim \mathcal{U}(0,3)$.
 - ii. Atribuir $w = 6.6613339(3 - v)^2$ e $s = 0$.
 - iii. Se $v < \frac{3}{2}$ então atribuir $s = 6.0432809(\frac{3}{2} - v)$.
 - iv. Se $v < 1$ então atribuir $s = s + 13.2626678(3 - v)^2 - w$.
6. Retornar x .

6.2 Variáveis Aleatórias Gamma

Dizemos que a variável aleatória $X: \Omega \rightarrow \mathbb{R}^+$ tem distribuição Gamma com parâmetros (α, β) ambos reais positivos (em símbolos $X \sim \mathcal{G}(\alpha, \beta)$) se a sua densidade é

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\{-x/\beta\} \mathbb{I}_{(0,+\infty)}(x).$$

O leitor já deve ter sentido que a geração de muitas classes de variáveis aleatórias depende criticamente da geração de variáveis aleatórias Gamma. Historicamente, foi o problema de geração de amostras de variáveis aleatórias univariadas que mais atenção tem recebido.

Outras aplicações onde elas são interessantes (para mencionar apenas algumas) são as relacionadas a Modelos de Regressão Generalizados. Nos Exercícios deste Capítulo ilustraremos melhor esta afirmação. Para isto, veremos alguns algoritmos especializados que dependem das propriedades enunciadas a seguir, e cuja prova deixamos como Exercício:

- Sejam X_1 e X_2 variáveis aleatórias independentes com distribuições $\mathcal{G}(\alpha_1, 1)$ e $\mathcal{G}(\alpha_2, 1)$ respectivamente. Então as variáveis aleatórias $\frac{X_1}{X_1+X_2}$ e X_1+X_2 são independentes com distribuições $\text{Beta}(\alpha_1, \alpha_2)$ e $\mathcal{G}(\alpha_1 + \alpha_2, 1)$ resp.
- Sejam as variáveis aleatórias $X \sim \mathcal{G}(\alpha, 1)$ e $Y \sim \mathcal{B}(\beta, \beta - \alpha)$ (com $0 < \alpha < \beta$). Então XY e $X(1 - Y)$ são independentes com distribuições $\mathcal{G}(\beta, 1)$ e $\mathcal{G}(\beta - \alpha, 1)$ resp.
- Se $X \sim \mathcal{G}(\alpha, \beta)$ então $cX \sim \mathcal{G}(\alpha, c\beta)$.

O último resultado diz que, para o problema de geração de amostras da $X \sim \mathcal{G}(\alpha, \beta)$ basta considerar o problema de gerar amostras da $X' \sim \mathcal{G}(\alpha, 1)$. O leitor verá que o valor do parâmetro α é o *divisor de águas* entre algoritmos.

6.2.1 Casos Simples de Geração de Gammas

Da densidade f_X acima, o leitor poderá concluir a validade dos seguintes algoritmos: gerar x amostra da $X \sim \mathcal{G}(\alpha, \beta)$ com

- se $\alpha = k \in \mathbb{N} \setminus \{0\}$ e $\beta = 1$ então retornar $x = -\ln(\prod_{i=1}^k u_i)$, onde u_1, \dots, u_k são amostras independentes da $U \sim \mathcal{U}(0, 1)$;
- se $\alpha = \frac{1}{2} + k$, com k inteiro, e $\beta = 1$ então retornar $x = z + \frac{1}{2}y^2$, onde z é amostra de $Z \sim \mathcal{G}(k, 1)$ independente de $Y \sim \mathcal{N}(0, 1)$, cuja amostra é y .

6.2.2 Geração de Gammas com $\alpha > 1$ e $\beta = 1$

Veremos dois métodos básicos, ambos os dois baseados no Método de Rejeição Geral.

6.2.2.1 Algoritmo XG

Deve-se a Best [18,19]:

1. Atribuir $b = \alpha - 1$, $c = 3\alpha - \frac{3}{4}$ e $C = 0$.
2. Repetir enquanto $C = 0$
 - (a) Gerar u_1 e u_2 amostras independentes da $U \sim \mathcal{U}(0,1)$.
 - (b) Atribuir $w = u_1(1 - u_1)$, $y = \sqrt{c/w}(u_1 - 1/2)$, $x = b + y$.
 - (c) Se $x \geq 0$ então
 - i. Atribuir $z = 64w^3u_2^2$.
 - ii. Atribuir $C = \mathbb{I}_{(\cdot,1]}(z^{-1}(1 - 2y^2/x))$.
 - iii. Se $C = 0$ atribuir $C = \mathbb{I}_{(\cdot,1]}(2(b \ln(x/b) - y)/\ln(z))$.
3. Retornar x .

6.2.2.2 Algoritmo GB

Deve-se a Cheng [31]:

1. Atribuir $b = \alpha - \ln(4)$, $c = \alpha + \sqrt{2\alpha - 1}$, $C = 0$.
2. Repetir enquanto $C = 0$
 - (a) Gerar u_1 e u_2 amostras independentes da $U \sim \mathcal{U}(0,1)$.
 - (b) Atribuir $y = \alpha \ln(u_2/(1 - u_2))$, $x = \alpha \exp\{u_2\}$, $z = u_1u_2^2$, $r = b + cy - x$.
 - (c) $C = \mathbb{I}_{(\cdot,1]}(r^{-1}\frac{9}{2}z - \ln(9/2))$.
 - (d) Se $C = 0$ então atribuir $C = \mathbb{I}_{(\cdot,1]}(r^{-1} \ln(z))$.
3. Retornar x .

6.2.3 Geração de Gammas com $\alpha < 1$ e $\beta = 1$

6.2.3.1 Mistura de Gamma com Exponencial

Talvez o caso mais simples de geração de amostras x da $X \sim \mathcal{G}(\alpha, 1)$ com $\alpha < 1$ seja o que usa uma amostra y da $Y \sim \mathcal{G}(\alpha + 1, 1)$ e uma Exponencial:

1. Gerar y da $Y \sim \mathcal{G}(\alpha + 1, 1)$ e u da $U \sim \mathcal{U}(0,1)$.
2. Atribuir $z = -\ln(u)$.
3. Retornar $x = y \exp\{-z/\alpha\}$.

6.2.3.2 Mistura de Gamma com Beta

Veremos dois algoritmos devidos a Johnk [67], um para gerar uma x amostra da $X \sim \text{Beta}(\alpha, 1 - \alpha)$, com $\alpha < 1$ usado para gerar uma y amostra da $Y \sim \mathcal{G}(\alpha, 1)$. O método baseia-se no fato que, com X nas condições acima, e com $E \sim \mathcal{E}(1)$ independente de X , vale que $XE \sim \mathcal{G}(\alpha, 1)$

Gerador de Beta Para obter uma amostra x da $X \sim \mathcal{B}(\alpha, \beta)$ faça

1. Atribuir $y_1 = y_2 = +\infty$.
2. Repetir enquanto $y_1 + y_2 \leq 1$
 - (a) Gerar u_1 e u_2 amostras independentes da $U \sim \mathcal{U}(0, 1)$.
 - (b) Atribuir $y_1 = \sqrt[3]{u_1}$ e $y_2 = \sqrt[3]{u_2}$.
3. Retornar $x = \frac{y_1}{y_1 + y_2}$.

Gerador de Gamma No algoritmo a seguir repetiremos, para maior clareza, a parte de geração da Beta.

1. Atribuir $y_1 = y_2 = +\infty$.
2. Repetir enquanto $y_1 + y_2 \leq 1$
 - (a) Gerar u_1 e u_2 amostras independentes da $U \sim \mathcal{U}(0, 1)$.
 - (b) Atribuir $y_1 = \sqrt[3]{u_1}$ e $y_2 = \sqrt[3]{u_2}$.
3. Gerar u_3 amostra da $U \sim \mathcal{U}(0, 1)$.
4. Retornar $x = \frac{-\ln(u_3)y_1}{y_1 + y_2}$.

Algoritmo de Rejeição Geral com Weibull

Lembremos que dizemos que $X \sim \mathcal{W}(\alpha)$, $\alpha > 0$ se a sua densidade é

$$f_W(w) = \alpha w^{\alpha-1} \exp\{-w^\alpha\} \mathbb{I}_{(0, \infty)}(w).$$

É fácil ver que a sua função de distribuição acumulada é dada por $(1 - \exp\{-x^\alpha\})\mathbb{I}_{(0, \infty)}(x)$, e daí resulta elementar obter amostras independentes w_1, \dots, w_n da $X \sim \mathcal{W}(\alpha)$, por exemplo pelo Método de Inversão.

Esta distribuição resulta interessante nesta Secção pois pode ser utilizada no método de Rejeição Geral para obter amostras da $\mathcal{G}(\alpha, 1)$, com $\alpha < 1$ usando o algoritmo a seguir:

1. Atribua $c = \alpha^{-1}$, $d = \alpha^{\alpha/(1-\alpha)}(1 - \alpha)$, $y_1 = y_2 = +\infty$ e $x = -\infty$.
2. Repetir enquanto $y_1 + y_2 \leq d + x$
 - (a) Gerar y_1 e y_2 amostras independentes da $Y \sim \mathcal{E}(1)$.
 - (b) Atribuir $x = y_1^c$.
3. Retorne x .

6.3 Variáveis Aleatórias von Mises

Estas variáveis aleatórias são interessante para modelar dados direcionais [95]. Diremos que a variável aleatória $\Theta: \Omega \rightarrow [-\pi, +\pi]$ tem distribuição de von Mises se a sua densidade é dada por

$$f_\Theta(\theta) = \frac{\exp\{\kappa \cos(\theta)\}}{2\pi I_0(\kappa)} \mathbb{I}_{[-\pi, +\pi]}(\theta),$$

onde $\kappa > 0$ é o parâmetro de forma e I_0 é a função de Bessel modificada de primeira classe e ordem zero, isto é

$$I_0(x) = \sum_{0 \leq j < \infty} \frac{1}{j!} (x/2)^{2j}.$$

Dado que não se conhece forma fechada para esta densidade nem para a função de distribuição acumulada, deveremos lançar mão do Método de Rejeição. O algoritmo a seguir deriva dele, e o leitor interessado na prova da validade do mesmo pode consultar [36, págs. 473–476].

1. Atribuir $r = 1 + \sqrt{1 + 4\kappa^2}$, $\rho = \frac{r - \sqrt{2r}}{2\kappa}$, $s = \frac{1 + \rho^2}{2\rho}$ e $C = 0$.
2. Repetir enquanto $C = 0$
 - (a) Gerar u_1 e u_2 amostras independentes da $U \sim \mathcal{U}(0, 1)$.
 - (b) Atribuir $z = \cos(\pi u_1)$, $w = \frac{1 + sz}{s + z}$ e $y = \kappa(s - w)$.
 - (c) Atribuir $C = \mathbb{I}_{[0, \cdot)}(w(2 - w) - v)$.
 - (d) Se $C = 0$ então atribuir $C = \mathbb{I}_{[0, \cdot)}(\ln(w/v) + 1 - w)$.
3. Atribuir $\theta = \frac{u}{|u| \cos(w)}$.
4. Retornar θ .

6.4 Distribuição de Poisson

Sabemos já que a variável aleatória $X: \Omega \rightarrow \mathbb{N}$ é dita com distribuição de Poisson e parâmetro $\lambda > 0$ (e escreveremos $X \sim \mathcal{P}(\lambda)$) se o seu vetor probabilidades $\mathbf{p} = (p_j)_{j \geq 0}$ está dado pelos valores

$$p_j = \Pr(X = j) = \frac{\lambda^j}{j!} \exp\{-\lambda\} \text{ para todo } j \in \mathbb{N}.$$

Esta distribuição também é chamada *lei de eventos raros*, pois pode ser obtida como a lei do número de sucessos em n ensaios binomiais com probabilidade p de sucesso, quando $n \rightarrow \infty$ e $p \rightarrow 0$.

Dado que no Capítulo 5 já vimos alguns dos principais algoritmos para gerar amostras de variáveis aleatórias com distribuição $\mathcal{P}(\lambda)$, neste apenas consideraremos àqueles que são particularmente interessantes.

6.4.1 Algoritmo dos Tempos de Chegada

O leitor é convidado a consultar o livro de James [65, págs. 22–27, 37–41] para ver a construção rigorosa de um dos processos estocásticos mais importantes: o Processo de Poisson. Das propriedades dele, que resumiremos na Proposição a seguir, deriva o algoritmo PTC.

Proposição 6.4.1 *Seja E_1, E_2, \dots uma sequência de variáveis com distribuição comum $\mathcal{E}(1)$ e seja a variável aleatória X definida para todo $\lambda > 0$ por*

$$X = \inf \left\{ k \in \mathbb{N} : \sum_{1 \leq i \leq k+1} E_i < \lambda \right\};$$

então $X \sim \mathcal{P}(\lambda)$.

Algoritmo PTC

1. Atribuir $x = 0, s = 0$.
2. Repetir
 - (a) Gerar y amostra da $Y \sim \mathcal{E}(1)$.
 - (b) Atribuir $s = s + y$.
 - (c) Se $s < \lambda$ então atribuir $x = x + 1$. Se não retornar x .

6.4.2 Algoritmo do Produto de Uniformes

Usando o fato já conhecido da relação entre variáveis aleatórias Uniformes e Exponenciais (isto é, que se $Y \sim \mathcal{E}(1)$ então $U = \exp\{-Y\} \sim \mathcal{U}(0, 1)$) podemos enunciar a seguinte Proposição

Proposição 6.4.2 *Seja U_1, U_2, \dots uma sequência de variáveis com distribuição comum $\mathcal{U}(0, 1)$ e seja a variável aleatória X definida para todo $\lambda > 0$ por*

$$X = \inf \left\{ k \in \mathbb{N} : \prod_{1 \leq i \leq k+1} U_i < \exp\{-\lambda\} \right\};$$

então $X \sim \mathcal{P}(\lambda)$.

O algoritmo que decorre da Proposição acima é:

Algoritmo PPU

1. Atribuir $x = 0, p = 1$.
2. Repetir
 - (a) Gerar u amostra da $U \sim \mathcal{U}(0, 1)$.
 - (b) Atribuir $p = up$.
 - (c) Se $p > \exp\{-\lambda\}$ então atribuir $x = x + 1$. Se não retornar x .

Ambos os algoritmos apresentados podem ser melhorados, principalmente no que diz respeito ao tempo de execução para valores “grandes” do parâmetro λ . Caso este ponto seja de especial interesse, o leitor pode ver as implementações apresentadas em [36, págs. 511–512].

6.5 Distribuição Binomial

Do já exposto, o leitor deve saber que dizemos que a variável aleatória X tem distribuição Binomial com parâmetro $p \in [0, 1]$ se $\text{IPr}(X = 1) = p = 1 - \text{IPr}(X = 0)$. Generalizando, diremos que a variável aleatória Y tem distribuição Binomial com parâmetros $p \in [0, 1]$ e $n \geq 1$ inteiro se ela distribui-se como a soma de n variáveis aleatórias independentes X como acima. Neste caso escreveremos $Y \sim \mathcal{B}(p, n)$, e é evidente que entenderemos $\mathcal{B}(p, 1) \equiv \mathcal{B}(p)$.

Nos primórdios destas notas vimos como gerar amostras da $X \sim \mathcal{B}(p)$ pelo Método de Inversão: para ter x amostra da $X \sim \mathcal{B}(p)$ gerar u amostra da $U \sim \mathcal{U}(0, 1)$ e atribuir $x = \Pi_{(0,p)}(u)$. Daí é evidente que, para gerar y amostra da $Y \sim \mathcal{B}(p, n)$ bastaria somar n amostras independentes x_1, \dots, x_n da $X \sim \mathcal{B}(p)$. Este método direto usa, então, n amostras de Uniformes para uma amostra da $\mathcal{B}(p, n)$.

6.5.1 Geração com Uma Uniforme

A técnica acima pode ser melhorada “aproveitando” a porção da Uniforme que não foi utilizada; note-se que as variáveis aleatórias $\Pi_{(0,p)}(U)$ e $\min(U/p, (1-U)/(1-p))$ são independentes, portanto, para gerar y amostra da $Y \sim B(p, n)$ pode proceder-se assim (suponhamos $p \leq \frac{1}{2}$, o outro caso é imediato):

1. Atribuir $x = 0$
2. Gerar u amostra da $U \sim U(0, 1)$.
3. Desde $i = 1$ até $i = n$ fazer
 - (a) Atribuir $b = \Pi_{(1-p,1)}(u)$ e $u = \frac{u-(1-p)b}{pb+(1-p)(1-b)}$.
 - (b) Atribuir $x = x + b$
4. Retornar x .

6.5.2 Busca Sequencial

O método geral de Busca Sequencial pode ser acelerado aproveitando a relação recursiva dos valores do vetor p :

$$p_{j+1} = p_j \frac{(n-1)p}{(j+1)(1-p)} \text{ para todo } 0 \leq j < n,$$

desta forma não são necessárias as tabelas de probabilidade... mas deve-se ter cuidado com uma possível acumulação de erros para valores grandes de n .

6.5.3 Tempos de Espera

O algoritmo a seguir utiliza variáveis aleatórias exponenciais para gerar uma amostra x da $X \sim B(p, n)$ aproveitando o resultado a seguir:

Proposição 6.5.1 *Sejam E_1, E_2, \dots i.i.d com distribuição $\mathcal{E}(1)$ e seja a variável aleatória X definida por*

$$X = \inf \left\{ k \in \mathbb{R}: \sum_{1 \leq i \leq k+1} \frac{E_i}{n-i+1} > -\ln(1-p) \right\}.$$

Então $X \sim B(n, p)$.

Dele decorre a validade do algoritmo a seguir:

1. Atribuir $q = -\ln(1-p)$, $x = 0$, $s = 0$.
2. Repetir enquanto $s > q$.
 - (a) Gerar z amostra da $Z \sim \mathcal{E}(1)$.
 - (b) Atribuir $s = s + \frac{z}{n-x}$, $x = x + 1$.
3. Atribuir $x = x - 1$.
4. Retornar x .

Deixamos ao leitor a tarefa de provar a Proposição acima, bem como todos os resultados enunciados sem prova desta Secção.

6.6 Exercícios

Nos exercícios onde pede-se realizar um programa de computador, pode-se utilizar qualquer linguagem que lhe seja familiar.

Exercício 6.1: Escreva o programa que implementa o Algoritmo N1 e gere uma amostra de tamanho 1000 de variável com distribuição $\mathcal{N}(0,1)$. Meça o tempo de CPU necessário para a geração da amostra. Submeta dita amostra a testes de bondade de ajuste (χ^2 , Kolmogorov-Smirnov, etc.) Desenhe os pares $(x_1, x_2), (x_3, x_4), \dots, (x_{999}, x_{1000})$ e verifique se aparece algum padrão que invalide a hipótese de independência das amostras.

Exercício 6.2: Idem ao Exercício 6.1 para o Algoritmo N2.

Exercício 6.3: Idem ao Exercício 6.1 para o Algoritmo N3.

Exercício 6.4: Idem ao Exercício 6.1 para o Algoritmo N4.

Exercício 6.5: Idem ao Exercício 6.1 para o Algoritmo JR I.

Exercício 6.6: Idem ao Exercício 6.1 para o Algoritmo JR II.

Exercício 6.7: Idem ao Exercício 6.1 para o Algoritmo de von Neumann.

Exercício 6.8: Idem ao Exercício 6.1 para o Algoritmo de Marsaglia-Bray.

Exercício 6.9: Construa uma tabela de dupla entrada; na primeira os oito algoritmos acima mais os dos exemplos 4.2.1 e 4.5.2 e o do exercício 4.9, na segunda os tempos de CPU e o número de linhas de programa fonte para cada um deles. Compare os resultados e tire as suas conclusões.

Exercício 6.10: Para o programa em linguagem C abaixo (onde `unif` é uma função que gera amostras independentes da $U \sim \mathcal{U}(0,1)$ com parâmetros `param_unif`) responda qual é a distribuição das amostras obtidas com sucessivas chamadas à função `saida(param_unif)`? Implemente-o com o seu gerador de uniformes e submeta seqüências obtidas com a função `saida(params_unif)` aos testes mencionados nos exercícios anteriores, tomando o cuidado de medir os tempos de CPU.

```
double saida( params_unif ){
    static char condicao = 'n';
    static float gset;
    double unif( params_unif ), rtemp, v1, v2, fac;
    if (condicao == 'n'){
        do{
            v1 = 2. * unif( params_unif ) - 1.;
            v2 = 2. * unif( params_unif ) - 1.;
            rtemp = v1 * v1 + v2 * v2;
        } while ( rtemp > 1 );
        condicao = 's';
        fac = sqrt( -2. * log( rtemp ) / rtemp );
        gset = v1 * fac;
        return( v2 * fac );
    }
    else{
        condicao='n';
        return( gset );
    }
}
```

}
}

Exercício 6.11: Compare os tempos de CPU necessários para gerar as amostras com o algoritmo acima e com as suas conclusões da tabela construída no Exercício 6.9.

Exercício 6.12: Prove os Teoremas 6.1.1, 6.1.2, 6.1.3 e 6.1.4.

Exercício 6.13: Prove que a validade dos Algoritmos do Método Polar Geral e dos N1, ..., N4 decorrem dos Teoremas acima.

Exercício 6.14: Prove as três afirmações do início da Seção 6.2.

Exercício 6.15: Prove a validade dos dois Algoritmos Simples da Seção 6.2.1.

Exercício 6.16: Prove que se $X \sim \text{Beta}(\alpha, 1 - \alpha)$, com $\alpha < 1$ e $U \sim \mathcal{U}(0, 1)$ então, para todo $\beta > 0$, $-\beta X \ln(U) \sim \mathcal{G}(\alpha, \beta)$ (provando assim a validade do Algoritmo de Johnk).

Exercício 6.17: Para resolver este exercício, o leitor deverá já ter implementado os geradores de amostras da $X \sim \mathcal{G}(\alpha, \beta)$ (e, idealmente, deveria ter escolhido os mais rápidos). Sejam os reais $\eta_1 = 2$ e $\eta_2 = 6$; para $\alpha = .5$, $\alpha = 1.3$ e $\alpha = 2$ faça: (a) Gerar a amostra u_1, \dots, u_{100} de ocorrências independentes da $U \sim \mathcal{U}(0, 20)$ e a amostra v_1, \dots, v_{100} de ocorrências independentes da $V \sim \mathcal{U}(0, 50)$. (b) Definir z_i amostra da $Z \sim \mathcal{G}(\alpha, \eta_1 u_i + \eta_2 v_i)$. (c) Ajustar um modelo de regressão linear para a variável dependente $z = (z_1, \dots, z_{100})$ e as variáveis independentes $u = (u_1, \dots, u_{100})$ e $v = (v_1, \dots, v_{100})$. (d) Detetar e explicar os problemas do ajuste acima. (e) Procurar transformações que estabilizem a variância dos erros e ajustar o novo modelo. (f) (Opcional) Ajuste o modelo adequado ao caso (b) usando o pacote GLIM. (g) Comente e compare os resultados.

Exercício 6.18: Para resolver este exercício, o leitor deverá já ter implementado os geradores de amostras da $X \sim \mathcal{P}(\lambda)$ (e, idealmente, deveria ter escolhido os mais rápidos). Sejam os reais $\eta_1 = 2$ e $\eta_2 = 6$, faça: (a) Gerar a amostra u_1, \dots, u_{100} de ocorrências independentes da $U \sim \mathcal{U}(0, 20)$ e a amostra v_1, \dots, v_{100} de ocorrências independentes da $V \sim \mathcal{U}(0, 50)$. (b) Definir z_i amostra da $Z \sim \mathcal{P}(\eta_1 u_i + \eta_2 v_i)$. (c) Ajustar um modelo de regressão linear para a variável dependente $z = (z_1, \dots, z_{100})$ e as variáveis independentes $u = (u_1, \dots, u_{100})$ e $v = (v_1, \dots, v_{100})$. (d) Detetar e explicar os problemas do ajuste acima. (e) Procurar transformações que estabilizem a variância dos erros e ajustar o novo modelo. (f) (Opcional) Ajuste o modelo adequado ao caso (b) usando o pacote GLIM. (g) Comente e compare os resultados.

Exercício 6.19: Sejam os reais $\eta_1 = 3\eta_2 = 5$; faça: (a) Gerar a amostra u_1, \dots, u_{100} de ocorrências independentes da $U \sim \mathcal{U}(0, 2)$ e a amostra v_1, \dots, v_{100} de ocorrências independentes da $V \sim \mathcal{U}(0, 7)$. (b) Definir z_i amostra da $Z \sim \mathcal{B}(\phi(\eta_1 u_i + \eta_2 v_i))$, onde $\phi(w) = \frac{e^w}{1+e^w}$. (c) Ajustar um modelo de regressão linear para a variável dependente $z = (z_1, \dots, z_{100})$ e as variáveis independentes $u = (u_1, \dots, u_{100})$ e $v = (v_1, \dots, v_{100})$. (d) Detetar e explicar os problemas do ajuste acima. (e) Procurar transformações que estabilizem a variância dos erros e ajustar o novo modelo. (f) (Opcional) Ajuste o modelo adequado ao caso (b) usando o pacote GLIM. (g) Comente e compare os resultados.

Capítulo 7

Tópicos e aplicações especiais de Simulação Estocástica

7.1 Uso da Simulação Estocástica em Inferência Estatística

Um dos usos mais evidentes da simulação estocástica em Estatística consiste na *aleatorização*, por exemplo em Planejamento de Experimentos e em Amostragem. Técnicas mais modernas que usam da simulação estocástica estão sendo sugeridas na medida do crescimento do potencial da computação. Entre essas últimas, podem destacar-se: os testes Monte Carlo, o *bootstrap* e os intervalos de confiança Monte Carlo.

7.1.1 Testes Monte Carlo

Consideremos uma experiência de teste de hipóteses onde H_0 é a hipótese nula, T é a estatística na qual baseia-se o teste, $0 < \alpha < 1$ é o nível de significância do teste e $C(\alpha)$ é o valor crítico (mais precisamente, diremos que o teste rejeita H_0 se $t \leq C(\alpha)$, onde t é o valor observado de T) e F é a distribuição de T sob H_0 . Então $\text{IPr}(T \leq C(\alpha)) = F((-\infty, C(\alpha)]) = \alpha$ e $F(C(\alpha)) = \alpha$. Chamaremos *teste convencional* a esta técnica.

Um dos problemas mais comuns nestes casos é o desconhecimento de $C(\alpha)$ e/ou de F . Para simplificar a discussão assumiremos que F é conhecida e contínua.

Para se ter um valor aproximado de $C(\alpha)$ poderíamos proceder assim: sejam m e R “grandes” (na Secção 3 deste Capítulo iremos precisar o que é “grande”).

1. Fazer $r = 1$.
2. Gerar uma amostra t_{r1}, \dots, t_{rm} de tamanho m da F .
3. Ordenar o vetor (t_{r1}, \dots, t_{rm}) obtendo $(t_{r(1)}, \dots, t_{r(m)})$ com $t_{r(1)} \leq \dots \leq t_{r(m)}$.
4. Definir $C_r = t_{r(k)}$ tal que $\frac{k}{m} \leq \alpha < \frac{k+1}{m}$.
5. Se $r = R$ seguir em 6. Se não $r := r + 1$ e voltar ao 2.

6. Definir como valor aproximado de $C(\alpha)$ o valor

$$\widehat{C(\alpha)} = \frac{1}{R} \sum_{1 \leq r \leq R} C_r.$$

Embora simples, este procedimento é muito ingênuo e não produz boas estimativas do $C(\alpha)$ na maioria dos casos, particularmente quando α é pequeno, exceto ao custo de um R e/ou m muito grandes. Voltaremos sobre isto mais adiante.

Na verdade, o procedimento descrito nada (ou quase nada) tem a ver com o nosso problema de teste. Simplesmente foi uma forma de estimar o valor de $C(\alpha)$ tal que

$$\int_{(-\infty, C(\alpha)]} dF = \alpha, \quad (7.1)$$

quer dizer, o α -ésimo quantil da F .

O procedimento a seguir baseia-se em idéias estreitamente relacionadas às que inspiraram o anterior, embora diferentes. Nesta nova técnica não estimamos o valor $C(\alpha)$; consitui-se, na verdade, num procedimento de teste da hipótese H_0 baseado em simulação. Por isto o procedimento recebeu o nome de *Teste Monte Carlo*.

Seja m "grande" e t o valor observado da estatística T .

1. Gerar uma amostra (t_1, \dots, t_m) da F .
2. Ordenar o vetor da etapa 1 obtendo $(t_{(1)}, \dots, t_{(m)})$ com $t_{(1)} \leq \dots \leq t_{(m)}$.
3. Seja $0 \leq k \leq m$ tal que

$$\frac{k}{m+1} \leq \alpha < \frac{k+1}{m+1}.$$

4. Se $t < t_{(k)}$ rejeitar H_0 .

Vejam agora a teoria que sustenta os Testes Monte Carlo e então daremos algumas dicas para as suas aplicações e referências bibliográficas complementares. Essencialmente, veremos alguns resultados que serão úteis para comparar as potências dos testes convencionais e Monte Carlo.

Lema 7.1.1: Seja $m \geq 2$ inteiro, (U_1, \dots, U_m) variável com distribuição comum $\mathcal{U}(0, 1)$, e sejam $(U_{1:m}, \dots, U_{m:m})$ as estatísticas de ordem associadas ao (U_1, \dots, U_m) . Poremos também $U_{(k)} = U_{k:m}$. Para cada $k = 1, 2, \dots, m$ a densidade da $U_{k:m}$ é dada por

$$\begin{aligned} f_{k:m}(\xi) &= \frac{m!}{(k-1)!(m-k)!} \xi^{k-1} (1-\xi)^{m-k} \Pi_{(0,1)}(\xi) \\ &= \frac{\Gamma(m+1)}{\Gamma(k)\Gamma(m+1-k)} \xi^{k-1} (1-\xi)^{m-k} \Pi_{(0,1)}(\xi), \end{aligned}$$

isto é $U_{k:m}$ tem a distribuição $\text{Beta}(k, m+1-k)$, e a função de distribuição acumulada é

$$F_{k:m}(u) = \begin{cases} 0 & \text{se } u \leq 0 \\ \sum_{k \leq j \leq m} \binom{m}{j} u^j (1-u)^{m-j} & \text{se } 0 < u < 1 \\ 1 & \text{se } u \geq 1 \end{cases}$$

A demonstração pode ver-se em algum textos básico de Probabilidade, por exemplo o de Hoel e outros [60].

Lembrando que se T tem distribuição F sob H_0 , então $U = F(T) \sim \mathcal{U}(0, 1)$, e é fácil ver que:

Nota 7.1.1: Observado t , o valor da estatística T , a probabilidade de rejeitar H_0 , usando o procedimento chamado Teste Monte Carlo, é dada por

$$p(u) = \sum_{0 \leq j \leq k-1} \binom{m}{j} u^j (1-u)^{m-j}, \text{ onde } u = F(t).$$

Em outras palavras, $p(u)$ é a proporção de vezes nas que o Teste Monte Carlo rejeitará H_0 quando $t = F^{-1}(u)$ foi o valor observado da estatística T .

Com efeito, sejam $(T_{(1)}, \dots, T_{(m)})$ as estatísticas de ordem de (T_1, \dots, T_m) , onde as variáveis T_1, \dots, T_m são i.i.d com distribuição F (a distribuição de T sob H_0). Então

$$\begin{aligned} \mathbb{P}\mathbb{R}(t < T_{(k)}) &= \mathbb{P}\mathbb{R}(F(t) < F(T_{(k)})) = \mathbb{P}\mathbb{R}(u < U_{(k)}) \\ &= 1 - \mathbb{P}\mathbb{R}(U_{(k)} \leq u) = 1 - F_{k:m}(u) = \sum_{0 \leq j \leq k-1} \binom{m}{j} u^j (1-u)^{m-j}. \end{aligned}$$

Vários valores do $p(u)$ acham-se tabelados em [96], onde o autor mostra que para k pequeno as distorções computacionais podem produzir resultados significativamente afastados dos verdadeiros.

A seguinte Proposição será de utilidade para se ter uma idéia da potência do Teste Monte Carlo.

Proposição 7.1.1 *Sejam F e G funções de distribuição acumuladas contínuas sobre \mathbb{R} ; (Ω, \mathcal{A}) um espaço mensurável; $m \geq 2$ inteiro; $U_j: \Omega \rightarrow [0, 1]$ mensurável para cada $j = 1, \dots, m$; $T: \Omega \rightarrow \mathbb{R}$ mensurável; $\mathbb{P}\mathbb{R}^{(0)}$ a única probabilidade sobre (Ω, \mathcal{A}) tal que*

- (U_1, \dots, U_m, T) são $\mathbb{P}\mathbb{R}^{(0)}$ -independentes,
- (U_1, \dots, U_m) são $\mathbb{P}\mathbb{R}^{(0)}$ -identicamente distribuídas com distribuição $\mathcal{U}(0, 1)$ e
- $\mathbb{P}\mathbb{R}_T^{(0)} = F$ (a distribuição da T sob $\mathbb{P}\mathbb{R}^{(0)}$);

$\mathbb{P}\mathbb{R}^{(1)}$ a única probabilidade sobre (Ω, \mathcal{A}) tal que

- (U_1, \dots, U_m, T) são $\mathbb{P}\mathbb{R}^{(1)}$ -independentes,
- (U_1, \dots, U_m) são $\mathbb{P}\mathbb{R}^{(1)}$ -identicamente distribuídas com distribuição $\mathcal{U}(0, 1)$ e
- $\mathbb{P}\mathbb{R}_T^{(1)} = G$.

Seja $1 \leq k \leq m$. $\mathcal{R} = \{\omega: T(\omega) < T_{(k)}(\omega)\}$, onde $(T_{(1)}, \dots, T_{(m)})$ é o vetor (T_1, \dots, T_m) ordenado de menor a maior e $T_i = F^{-1}(U_i)$ para $i = 1, \dots, m$. Então

$$1. \mathbb{P}\mathbb{R}^{(0)}(\mathcal{R}) = \frac{k}{m+1}.$$

2. $\mathbb{P}r^{(1)}(\mathcal{R}) = \int_0^1 \beta(\xi) f_{k:m}(\xi) d\xi$, onde

$$\beta(\xi) = \mathbb{P}r^{(1)}(T < F^-(\xi)) = G(T < F^-(\xi)) = G(F^-(\xi)), \quad (7.2)$$

isto é $\beta(\xi)$ é a função potência do teste convencional.

Demonstração de 1:

$$\begin{aligned} \mathbb{P}r^{(0)}(\mathcal{R}) &= \mathbb{P}r^{(0)}(T < T_{(k)}) = \int_{-\infty}^{+\infty} \mathbb{P}r^{(0)}(T_{(k)} > T | T = t) d\mathbb{P}r^{(0)}(t) \\ &= \int_{-\infty}^{+\infty} \mathbb{P}r^{(0)}(T_{(k)} > t) dF(t) \end{aligned} \quad (7.3)$$

Ora: $\mathbb{P}r^{(0)}(T_{(k)} > t) = \mathbb{P}r^{(0)}(U_{(k)} > F(t)) = \int_{F(t)}^1 f_{k:m}(\xi) d\xi$. Substituindo em 7.3 temos

$$\begin{aligned} \mathbb{P}r^{(0)}(\mathcal{R}) &= \int_{-\infty}^{+\infty} \int_{F(t)}^1 f_{k:m}(\xi) dF(t) \\ &= \int_{-\infty}^{+\infty} \int_0^1 g(\xi, t) f_{k:m}(\xi) d\xi dF(t) \end{aligned} \quad (7.4)$$

onde $g : (0, 1) \times \mathbb{R} \rightarrow \{0, 1\}$ é dada por

$$g(\xi, t) = \begin{cases} 1 & \text{se } F(t) \leq \xi < 1 \\ 0 & \text{caso contrário.} \end{cases}$$

Aplicando o Teorema de Fubini na equação 7.4 temos

$$\mathbb{P}r^{(0)}(\mathcal{R}) = \int_0^1 \left[\int_{-\infty}^{+\infty} g(\xi, t) dF(t) \right] f_{k:m}(\xi) d\xi \quad (7.5)$$

Usando na equação 7.5 o resultado a seguir

$$\int_{-\infty}^{+\infty} g(\xi, t) dF(t) = \int_{-\infty}^{F^-(\xi)} dF(t) = F(F^-(\xi)) = \xi$$

fica

$$\mathbb{P}r^{(0)}(\mathcal{R}) = \int_0^1 \xi f_{k:m}(\xi) d\xi = \frac{k}{m+1},$$

levando em consideração que o valor esperado de uma variável aleatória com distribuição $\text{Beta}(a, b)$ é $a/(a+b)$ e o Lema 7.1.1.

A prova da parte 2 é análoga à anterior estabelecendo que

$$\begin{aligned} \mathbb{P}r^{(1)}(\mathcal{R}) &= \int_0^1 \left[\int_{-\infty}^{+\infty} g(\xi, t) dG(t) \right] f_{k:m}(\xi) d\xi \\ &= \int_0^1 G(F^-(\xi)) f_{k:m}(\xi) d\xi. \end{aligned}$$

O resultado anterior sugere definir a função potência do Teste Monte Carlo descrito anteriormente, para todo $0 < \alpha < 1$, por

$$\beta^{(m)}(\alpha) = \int_0^1 \beta(\xi) b(\alpha, m, \xi) d\xi \quad (7.6)$$

onde $\xi \mapsto b(\alpha, m, \xi)$ é a função densidade da $\text{Beta}(\alpha(m+1), (1-\alpha)(m+1))$, isto é

$$b(\alpha, m, \xi) = \frac{\Gamma(m+1)}{\Gamma(\alpha(m+1))\Gamma((1-\alpha)(m+1))} \xi^{\alpha(m+1)-1} (1-\xi)^{(1-\alpha)(m+1)-1} \mathbb{I}_{(0,1)}(\xi).$$

Os resultados a seguir, cujas demonstrações podem ver-se nas referências [23,61,66,110], estabelecem relações úteis para comparações entre $\beta^{(m)}$ e β : as funções potência dos Testes Monte Carlo e convencional, respectivamente.

Proposição 7.1.2 *Sejam F e G duas funções de distribuição acumulativas contínuas sobre \mathbb{R} ; $m \geq 2$ inteiro; $\beta: (0,1) \rightarrow \mathbb{R}$ e $\beta^{(m)}: (0,1) \rightarrow \mathbb{R}$ definidas por*

$$\begin{aligned} \beta(\xi) &= G(F^{-}(\xi)), \quad \xi \in (0,1) \\ \beta^{(m)}(\xi) &= \int_0^1 \beta(\zeta) b(\xi, m, \zeta) d\zeta. \end{aligned}$$

Se β é côncava, então $\beta^{(m)}(\alpha) \uparrow \beta(\alpha)$ quando $m \rightarrow \infty$ para todo $0 < \alpha < 1$.

Proposição 7.1.3 *Na situação da Proposição anterior, se $\beta(0) = 0$, $\beta(1) = 1$ e β é côncava sobre $[0,1]$ então*

$$\frac{\beta^{(m)}(\alpha)}{\beta(\alpha)} \geq 1 - \frac{\mathbb{E}\{|Z - \alpha|\}}{2\alpha},$$

onde Z é uma variável aleatória com distribuição $\text{Beta}(\alpha(m+1), (1-\alpha)(m+1))$.

Nota: Levando em conta que a variável aleatória Z da Proposição anterior pode aproximar-se por uma variável aleatória com distribuição Normal de média $\alpha(1-\alpha)/m$, temos que na situação da Proposição 7.1.2 (com $\beta: [0,1] \rightarrow \mathbb{R}$ côncava):

$$\frac{\beta^{(m)}(\alpha)}{\beta(\alpha)} \geq 1 - \sqrt{\frac{1-\alpha}{2\pi m\alpha}},$$

para m “grande”.

Os resultados anteriores confirmam as idéias intuitivas sobre os Testes Monte Carlo. No que diz respeito à significância do Teste Monte Carlo, devem evitar-se valores pequenos de k e, como deve ser

$$\frac{k}{m+1} \leq \alpha < \frac{k+1}{m+1},$$

então m deve ser “grande”. Se não se for muito estrito no tocante ao nível de significância, a prática aconselha tomar $m = 99$.

7.1.2 Estimação do desvio padrão de um estimador usando "bootstrap"

Consideremos a seguinte situação: $k \geq 1$ inteiro; \mathbf{X} um vetor aleatório com distribuição F ; $\mathbf{X}_1, \dots, \mathbf{X}_n$ uma amostra de tamanho n da \mathbf{X} ; $t: (\mathbb{R}^k)^n \rightarrow \mathbb{R}^r$ ($r \geq 1$) uma função mensurável; $T = t(\mathbf{X}_1, \dots, \mathbf{X}_n)$. Do ponto de vista das aplicações, pensemos que T é um estimador de uma certa característica $\theta \in \mathbb{R}^r$ da distribuição de \mathbf{X} (veja exemplo mais adiante). O problema é obter-se um estimador do desvio padrão da T quando temos obtido observações $\mathbf{x}_1, \dots, \mathbf{x}_n$ das $\mathbf{X}_1, \dots, \mathbf{X}_n$.

O método que veremos brevemente a seguir baseia-se em simulações. É recomendável quando $\sigma(T)$ (o desvio padrão de T) não pode conhecer-se por métodos analíticos; por exemplo porque a distribuição exata de T não é conhecida.

As referências básicas sobre esta técnica chamada *bootstrap* são [40,41]. O algoritmo que se segue foi sugerido no trabalho de Efron e Gong referenciado acima. Para simplificar, suponhamos que $k = r = 1$. Sejam x_1, \dots, x_n os valores observados de X_1, \dots, X_n e seja $\xi \mapsto \hat{\mu}((x_1, \dots, x_n), \xi)$ a função de distribuição empírica de (X_1, \dots, X_n) , isto é

$$\begin{aligned}\hat{\mu}((x_1, \dots, x_n), \xi) &= \frac{1}{n} \# \{i: x_i \leq \xi\} \\ &= \frac{1}{n} \sum_{1 \leq i \leq n} \mathbb{I}_{(-\infty, \xi]}(x_i).\end{aligned}$$

Seja B um inteiro positivo.

1. Seja $b = 1$.
2. Gerar uma amostra $x_{b1}^*, \dots, x_{bn}^*$ de tamanho n de $\hat{\mu}((x_1, \dots, x_n), \cdot)$ (amostra de *bootstrap*). Isto é, fazer n estimações com reposição do conjunto $\{x_1, \dots, x_n\}$. Assim, x_{bj}^* será um dado valor dos x 's com probabilidade $1/n$.
3. Calcular $t_b^* = t(x_{b1}^*, \dots, x_{bn}^*)$, chamada *estimativa bootstrap* de θ associada à amostra (x_1, \dots, x_n) .
4. Se $b = B$ seguir em 5. Se não fazer $b = b + 1$ e voltar ao 2.
5. Definir $\bar{t}_B^* = B^{-1} \sum_{b=1}^B t_b^*$, chamado *valor médio* do estimador bootstrap baseado em B replicações. Finalmente, definir

$$S_B = \sqrt{\frac{1}{B-1} \sum_{1 \leq b \leq B} (t_b^* - \bar{t}_B^*)^2}.$$

Assim, S_B é por definição o *estimador bootstrap* do desvio padrão de T baseado na amostra (x_1, \dots, x_n) e B replicações.

A idéia intuitiva que sustenta esta técnica é que, se B é suficientemente grande, S_B estará próximo ao desvio padrão verdadeiro de $T^* = t(X_1^*, \dots, X_n^*)$, onde X_1^*, \dots, X_n^* são variáveis com distribuição $\hat{\mu}((x_1, \dots, x_n), \cdot)$; e se n também é "grande" então $\sigma(T^*) \approx \sigma(T)$.

Exemplo: O desvio padrão do coeficiente de correlação amostral. Neste exemplo consideraremos (X, Y) um vetor aleatório bi-dimensional com distribuição F contínua tal que esteja

definido

$$\begin{aligned}\rho(X, Y) &= \text{Correlação entre } X \text{ e } Y \\ &= \frac{\mathbb{E}\{(X - \mathbb{E}\{X\})(Y - \mathbb{E}\{Y\})\}}{\sqrt{\text{Var}X \cdot \text{Var}Y}},\end{aligned}$$

chamado o *coeficiente de correlação de Pearson entre X e Y*. Chama-se *coeficiente de Pearson amostral* à estatística t definida sobre um sub-conjunto conveniente de $(\mathbb{R}^2)^n$ por

$$t((x_1, y_1), \dots, (x_n, y_n)) = \frac{\sum_{1 \leq i \leq n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left\{ \sum_{1 \leq i \leq n} (x_i - \bar{x})^2 \right\} \left\{ \sum_{1 \leq i \leq n} (y_i - \bar{y})^2 \right\}}},$$

onde $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ e $\bar{y} = n^{-1} \sum_{i=1}^n y_i$.

Usualmente esta estatística é denotada por $\hat{\rho}$ (ou $\hat{\rho}_n$, se o tamanho da amostra for relevante).

No caso $F =$ normal bi-variada com média (μ_x, μ_y) e matriz de covariância

$$\begin{bmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix},$$

Efron e Gong [41] mostram que para n “grande”

$$\sigma(\hat{\rho}_n) \approx \frac{(1 - \rho^2)}{\sqrt{n - 3}},$$

pelo que um estimador razoável do desvio padrão do $\hat{\rho}_n$ nesta situação é

$$\hat{\sigma}(\hat{\rho}_n) \approx \frac{(1 - \hat{\rho}^2)}{\sqrt{n - 3}},$$

chamado *estimador do desvio padrão de $\hat{\rho}_n$ da Teoria Normal*.

Denotemos com $\hat{\sigma}_B(\hat{\rho}_n)((x_1, y_1), \dots, (x_n, y_n))$ o estimador bootstrap do desvio padrão de $\hat{\rho}_n$ baseado em B replicações e numa amostra $(x_1, y_1), \dots, (x_n, y_n)$ de (X, Y) . Em [88] descrevem-se brevemente os resultados da seguinte experiência Monte Carlo

1. Geraram 14.000 pontos $(x_1, y_1), \dots, (x_{14000}, y_{14000})$ com distribuição

$$\mathcal{N}_2 \left((0, 0), \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right) \quad \text{i. e. } \rho = 0.5.$$

2. Para cada $n = 8, 14, 20, 26, 32, 38, 44$ e 50:

(a) Para cada $r = 1, \dots, [14000/n]$ calcularam:

- i. $\hat{\rho}_n(r)$: o coeficiente de correlação amostral baseado na amostra (de tamanho n) $(x_{a+1}, y_{a+1}), \dots, (x_{a+n}, y_{a+n})$, onde $a = (r - 1)n$.

- ii. $\hat{\sigma}(\hat{\rho}_n)(r)$: o estimador do desvio padrão de $\hat{\rho}_n$ da Teoria Normal a amostra anterior, isto é $\hat{\sigma}(\hat{\rho}_n)(r) = (n-3)^{-1/2}(1 - (\hat{\rho}_n(r))^2)$.
- iii. Para $B = 128$ e $B = 152$ calcularam $\hat{\sigma}_B(\hat{\rho}_n)(r)$: o estimador bootstrap do desvio padrão de $\hat{\rho}_n$ baseado na amostra anterior e B repetições.

Entre as principais conclusões obtidas pelos autores de [88] destacamos duas:

1. O estimador $\hat{\sigma}(\hat{\rho}_n)$ resultou evidentemente viciado.
2. O $\hat{\sigma}(\hat{\rho}_n)$ foi melhor que $\hat{\sigma}_B(\hat{\rho}_n)$ no que diz respeito às variâncias Monte Carlo. Mas deve ter-se presente que isso foi devido, possivelmente, a que a distribuição F era uma Gaussiana; deve ser levado em conta que o estimador bootstrap é não paramétrico.

O leitor é convidado a fazer o Exercício 7.3 para se ter uma idéia das performances do $\hat{\sigma}(\hat{\rho}_n)$ e do $\hat{\sigma}_B(\hat{\rho}_n)$ em um caso não Gaussiano.

Estudos que destacam as vantagens e desvantagens destas técnicas, junto com sugestões para melhorá-las, têm aparecido em diversos trabalhos. Entre eles os de Efron e Tibshirani [42] e de Banks [15].

7.1.3 Intervalos de Confiança Monte Carlo

Consideremos a seguinte situação: $k \geq 1$ inteiro; \mathbf{X} um vetor aleatório k -dimensional com distribuição F ; $\mathbf{X}_1, \dots, \mathbf{X}_n$ uma amostra de tamanho n da \mathbf{X} ; $t: (\mathbb{R}^k)^n \rightarrow \mathbb{R}^r$ ($r \geq 1$) uma função mensurável; $T = t(\mathbf{X}_1, \dots, \mathbf{X}_n)$. Como na Secção 7.1.2 pensaremos que T é um estimador de uma certa característica $\theta \in \mathbb{R}^r$ da distribuição de \mathbf{X} . O problema que vamos agora considerar é o de obter um intervalo de confiança r -dimensional para θ baseados nas observações $\mathbf{x}_1, \dots, \mathbf{x}_n$ das $\mathbf{X}_1, \dots, \mathbf{X}_n$.

Situação 1: Suponhamos que $r = 1$; que existem F_θ , distribuição sobre \mathbb{R} ; $\Theta \subset \mathbb{R}$ intervalo aberto; $\{F_\nu: \nu \in \Theta\}$ família de distribuições sobre \mathbb{R} tal que

1. $\theta \in \Theta$.
2. Se Y é uma variável aleatória com distribuição F_ν , $\nu \in \Theta$, então $Y - \nu$ tem distribuição F_0 .
3. T tem distribuição F_θ ; logo (por 2)

$$T - \theta \sim F_0. \quad (7.7)$$

Seja $\hat{\theta} = t(\mathbf{x}_1, \dots, \mathbf{x}_n)$. Se T é estimador consistente de θ é razoável supor que $\hat{\theta} \in \Theta$. Então, se θ^* é uma variável aleatória com distribuição $F_{\hat{\theta}}$, e por 2 acima temos que $\theta^* - \hat{\theta} \sim F_0$. Daí, e pelo resultado 7.7 faz sentido usar a variabilidade da θ^* em torno de $\hat{\theta}$ para inferir a variabilidade da T em torno de θ .

O procedimento a seguir define um Intervalo de Confiança Monte Carlo com nível de confiança aproximado $1 - \alpha$, com $0 < \alpha < 1$. Sejam m "grande" e $\hat{\theta}$ o valor observado do estimador T :

1. Gerar uma amostra $(\theta_1^*, \dots, \theta_m^*)$ da $F_{\hat{\theta}}$.
2. Ordenar o vetor do passo 1 obtendo $\theta_{(1)}^*, \dots, \theta_{(m)}^*$ com $\theta_{(1)}^* \leq \dots \leq \theta_{(m)}^*$.
3. Sejam $1 \leq j < k \leq m$ tais que

$$\frac{j}{m+1} \leq \frac{1}{2}\alpha < \frac{j+1}{m+1} \leq \frac{k}{m+1} \leq 1 - \frac{1}{2}\alpha < \frac{k+1}{m+1}.$$

4. Definir $L = \theta_{(j)}^*$ e $U = \theta_{(k)}^*$.
5. Definir como *intervalo de confiança Monte Carlo* a

$$(2\hat{\theta} - U, 2\hat{\theta} - L). \quad (7.8)$$

Se F_0 for simétrica, o intervalo de confiança Monte Carlo pode ser definido como

$$(L, U). \quad (7.9)$$

Situação 2: Suponhamos que existem F_1 distribuição sobre $(0, +\infty)$; $\Theta \subset (0, +\infty)$ intervalo aberto; $\{F_\nu: \nu \in \Theta\}$ família de distribuições sobre $(0, +\infty)$ tais que

1. $\theta \in \Theta$.
2. Se Y é uma variável aleatória positiva com distribuição F_ν e $\nu \in \Theta$, então Y/ν tem distribuição F_1 .
3. T tem distribuição F_θ ; daí $T/\theta \sim F_1$.

Seja $\hat{\theta} = t(\mathbf{x}_1, \dots, \mathbf{x}_n)$. Se T é um estimador consistente de θ é razoável supor que $\hat{\theta} \in \Theta$. Então, se θ^* é uma variável aleatória com distribuição $F_{\hat{\theta}}$, por 2 acima temos que

$$\frac{\theta^*}{\hat{\theta}} \sim F_1.$$

isto é, T/θ e $\theta^*/\hat{\theta}$ têm a mesma distribuição F_1 .

Agora, o procedimento para definir um intervalo de confiança Monte Carlo para θ de nível aproximado de confiança $1 - \alpha$, com $0 < \alpha < 1$, é similar ao da Situação 1: definir as etapas 1, 2, 3 e 4 como anteriormente e

5. Definir como intervalo de confiança Monte Carlo a

$$\left(\frac{\hat{\theta}^2}{U}, \frac{\hat{\theta}^2}{L} \right). \quad (7.10)$$

Se $\ln(T/\theta)$ tem distribuição simétrica em torno de 0, novamente podemos substituir 7.10 por

$$(L, U) \quad (7.11)$$

O leitor interessado em saber mais sobre este tema pode consultar os trabalhos de Ripley [110], Efron [39, Capítulo 10] e Buckland [26].

Finalmente, para afixar algumas das idéias apresentadas nesta Secção, é aconselhável fazer o Exercício 7.4 no final deste Capítulo.

7.2 Otimização combinatória usando relaxação estocástica (simulated annealing)

O Simulated Annealing é uma forma de achar (em forma aproximada, em princípio) mínimos de funções tipicamente muito complicadas usando técnicas de amostragem estocástica. Desde a sua aparição, por volta dos anos '85 [75], vem sendo uma área de ativa pesquisa em Simulação Estocástica (é claramente uma experiência Monte Carlo), e continua a dar bons resultados práticos.

As suas principais qualidades são a simplicidade conceitual, a sua aplicabilidade a uma classe grande de problemas e a sua boa performance quando comparada com técnicas determinísticas. Também permite ver a conexão que existe entre processos estocásticos, redes neuronais, máquinas de Boltzman e Otimização Combinatória.

Dado que é um assunto com fronteiras ainda difusas e com bastantes problemas em aberto, faremos apenas uma apresentação das idéias centrais, colocadas dentro do contexto deste trabalho e apelando menos para o formalismo que para a intuição do leitor. Excelentes introduções ao assunto, menos breves que a aqui exposta, podem ver-se nos artigos de Barbosa [16] e de Aarts e Laarhoven [8].

O nome provém de uma analogia com o processo físico de *recozido* de materiais, isto é se queremos obter um cristal de alto teor de pureza, um método bom será derreter o material e esfriá-lo **lentamente** para que, aos poucos, as estruturas indesejáveis desapareçam. Sob certas condições, é possível fazer o análogo com funções, onde o *cristal* desejado é um valor de mínimo.

Sem perda de generalidade, definiremos um problema combinatório como o problema de achar os argumentos para o qual uma função atinge o seu valor mínimo, e quando o domínio da função é "complicado".

Formalizando, seja a função $f: \Xi \rightarrow \mathbb{R}^+$ (que chamaremos de custo), com domínio tal que $\#\Xi < +\infty$ (típicamente $\#\Xi$ é muito grande, por exemplo $\#\Xi = (n-1)!$ com $n \approx 100$: o problema do Caixeiro viajante [7]; ou $\#\Xi = 2^n$ com $n \approx 1024$: o problema da reconstrução de uma imagem binária [28]). Seja ainda o conjunto $\Xi_0 = \{\xi \in \Xi: f(\xi) \leq f(\eta) \forall \eta \in \Xi, \eta \neq \xi\}$. Então o problema consiste em achar (em geral, pelo menos um elemento de) Ξ_0 .

Uma noção importante será a de *vizinhança* (em algum sentido, que dependerá de cada problema (f, Ξ)) de um argumento. Diremos que o conjunto $\partial_\xi \subset \Xi$ é a vizinhança do argumento $\xi \in \Xi$ se (i) $\#\partial_\xi \ll \#\Xi \forall \xi \in \Xi$, (ii) $\eta \in \partial_\xi \iff \xi \in \partial_\eta \forall (\xi, \eta) \in \Xi \times \Xi$ e (iii) se o grafo com vértices Ξ e arestas induzidas por $\{\partial_\xi: \xi \in \Xi\}$ tem uma única componente conexa.

Para fixar idéias, vejamos um exemplo clássico:

O Caixeiro Viajante Sejam $(x_1, y_1), \dots, (x_n, y_n)$ as coordenadas geográficas de n cidades. Procura-se a seqüência de n cidades tal que (a) todas elas estejam na seqüência (b) a distância total do caminho seja mínima. Este é um dos problemas mais clássicos da área de Pesquisa Operacional; de fato, é um dos que caracteriza uma classe importantíssima de problemas: os *Problemas NP-completos* [47]. Dado que o número de soluções possíveis é $\#\Xi = (n-1)!$, é inviável fazer uma procura direta de Ξ_0 no conjunto Ξ . A literatura sobre algoritmos que, de uma ou outra forma, procuram soluções aproximadas para valores de n "grandes" é enorme. Algumas técnicas já são tão clássicas como o mesmo problema, e citaremos uma que

é particularmente adequada aos nossos propósitos: a Busca Local.

Para tanto, seja um elemento qualquer $\xi_0 \in \Xi$ e o custo inicial $f(\xi_0)$, seja $i = 0$ e $f_i^* = f(\xi_0)$. O algoritmo de Busca Local, definido o sistema de vizinhanças $\{\partial_\xi: \xi \in \Xi\}$, funciona da seguinte maneira:

1. Avaliar $f(\eta)$ para todo $\eta \in \partial_{\xi_i}$.
2. Atribuir $\xi_{i+1} = \min_{\eta \in \partial_{\xi_i}}^{-1} f(\eta)$ e $f_{i+1}^* = \min_{\eta \in \partial_{\xi_i}} f(\eta)$.
3. Se $f_{i+1}^* < f_i^*$ então voltar ao passo 1 com $i = i + 1$. Caso contrário parar e retornar como solução o par (ξ_{i+1}, f_{i+1}^*) .

Evidentemente, o algoritmo parará no primeiro mínimo local que achar o que, para problemas de Otimização Combinatória em geral, pode ser muito ruim: a solução ótima pode estar bastante afastada dele. Uma forma de melhorar esta técnica é a de rodá-la para várias configurações iniciais ξ_0 , e ficar com a melhor solução obtida. Outra forma de impedir que o algoritmo pare no primeiro mínimo local, é dotar ao processo de busca de uma dinâmica estocástica: toda vez que ele achar um mínimo local ele será aceito como nova configuração, já se todas as opções forem de maior custo que a configuração atual haverá uma probabilidade positiva de aceitar alguma delas. Desta forma, aceitando de vez em vez alguma configuração ruim, o algoritmo poderá escapar dos mínimos locais e, eventualmente, atingir ou o mínimo global ou alguma configuração perto dele. Espera-se que quanto mais tempo o algoritmo andar à procura do mínimo global, mais perto dele estará e, para que o passeio aleatório não o afaste dele, a probabilidade de aceitar configurações ruins deve diminuir: para isto existe o parâmetro T , a “temperatura” do sistema.

Utilizando a notação acima, uma versão do algoritmo de **Simulated Annealing** fica definida como: sejam a configuração inicial $\xi_0 \in \Xi$, a temperatura inicial $T_0 < +\infty$, o custo inicial $f(\xi_0)$ e o sistema de vizinhanças $\{\partial_\xi: \xi \in \Xi\}$ para o problema, seja ainda $i = 0$. Enquanto alguma condição não seja satisfeita faça:

1. Escolha um $\eta \in \partial_{\xi_i}$ segundo alguma regra predeterminada.
2. Observar u amostra da $U \sim \mathcal{U}(0, 1)$.
3. Atribuir $\xi_{i+1} = \eta$ se $u \leq \min\{1, \exp\{T_i^{-1}(f(\xi_i) - f(\eta))\}\}$, se não atribuir $\xi_{i+1} = \xi_i$.
4. Fazer $i = i + 1$.

Para o algoritmo funcionar, a seqüência de temperaturas $\{T_i\}$ deve ser decrescente, e segundo uma lei logarítmica; o valor inicial T_0 depende do problema em questão. Prova-se que $\xi_i \rightarrow \tilde{\xi} \in \Xi_0$ para todo ξ_0 quando $i \rightarrow \infty$. No livro de Press e outros [108, Cap. 10] pode ver-se uma outra versão deste método para o problema do Caixeiro Viajante.

Existem vários estudos apresentando outras implementações, e o leitor interessado neste assunto pode consultar os livros de Aarts e Korst [7] ou de Aarts e Laarhoven [8], e as referências neles. Já para aplicações ao processamento de imagens, os trabalhos seminais são os de Geman and Geman [49] e de Carnevalli e outros [28]

7.3 Integração Estocástica ou Monte Carlo

Na verdade, o objetivo de quase todo trabalho de simulação acaba sendo a estimação de uma ou mais esperanças da forma $\mathbb{E}\{\phi(\mathbf{X})\}$, onde \mathbf{X} é um vetor aleatório $k \geq 1$ -dimensional e ϕ é uma função a valores em \mathbb{R}^r , com $r \geq 1$. Claro está que isto reduz-se a calcular uma integral que, freqüentemente, não tem uma forma analítica fechada, e que os métodos numéricos para a solução inexistem ou são pouco confiáveis devido a problemas de instabilidade. Não é surpreendente, então, que a simulação estocástica tenha sido usada para obter estimativas dessas integrais.

Nesta Secção veremos uma técnica que usa simulação com a finalidade de estimar integrais definidas. É bem simples e tem sido alvo de experimentação para as chamadas *técnicas de redução da variância* que veremos no próximo Capítulo.

Sejam $k \geq 1$ inteiro; f uma densidade de probabilidade sobre \mathbb{R}^k ; ϕ uma função mensurável definida sobre \mathbb{R}^k . Consideremos o problema de calcular

$$\theta = \int \phi(x) f(x) dx. \quad (7.12)$$

Usualmente, se $k = 1$ calcula-se θ analiticamente ou por métodos de integração numérica. Se $k = 2$ pode usar-se integração iterada, mas isto será bem difícil se o conjunto $\{x: f(x) > 0\} \cap \{x: \phi(x) \neq 0\}$ tem uma forma complicada. Já para $k > 2$ os problemas que se apresentam quando tenta-se aplicar integração numérica fazem aconselhável aplicar o procedimento que veremos a seguir, chamado *Integração Monte Carlo*. Este método é aplicável, em princípio, qualquer que seja k , a dimensão do problema. Seja n "grande".

1. Gerar uma amostra $\mathbf{x}_1, \dots, \mathbf{x}_n$ da f .
2. Estimar θ por

$$\hat{\theta} = \frac{1}{n} \sum_{1 \leq i \leq n} \phi(\mathbf{x}_i).$$

Em outras palavras, usar como aproximação do θ um valor da variável aleatória

$$T = \frac{1}{n} \sum_{1 \leq i \leq n} \phi(\mathbf{X}_i),$$

onde $\mathbf{X}_1, \dots, \mathbf{X}_n$ são i.i.d com densidade comum f .

Agora, qual é a precisão deste método? Como sabemos, neste caso

$$\text{Var}(T) = \frac{1}{n} \int (\phi(x) - \theta)^2 f(x) dx = \frac{c}{n} \quad (7.13)$$

(por definição $c = \int (\phi(x) - \theta)^2 f(x) dx$). Assim, a precisão do $\hat{\theta}$ é proporcional a $1/\sqrt{n}$.

Isto é bem pior do que a precisão de métodos numéricos que, usando n pontos no domínio da integração, chegam a obter precisões da ordem de n^{-4} ou ainda melhores. Mas já dissemos que o método de integração Monte Carlo deve ser usado quando não se tem mais nada para fazer o cálculo de θ . Por outra parte, usando simulação convenientemente, podemos tentar reduzir o valor de c na equação 7.13, embora não possamos alterar a forma desta equação.

Exemplo 7.3.1: Diversas formas de calcular $\theta = \text{IPr}(C > 2)$, onde $C \sim \text{Cauchy padrão}$, via integração Monte Carlo.

Neste caso, tomando f como a densidade da Cauchy padrão —isto é $f(x) = (\pi(1+x^2))^{-1}$ — queremos calcular

$$\theta = \int \phi(x)f(x) dx = \int \mathbb{I}_{(2,+\infty)}(x) \frac{1}{\pi(1+x^2)} dx. \quad (7.14)$$

Evidentemente θ pode calcular-se analiticamente, mas o intuito deste Exemplo é ilustrar a performance do método aqui sugerido, bem como várias formas para melhorá-la.

Cálculo Analítico:

$$\begin{aligned} \theta &= \int_2^{+\infty} \frac{1}{\pi(1+x^2)} dx = \frac{1}{\pi} \arctan(x) \Big|_2^{+\infty} = \frac{1}{\pi} \left(\frac{\pi}{2} - \arctan(2) \right) \\ &= \frac{1}{2} - \frac{\arctan(2)}{\pi} \approx 0.1476 \end{aligned}$$

Método 1 (Monte Carlo à força bruta): Seja $\phi_1(x) = \mathbb{I}_{(2,+\infty)}(x)$. Gerar x_1, \dots, x_n com densidade Cauchy padrão e estimar θ por

$$T_1 = \frac{1}{n} \sum_{1 \leq i \leq n} \phi_1(x_i),$$

onde $\phi_1(x) = \mathbb{I}_{(2,+\infty)}(x)$.

Ora, se X_1, \dots, X_n são variáveis com distribuição Cauchy padrão é fácil ver que

$$nT_1 = n \left(\frac{1}{n} \sum_{1 \leq i \leq n} \phi_1(X_i) \right)$$

tem distribuição Binomial(n, θ), daí que

$$\text{Var}(T_1) = \frac{\theta(1-\theta)}{n} \approx \frac{0.126}{n}.$$

Método 2: Observemos que também vale que

$$\theta = \frac{1}{2} \text{IPr}(|C| > 2).$$

Podemos então definir como estimador de θ :

$$T_2 = \frac{1}{n} \sum_{1 \leq i \leq n} \phi_2(X_i),$$

onde $\phi_2(x) = \frac{1}{2} \mathbb{I}_{(2,+\infty)}(|x|)$, e X_1, \dots, X_n são variáveis com distribuição Cauchy padrão.

Neste caso $2nT_2$ tem distribuição $\text{Bi}(n, 2\theta)$ e portanto

$$\text{Var}(T_2) = \frac{2\theta(1-2\theta)}{n} \approx \frac{0.052}{n}.$$

Método 3: É fácil ver que

$$1 - 2\theta = \int_{-2}^2 f(x) dx = 2 \int_0^2 f(x) dx = 4 \int f(x)g(x) dx,$$

onde $g(x) = \frac{1}{2} \mathbb{I}_{(0,2)}(x)$ é a densidade da $\mathcal{U}(0,2)$. Logo,

$$\theta = \frac{1}{2} - \int 2f(x)g(x) dx.$$

Podemos definir então como estimador de θ :

$$T_3 = \frac{1}{2} - \frac{1}{n} \sum_{1 \leq i \leq n} \phi_3(X_i),$$

onde $\phi_3(x) = 2f(x)$ e X_1, \dots, X_n são variáveis com distribuição $\mathcal{U}(0,2)$.

Pode provar-se que

$$\text{Var}(T_3) \approx \frac{0.028}{n}.$$

Método 4: Usando a transformação $y = x^{-1}$ temos

$$\theta = \int_2^{+\infty} \frac{dx}{\pi(1+x^2)} = \int_0^{\frac{1}{2}} \frac{y^{-2} dy}{\pi(1+y^{-2})} = \int_0^{\frac{1}{2}} \frac{dy}{\pi(1+y^2)} = \int \frac{1}{2\pi(1+y^2)} h(y) dy,$$

onde $h(y) = 2\mathbb{I}_{(0,1/2)}(y)$ é a densidade da $\mathcal{U}(0,1/2)$.

Daí que possamos definir como estimador de θ :

$$T_4 = \frac{1}{n} \sum_{1 \leq i \leq n} \phi_4(X_i),$$

onde $\phi_4(x) = \frac{1}{2}f(x) = (2\pi(1+x^2))^{-1}$ e X_1, \dots, X_n são variáveis com distribuição $\mathcal{U}(0,1/2)$.

É possível ver que $\text{Var}(T_4) \approx 9.3 \cdot 10^{-5}/n$, o que reduz a variância dos estimador de θ em 1350 vezes respeito do estimador *força bruta*.

No próximo Capítulo veremos uma técnica de redução da variância aplicável em várias situações. Ela dará uma precisão ainda superior à obtida com T_4 .

É claro que o exemplo acima é artificial, porém ilustra o fato de que antes de fazer uma integração Monte Carlo é bom pensar um pouquinho.

7.4 Elementos de Simulação Estocástica Multivariada

O uso de técnicas de simulação no estudo de fenômenos nos quais aparecem vetores aleatórios é muito mais importante do que nos daqueles onde aparecem variáveis aleatórias uni-dimensionais. Isto deve-se à grande dificuldade de se ter técnicas analíticas para o tratamento de vetores aleatórios. Apesar disso, nestas notas só trataremos brevemente o tema de simulação de vetores aleatórios, indicando alguns métodos para a geração de certos vetores típicos (v. g. normais e multinomiais) apenas para o caso bi-dimensional. Também o leitor poderá ver

na lista de referências alguma bibliografia complementar. Sem dúvida alguma, este assunto ainda está muito aberto às pesquisas, tanto de caráter teórico como aplicado.

Começaremos com um exemplo interessante de um vetor aleatório (X, Y) tal que a distribuição conjunta não é absolutamente contínua, sendo que as distribuições marginais o são.

Exemplo 7.4.1: O vetor bivariado exponencial de Marshall-Olkin. Sejam E_1 , E_2 e E_3 variáveis com distribuição $\mathcal{E}(\lambda^{-1})$, isto é com densidade $x \mapsto \lambda \exp\{-\lambda x\} \mathbb{I}_{(0,+\infty)}(x)$; $X = \min(E_1, E_3)$, $Y = \min(E_2, E_3)$. Então

1. X e Y têm distribuição $\mathcal{E}((2\lambda)^{-1})$.
2. $\mathbb{P}(X = Y) > 0$.

Provemos 1: $\mathbb{P}(X > x) = \mathbb{P}(E_1 > x, E_3 > x) = \mathbb{P}(E_1 > x)\mathbb{P}(E_3 > x) = \exp\{-2\lambda x\}$. Logo $X \sim \mathcal{E}((2\lambda)^{-1})$; da mesma forma obtém-se a distribuição de Y .

Para ver 2 note-se que $\mathbb{P}(X = Y) = \mathbb{P}(E_1 \geq E_3, E_2 \geq E_3) = \lambda \int_0^{+\infty} \mathbb{P}(E_1 \geq x, E_2 \geq x) \exp\{-\lambda x\} dx$. Daí $\mathbb{P}(X = Y) = \lambda \int_0^{+\infty} \exp\{-3\lambda x\} dx = \frac{1}{3}$.

O leitor pode ver a que geração de uma amostra de (X, Y) é bem simples (veja o Exercício 7.6).

Este modelo tem uma interpretação física bastante intuitiva: suponhamos um mecanismo constituído por duas componentes com tempos de falha independentes e idênticos (E_1, E_2) , mas que podem falhar simultaneamente devido a um certo “shock” que ocorre aleatoriamente em instantes de tempos dados pela variável E_3 . Uma generalização deste modelo pode ver-se na referência [81].

Sem dúvida, a distribuição bi-variada mais conhecida e usada é a distribuição Gaussiana (ou Normal) bi-variada. Vejamos então dois métodos de geração de amostras dessa distribuição.

7.4.1 Geração de Normal Bi-variada

Lembremos que o vetor aleatório (X, Y) é dito normal bi-variado com parâmetros μ_X , σ_X , μ_Y , σ_Y e ρ se a sua distribuição é dada pela densidade (sobre \mathbb{R}^2)

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{\frac{1}{2}Q^{-1}(x-\mu_X, y-\mu_Y)\right\}, \quad (7.15)$$

onde

$$Q^{-1}(u,v) = \frac{1}{1-\rho^2} \left(\frac{u^2}{\sigma_X^2} - \frac{2\rho uv}{\sigma_X\sigma_Y} + \frac{v^2}{\sigma_Y^2} \right).$$

O leitor interessado pode ver em Johnson e Kotz [69] uma discussão completa sobre esta distribuição, junto com tabelas da função de distribuição acumulada para $\mu_X = \mu_Y = 0$ e $\sigma_X = \sigma_Y = 1$.

O resultado a seguir é ao nível de uma prova corriqueira nos cursos de Estatística, e resulta muito útil para a geração de vetores normais bi-variados.

Proposição 7.4.1 *Seja (X, Y) um vetor aleatório com a densidade dada na equação 7.15. Então*

1. $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ e $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.
2. A distribuição de X condicional a $Y = y$ é $\mathcal{N}(\mathbb{E}\{X|y\}, \text{Var}(X|y))$, com $\mathbb{E}\{X|y\} = \mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y)$ e $\text{Var}(X|y) = \sigma_X^2(1 - \rho^2)$.

O resultado a seguir é a chave para um dos métodos de geração de amostras normais bi-variadas.

Proposição 7.4.2 *Sejam X_1, X_2 variáveis com distribuição $\mathcal{N}(0, 1)$; a, b, c e d números reais; sejam X e Y dados por*

$$\begin{cases} X &= aX_1 + bX_2 \\ Y &= cX_1 + dX_2, \end{cases}$$

então (X, Y) é normal bi-variada com $\mu_X = \mu_Y = 0$, $\sigma_X^2 = a^2 + b^2$, $\sigma_Y^2 = c^2 + d^2$ e $\rho = \text{Correlação}(X, Y) = \frac{ac+bd}{\sqrt{(a^2+b^2)(c^2+d^2)}}$.

Reciprocamente, seja (X, Y) normal bi-variado com $\mu_X = \mu_Y = 0$ e definindo $a = \sigma_X\sqrt{1-\rho^2}$, $b = \rho\sigma_X$, $c = 0$, $d = \sigma_Y$:

$$\begin{cases} X' &= aX_1 + bX_2 \\ Y' &= cX_1 + dX_2, \end{cases}$$

onde X_1 e X_2 são como na parte 1, então (X', Y') tem a mesma distribuição que (X, Y) .

Agora usaremos esses resultados para gerar uma amostra $(x_1, y_1), \dots, (x_n, y_n)$ de tamanho n do (X, Y) com distribuição normal bi-variada de parâmetros $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$ e $\rho \neq 0$ qualquer número no intervalo $(-1, 1)$.

Método 1: (ou condicional)

1. $i = 1$.
2. Gerar y_i da $\mathcal{N}(0, 1)$.
3. Gerar x_i da $\mathcal{N}(\rho y_i, 1 - \rho^2)$.
4. Se $i < n$ fazer $i = i + 1$ e voltar a 2. Se não, terminar.

Método 2: a partir de independentes

1. $i = 1$.
2. Gerar z_1 e z_2 , amostra de tamanho 2 da $\mathcal{N}(0, 1)$.
3. Definir $x_i = z_1\sqrt{1-\rho^2} + \rho z_2$ e $y_i = z_2$.
4. Se $i < n$ fazer $i = i + 1$ e voltar a 2. Se não, terminar.

7.4.2 Geração de Bernoulli bi-variada e de Tabelas de Contingência 2×2

Diz-se que um vetor aleatório (X, Y) tem distribuição Bernoulli bi-variada com parâmetros p_{00} , p_{01} , p_{10} e p_{11} (não negativos e tais que $p_{00} + p_{01} + p_{10} + p_{11} = 1$) se

$$\begin{aligned}\text{IPr}((X, Y) = (0, 0)) &= p_{00} \\ \text{IPr}((X, Y) = (0, 1)) &= p_{01} \\ \text{IPr}((X, Y) = (1, 0)) &= p_{10} \\ \text{IPr}((X, Y) = (1, 1)) &= p_{11}\end{aligned}\tag{7.16}$$

Ora, se $(x_1, y_1), \dots, (x_n, y_n)$ é uma amostra de tamanho n de um tal (X, Y) e definimos

$$\begin{aligned}n_{00} &= \#\{i: (x_i, y_i) = (0, 0)\} \\ n_{01} &= \#\{i: (x_i, y_i) = (0, 1)\} \\ n_{10} &= \#\{i: (x_i, y_i) = (1, 0)\} \text{ e} \\ n_{11} &= \#\{i: (x_i, y_i) = (1, 1)\},\end{aligned}$$

é costume mostrar os resultados na forma de uma tabela com o seguinte aspecto

		Y		
		1	0	
X	1	n_{11}	n_{10}	n_{1+}
	0	n_{01}	n_{00}	n_{0+}
		n_{+1}	n_{+0}	

chamada *tabela de contingência 2×2* , onde

$$\begin{aligned}n_{0+} &= n_{00} + n_{01} & n_{+0} &= n_{00} + n_{10} \\ n_{1+} &= n_{10} + n_{11} & n_{+1} &= n_{01} + n_{11},\end{aligned}$$

verificando-se, obviamente, que $n_{00} + n_{01} + n_{10} + n_{11} = n$.

Assim, o problema de gerar uma amostra de tamanho n de um vetor aleatório com distribuição de Bernoulli bi-variada é equivalente ao de gerar uma tabela de contingência 2×2 . Consideraremos agora dois métodos de geração de uma tabela de contingência 2×2 .

Método 1: marginais não pré-fixadas

1. $i = 1$
2. Gerar u da $\mathcal{U}(0, 1)$.
 3. (a) Se $u \leq p_{11}$ definir $(x_i, y_i) = (1, 1)$.
 - (b) Se $p_{11} < u \leq (p_{10} + p_{11})$ definir $(x_i, y_i) = (1, 0)$.
 - (c) Se $(p_{10} + p_{11}) < u \leq (p_{01} + p_{10} + p_{11})$ definir $(x_i, y_i) = (0, 1)$.
 - (d) Caso contrário definir $(x_i, y_i) = (0, 0)$.
4. Se $i < n$ fazer $i = i + 1$ e seguir em 2. Se não, terminar.

Método 2: marginais pré-fixadas. Existem situações experimentais onde os valores de n_{1+} e de n_{0+} (ou de n_{+1} e n_{+0}) estão pré-fixados. Por exemplo, podemos querer estudar a incidência do hábito de fumar no câncer de pulmão comparando amostras de $n_{+1} = 500$ pacientes afetados desse câncer e $n_{+0} = 500$ indivíduos de controle.

Nesta situação os valores já dados para a geração são p_{00} , p_{01} , p_{10} , p_{11} , n_{+0} e n_{+1} ; isto é devemos gerar os valores n_{00} , n_{01} , n_{10} e n_{11} . Para isso

1. Gerar z_1 , z_2 das variáveis independentes Z_1 e Z_2 com $Z_1 \sim Bi(n_{+1}, \frac{p_{11}}{p_{11}+p_{01}})$ e $Z_2 \sim Bi(n_{+0}, \frac{p_{10}}{p_{10}+p_{00}})$.
2. Definir $n_{11} = z_1$, $n_{01} = n_{+1} - n_{11}$, $n_{10} = z_2$ e $n_{00} = n_{+0} - n_{10}$.

Procedendo desta forma vemos que $\text{IPr}(X = 1|Y = 1) = \frac{p_{11}}{p_{11}+p_{01}}$ e $\text{IPr}(X = 1|Y = 0) = \frac{p_{10}}{p_{10}+p_{00}}$; e daí que considerando n_{00} = pares iguais à $(0, 0)$, n_{01} = pares iguais à $(0, 1)$, n_{10} = pares iguais à $(1, 0)$ e n_{11} = pares iguais à $(1, 1)$, teremos obtido uma amostra de tamanho $n = n_{+0} + n_{+1}$ do vetor (X, Y) com distribuição dada pela equação 7.16.

Para finalizar esta Secção indicamos ao leitor interessado em saber mais em torno da geração de vetores aleatórios dar uma olhada em De Vroye [36] e Johnson [68].

7.5 Simulação de processos físicos

Por “simulação de processos físicos” entendemos o uso da Simulação Estocástica para gerar ocorrências de processos estocásticos que, em algum sentido, são bons modelos de *coisas que acontecem na natureza*. Isto é, deixaremos um pouco de lado a discussão de como gerar ocorrências de variáveis aleatórias, para considerar problemas práticos onde elas estão envolvidas.

Consideraremos alguns poucos exemplos, para ilustrar o uso das técnicas vistas até agora, e sem o intuito de esgotar o assunto. Os exemplos 1 e 2 do Capítulo 1 ilustram os casos onde é necessário responder perguntas que envolvem filas e moléculas, respectivamente.

Salientemos que os exemplos aqui considerados foram escolhidos em forma mais ou menos arbitrária, dependendo apenas das preferências dos autores, e que seremos muito breves na exposição dos mesmos. O leitor interessado em aprofundar-se nestes assuntos poderá consultar a bibliografia indicada.

7.5.1 Um Modelo para o Ferromagnetismo

Este exemplo poderia muito bem ter sido incluído na Secção 7.4, pois estaremos interessados em gerar uma amostra de um vetor aleatório (de dimensão bastante grande), mais este caso aparece aqui por ser um caso bastante particular: é o célebre **Modelo de Ising**.

Este modelo foi proposto por Ising na década dos '20 como uma forma de explicar, com um modelo estocástico, o aparecimento da magnetização espontânea nos materiais ferromagnéticos. Define-se, para o caso finito, da seguinte forma: seja $X: \Omega \rightarrow \{-1, +1\}^S$, com $S = \{1, \dots, N\} \times \{1, \dots, N\}$ um vetor aleatório com distribuição dada pela expressão

$$\text{IPr}(X = x) = Z_\beta^{-1} \exp \left\{ 2\beta \sum_{\|s-t\|=1} x_s x_t \right\}, \quad \text{onde } x = (x_s)_{s \in S}$$

e $Z_\beta = \sum_{x \in \Xi} \exp\{2\beta \sum_{\|s-t\|=1} x_s x_t\}$ é a função partição

O problema consiste em simular x_1, \dots, x_n ocorrências de X acima, e medir uma função qualquer $\phi(x_i)$. Vários algoritmos foram propostos para isto, notadamente os mais usados os de Metrópolis e de Geman e Geman [48, págs. 117–193] e, mais recentemente, apareceram os algoritmos de Swendsen-Wang [100,101,119] e de Wolff [130].

Estes últimos baseiam-se numa analogia que existe entre a distribuição de X acima e um problema de percolação. Podemos resumí-los da seguinte forma: seja $x(0) = [x_s(0)]_{s \in S}$ um elemento qualquer de $\{-1, +1\}^S$ (a configuração inicial). Ambos os algoritmos atuam iterativamente sobre vetores sucessivos $x(k) = [x_s(k)]_{s \in S}$ de tal forma que há convergência $\Pr(X(k) = x) \rightarrow \Pr(X = x)$ quando $k \rightarrow \infty$ para todo $x(0)$. Para $\beta \geq 0$ podemos descrever cada iteração como: seja $k = 0$ e troque $x(k)$ por $x(k+1)$ fazendo

1. Coloque um arco, com probabilidade $1 - \exp\{-\beta\}$, entre cada par de sítios de $x(k)$ que tem o mesmo valor e estão a distância 1 em S .
2. Construa \mathcal{C} , o conjunto de todas as componentes conexas do grafo obtido no passo anterior.
3. Forme $\mathcal{C}' \subset \mathcal{C}$, o conjunto das componentes conexas que trocarão de valor.
4. Troque $x_s(k+1) = (-1)^{\mathbb{1}_{\mathcal{C}'(s)}} x_s(k)$ para todo $s \in S$.
5. Faça $k = k+1$ e retorne ao passo 1 se alguma condição não for satisfeita.

Os algoritmos diferem no passo 3: como formar \mathcal{C}' , o conjunto das componentes conexas que trocarão de valor. As especificações são:

Suponha que na iteração k tem-se $L(k)$ componentes conexas no grafo induzido pelo passo 1 acima, i. e. $\mathcal{C} = \{\lambda_1, \dots, \lambda_{L(k)}\}$; agora observe $u_1, \dots, u_{L(k)}$, uma seqüência de $L(k)$ amostras independentes da $U \sim \mathcal{U}(0,1)$. Swendsen-Wang forma \mathcal{C}' como

$$\mathcal{C}'_{\text{SW}} = \{\lambda_j \in \mathcal{C} : u_j < 1/2, 1 \leq j \leq L(k)\}.$$

Para o outro algoritmo observe $u_1, \dots, u_{\#S}$, uma seqüência de $\#S$ amostras independentes da $U \sim \mathcal{U}(0,1)$. Forme então \mathcal{C}' como

$$\mathcal{C}'_{\text{Wolff}} = \{\lambda_j \in \mathcal{C} : x_m(k) \in \lambda_j \text{ tal que } u_m = \sup_{1 \leq i \leq \#S} \{u_i\}\}.$$

Procedendo com estes algoritmos, podemos simular as ocorrências desejadas e obter em forma aproximada $\mathbb{E}\{\phi(X)\}$, bem como comparar estas dinâmicas com as de Metropolis e de Geman e Geman [45].

O leitor interessado em aplicações de Simulação Estocástica à Física, pode ver os textos de Binder [22,21], de Hammersley e Handscomb [58] e o de Spanier e Gelbard [116], entre outros.

7.5.2 Barcos em Tempos de Guerra

O exemplo a seguir, apresentado em [88], ilustra muito bem o tipo de perguntas que podemos responder a respeito de um processo complexo (mais precisamente, um problema de geometria estocástica), usando Simulação Estocástica.

Suponhamos que uma frota de barcos pretende atravessar um canal protegido com minas. Deseja-se simular (em forma bastante simples, para começar) um processo que descreva isto sob as hipóteses a seguir:

1. o canal pode ser considerado unidimensional de largura 10000 e extremos em -5000 e $+5000$; isto é as minas são colocadas ao longo desse segmento, e explodem toda vez que um barco as ativa;
2. o número de minas tem uma distribuição de Poisson com parâmetro $\lambda = 100$;
3. as minas são colocadas segundo duas estratégias possíveis:
 - (a) igualmente espaçadas no intervalo $(-5000, +5000)$, ou
 - (b) em pontos escolhidos independentemente segundo uma distribuição uniforme nesse intervalo;
4. Se x é a distância entre um dado barco e uma dada mina, então a probabilidade da mina explodir é dada pela função $A(x) = p \exp\{-\pi x/a^2\}$, com $p \in [0, 1]$ e $a > 0$.
5. Se x é a distância entre um dado barco e uma dada mina, então a probabilidade do barco ser danificado pela explosão é dada pela função $D(x) = q \exp\{-\pi x/d^2\}$, com $q \in [0, 1]$ e $d > 0$. Um barco ou está danificado (inútil) ou não o está.
6. A frota está composta por 6 barcos, que atravessarão o canal sucessivamente em pontos denotados y_1, \dots, y_6 segundo as estratégias possíveis:
 - (a) o primeiro barco escolhe y_1 segundo uma distribuição $\mathcal{U}(-5000, +5000)$ e os outros o seguem, isto é $y_2 = \dots = y_6 = y_1$.
 - (b) y_1, \dots, y_6 são amostras independentes da $U \sim \mathcal{U}(-5000, +5000)$.
 - (c) o primeiro barco escolhe y_1 segundo uma distribuição $\mathcal{N}(0, \sigma^2)$ truncada e os outros o seguem, isto é $y_2 = \dots = y_6 = y_1$.
 - (d) y_1, \dots, y_6 são amostras independentes da $Y \sim \mathcal{N}(0, \sigma^2)$ truncada.

Problemas:

1. O leitor é o inimigo da frota, e quer maximizar o número esperado de barcos danificados para todas as estratégias de passagem, escolhendo uma das estratégias de localização de minas.
2. O leitor é o almirante da frota, e quer achar a estratégia de passagem que minimize o número de barcos danificados para todas as estratégias de localização de minas.

O leitor já deve ser capaz de montar o experimento Monte Carlo que responda estas perguntas...

7.5.3 Projéteis

Prosseguindo com os problemas bélicos, suponhamos o caso de quisermos estimar o *intervalo* onde, com alta probabilidade, cairá um projétil cuja trajetória obedece as seguintes hipóteses (simplificação bidimensional do problema):

1. ele é lançado com uma velocidade inicial $v_0 \varepsilon_v$ e com uma elevação de $\theta_0 + \varepsilon_e$, onde $\varepsilon_v \sim \mathcal{G}(\alpha, 1)$ é independente de $\varepsilon_e \sim \text{von Mises}(\kappa)$;
2. ele atravessará duas capas diferentes de ar, ao ingressar na primeira (que inicia-se à metade da altura atingível) sofrerá uma única desviação d_1 na direção vertical com $D_1 \sim \mathcal{U}(0, 5)$; ao sair da segunda (que acaba no instante em que começa a queda) sofrerá uma única desviação d_2 na direção vertical, com $D_2 \sim \mathcal{U}(0, 10)$.
3. com probabilidade $p \in [0, 1]$ ele atingirá algum objeto voador (pássaro, anti-míssil) no instante de estar a $1/4$ da altura total que o desviará na vertical uma distância d_3 , com $D_3 \sim \text{Cauchy padrão}$.

Problema: calcular o intervalo de confiança (Monte Carlo) de 95% para o ponto atingido pelo projétil no solo para valores conhecidos (estimados) de α , κ e p . Como no caso anterior, o leitor já deve ser capaz de montar o experimento Monte Carlo que responda esta e outras perguntas pertinentes.

7.5.4 Localização de Árvores

Voltemo-nos a assuntos mais inofensivos. Uma das áreas mais ativas da Estatística é a chamada *Estatística Espacial*, que estuda a inferência sobre processos estocásticos com contradomínio em dimensões ≥ 2 . O assunto apresenta complicações notáveis, principalmente desde o ponto de vista teórico, dado que nem sempre é fácil definir os tais processos estocásticos.

Exemplos onde é necessária uma modelagem de processos espaciais são os estudos de proliferação de células, de queda de meteoritos, de ocorrência de doenças em plantas e, o que consideraremos nestas notas, de localização de árvores. Um bom livro para o interessado nestes assuntos é o de Diggle [37].

O problema pode ser colocado nos seguintes termos: suponhamos estarmos interessados na forma em que as árvores distribuem-se numa determinada região, e que tenhamos registrado como pontos no plano a localização de, digamos, 100 árvores numa região de $1000 \times 1000 m^2$. As perguntas são: a localização de uma árvore influe na localização das outras? isto é, as árvores têm tendência a agrupar-se, a afastar-se ou são indiferentes?

A resposta não é simples de ser respondida. Precisaremos comparar os dados colhidos com simulações de processos teoricamente tratáveis. Em particular, será de interesse o *Processo Generalizado de Poisson* que podemos resumir em forma simplificada dizendo que, dado um número N de ocorrências de eventos numa região retangular $R = (R_h \times R_v) \subset \mathbb{R}^2$, *indiferência de localização* é equivalente a ter as coordenadas das ocorrências distribuídas uniformemente. Isto é, se as N coordenadas observadas são $(x_1, y_1), \dots, (x_N, y_N)$, deveremos testar se (x_1, \dots, x_N) e (y_1, \dots, y_N) são, respectivamente, amostras independentes das $X \sim \mathcal{U}(R_h)$ e $Y \sim \mathcal{U}(R_v)$.

Para fazer dito teste, lança-se mão de estatísticas mais ou menos complexas, que por sua vez dependem dos valores observados de um número relativamente grande de simulações do Processo Generalizado de Poisson.

7.5.5 Modelos de Crescimento

A modelagem do crescimento (de plantas, cristais, linhas de fratura, etc.) provou ser uma fonte muito rica de modelos determinísticos e estocásticos. Os objetos sob estudo não podem ser descritos facilmente com modelos simples ou clássicos: convivem neles uma certa *regularidade* com uma enorme *diversidade*. Se o leitor já teve contato com algum trabalho de divulgação sobre fractais, poderá estar imaginando que o nosso discurso está indo nessa direção.

Nesta Secção descreveremos dois algoritmos para a modelagem de crescimento. Resultados da aplicação de algoritmos deste tipo podem ver-se em [118], e esses resultados podem compararse com os estudos de crescimento em deposição eletroquímica apresentados em [113].

7.5.5.1 Crescimento Balístico

Sejam L e M inteiros positivos (grandes). Seja $\mathbf{p} = (p_1, \dots, p_L)$ um vetor de probabilidades e I uma variável aleatória tal que $\text{IPr}(I = i) = p_i$. Seja $X: \{1, \dots, L\} \times \{1, \dots, M\}$ dada por $x(i, j) = 0$ para todo $1 \leq i \leq L$ e $1 \leq j \leq M$. O algoritmo a seguir consiste em modificar o valor de x em um certo número N (aleatório) de pontos $(i, j) \in \{1, \dots, L\} \times \{1, \dots, M\}$. Para simplificar a notação poremos $x(0, j) = 0$ para todo $0 \leq j \leq M$, $x(i, 0) = 1$ para todo $1 \leq i \leq L$ e $x(L+1, j) = 0$ para todo $0 \leq j \leq M$. Depois

1. Gerar uma amostra i de I .
2. Calcular $m(i) = \max\{0 \leq m \leq M: x(i-1, m) + x(i, m) + x(i+1, m) \neq 0\}$.
3. Se $m(i) < M$ fazer $x(i, m(i)+1) = 1$ e voltar ao passo 1; se não, parar.

De posse dos valores de x podemos desenhar em um retângulo de tamanho $L \times M$ o resultado de rodar o algoritmo, colocando “preto” nas coordenadas (i, j) tais que $x(i, j) = 0$ e colocando “branco” nas (i, j) tais que $x(i, j) = 1$.

7.5.5.2 Crescimento por Agregação

Seja agora $\Lambda = \{1, \dots, L\} \times \{1, \dots, M\}$, com L e M grandes; B denotará a borda de Λ , isto é $B = \{(1, j), (L, j), (i, 1), (i, M)\}$ para todo $1 \leq i \leq L$ e todo $1 \leq j \leq M$, e suporemos dado um vetor de probabilidade \mathbf{p}_B que associa um valor a cada ponto de B . Diremos que dois pontos $i = (i_1, i_2)$ e $j = (j_1, j_2)$ de Λ estão a distância $\rho = \sqrt{(i_1 - j_1)^2 + (i_2 - j_2)^2}$. Para todo conjunto não vazio $A \subset \Lambda$ definiremos a sua vizinhança como $\partial_A = \{(i, j) \in \Lambda: \exists a \in A \text{ tal que } \rho((i, j), a) = 1\}$. Usaremos a variável aleatória “orientação” (do passeio aleatório) $O: \Omega \rightarrow \{(-1, 0), (1, 0), (0, 1), (0, -1)\}$ e a soma de coordenadas definida como $a \oplus b = (a_1 + b_1 \bmod L, a_2 + b_2 \bmod M)$ para todo par de pontos a, b em Λ . O algoritmo é: seja $A = \emptyset$

1. Sejam a_1 uma amostra da distribuição de $U \sim \mathcal{U}\{1, \dots, L\}$ e a_2 uma amostra da distribuição de $U \sim \mathcal{U}\{1, \dots, M\}$, independentes.
2. Atribuir $A = A \cup \{(a_1, a_2)\}$.
3. Enquanto $\partial_A \cap B \neq \emptyset$ fazer

- (a) Gerar b amostra de p_B .
- (b) Gerar o amostra de $O \sim (1/4, \dots, 1/4)$.
- (c) Atribuir $b = b \oplus o$.
- (d) Se $b \notin \partial_A$ então volver ao passo (b). Se $b \in \partial_A$ atribuir $A = A \cup \{b\}$.

Desta maneira, uma vez satisfeita a condição $\partial_A \cap B \neq \emptyset$, podemos desenhar o objeto obtido atribuindo a cor “preta” às posições $a \in A$, e a cor “branca” àquelas $a \notin A$ em um retângulo de tamanho $L \times M$.

7.6 Exercícios

Nos exercícios onde pede-se realizar um programa de computador, pode-se utilizar qualquer linguagem que lhe seja familiar.

Exercício 7.1: Tabular $p(u)$ (da Nota 7.1.1) para $u = 0.001, 0.005, 0.01, 0.05$ e 0.1 e para os valores do para (k, m) para os testes Monte Carlo mais comuns: $(k = 1, m = 19)$, $(5, 99)$, $(1, 99)$, $(10, 199)$ e $(2, 199)$.

Exercício 7.2: Seja Φ a função de distribuição acumulada da $\mathcal{N}(0, 1)$; para cada $\theta > 0$ seja G_θ a função de distribuição acumulada da $\mathcal{N}(\theta, 1)$. Também

$$\begin{aligned}\beta_\theta(\xi) &= G_\theta(\Phi^{-1}(\xi)), \quad \xi \in (0, 1) \\ \beta_\theta^{(m)}(\alpha) &= \int_0^1 \beta_\theta(\xi) b(\alpha, m, \xi), \text{ para } \alpha \in (0, 1);\end{aligned}$$

com

$$b(\alpha, m, \xi) = \frac{\Gamma(m+1)}{\Gamma(\alpha(m+1))\Gamma((1-\alpha)(m+1))} \xi^{\alpha(m+1)-1} (1-\xi)^{(1-\alpha)(m+1)-1} \mathbb{I}_{(0,1)}(\xi).$$

Para $\alpha = 0.05$ e 0.01 , e para $m = 99$:

1. calcular $\beta_\theta^{(m)}(\alpha)$ por métodos numéricos ou pelos sugeridos na Secção 7.3.
2. Fazer desenhos simultâneos de

$$\theta \mapsto \beta_\theta(\alpha) \text{ e } \theta \mapsto \beta_\theta^{(m)}(\alpha).$$

3. Fazer desenhos simultâneos de

$$\theta \mapsto \frac{\beta_\theta^{(m)}(\alpha)}{\beta_\theta(\alpha)} \text{ e } \theta \mapsto 1 - \left[\frac{1-\alpha}{2\pi m\alpha} \right]^{\frac{1}{2}}.$$

Exercício 7.3: Fazer a experiência Monte Carlo do Lewis e Orav [88] descrita na Secção 7.1.2, somente para $n = 14$ e $B = 128$ com

1. F como nessa experiência (isto é, F a distribuição da $\mathcal{N}_2 \left((0, 0), \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right)$).

2. F a distribuição de (X, Y) onde $X = E_1 + E_2$ e $Y = E_2 + E_3$, sendo E_1, E_2 e E_3 variáveis com distribuição $\mathcal{E}(1)$. É simples provar que neste caso $\rho(X, Y) = \frac{1}{2}$.
3. Guardar num arquivo os valores $\hat{\rho}_n(r)$, $\hat{\sigma}(\hat{\rho}_n)(r)$ e $\hat{\sigma}_B(\hat{\rho}_n)(r)$, $r = 1, \dots, 1000$ para continuar com o Problema no Capítulo 8.

Exercício 7.4: Seja x_1, \dots, x_n uma amostra de tamanho $n \geq 2$ da população com densidade $\mathcal{E}(\mu^{-1})$, $\mu > 0$ (isto é, as variáveis X_1, \dots, X_n são variáveis com densidade comum $g_\mu(x) = \mu \exp\{-\mu x\} \mathbb{I}_{(0, +\infty)}(x)$). Sejam $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ e $T = \frac{n-1}{n\bar{X}}$. Pode ver-se que T é o estimador não viciado de mínima variância para μ .

Este Exercício, e outros relacionados que virão no Capítulo 8, visam realizar um estudo por simulação da performance do método “Intervalo de Confiança Monte Carlo” para μ baseado em T .

1. Provar que \bar{X} tem distribuição $\text{Gamma}(n, (n\mu)^{-1})$; isto é, \bar{X} tem uma densidade dada por

$$\gamma(x; n, (n\mu)^{-1}) = \frac{(n\mu)^n}{(n-1)!} x^{n-1} \exp\{-n\mu x\} \mathbb{I}_{(0, +\infty)}(x).$$

2. Provar que T tem densidade dada por

$$f_\mu(x) = \frac{(n-1)^{n-1}}{(n-2)!} \mu^n x^{-(n+1)} \exp\left\{-(n-1)\mu \frac{1}{x}\right\} \mathbb{I}_{(0, +\infty)}(x).$$

Agora aplicaremos o visto na Situação 2 da Secção 7.1.3: seja F_1 a distribuição sobre $(0, +\infty)$ dada pela densidade

$$f_1(x) = \frac{(n-1)^{n-1}}{(n-2)!} x^{-(n+1)} \exp\left\{-(n-1)\mu \frac{1}{x}\right\} \mathbb{I}_{(0, +\infty)}(x).$$

Para cada $\nu > 0$ seja F_ν a distribuição sobre $(0, +\infty)$ dada pela densidade

$$f_\nu(x) = \frac{(n-1)^{n-1}}{(n-2)!} \nu^n x^{-(n+1)} \exp\left\{-(n-1)\nu \frac{1}{x}\right\} \mathbb{I}_{(0, +\infty)}(x).$$

3. Provar que se Y é uma variável aleatória com distribuição F_ν então Y/ν tem distribuição F_1 .

Pelos itens 2 e 3 temos que $\frac{T}{\mu}$ tem distribuição F_1 . Realizar agora a seguinte experiência Monte Carlo:

1. Gerar 30.000 pontos x_1, \dots, x_{30000} com distribuição $\mathcal{E}(1/2)$.
2. Para $n = 10$ e 100 , e para cada $r = 1, \dots, 30000/n$
 - (a) Definir $\hat{\mu}(r) = \frac{n-1}{n\bar{X}_r}$, onde $\bar{X}_r = n^{-1} \sum_{i=1}^n X_{a+i}$, $a = (r-1)n$. Seja $m = 1000$.
 - (b) Gerar uma amostra $(\theta_1^*, \dots, \theta_m^*)$ de $F_{\hat{\mu}(r)}$.

- (c) Ordenar $(\theta_1^*, \dots, \theta_m^*)$ obtendo $(\theta_{(1)}^*, \dots, \theta_{(m)}^*)$ com $\theta_{(1)}^* \leq \dots \leq \theta_{(m)}^*$.
- (d) Sejam $1 \leq j < k \leq m$ tais que $\frac{j}{m+1} \leq \frac{\alpha}{2} < \frac{j+1}{m+1} \leq \frac{k}{m+1} \leq 1 - \frac{\alpha}{2} < \frac{k+1}{m+1}$ com $\alpha = 0.05$.
- (e) Definir $L(r) = \theta_{(j)}^*$ e $U(r) = \theta_{(k)}^*$.
- (f) Definir o intervalo de confiança Monte Carlo $\left(\frac{\hat{\mu}(r)^2}{U(r)}, \frac{\hat{\mu}(r)^2}{L(r)} \right)$.
3. Para $n = 10$ e 100 guardar em arquivos diferentes os valores $\hat{\mu}(r)$, $L(r)$, $U(r)$, $\frac{\hat{\mu}(r)^2}{U(r)}$ e $\frac{\hat{\mu}(r)^2}{L(r)}$ para $r = 1, \dots, 30000/n$.

Ajuda para 2b: pode ver-se facilmente que se Y é uma variável aleatória com distribuição F_ν então

$$\frac{2\nu(n-1)}{Y} \sim \chi_{2n}^2;$$

daí que para obter $\theta_1^*, \dots, \theta_m^*$ basta obter z_1, \dots, z_m com distribuição χ_{2n}^2 e definir

$$\theta_i^* = \frac{2\hat{\mu}(r)(n-1)}{z_i}, \quad i = 1, \dots, m.$$

Exercício 7.5: Completar o Exemplo 7.3.1 provando as afirmações que no texto ficaram sem demonstração.

Exercício 7.6: Continuação do Exemplo 7.4.1. Faça um programa para gerar amostras de tamanho n : $(x_1, y_1), \dots, (x_n, y_n)$ do vetor aleatório (X, Y) e compare, para vários n 's, os valores de $n^{-1} \# \{i: x_i = y_i\}$ com $1/3$, valor teórico de $\text{IPr}(X = Y)$.

Exercício 7.7: Uma generalização do Exemplo 7.4.1. Sejam $\lambda_1 > 0$, $\lambda_2 > 0$ e $\lambda_3 > 0$; E_1 , E_2 e E_3 independentes com distribuições $\mathcal{E}(\lambda_1^{-1})$, $\mathcal{E}(\lambda_2^{-1})$ e $\mathcal{E}(\lambda_3^{-1})$ respectivamente. Defina $X = \min(E_1, E_2)$.

1. Encontrar a correlação entre X e Y .
2. Estudar a função de distribuição do vetor (X, Y) : ela é absolutamente contínua (isto é existe uma densidade de probabilidade sobre \mathbb{R}^2 para (X, Y))? Se não o for, descrever as descontinuidades ou concentrações.

Exercício 7.8: Fazer dois programas usando qualquer gerador de normais univariadas, implementando os métodos de geração de vetores aleatórios normais bi-variados vistos na Secção 7.4.1.

Exercício 7.9: Seja (X, Y) com distribuição Bernoulli bi-variada com parâmetros p_{00} , p_{01} , p_{10} e p_{11} . Provar que X e Y são independentes se e somente se $p_{11} = (p_{10} + p_{11})(p_{01} + p_{11})$.

Exercício 7.10: Fazer dois programas implementando os métodos de geração de tabelas de contingência 2×2 expostos na Secção 7.4.2.

Exercício 7.11: Para o Exemplo 2 do Capítulo 1 considere os seguintes valores (adimensionais)

- intensidade do processo: $\lambda = 500$;
- comprimentos dos elementos base: $c_1 = 1$, $c_2 = 2$, $c_3 = 2.5$, $c_4 = 3$ e $c_5 = 4$;
- matriz de preferências: $p_{k\ell} = 1/4$ se $k \neq \ell$, 0 caso contrário.

Gere uma amostra de tamanho 100 de moléculas aceitáveis, desenhe algumas delas e responda as seguintes perguntas:

1. Quais são a média e o desvio padrão do número de componentes das moléculas aceitáveis? A rejeição de algumas moléculas afeta os valores teóricos esperados para uma amostra n_1, \dots, n_{100} da $\mathcal{P}(500)$? Por que?
2. Qual a taxa de rejeição de moléculas, isto é quantas moléculas foram destruídas pelo processo de formação até conseguir as 100 desejadas?
3. Quais são a média e o desvio padrão do raio das moléculas aceitáveis? Faça um histograma destes valores e ajuste algumas distribuições (*Weibull*, etc.) procurando a que melhor explica os dados.

Exercício 7.12: Responda às perguntas dos exemplos apresentados na Secção 7.5.

Capítulo 8

Análise dos Resultados obtidos em uma Simulação por Computador

8.1 Introdução

Existem muitas formas de fazer uma síntese do comportamento de uma amostra de dados uni- ou multidimensionais, sejam eles reais ou obtidos em uma simulação. Cada uma dessas técnicas serve para salientar certas estruturas nos dados: médias, variâncias, quantis, etc. e poder construir testes de hipóteses sobre a população da variável ou vetor aleatório de interesse.

Freqüentemente, nos estudos de simulação, teremos “muitas” amostras para serem analisadas. Sendo que cada uma delas pode ter sido obtida por simulações similares do sistema sob estudo, mudando apenas os valores de alguns parâmetros. O que se deseja é saber, por exemplo, quais são os valores desses parâmetros que fazem o sistema simulado mais adequado para representar um certo sistema real ou, talvez, como o sistema se comporta quando os valores desses parâmetros são mudados.

Para obter resultados relevantes, um estudo de simulação deve ser planejado com cuidado. Devemos ter sempre presente que um tal estudo é uma experiência com números, mas parecido com as experiências que poderiam ser feitas com outros tipos de unidades experimentais (parcelas de terrenos, animais de laboratório, indivíduos de uma população, etc.) Desse ponto de vista, em geral as *experiências Monte Carlo* ou de *simulação estocástica* têm a vantagem de que as componentes sob estudo (amostras obtidas por simulação) podem ser totalmente controladas pelo experimentador, o que não acontece habitualmente com as experiências das chamadas **Ciências Aplicadas**.

Assim, o planejamento de um estudo de simulação passa a ser um problema mais dos estudados em Planejamento de Experimentos. Muitas das idéias aí usadas podem servir muito bem à Simulação Estocástica (veja, por exemplo, o livro de Box e outros [24] para uma boa introdução ao assunto). Devemos identificar as hipóteses críticas no nosso modelo de simulação e produzir todas as *saídas* que sejam necessárias para isolar os efeitos de cada fator, além de escolher as combinações dos fatores mais relevantes.

Em geral, um modelo de simulação pode ser dividido em duas componentes. Uma delas

está formada pelos parâmetros e estruturas de relação das diversas variáveis aleatórias (digamos os **inputs**). A outra é a resposta ou saída (**output**) da simulação. Diversas são as questões a serem levadas em conta na hora de planejar a experiência. Por exemplo

1. Quais são as combinações de níveis dos fatores que devem ser analisadas em cada rodada da simulação? Certas técnicas que podem ajudar na identificação dessas combinações podem verse em [102].
2. Como organizar e mostrar as saídas?
3. Como deve ser construído o modelo a ser simulado para estudar a influência dos diversos fatores nas respostas?
4. Como usar o controle, disponível em toda simulação, sobre os inputs para fazer os outputs mais precisos?

As três primeiras questões são de caráter geral, aplicáveis a qualquer problema de planejamento de experiências ou análise de dados. Uma discussão detalhada sobre isto, aplicada especificamente à simulação, aparece no trabalho de Kleijnen [77], e mais concretamente ainda ao caso de comparação de estimadores em [56]. A questão 4 está diretamente relacionada à simulação. Quase tudo o que veremos neste Capítulo tem a ver com diferentes respostas a essa pergunta. Na Seção final deste Capítulo, indicaremos sugestões sobre a questão 2.

No que diz respeito especificamente à Simulação Estocástica como metodologia de estudo em Estatística, dois são os fatores mais relevantes: o tamanho das amostras e as distribuições das quais serão geradas as *observações*. A maioria dos trabalhos que usam Simulação Estocástica (em Estatística) são de comparação de diversas técnicas sob várias distribuições de dados. Dentre esses, estão incluídos os referentes à determinação (ou aproximação) da distribuição exata de certas estatísticas.

Exemplos de situações nas quais podemos querer comparar as performances de diferentes estatísticas são

- a) A estatística *t*-Student com outras alternativas para obter procedimentos de teste e de intervalos de confiança, confiáveis mesmo quando as distribuições dos dados não são normais.
- b) Estimação do centro de simetria de uma distribuição simétrica usando diversos estimadores: média amostral, mediana amostral, “médias aparadas”, estimadores robustos, etc.
- c) Estimadores robustos versus mínimos quadrados em regressão.
- d) Estimação paramétrica versus não paramétrica.

Para avaliar a performance poderiam ser respondidas questões tais como

- q1) Como varia o vício de um estimador respeito à variação do tamanho das amostras?
- q2) Como varia a variância de um estimador respeito à variação do tamanho das amostras?

- q3) Como são as distribuições assintóticas (se existirem) dos estimadores sob comparação?
- q4) Pode-se saber alguma coisa a respeito da distribuição exata dos estimadores? Por exemplo: estimar a média, a variância, os quantis; ou, melhor ainda, existem resultados analíticos sobre o assunto?
- q5) Se os estimadores são assintoticamente normais, qual é a velocidade de convergência respeito ao tamanho amostral?

Um esquema geral de um estudo de simulação para tentar responder essas questões poderia ser:

- p1) Gerar o modelo sob simulação, digamos n vezes.
- p2) Calcular (e guardar) os valores das diversas estatísticas sob comparação para a amostra de tamanho n gerada em p1.
- p3) Repetir (*replicação*) p1) e p2) um certo número de vezes, digamos m .
- p4) Analisar as distribuições empíricas das estatísticas baseadas nas “amostras” obtidas na etapa p3).

Para completar o estudo, eventualmente deverá repetir-se ou modificar-se o esquema acima para:

- f1) Vários valores de n ; os que dão informação sobre “pequenas amostras” e sobre o grau de aproximação dos resultados assintóticos.
- f2) Diferentes distribuições nos dados de entrada, incluindo
- várias famílias de distribuições;
 - diferentes valores dos parâmetros em cada uma dessas famílias, e
 - possível dependência entre observações.

Como já foi dito, a ênfase neste Capítulo será sobre como fazer os outputs de um trabalho de simulação o mais preciso possível.

Um primeiro passo no sentido de aumentar a precisão de um certo estimador, poderia ser procurar técnicas eficientes para o cálculo das quantidades envolvidas; ou aumentar o tamanho (m) das amostras sem aumentar o custo computacional, usando métodos de geração ou de implementação mais rápidos. Por exemplo, se os programas de computação foram escritos numa linguagem de alto nível como FORTRAN, tentar escrevê-los **sem muito esforço** em uma linguagem de menor nível, por exemplo em C ou ainda em Assembler. Isto pode fazer as gerações e/ou as contas mais rápidas. Porém, podem aparecer problemas de programação formidáveis e, com eles, o risco de erros. Como em tudo, é preciso bom senso...

As técnicas para acrescentar a precisão dos resultados de uma simulação tem recebido o nome de *técnicas de redução da variância*. Esse nome pode justificar-se levando em consideração o que habitualmente se faz nos estudos Monte Carlo:

Situação habitual: seja $F: \mathbb{R} \rightarrow [0, 1]$ uma função de distribuição acumulada; X uma variável aleatória com distribuição dada pela F acima e seja $g: \mathbb{R}^n \rightarrow \mathbb{R}$ uma função mensurável. **Problema:** estimar $\theta = \mathbb{E}_F\{g(X)\}$. Estimador Monte Carlo à força bruta:

1. Escolher o número de replicações mais ou menos arbitrariamente grande (digamos $M = 500, 1000$ ou mais).
2. Para cada $i = 1, 2, \dots, M$ gerar uma amostra $x_{m,1}, \dots, x_{m,n}$ da X (com distribuição F) e calcular

$$g_n(m) = g_n(x_{m,1}, \dots, x_{m,n}). \quad (8.1)$$

3. Definir como estimador de θ a

$$\bar{g}_n = \frac{1}{M} \sum_{1 \leq m \leq M} g_n(m).$$

4. Definir como estimador da variância de \bar{g}_n a

$$S^2(\bar{g}_n) = \frac{1}{M(M-1)} \sum_{1 \leq m \leq M} (g_n(m) - \bar{g}_n)^2.$$

As técnicas de redução da variância consistem em modificar esse esquema para obter uma diminuição da variância do estimador de θ . Por exemplo: modificando a forma de geração das diversas variáveis aleatórias envolvidas, ou incorporando conhecimento analítico sobre parte dos cálculos sendo realizados, etc.

O leitor interessado em uma das primeiras exposições na literatura estatísticas sobre “redução da variância” pode ver o trabalho de Kahn [70]. Para uma resenha sobre as aplicações o texto de Kleijnen [76] e, para para uma exposição mais atualizada (embora muito técnica) Wilson [128]. Já o trabalho de Nelson e Schmeiser [106] é uma tentativa de fornecer um tratamento sistemático e uma base teórica para a “redução da variância”.

Sem o intuito de fornecer uma definição formal, podemos dizer que uma técnica é de redução da variância para o problema de estimar $\theta = E_F\{g(X)\}$ por simulação, se ela produz para um certo número M de replicações um estimador de θ com uma variância menor que a do estimador 8.1 do método de simulação à força bruta.

Em geral, θ poderia ser um vetor, g poderia ter uma forma bem complicada, e a distribuição F também. Nesse caso, o uso de alguma técnica de redução da variância é essencial se queremos respostas com uma precisão razoável.

As técnicas em questão poderiam ser agrupadas em duas categorias: as que usam outras variáveis além da $g(X)$ (variável objetivo) e as que modificam o esquema de amostragem das observações de $g(X)$.

Entre as primeiras, veremos *variáveis negativamente correlacionas* (anthitetic variables), *variáveis de controle* e *variáveis de controle com regressão*. Entre as da segunda categoria, *amostragem condicional*, *amostragem seletiva* (importance sampling) e *amostragem estratificada*.

Como tudo na vida... alguma desvantagem têm essas técnicas. Vejamos:

- a) A variância é usada como medida de dispersão razoável quando a variável em questão parece-se a (ou é) uma normal. No entanto, uma certa técnica de RV (redução da variância) poderia induzir à não normalidade dos dados e, ainda mais,

muito afastadas que poderiam afetar severamente a performance da T_n dada em 8.2. Essa alternativa tem sido chamada *estatística t-Student aparada*.

Vejamos a sua definição: sejam os valores x_1, \dots, x_n como antes, $(x_{(1)}, \dots, x_{(n)})$ o vetor (x_1, \dots, x_n) ordenado de maneira que $x_{(1)} \leq \dots \leq x_{(n)}$. Definiremos

$$\bar{x}_{n,1} = \frac{1}{n-2} \sum_{2 \leq i \leq n-1} x_{(i)}.$$

Então, a *t-Student aparada* está definida por

$$t_{n,1}(x_{(1)}, \dots, x_{(n)}) = \frac{\frac{1}{n-2} \sum_{i=2}^{n-1} x_{(i)} - \mu_0}{\sqrt{\frac{1}{n(n-2)} \sum_{i=2}^{n-1} (x_{(i)} - \bar{x}_{n,1})^2}}.$$

Algumas questões que poderiam tentar-se resolver usando simulação são:

- T1) Se X_1, \dots, X_n são *valid* com uma certa distribuição F (por exemplo F normal $\mathcal{N}(\mu_0, \sigma^2)$) qual é a distribuição da $T_{n,1} = t(X_1, \dots, X_n)$ para vários valores de n e várias F ? Em particular, quais são os quantis da $T_{n,1}$ também para vários n e F ? Lembremos que dado $0 < \alpha < 1$ chama-se “ α -quantil de uma variável Y a um valor, digamos $y(\alpha)$, tal que $\Pr(Y \leq y(\alpha)) = \alpha$.
- T2) Na situação anterior, para quais valores de n pode-se considerar que $T_{n,1}$ é aproximadamente normal? E, nesses casos, quais são a média e o desvio padrão dessa normal aproximadamente?
- T3) Comparar as distribuições de T_n e $T_{n,1}$ para vários valores de n e F (função de distribuição das “observações” X_1, \dots, X_n). Isto, visando ter uma idéia das potências dos testes baseados em T_n e $T_{n,1}$.

8.2 Variáveis Negativamente Correlacionadas

Também chamada técnica RVAV (redução da variância usando variáveis *antithetic*). Considere a situação habitual simplificada, isto é sejam F uma função de distribuição acumulada; uma função $g: \mathbb{R} \rightarrow \mathbb{R}$ (mensurável e “fácil” de ser calculada); X uma variável aleatória com distribuição F . Estimar $\theta = \mathbb{E}\{g(X)\}$ (ou $\mathbb{E}_F\{g(X)\}$) para destacar a distribuição da X .

A RVAV está baseada na seguinte observação: se X_1 e X_2 são variáveis aleatórias tais que $\text{Var}(X_1) = \text{Var}(X_2) = \sigma^2$ e $\rho = \text{Correlação}(X_1, X_2)$, então

$$\text{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2}\sigma^2(1 + \rho).$$

Daí, se $\rho < 0$, então $\text{Var}(X_1 + X_2)/2 < \text{Var}(X_1)$. Ora, então, se também $\mathbb{E}\{X_1\} = \mathbb{E}\{X_2\} = \mu$, temos $\mathbb{E}\{(X_1 + X_2)/2\} = \mu$ e assim $(X_1 + X_2)/2$ é um estimador de μ (mesmo que X_1 ou que X_2), porém com menor variância num fator $1/2$. O seguinte resultado vai ajudar na busca de variáveis aleatórias como X_1 e X_2 negativamente correlacionadas [110]:

Teorema 8.2.1 *Sejam $g: (0,1) \rightarrow \mathbb{R}$ uma função estritamente crescente (ou decrescente) e U uma variável aleatória com distribuição $\mathcal{U}(0,1)$, então*

$$\rho(g(U), g(1-U)) < 0.$$

Demonstração: Seja $\theta = \mathbb{E}\{g(U)\}$. É fácil ver que $\mathbb{E}\{g(1-U)\} = \theta$. Daí

$$\text{Cov}(g(U), g(1-U)) = \mathbb{E}\{g(U)(g(1-U) - \theta)\} \quad (8.3)$$

Seja $t = \inf\{u: g(u) > \theta\}$, é fácil ver que

$$\begin{aligned} 0 < u < t &\implies g(u)(g(1-u) - \theta) < g(t)(g(1-u) - \theta) \\ \text{e que } t < u < 1 &\implies g(u)(g(1-u) - \theta) < g(t)(g(1-u) - \theta). \end{aligned}$$

Daí, pelo resultado 8.3 temos:

$$\begin{aligned} \text{Cov}(g(U), g(1-U)) &< t \int_0^t g((1-u) - \theta) du + t \int_t^1 g((1-u) - \theta) du \\ &= t \int_0^1 g((1-u) - \theta) du = t(\mathbb{E}\{g(1-U)\} - \theta) = 0 \square \end{aligned}$$

Voltemos à nossa situação habitual e suponhamos que g é estritamente crescente ou decrescente. Sejam M inteiro par e $(U_1, \dots, U_{M/2})$ variável com distribuição $\mathcal{U}(0,1)$; definamos par todo $i = 1, \dots, M/2$:

$$X_i = F^-(U_i) \text{ e } X_i^a = F^-(1 - U_i).$$

Então $(X_1, \dots, X_{M/2})$ são variáveis com distribuição F ; analogamente $X_1^a, \dots, X_{M/2}^a$. Ora, para cada $j = 1, \dots, M/2$ seja

$$g_j^A = \frac{g(X_j) + g(X_j^a)}{2}.$$

Definamos então *estimador de θ baseado em variáveis negativamente correlacionadas* a

$$\bar{g}^A = \frac{1}{M/2} \sum_{1 \leq j \leq M/2} g_j^A = \frac{\bar{g} + \bar{g}_a}{2}, \quad \text{onde } \bar{g} = \frac{1}{M/2} \sum_{1 \leq i \leq M/2} g(X_i), \bar{g}_a = \frac{1}{M/2} \sum_{1 \leq i \leq M/2} g(X_i^a).$$

Logo,

$$\text{Var}(\bar{g}^A) = \frac{\text{Var}(g(X))}{M} (1 + \rho_A) \quad \text{onde } \rho_A = \text{Correlação}(g(X_1), g(X_1^a)).$$

Notemos que

1. Pelo Teorema 8.2.1, $\rho_A < 0$, e
2. $\text{Var}(g(X))/M$ é a variância do estimador Monte Carlo à força bruta, isto é a já vista $M^{-1} \sum_{i=1}^M g(X_i^*)$, onde X_1^*, \dots, X_M^* são variáveis com distribuição F .

Ora, pela definição de \bar{g}^A temos que um estimador da variância do estimador \bar{g}^A é dado por

$$\widehat{\text{Var}}(g^A) = \frac{2}{M} \frac{2}{M-2} \sum_{1 \leq j \leq M/2} (g_j^A - \bar{g}^A).$$

Daí que podem ser definidos intervalos de confiança; em particular, se os testes sobre a amostra $g_1^A, \dots, g_{M/2}^A$ permitem supor que elas estão (aproximadamente) distribuídas como uma normal, então poderiam ser usados os intervalos baseados na estatística t -Student.

Vantagens e Desvantagens da RVAV:

1. (Vantagem notável) Se X é simétrica em torno de 0, qualquer método de geração de X serve (não apenas o Método de Inversão). Com efeito, nesse caso, é suficiente definir como $X^a = -X$.
2. A técnica RVAV, como foi definida, requer que a geração das X 's seja o Método de Inversão (exceto para X simétrica, como vimos acima). Isso pode ser muito pesado do ponto de vista computacional, fazendo com que seja mais rápido fazer Monte Carlo à força bruta com M grande.
3. Às vezes a correlação negativa obtida é baixa, e aí não se justifica usar RVAV.
4. É freqüente que a situação habitual não seja tão simples como a considerada até agora. Por exemplo, poderia acontecer que a quantidade a estimar fosse função de várias variáveis. Pode ser bem difícil definir pares negativamente correlacionados para cada uma das variáveis envolvidas.

Finalizaremos esta Secção dando dois resultados que indicam as limitações desta técnica, e também por que é conveniente fazer o já sugerido na situação habitual.

Definição 8.2.1 *Seja X uma variável aleatória com fda F . Diz-se que um vetor aleatório (Z, Z^*) com fda $G: \mathbb{R}^2 \rightarrow [0, 1]$ é um par AV para F se*

1. $G_Z = G_{Z^*} = F$, isto é as marginais de Z e de Z^* são iguais à F .
2. $\rho(Z, Z^*) < 0$.

Teorema 8.2.2 *Ripley [110, págs. 130-131]: seja X uma va com fda F . Vale que*

- a) *seja (Z, Z^*) um par AV para F , então*
 - a1) $\max\{0, F(z) + F(z^*) - 1\} \leq \mathbb{P}(Z \leq z, Z^* \leq z^*) \leq F(\min\{z, z^*\})$.
 - a2) $\text{Cov}(Z, Z^*) = \iint [\mathbb{P}(Z \leq z, Z^* \leq z^*) - F(z)F(z^*)] dz dz^*$.
- b) *Seja $Z = Z^*$, então $\mathbb{P}(Z \leq z, Z^* < z^*) = F(\min\{z, z^*\})$.*
- c) *Sejam $Z = F^-(U)$ e $Z^* = F^-(1 - U)$, $U \sim \mathcal{U}(0, 1)$; então $\mathbb{P}(Z \leq z, Z^* \leq z^*) = \max\{0, F(z) + F(z^*) - 1\}$*

O importante deste Teorema é que dele resulta que o par (Z, Z^*) AV para F com menor correlação possível é o formado por $Z = F^-(U)$ e $Z^* = F^-(1 - U)$ com $U \sim \mathcal{U}(0, 1)$.

O resultado a seguir será muito útil para ver a redução da variância que pode atingir-se usando pares AV, particularmente o indicado acima, quando se usa integração Monte Carlo.

Proposição 8.2.1 *Seja a função $\phi: (a, b) \rightarrow \mathbb{R}$ integrável; $\theta = \int_a^b \phi(x) dx$; U_1, \dots, U_n i.i.d com distribuição $\mathcal{U}(0, 1)$ e sejam*

$$\begin{aligned}\tilde{\theta} &= \frac{b-a}{2n} \sum_{1 \leq i \leq n} \left\{ \phi(a + (b-a)U_i) + \phi(b - (b-a)U_i) \right\} \\ \hat{\theta} &= \frac{b-a}{n} \sum_{1 \leq i \leq n} \left\{ \phi(a + (b-a)U_i) \right\}.\end{aligned}$$

Então

$$\begin{aligned}i) \text{ Var}(\tilde{\theta}) &= \frac{b-a}{4n} \int_0^1 \{ \psi(u) + \psi(1-u) \}^2 du - \frac{\theta^2}{n}, \text{ e} \\ ii) \text{ Var}(\hat{\theta}) &= \frac{b-a}{n} \int_0^1 \{ \psi(u) \}^2 du - \frac{\theta^2}{n},\end{aligned}$$

onde $\psi(u) = \phi(a + (b-a)u)$.

A demonstração fica como exercício.

Deste resultado deduz-se que $\text{Var}(\tilde{\theta})$ será pequena se $\{ \psi(u) + \psi(1-u) \}$ é bem próxima de uma constante; em particular, se ψ (e daí ϕ) são aproximadamente lineares.

Para completar esta Secção, é recomendável tentar resolver os Exercícios 8.1, 8.2 e 8.3.

8.3 Variáveis de Controle

Também conhecida como *técnica RVCV*. Consideremos a situação habitual generalizada: $n \geq 1$, $F: \mathbb{R}^n \rightarrow [0, 1]$ uma fda, \mathbf{X} um vetor aleatório n -dimensional com fda F e g uma função mensurável. Problema: estimar $\theta = \mathbb{E}_F\{g(\mathbf{X})\}$.

Definição 8.3.1 *Diz-se que $C = \psi(\mathbf{X})$ é uma variável de controle simples para o Problema acima se $\mu_C = \mathbb{E}_F\{\psi(\mathbf{X})\} = \mathbb{E}_F\{C\}$ é conhecida analiticamente, seja em forma exata ou aproximada, e se sabe-se (ou suspeita-se) que $g(\mathbf{X})$ e C são positivamente correlacionadas.*

Suponhamos que C é uma tal variável. Definimos $Y = g(\mathbf{X})$ e $Y' = Y - (C - \mu_C)$. Suponhamos ainda que $(Y_1, C_1), \dots, (Y_M, C_M)$ são vetores aleatórios i.i.d com distribuição (Y, C) . Então definimos $Y'_i = Y_i - (C_i - \mu_C)$. Se

$$\bar{Y}' = \frac{1}{M} \sum_{1 \leq i \leq M} Y'_i, \quad \bar{Y} = \frac{1}{M} \sum_{1 \leq i \leq M} Y_i, \quad \bar{C} = \frac{1}{M} \sum_{1 \leq i \leq M} C_i,$$

temos que

$$\mathbb{E}\{\bar{Y}'\} = \mathbb{E}\{\bar{Y}\} = \mathbb{E}\{Y\} = \theta \text{ e } \sigma^2(\bar{Y}') = \sigma^2(\bar{Y}) + \sigma^2(\bar{C} - \mu_C) - 2\text{Cov}(\bar{Y}, \bar{C} - \mu_C).$$

Daí vemos que se

$$\sigma^2(\bar{C} - \mu_C) - 2\text{Cov}(\bar{Y}, \bar{C} - \mu_C) < 0, \quad (8.4)$$

então \bar{Y}' é um bom estimador de θ com menor variância que \bar{Y} , que é o estimador Monte Carlo à força bruta. Mas é claro que para que isto funcione devemos conhecer exatamente as

quantidades que aparecem na equação 8.4, ou então estimá-las... com maior precisão e mais facilmente que θ (!!).

Suponhamos primeiramente que $\sigma^2(Y) = \sigma^2(C) = \sigma^2$, e seja $\rho_{CY} = \text{Correlação}(Y, C)$. Então

$$\sigma^2(\bar{Y}') = \frac{2\sigma^2(1 - \rho_{CY})}{M} = \sigma^2(\bar{Y})(2(1 - \rho_{CY})); \quad (8.5)$$

donde deduz-se que para obter redução da variância respeito da \bar{Y} deve ter-se $\rho_{CY} > 1/2$. Veja o Exercício 8.4 para o caso $\sigma^2(Y) \neq \sigma^2(C)$.

Notemos que, dado que \bar{Y}' é uma média amostral, sempre podemos obter uma estimação da variância de \bar{Y}' como

$$\widehat{\sigma^2}(\bar{Y}') = \frac{1}{M(M-1)} \sum_{1 \leq i \leq M} (Y'_i - \bar{Y}')^2.$$

Exemplo 8.3.1: Continuação do Exemplo 8.1.1 (a estatística t -Student aparada) Neste caso sejam $n \geq 2$; $F: \mathbb{R} \rightarrow [0, 1]$ uma fda; $X_n = (X_{n,1}, \dots, X_{n,n})$ variad com distribuição F ; $t_n: \mathbb{R}_*^n \rightarrow \mathbb{R}$ e $t_{n,1}: \mathbb{R}_*^n \rightarrow \mathbb{R}$, onde $\mathbb{R}_*^n = \{(x_1, \dots, x_n) \in \mathbb{R}^n: x_i \neq x_j \text{ se } i \neq j\}$ como foram definidas no Exemplo 8.1.1 com $\mu_0 = 0$ (para simplificar). Problema: estimar $\theta_n = \mathbb{E}_F\{t_{n,1}(X_{n,1}, \dots, X_{n,n})\}$ para $F = \mathcal{N}(0, 1)$ (poderia tentar-se com outras F também mas, para aplicar RVCV do jeito que explicaremos a seguir, deve ser conhecida $\mathbb{E}_F\{t_{n,1}(X_{n,1}, \dots, X_{n,n})\}$). De fato, já sabemos analiticamente que $\theta_n = 0$, mas é bom saber como funciona RVCV. Também sabemos que no caso normal, ou seja quando $F = \mathcal{N}(0, 1)$, vale que $\mathbb{E}_F\{t_{n,1}(X_{n,1}, \dots, X_{n,n})\} = 0$, o que sugere usar como variável de controle simples a $C = t_n(X_{n,1}, \dots, X_{n,n})$. Sejam agora $Y = t_{n,1}(X_{n,1}, \dots, X_{n,n})$ e $Y' = Y - C$. É recomendável tentar fazer a seguinte experiência de simulação (Exercício 8.5): seja M inteiro.

1. Para cada $m = 1, 2, \dots, M$ gerar $(x_{m,1}, \dots, x_{m,n})$, amostra da $(X_{n,1}, \dots, X_{n,n})$ e calcular $y(m) = t_{n,1}(X_{m,1}, \dots, X_{m,n})$, $a(m) = t_n(X_{m,1}, \dots, X_{m,n})$ e $y'(m) = y(m) - a(m)$.
2. Definir os estimadores de θ_n :

$$\hat{\theta}_n = \frac{1}{M} \sum_{1 \leq m \leq M} y(m) \text{ e } \tilde{\theta}_n = \frac{1}{M} \sum_{1 \leq m \leq M} y'(m).$$

3. Calcular os estimadores da variância de $\hat{\theta}_n$ e $\tilde{\theta}_n$, isto é

$$S^2(\hat{\theta}_n) = \frac{1}{M} S^2(Y) = \frac{1}{M(M-1)} \sum_{1 \leq m \leq M} (y(m) - \hat{\theta}_n)^2$$

$$S^2(\tilde{\theta}_n) = \frac{1}{M} S^2(Y') = \frac{1}{M(M-1)} \sum_{1 \leq m \leq M} (y'(m) - \tilde{\theta}_n)^2.$$

4. Também calcular

$$\bar{a}_n = \frac{1}{M} \sum_{1 \leq m \leq M} a(m)$$

$$S^2(\bar{a}_n) = \frac{1}{M} S^2(C) = \frac{1}{M(M-1)} \sum_{1 \leq m \leq M} (a(m) - \bar{a}_n)^2,$$

e o estimador da ρ_{CY} dado por (coeficiente de correlação amostral de Pearson):

$$\rho_{CY} = \frac{\sum_{1 \leq m \leq M} (y(m) - \bar{\theta}) (a(m) - \bar{a}_n)}{S(Y) S(C)}$$

Então a redução da variância obtida usando $\tilde{\theta}_n$ sobre $\hat{\theta}_n$ pela equação 8.5 deve ser aproximadamente $2(1 - \hat{\rho}_{CY})$. Para verificar isso, comparar

$$\frac{S^2(\tilde{\theta}_n)}{S^2(\hat{\theta}_n)} \text{ com } 2(1 - \hat{\rho}_{CY}).$$

Vejam agora uma interpretação prática do fator $2(1 - \rho_{CY})$ da fórmula 8.5. Veremos em quanto deveríamos aumentar o número de replicações (M) para obter a mesma redução da variância obtida com \bar{Y}' sobre \bar{Y} . Seja $M^* \geq M$ e definamos $\bar{Y}^* = M^{-1} \sum_{i=1}^M Y_i$ (baseado em M replicações); temos

$$\frac{\sigma^2(\bar{Y})}{\sigma^2(\bar{Y}^*)} = \frac{M^{-1} \sigma^2(Y)}{(M^*)^{-1} \sigma^2(Y)} = \frac{M^*}{M},$$

e pela equação 8.5

$$\frac{\sigma^2(\bar{Y})}{\sigma^2(\bar{Y}')} = (2(1 - \rho_{CY}))^{-1},$$

donde deduz-se que deveríamos ter M^* tal que

$$\frac{M^*}{M} \approx (2(1 - \rho_{CY}))^{-1};$$

isto é $M^* \approx M(2(1 - \rho_{CY}))^{-1}$. Quer dizer, $(2(1 - \rho_{CY}))^{-1}$ é o fator que indica quanto deve aumentar-se o número de replicações para obter com Monte Carlo à força bruta a mesma redução da variância que usando RVCV.

Para finalizar esta Secção, destacaremos algumas vantagens e desvantagens desta técnica RVCV. Entre as vantagens:

- v1) Aplica-se mais geralmente que a RVAV.
- v2) É bem simples na maioria dos casos em que é aplicável.

Já entre as desvantagens:

- d1) É preciso conhecer $\text{IE}\{C\}$.
- d2) Nem sempre, com um controle simples, é fácil de se obter uma redução da variância que justifique a complicação dos programas de computador para implementá-la.

Para tentar resolver d2), embora parcialmente, tem-se sugerido a técnica que passaremos a estudar a seguir.

8.4 Variáveis de Controle com Regressão

Também conhecida como técnica RVCVR. Na verdade, na Definição 8.3.1 foi usada como estimador de θ a média amostral de $Y' = Y - (C - \mu_C)$ na crença de que $\rho_{YC} > 0$; caso tivéssemos pensado que $\rho_{YC} < 0$, deveríamos ter usado $Y' = Y + (C - \mu_C)$. Mais geralmente, poderíamos ter usado um *estimador ajustado por regressão* da forma

$$\bar{Y}^* = \bar{Y} + \beta(\bar{C} - \mu_C)$$

e escolher β tal que minimize $\sigma^2(\bar{Y}^*)$. A seguinte Proposição nos diz como deve ser esse β

Proposição 8.4.1 *Sejam $(Y_1, C_1), \dots, (Y_M, C_M)$ vetores aleatórios iid com distribuição (Y, C) (um vetor aleatório bidimensional qualquer) e $\mu_C = \mathbb{E}\{C\}$. Para cada $b \in \mathbb{R}$ seja Y^b a variável aleatória definida por $Y^b = Y + b(C - \mu_C)$ e sejam*

$$\bar{Y} = \frac{1}{M} \sum_{1 \leq m \leq M} Y_m, \quad \bar{C} = \frac{1}{M} \sum_{1 \leq m \leq M} C_m, \quad \text{e} \quad \bar{Y}^b = \frac{1}{M} \sum_{1 \leq m \leq M} Y_m^b,$$

onde $Y_m^b = Y_m + b(C_m - \mu_C)$. Então

$$\text{Var}(\bar{Y}^b) \leq \text{Var}(\bar{Y}) \quad \forall b \in \mathbb{R} \iff \beta = \frac{-\text{Cov}(\bar{Y}, \bar{C})}{\text{Var}(\bar{C})} = \frac{-\text{Cov}(Y, C)}{\text{Var}(C)} = -\rho_{CY} \frac{\sigma(Y)}{\sigma(C)}. \quad (8.6)$$

A demonstração fica como exercício.

Esse resultado sugere que, na situação habitual generalizada, definamos a técnica de redução da variância pelo método de variável de controle com ajuste por regressão (ou técnica RVCVR) da seguinte maneira, havendo escolhido $C = \psi(\mathbf{X})$ tal que $\mathbb{E}(C) = \mu_C$ é conhecida e chamando Y à $g(\mathbf{X})$.

Se $\sigma(Y)$, $\sigma(C)$ e ρ_{CY} são conhecidos, definir $\beta = -\rho_{CY} \frac{\sigma(Y)}{\sigma(C)}$. Se algum dos parâmetros (ou todos eles) são desconhecidos, então estimá-los; digamos que $\hat{\sigma}(Y)$, $\hat{\sigma}(C)$ e $\hat{\rho}(CY)$ são esses estimadores, então definir $\beta = -\hat{\rho}_{CY} \frac{\hat{\sigma}(Y)}{\hat{\sigma}(C)}$. A seguir, definir $Y^* = Y + \beta(C - \mu_C)$. Suponhamos agora que $(Y_1, C_1), \dots, (Y_M, C_M)$ vetores aleatórios iid com distribuição (Y, C) ; definimos $Y_i^* = Y + \beta(C_i - \mu_C)$. E agora continuamos como na Seção 8.3, isto é definimos

$$\bar{Y}^* = \frac{1}{M} \sum_{1 \leq m \leq M} Y_m^*, \quad \bar{Y} = \frac{1}{M} \sum_{1 \leq m \leq M} Y_m, \quad \text{e} \quad \bar{C} = \frac{1}{M} \sum_{1 \leq m \leq M} C_m.$$

Pela Proposição 8.4.1, se $\sigma(Y)$, $\sigma(C)$ e ρ_{CY} são conhecidos temos

$$\text{Var}(\bar{Y}^*) = \frac{1}{M} \sigma^2(Y) (1 - \rho_{CY}^2),$$

e também sabemos que

$$\text{Var}(\bar{Y}) = \frac{1}{M} \sigma^2(Y).$$

Logo,

$$\frac{\text{Var}(\bar{Y}^*)}{\text{Var}(\bar{Y})} = 1 - \rho_{CY}^2. \quad (8.7)$$

Da mesma maneira como fizemos na Secção anterior, podemos interpretar $(1 - \rho_{CY}^2)^{-1}$ como o fator que indica quanto deve aumentar-se o número de replicações par obter com Monte Carlo à força bruta a mesma redução da variância que usando a técnica RVCVR.

Da equação 8.7 deduz-se também que, no caso $\sigma(Y)$, $\sigma(C)$ e ρ_{CY} conhecidos sempre consegue-se redução da variância, exceto quando $\rho_{CY} = 0$: nessa situação $\text{Var}(\bar{Y}^*) = \text{Var}(\bar{Y})$.

Infelizmente, na maioria dos casos de interesse, $\sigma(Y)$ e ρ_{CY} devem ser estimados a partir da amostra $(Y_1, C_1), \dots, (Y_M, C_M)$, o que pode introduzir um vício no \bar{Y}^* que pode comprometer a validade do afirmado até agora, em particular da equação 8.7. Mesmo assim, é possível estimar $\text{Var}(\bar{Y}^*)$ por

$$\widehat{\sigma^2}(\bar{Y}^*) = \frac{1}{M(M-1)} \sum_{1 \leq m \leq M} (Y_m^* - \bar{Y}^*)^2,$$

e comparar esse valor com

$$\widehat{\sigma^2}(\bar{Y}) = \frac{1}{M(M-1)} \sum_{1 \leq m \leq M} (Y_m - \bar{Y})^2,$$

mas $\widehat{\sigma^2}(\bar{Y}^*)$ será viciado.

Também podem usar-se outras técnicas para estimar $\sigma^2(\bar{Y}^*)$, como a já vista bootstrap ou ainda a chamada de *batching*, que pode ver-se na referência [88] com detalhe ou no Apêndice B.

Novamente, se $\sigma(Y)$, $\sigma(C)$ e ρ_{CY} são conhecidos, temos que β não será uma variável aleatória e $\widehat{\sigma^2}(\bar{Y}^*)$ será um estimador não viciado da $\text{Var}(\bar{Y}^*)$. Se, além disso, Y_1^*, \dots, Y_n^* podem considerar-se normalmente distribuídas, então um intervalo de confiança para $\theta = \mu_Y$ de nível $1 - \alpha$ será

$$\left[\bar{Y}^* - t_{\alpha/2, M-1} \widehat{\sigma}(\bar{Y}^*), \bar{Y}^* + t_{\alpha/2, M-1} \widehat{\sigma}(\bar{Y}^*) \right],$$

onde $t_{\alpha/2, M-1}$ é dada por $\Pr(T \leq t_{\alpha/2, M-1}) = 1 - \alpha/2$ e $T \sim t\text{-Student}$ com $M - 1$ graus de liberdade.

Ora, também é possível usar variáveis de controle (simples ou com regressão) para construir estimadores de θ que não são médias amostrais como \bar{Y}^* . Para ilustrar o tipo de questões que podem aparecer, vejamos o seguinte

Exemplo 8.4.1: Continuação dos Exemplos 8.1.1 e 8.3.1 (estimação dos quantis da estatística $t\text{-Student}$ aparada) Consideremos a situação colocada no início do Exemplo 8.3.1, mas neste caso o problema será estimar $\theta_{n,\alpha}$ tal que $\Pr(t_{n,1}(X_{n,1}, \dots, X_{n,n}) \leq \theta_{n,\alpha}) = 1 - \alpha$. Para revisar o que diz respeito à estimação de quantis, o leitor interessado pode ler alguma coisa (muito breve) no Apêndice A.

• Método 1: Usando o estimador do Apêndice A: seja K inteiro

1. Para cada $k = 1, \dots, K$ gerar $(x_{k,1}, \dots, x_{k,n})$ amostra da $(X_{k,1}, \dots, X_{k,n})$ (suporemos que a F é normalmente distribuída: $F = \mathcal{N}(0, 1)$, salvo indicação em contrário). Calcular $t_{n,1}(k) = t_{n,1}(x_{k,1}, \dots, x_{k,n})$

2. Ordenar $(t_{n,1}(1), \dots, t_{n,1}(K))$ obtendo $t_{n,1}(K : 1) < t_{n,1}(K : 2) < \dots < t_{n,1}(K : K)$.
3. Definir como "valor observado do $Y = \hat{x}_{\alpha,K}$ " a $y_\alpha = t_{n,1}(K : k_0)$, onde

$$k_0 = \begin{cases} \alpha K & \text{se } \alpha K \text{ é inteiro,} \\ [\alpha K + 1] & \text{se não.} \end{cases}$$

4. Como estimador da $\text{Var}(Y)$ poderíamos usar o sugerido pela fórmula A.2:

$$\widehat{\sigma^2}(Y) = \frac{\alpha(1 - \alpha)}{K\phi(y_\alpha)},$$

e onde ϕ é a densidade da $\mathcal{N}(0, 1)$. Também poderíamos calcular $S_B^2(Y)$, o estimador bootstrap da $\text{Var}(Y)$ baseado em $(t_{n,1}(1), \dots, t_{n,1}(K))$ e B (por exemplo $B = 512$) replicações — veja a Seção 7.1.2. Outra alternativa seria aplicar a técnica de *batching* (veja o Apêndice B) e que vamos ver explicitamente para este mesmo caso, daqui a pouco.

• **Método 2:** Usando como variável de controle simples o estimador do α -quantil da t -Student com $n - 1$ graus de liberdade. A idéia é aplicar o já visto na Seção 8.3 usando como C o estimador do α -quantil da t -Student com $n - 1$ graus de liberdade. Para tanto, deveremos conhecer μ_C isto é, caso estejamos trabalhando com uma amostra de tamanho K da $T_n = t_n(X_{n,1}, \dots, X_{n,n})$ a esperança da estatística de ordem αK (ou $[\alpha K] + 1$) de uma t -Student baseada em amostras de tamanho K . Se n é "grande" (digamos $n \geq 20$), poderíamos substituir a t -Student pela $\mathcal{N}(0, 1)$, e assim usar como μ_C o valor da esperança da estatística de ordem αK (ou $[\alpha K] + 1$) de uma $\mathcal{N}(0, 1)$ baseada em amostras de tamanho K . Esses valores podem achar-se tabelados, por exemplo, em [107].

Assim, o algoritmo correspondente a este método é: seja K inteiro.

1. Para cada $k = 1, \dots, K$ gerar $(x_{k,1}, \dots, x_{k,n})$ amostra de $(X_{k,1}, \dots, X_{k,n})$. Calcular $t_n(k) = t_n(x_{k,1}, \dots, x_{k,n})$ e $t_{n,1}(k) = t_{n,1}(x_{k,1}, \dots, x_{k,n})$.
2. Ordenar $t(1), \dots, t_n(K)$ e $t_{n,1}(1), \dots, t_{n,1}(K)$ obtendo $t(K : 1), \dots, t_n(K : K)$ e $t_{n,1}(K : 1), \dots, t_{n,1}(K : K)$.
3. Definir como valor observado de Y (estimador do α -quantil da $t_{n,1}(X_{k,1}, \dots, X_{n,n})$) e como valor observado de C (estimador do α -quantil da $t_n(X_{k,1}, \dots, X_{n,n})$) os valores $y_\alpha = t_{n,1}(K : k_0)$ e $C_\alpha = t_n(K : k_0)$, sendo k_0 como no Método 1. Definir como valor observado de $Y' = Y - (C - \mu_C)$ a $y'_\alpha = y_\alpha - (C_\alpha - \mu_C)$ (o μ_C para $n = 30$ e $K = 20$ é 2.09).

Sobre as estimações das variâncias dos estimadores envolvidos, vale o já indicado no ponto 4 do Método 1.

• **Método 3:** Usando como variável de controle com regressão o mesmo estimador que no Método 2. O Método 2 está baseado na crença (falsa!) de que se as médias amostrais

de amostras de certo tamanho de duas variáveis aleatórias estão altamente correlacionadas, então as estatísticas de ordem dessas amostras também o estarão.

Para corrigir esse defeito, bem como para mostrar a aplicação do método de *batching* (ver o Apêndice B) para calcular as variâncias dos estimadores do $\theta_{n,\alpha}$ é que sugerimos a realização da experiência Monte Carlo proposta mais adiante.

A idéia é agora usar Y e C como definidos no Método 2 e definir como estimador de $\theta_{n,\alpha}$ a

$$Y^* = Y + \beta(C - \mu_C)$$

com β tal que $\beta = (\text{ou } \approx) -\rho_{CY} \frac{\sigma(Y)}{\sigma(C)}$. Mas, não temos os valores de ρ_{CY} , $\sigma(Y)$ nem de $\sigma(C)$; assim que deveremos estimá-los.

A seguinte experiência Monte Carlo visa analisar o comportamento dos estimadores definidos por esta técnica (RVCVR), para n e K fixos, para os valores de $\alpha \in \{\frac{1}{K}, \frac{2}{K}, \dots, \frac{K}{K}\}$ e para β estimado. Note-se que, para cada α , teremos diferentes Y , β , C , μ_C e Y^* .

Sejam M e K inteiros

1. Para cada $m = 1, \dots, M$ gerar $(x_{m,1}, \dots, x_{m,n})$ uma amostra da $(X_{m,1}, \dots, X_{m,n})$.
2. Para cada $r = 1, \dots, R = [M/K]$ seja $a = (r-1)K$
 - (a) Para cada $k = 1, \dots, K$ calcular $t_n(k, r) = t_n(x_{k,1}, \dots, x_{k,n})$ e $t_{n,1} = t_{n,1}(x_{k,1}, \dots, x_{k,n})$.
 - (b) Ordenar $(t_n(1, r), \dots, t_n(K, r))$ e $(t_{n,1}(1, r), \dots, t_{n,1}(K, r))$ obtendo $t_n(K : 1, r) \leq \dots \leq t_n(K : K, r)$ e $t_{n,1}(K : 1, r) \leq \dots \leq t_{n,1}(K : K, r)$ (r está fixo).
3. (Estimação de β (para cada α)) Para cada $k = 1, \dots, K$ definir

$$\begin{aligned} \overline{t_n(K : k)} &= \frac{1}{R} \sum_{1 \leq r \leq R} t_n(K : k, r) \\ \widehat{\sigma^2}(C(k)) &= \frac{1}{R-1} \sum_{1 \leq r \leq R} (t_n(K : k, r) - \overline{t_n(K : k)})^2 \\ \overline{t_{n,1}(K : k)} &= \frac{1}{R} \sum_{1 \leq r \leq R} t_{n,1}(K : k, r) \\ \widehat{\sigma^2}(Y(k)) &= \frac{1}{R-1} \sum_{1 \leq r \leq R} (t_{n,1}(K : k, r) - \overline{t_{n,1}(K : k)})^2 \\ \widehat{\rho}(k) = \widehat{\rho}_{C(k), Y(k)} &= \frac{\frac{1}{R} \sum_{1 \leq r \leq R} (t_n(K : k, r) - \overline{t_n(K : k)}) (t_{n,1}(K : k, r) - \overline{t_{n,1}(K : k)})}{\widehat{\sigma}(C(k)) \widehat{\sigma}(Y(k))} \\ \beta(k) &= \widehat{\rho}_{C(k), Y(k)} \frac{\widehat{\sigma}(Y(k))}{\widehat{\sigma}(C(k))}. \end{aligned}$$

4. (Definição do estimador do k/K -quantil para cada $k = 1, \dots, K$ da distribuição $t_{n-1}(X_{n,1}, \dots, X_{n,n})$ usando RVCVR) Para cada $k = 1, \dots, K$ e cada $r = 1, \dots, R$ sejam $y(k, r) = t_{n,1}(K : k, r)$, $C(k, r) = t_n(K : k, r)$ e $y^*(k, r) = y(k, r) + \beta(k)(C(k, r) - \mu(k))$, onde $\mu(k)$ é como μ_C no Método 2 para $\alpha = k/K$. Note-se

que $(y(k, 1), \dots, y(k, R))$ é uma amostra de tamanho R do estimador $Y(k)$ = estatística de ordem k -ésima correspondente a amostras de tamanho K da $T_{n,1} = t_{n-1}(X_{n,1}, \dots, X_{n,n})$; $(C(k, 1), \dots, C(k, R))$ é uma amostra de tamanho R do estimador $C(k)$ = estatística de ordem k -ésima correspondente a amostras de tamanho K da $T_n = t_n(X_{n,1}, \dots, X_{n,n})$; e $(y^*(k, 1), \dots, y^*(k, R))$ é uma amostra de tamanho R do estimador $Y^*(k) = Y(k) + \beta(k)(C(k) - \mu(k))$, que é o estimador do k/K -quantil da $T_{n,1}$ baseado em $Y(k)$, com controle $C(k)$ ajustado por regressão. Então, tendo assim particionado (*sectioning*) o M estamos em condições de estimar os valores de $\mathbb{E}\{Y^*(k)\}$ e de $\sigma^2(Y^*(k))$ definindo, respectivamente,

$$\bar{y}^*(k) = \frac{1}{R} \sum_{1 \leq r \leq R} y^*(k, r) \text{ e } \widehat{\sigma}^2(Y^*(k)) = \frac{1}{R-1} \sum_{1 \leq r \leq R} (y^*(k, r) - \bar{y}^*(k))^2.$$

Notemos que na Etapa 3 já temos calculado os estimadores de $\mathbb{E}\{Y(k)\}$, $\text{Var}(Y(k))$, $\mathbb{E}\{C(k)\}$ e $\text{Var}(C(k))$ sendo, respectivamente

$$\bar{y}(k) = \overline{t_{n,1}(K : k)}, \widehat{\sigma}^2(Y(k)) = \overline{t_n(K : k)} \text{ e } \widehat{\sigma}^2(C(k)).$$

Também podemos obter estimadores para $\text{Var}(\bar{Y}^*(k))$, $\text{Var}(\bar{Y}(k))$ e $\text{Var}(\bar{C}(k))$, onde $\bar{Y}^*(k)$, $\bar{Y}(k)$ e $\bar{C}(k)$ são as médias amostrais sobre “amostras” de tamanho R de $Y^*(k)$, $Y(k)$ e $C(k)$ respectivamente; simplesmente

$$\widehat{\sigma}^2(\bar{Y}^*(k)) = \frac{1}{R} \widehat{\sigma}^2(Y^*(k)), \widehat{\sigma}^2(\bar{Y}(k)) = \frac{1}{R} \widehat{\sigma}^2(Y(k)) \text{ e } \widehat{\sigma}^2(\bar{C}(k)) = \frac{1}{R} \widehat{\sigma}^2(C(k)).$$

Embora... o valor “observado” $\bar{y}^*(k)$ do $\bar{Y}^*(k)$ não está formado pela média de valores independentes de $Y^*(k)$ dado que $y^*(k, 1), \dots, y^*(k, R)$ estão relacionados através do β .

5. Finalmente, para apresentar os resultados mais relevantes e ver se realmente obtemos redução da variância com $Y^*(k)$ e $\bar{Y}^*(k)$, poderíamos construir uma tabela como as Tabelas 8.1 e 8.2:

k	$\bar{C}(k)$	$\widehat{\sigma}(C(k))$	$\widehat{\sigma}(\bar{C}(k))$	$\bar{y}(k)$	$\widehat{\sigma}(Y(k))$	$\widehat{\sigma}(\bar{Y}(k))$
1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
K

Tabela 8.1: Apresentação de Resultados

Vamos finalizar esta Seção com alguns comentários sobre as técnicas RVCVR, sem dúvida a técnica de Redução da Variância que pode ser aplicada ao maior número de estudos de simulação.

Comentário 1: Uma das desvantagens da técnica RVCVR é que o desconhecimento de β conduz a ter que estimá-lo a partir dos dados da mesma simulação, o que produz \bar{Y}^* formado

k	$\hat{\rho}(k)$	$\beta(k)$	$\mu(k)$	$y^*(k)$	$\hat{\sigma}(Y^*(k))$	$\hat{\sigma}(\bar{Y}^*(k))$
1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
K

Tabela 8.2: Apresentação de Resultados

pela média de variáveis não independentes e que (o mais importante) não seja mais verdadeiro que $\mathbb{E}\{Y^*\} = \mathbb{E}\{Y\}$ dado que, rigorosamente, β é uma variável aleatória neste caso. Esse problema poderia tentar ser driblado de várias formas: uma por estimação préviado β ; dessa forma conseguiríamos Y^* com $\mathbb{E}\{Y^*\} = \mathbb{E}\{Y\}$, mas não poderíamos assegurar uma redução da variância. Outra maneira seria particionando a amostra de tamanho M da simulação em dois grupos e proceder como se indica no Exercício 8.7

Comentário 2: Naturalmente, pensa-se que pode usar-se não apenas uma variável de controle, mas sim várias, digamos C_1, \dots, C_s e tentar um ajuste por regressão

$$Y^{**} = Y + \sum_{1 \leq \ell \leq s} \beta_\ell (C_\ell - \mathbb{E}\{C_\ell\}).$$

As complicações na estimação dos β_ℓ e a necessidade de conhecer os valores $\mathbb{E}\{C_\ell\}$ para todo $1 \leq \ell \leq s$ limitam a aplicação da técnica. Por exemplo, para $s = 2$ as fórmulas envolvidas nos cálculos são *horríveis* (veja, por exemplo [88]). Como caso particular, resulta o ajuste não linear (por exemplo quadrático) com uma variável de controle só, isto é

$$Y^{**} = Y + \beta_1 (C - \mathbb{E}\{C\}) + \beta_2 (C^2 - \mathbb{E}\{C^2\}).$$

O leitor interessado na aplicação desta técnica pode tentar a resolução do Exercício 8.8.

Comentário 3: É de utilidade distinguir entre técnicas de controle com variáveis de controle *internas* e *externas*. Sem o intuito de sermos rigorosos, definimos às primeiras como àquelas que tem média conhecida e que aparecem como subproduto na hora de calcular a variável de interesse Y . Desse modo, o seu cálculo não envolve uso extra de computador. Já as variáveis de controle externas são as que devem ser calculadas por procedimentos além dos realizados para calcular Y . As técnicas vistas nesta Secção são do segundo tipo. Com respeito às do primeiro grupo: em muitos casos a média de uma estatística é conhecida, mas nenhuma outra propriedade, então é legal tentar usar a média amostral dessa estatística como uma variável de controle (interna, dado que em geral não vai demandar maior esforço computacional além do já necessário para os cálculos da estatística em questão). Para afixar idéias sobre isto, o leitor poderia tentar resolver os Exercícios 8.9 e 8.10.

Comentário 4: As técnicas RVCV e RVCVR parecem ter dado resultados satisfatórios em certos casos simples. Quando o sistema ou modelo a simular é complexo, os controles eficientes são muito difíceis de serem encontrados. Porém, o leitor interessado pode ver aplicações nesses modelos complexos nas referências [79,88,129]. Sobre uma avaliação do uso destas técnicas em estudos Monte Carlo poderia ver-se o trabalho [80].

Finalmente, um ponto importante. Checar a validade de um programa de simulação é essencial. Nesse sentido, é fortemente recomendado calcular efetivamente qualquer quantidade conhecida (teórica e analiticamente), mesmo que não seja usada para redução da variância.

8.5 Amostragem Condicional

(Ou técnica de redução da variância RVAC) Voltemos a considerar a situação habitual simplificada: sejam F uma fda; $g: \mathbb{R} \rightarrow \mathbb{R}$ uma função mensurável e X uma variável aleatória com distribuição F . Problema: estimar $\mathbb{E}_F\{g(X)\}$.

Suponhamos que X está correlacionada com uma outra variável aleatória Y tal que $\mathbb{E}\{g(X)|Y\}$ resulta conhecida ou fácil de calcular. Será possível usar essa informação para conseguir reduzir a variância do estimador Monte Carlo à força bruta dado pela expressão $\theta = M^{-1} \sum_{i=1}^M g(X_i)$, onde X_1, \dots, X_M são i.i.d com distribuição F ? O resultado a seguir fornece uma resposta positiva

Proposição 8.5.1 *Sejam (Z_1, Z_2) um vetor aleatório bidimensional; $h: \mathbb{R} \rightarrow \mathbb{R}$ e $v: \mathbb{R} \rightarrow \mathbb{R}$ funções mensuráveis tais que $h(Z_2) = \mathbb{E}\{Z_1|Z_2\}$ e $v(Z_2) = \mathbb{E}\{(Z_1 - h(Z_2))^2|Z_2\} = \mathbb{E}\{Z_1^2|Z_2\} - (h(Z_2))^2$. Então $\text{Var}(H(Z_2)) = \text{Var}(Z_1) - \mathbb{E}\{v(Z_2)\} \leq \text{Var}(Z_1)$.*

Demonstração:

$$\begin{aligned} \text{Var}(Z_1) &= \mathbb{E}\{Z_1^2\} - (\mathbb{E}\{Z_1\})^2 = \mathbb{E}\{\mathbb{E}\{Z_1^2|Z_2\}\} - (\mathbb{E}\{\mathbb{E}\{Z_1|Z_2\}\})^2 \\ &= \mathbb{E}\{\mathbb{E}\{Z_1^2|Z_2\}\} - (\mathbb{E}\{h(Z_2)\})^2 \\ &= \mathbb{E}\{\mathbb{E}\{Z_1^2|Z_2\}\} - \mathbb{E}\{(h(Z_2))^2\} + \mathbb{E}\{(h(Z_2))^2\} - (\mathbb{E}\{h(Z_2)\})^2 \\ &= \mathbb{E}\{\mathbb{E}\{Z_1^2|Z_2\}\} - (h(Z_2))^2 + \text{Var}(h(Z_2)). \end{aligned}$$

Aplicando este resultado a $Z_1 = g(X)$, $Z_2 = Y$ temos que $\text{Var}(h(Y)) \leq \text{Var}(g(X))$, onde $h(Y) = \mathbb{E}\{g(X)|Y\}$. Note-se que $\mathbb{E}\{h(Y)\} = \mathbb{E}\{\mathbb{E}\{g(X)|Y\}\} = \mathbb{E}\{g(X)\} = \theta$, pelo que $h(Y)$ é um estimador de θ com variância menor ou igual à variância de $g(X)$.

Do ponto de vista computacional temos o seguinte **Algoritmo para simulação usando “amostragem condicional”** no problema de estimar $\theta = \mathbb{E}_F\{g(X)\}$: seja M inteiro.

1. Gerar y_1, \dots, y_M amostra de tamanho M de Y .
2. Calcular $w_m = h(y_m) = \mathbb{E}\{g(X)|Y = y_m\}$ para $m = 1, \dots, M$.
3. Calcular como estimador de θ : $\theta^* = M^{-1} \sum_{m=1}^M w_m$; logo θ^* é um “valor observado” da variável aleatória $Y^* = M^{-1} \sum_{m=1}^M h(Y_m)$, com $h(Y_m) = \mathbb{E}\{X|Y_m\}$ e Y_1, \dots, Y_M i.i.d com a mesma distribuição de Y .
4. Estimar a variância de Y^* por

$$(s^*)^2 = \frac{1}{M(M-1)} \sum_{1 \leq m \leq M} (w_m - \theta^*)^2,$$

quer dizer, $(s^*)^2$ é um “valor observado” da variável aleatória

$$\widehat{\sigma^2}(s^*)^2 = \frac{1}{M(M-1)} \sum_{1 \leq m \leq M} (h(Y_m) - Y^*)^2.$$

Notemos que

$$\text{Var}(Y^*) = \frac{1}{M} \text{Var}(h(Y)) \leq \frac{1}{M} \text{Var}(g(X)) = \text{Var}(\bar{\theta}).$$

Obviamente, o problema com esta técnica é calcular $\mathbb{E}\{g(X)|Y\}$. As proposições seguintes são de utilidade para essa conta em muitas situações de interesse prático como veremos um pouco mais adiante.

Proposição 8.5.2 *Suponhamos que Y e Z são independentes e que $X = Z/Y$ ($Y > 0$ para simplificar). Então*

- a) $\mathbb{E}\{X|Y\} = Y^{-1}\mathbb{E}\{Z\}$.
- b) $\mathbb{E}\{X^2|Y\} = Y^{-2}\mathbb{E}\{Z\}$.
- c) *Sejam $x_0 \in \mathbb{R}$ e $g(x) = \mathbb{I}_{(-\infty, x_0]}(x)$ (logo $\mathbb{E}\{g(X)\} = \mathbb{P}\{X \leq x_0\}$); então $\mathbb{E}\{g(X)|Y\} = F_Z(x_0 Y)$, onde F_Z é a fda da Z .*

Demonstração: As partes a) e b) ficam como exercício. Vejamos c): seja $y > 0$ um possível valor de Y . Então

$$\begin{aligned} \mathbb{E}\{g(X)|Y = y\} &= \mathbb{E}\{g(Z/Y)|Y = y\} = \mathbb{E}\{g(Z/y)|Y = y\} \\ &= \mathbb{E}\{\mathbb{I}_{(-\infty, x_0]}(Z/y) \mid Y = y\} = \mathbb{E}\{\mathbb{I}_{(-\infty, x_0 y]}(Z) \mid Y = y\}, \end{aligned}$$

usando a independência entre Y e Z temos

$$\mathbb{E}\{\mathbb{I}_{(-\infty, x_0 y]}(Z)\} = \mathbb{P}\{Z \leq x_0 y\} \square$$

Proposição 8.5.3 *Suponhamos que (Y, Z) é normal bivariada com parâmetros $\mu_Z, \mu_Y, \sigma_Z^2, \sigma_Y^2$ e ρ_{YZ} , e que $X = Z/Y$ (como $\mathbb{P}\{Y = 0\} = 0$ podemos supor que Z/Y está bem definida). Então*

$$\mathbb{E}\{X|Y\} = \frac{\mu_Z}{Y} + \frac{\sigma_Z}{\sigma_Y} \rho_{YZ} \left(1 - \frac{\mu_Y}{Y}\right).$$

Demonstração: Seja $y \leq 0$ um valor de Y . Então

$$\mathbb{E}\{X|Y = y\} = \mathbb{E}\{Z/y|Y = y\} = \frac{1}{y} \mathbb{E}\{Z|Y = y\},$$

que, pela Proposição 7.4.1 resulta ser

$$= \frac{1}{y} \left(\mu_Z + \rho_{YZ} \frac{\sigma_Z}{\sigma_Y} (y - \mu_Y) \right) = \frac{\mu_Z}{y} + \frac{\sigma_Z}{\sigma_Y} \left(1 - \frac{\mu_Y}{y}\right).$$

Nota Importante: Os resultados de ambas as Proposições acima podem ser aplicados em vários casos, como pode ver-se levando em consideração o que apresentaremos a seguir (o leitor poderia conferir o resultado): Sejam (Z, Y) independentes com $Z \sim \mathcal{N}(0, 1)$ e $X = Z/Y$; então

- a) $X \sim t_{n-1}$ se $Y = \sqrt{S/(n-1)}$ com $S \sim \chi_{n-1}^2$.
 b) X tem distribuição normal contaminada, (isto é X tem densidade dada por $f(x) = (1-p)\phi(x) + \frac{p}{k}\phi(x/k)$ onde ϕ é a densidade da $\mathcal{N}(0,1)$, $k > 1$ e $0 < p < 1/2$) se $\text{IPr}(Y = 1/k) = p$ e $\text{IPr}(Y = 1) = 1 - p$.
 c) X tem a distribuição de Cauchy padrão se $Y = |N|$ onde $N \sim \mathcal{N}(0,1)$.

Ora, que pode fazer-se nas situações mais complexas do que a da situação habitual simplificada? Consideremos a situação habitual generalizada (início da Secção 8.3): $F: \mathbb{R} \rightarrow [0,1]$ uma fda; X uma va com distribuição F ; $n \geq 2$ e X_1, \dots, X_n vauid com distribuição igual à da X ; $g: \mathbb{R}^n \rightarrow \mathbb{R}$ mensurável. Problema: estimar $\theta = \mathbb{E}_F\{g(X_1, \dots, X_n)\}$.

O que pode ser feito depende, logicamente, de F e da g . Por exemplo, na referência [88] (veja exercício 8.11) é mostrada uma aplicação desta técnica para o caso de $g(x_1, \dots, x_n) = \min\{x_1, \dots, x_n\}$ quando F é uma Gamma com parâmetros β e $0 < \alpha < 1$, isto é F tem densidade dada por

$$f(x) = \frac{x^{\alpha-1} \exp\{-x/\beta\}}{\Gamma(\alpha)\beta^\alpha} \mathbb{I}_{(0,+\infty)}(x).$$

8.6 Amostragem Seletiva

(Ou *técnica* de redução da variância *RVAS*) Em inglês, esta técnica tem recebido o nome de *importance samplig*. A idéia por trás deste método está relacionada com as idéias de ponderação e de estratificação das amostras da Teoria de Amostragem. Em palavras

trata-se de simular mais “amostras” da parte que interessa para o objetivo da experiência de simulação a realizar.

Para fixar conceitos, consideremos a situação habitual simplificada, isto é sejam $f: \mathbb{R} \rightarrow [0, +\infty)$ uma função de densidade; $g: \mathbb{R} \rightarrow \mathbb{R}$ uma função mensurável e X uma variável aleatória com distribuição f . Problema: estimar $\theta = \mathbb{E}_f\{g(X)\}$.

Sejam ainda $h: \mathbb{R} \rightarrow [0, +\infty)$ uma outra densidade tal que $h(t) > 0 \iff g(t)f(t) > 0$, isto é $\{t: h(t) > 0\} = \{t: g(t)f(t) > 0\}$; $\psi: \mathbb{R} \rightarrow [0, +\infty)$ dada por

$$\psi(t) = \begin{cases} \frac{g(t)f(t)}{h(t)} & \text{se } h(t) > 0 \\ 0 & \text{caso contrário;} \end{cases}$$

$n \geq 2$ e Y_1, \dots, Y_n amostra da h (isto é Y_1, \dots, Y_n são vauid com distribuição h). Definimos

$$\hat{\theta}_k = \frac{1}{n} \sum_{1 \leq i \leq n} \psi(Y_i).$$

Proposição 8.6.1 $\mathbb{E}\{\hat{\theta}_k\} = \theta$, isto é $\hat{\theta}_k$ é um estimador não viciado de θ .

Demonstração:

$$\begin{aligned} \mathbb{E}\{\hat{\theta}_k\} &= \frac{1}{n} \sum_{1 \leq i \leq n} \mathbb{E}\{\psi(Y_i)\} = \frac{1}{n} \int \psi(t)h(t) dt \\ &= \frac{1}{n} \sum_{1 \leq i \leq n} \int \frac{g(t)f(t)}{h(t)} h(t) dt = \frac{1}{n} \sum_{1 \leq i \leq n} \int g(t)f(t) dt = \theta \square \end{aligned}$$

Seja agora

$$h^*(t) = \alpha |g(t)f(t)|, \text{ com } \alpha^{-1} = \int |g(t)f(t)| dt = \int |g(t)| |f(t)| dt. \quad (8.8)$$

Proposição 8.6.2

$$\text{Var}(\hat{\theta}_{h^*}) = \frac{1}{n}(\alpha^{-2} - \theta^2)$$

Demonstração:

$$\begin{aligned} \text{Var}(\hat{\theta}_{h^*}) &= \frac{1}{n} \int \left(\frac{g(t)f(t)}{h(t)} - \theta \right)^2 h^*(t) dt \\ &= \frac{1}{n} \int \left(\frac{g(t)f(t)}{\alpha |g(t)| |f(t)|} - \theta \right)^2 \alpha |g(t)| |f(t)| dt \\ &= \frac{1}{n} \int \left(\frac{1}{\alpha^2} - \frac{2\theta g(t)f(t)}{\alpha |g(t)| |f(t)|} + \theta^2 \right) \alpha |g(t)| |f(t)| dt \\ &= \frac{1}{n} \left[\frac{1}{\alpha^2} \int \alpha |g(t)| |f(t)| dt - 2\theta \int g(t)f(t) dt + \theta^2 \alpha \int |g(t)| |f(t)| dt \right] \\ &= \frac{1}{n} \left(\alpha^{-2} - 2\theta \int g(t)f(t) dt + \theta^2 \right) = \frac{1}{n} (\alpha^{-2} - 2\theta^2 + \theta^2) \\ &= \frac{1}{n} (\alpha^{-2} - \theta^2) \square \end{aligned}$$

O resultado a seguir diz que a h^* definida na Proposição 8.6.1 é a que produz o estimador $\hat{\theta}_h$ com a menor variância.

Proposição 8.6.3 *Seja $h: \mathbb{R} \rightarrow [0, +\infty)$ uma densidade qualquer tal que $\{t: h(t) > 0\} = \{t: g(t)f(t) > 0\}$. Então*

$$\text{Var}(\hat{\theta}_h) \geq \frac{1}{n}(\alpha^{-2} - \theta^2).$$

Demonstração:

$$\begin{aligned} \text{Var}(\hat{\theta}_h) &= \frac{1}{n} \int \left(\frac{g(t)f(t)}{h(t)} - \theta \right)^2 h(t) dt \\ &= \frac{1}{n} \left[\int \frac{g^2(t)f^2(t)}{h(t)} dt - \theta^2 \right]. \end{aligned} \quad (8.9)$$

Ora, usando a desigualdade de Cauchy-Schwartz temos

$$\begin{aligned} \int \frac{g^2(t)f^2(t)}{h(t)} dt &= \left(\int \left(\frac{|g(t)| |f(t)|}{\sqrt{h(t)}} \right)^2 dt \right) \left(\int (\sqrt{h(t)})^2 dt \right) \\ &\geq \left[\int \frac{|g(t)| |f(t)|}{\sqrt{h(t)}} \sqrt{h(t)} dt \right]^2 = \left[\int |g(t)| |f(t)| dt \right]^2 = \alpha^{-2}. \end{aligned}$$

Daí e da 8.9 fica provada a Proposição.

O método parece bom... mas a escolha de h^* é impossível do ponto de vista prático, pois para a sua definição deveríamos conhecer $\int |g(t)|f(t) dt$... que é o nosso problema!

Bom, mas a teoria anterior fornece alguma idéia do que fazer. Isto é, procurar que h seja o mais perto possível de $|g(t)|f(t)$. Vejamos um exemplo:

Exemplo 8.6.1 O Método 4 do Exemplo 7.3.1. Lembremos que nesse exemplo queríamos aplicar simulação para estimar $\theta = \text{IPr}(C > 2)$ onde $C \sim \text{Cauchy padrão}$. Em outros termos,

$$\theta = \int_2^{+\infty} f(t) dt = \int_{-\infty}^{+\infty} g(t)f(t) dt,$$

com $f(x) = (\pi(1+x^2))^{-1}$ e $g(x) = \mathbb{I}_{[2,+\infty)}(x)$. Ora, $f(x)$ para $x > 2$ tem um comportamento similar ao de $x \mapsto x^{-2}$, e mais acentuado quando $x \rightarrow \infty$. Isto sugere definir h por

$$h(x) = cx^{-2} \mathbb{I}_{[2,+\infty)}(x) \text{ com } c \text{ tal que } \int h(x) dx = 1.$$

Uma conta fácil mostra que $c = 2$. Então

$$\psi(t) = \begin{cases} \frac{(\pi(1+t^2))^{-1}}{2t^{-2}} = (2\pi(1+t^{-2}))^{-1} & \text{se } t \geq 2 \\ 0 & \text{caso contrário.} \end{cases}$$

Assim, o nosso estimador será

$$\hat{\theta}_h = \frac{1}{n} \sum_{1 \leq i \leq n} \psi(Y_i),$$

onde Y_1, \dots, Y_n é uma amostra de h . Mas, vejamos como é a distribuição H associada à h : uma conta simples estabelece que

$$H(y) = \begin{cases} 1 - \frac{2}{y} & \text{se } y \geq 2 \\ 0 & \text{caso contrário,} \end{cases}$$

daí que $H^{-1}(t) = 2(1-t)$ para $0 < t < 1$. Então para obter uma amostra Y_1, \dots, Y_n de H basta fazer $Y_i = H^{-1}(U_i) = 2/(1-U_i)$ com U_1, \dots, U_n amostra da $\mathcal{U}(0,1)$; ou também $Y_i = V_i^{-1}$ com V_1, \dots, V_n amostra da $\mathcal{U}(0,1/2)$. Resumindo,

$$\hat{\theta}_h = \frac{1}{n} \sum_{1 \leq i \leq n} (2\pi(1+Y_i^{-2}))^{-1} = \frac{1}{n} \sum_{1 \leq i \leq n} (2\pi(1+V_i^2))^{-1},$$

que é o estimador sugerido no Método 4 do Exemplo 7.3.1.

Para finalizar esta Secção, os casos mais realistas nos quais a aplicação da técnica RVAS poderia ser importante são

1. Na estimação de $\text{IPr}((X,Y) \in A)$ com A dado e (X,Y) normal bivariada. O leitor interessado neste caso pode ver o trabalho de Moran [104].
2. Na geração de certas "leis estáveis". No livro de Durrett [38] (ou em livros avançados de Probabilidade) temos a seguinte

Definição 8.6.1 Diz-se que uma v.a. X tem por distribuição uma lei estável com parâmetros $0 < \alpha \leq 1$, $-1 \leq \beta \leq 1$, $c \in \mathbb{R}$ e $b > 0$ se a função característica de X é definida por

$$C_X(t) = \exp\{itc - b|t|^\alpha(1 + i\beta \operatorname{sign}(t)w_\alpha(t))\},$$

$$\text{onde } w_\alpha(t) = \begin{cases} \tan(\pi\alpha/2) & \text{se } \alpha \neq 1 \\ \frac{2}{\pi} \ln(|t|) & \text{se } \alpha = 1 \end{cases}.$$

Exemplos:

1. Para $\alpha = 2$ e $\beta = 0$ temos $X \sim \mathcal{N}(c, 2b)$.
2. Para $\alpha = 1$, $\beta = 0$, $b = 1$ e $c = 0$ temos $X \sim$ Cauchy padrão.

Destas variáveis aleatórias, em geral, somente é conhecida analiticamente a sua função característica. Porém, podem ser geradas aplicando, por exemplo, um método devido a Chambers e outros [29]. Já que para $\beta = 0$ essas variáveis são simétricas, Lewis e Orav [88] conjecturam que a técnica RVAS poderia ser usada para estimar os percentis dessas variáveis aleatórias (isto é, dado x_0 estimar $\operatorname{IPr}(X \leq x_0)$).

8.7 Amostragem Estratificada

(Ou técnica de redução da variância RVAE) Nesta Secção trataremos o assunto muito brevemente. O leitor interessado em maiores detalhes pode consultar as referências [30, 70, 128].

Consideremos a seguinte situação: sejam $k \geq 1$; $I^k = (0, 1)^k$; $\mathbf{U} = (U_1, \dots, U_k)$ com distribuição uniforme sobre $I^k = (0, 1)^k$; $\psi: I^k \rightarrow \mathbb{R}$ mensurável; $Y = \psi(\mathbf{U})$. Problema: estimar $\theta = \mathbb{E}\{Y\} = \int_{I^k} \psi(\mathbf{u}) d\mathbf{u}$.

Cheng [30] introduziu a idéia de usar uma função intermediária, digamos $\beta: I^k \rightarrow \mathbb{R}$ tal que, se $X = \beta(\mathbf{U})$ então

1. A f.d.a. F_X de X seja conhecida.
2. Y e X estejam correlacionadas.
3. Dado um valor possível, x_0 , de X seja "simples" obter uma amostra da $f_{\mathbf{U}|\beta(\mathbf{U})=x_0}$ (densidade condicional de \mathbf{U} dada $\beta(\mathbf{U}) = x_0$).

A seguir escolhe-se uma certa partição (estratificação) do conjunto \mathcal{X} de todos os possíveis valores de X . Formalmente, seja $L \geq 2$ e sejam S_1, \dots, S_L subconjuntos não vazios e mensuráveis de \mathcal{X} tais que

$$\mathcal{X} = \bigcup_{\ell=1}^L S_\ell, \text{ e } S_\ell \cap S_{\ell'} = \emptyset \text{ se } \ell \neq \ell' \text{ para } \ell = 1, \dots, L.$$

Seja $p_\ell = \mathbb{P}\text{r}(X \in S_\ell)$ para $\ell = 1, \dots, L$. Para cada $\ell = 1, \dots, L$ seja n_ℓ inteiro, escolhido de alguma maneira (veja mais adiante); seja ainda $\mathbb{P}\text{r}_\ell$ a única probabilidade sobre $\mathcal{X} \times S_\ell$ tal que

$$\mathbb{P}\text{r}_\ell(B \cap (\mathcal{X} \times S_\ell)) = \frac{\mathbb{P}\text{r}((X, Y) \in B \cap (\mathcal{X} \times S_\ell))}{\mathbb{P}\text{r}(X \in S_\ell)}$$

para todo $B \subset \mathbb{R}^2$ Boreliano; isto é $\mathbb{P}\text{r}_\ell(B) = \mathbb{P}\text{r}((X, Y) \in B | X \in S_\ell)$ para todo B da forma $(\text{Boreliano de } \mathbb{R}^2) \cap (\mathcal{X} \times S_\ell)$.

Seja agora $(Y_{\ell,1}, X_{\ell,1}), \dots, (Y_{\ell,n_\ell}, X_{\ell,n_\ell})$ uma amostra de tamanho n_ℓ da $\mathbb{P}\text{r}_\ell$. Definimos

$$\bar{Y}_\ell = \frac{1}{n_\ell} \sum_{1 \leq j \leq n_\ell} Y_{\ell,j}.$$

É fácil verificar que $\mathbb{E}_{\mathbb{P}\text{r}_\ell}(\bar{Y}_\ell) = \mathbb{E}\{Y | X \in S_\ell\}$. Seja agora $n = n_1 + \dots + n_L$.

Definição 8.7.1 *Chama-se estimador estratificado de θ a*

$$\hat{\theta}_n = \sum_{1 \leq \ell \leq L} \mathbb{P}\text{r}_\ell \bar{Y}_\ell,$$

onde $\mathbf{n} = (n_1, \dots, n_L)$

É bem simples verificar que $\mathbb{E}_{\mathbb{P}\text{r}}\{\hat{\theta}_n\} = \theta$, daí que $\hat{\theta}_n$ é um estimador não viciado de θ . Usaremos a notação a seguir

$$\mu_{Y,\ell} = \mathbb{E}\{Y | X \in S_\ell\}, \quad \sigma_{Y,\ell}^2 = \mathbb{E}\{Y^2 | X \in S_\ell\} - (\mu_{Y,\ell})^2, \quad \text{e } \bar{\sigma}_Y = \sum_{1 \leq \ell \leq L} p_\ell \sigma_{Y,\ell}.$$

Para a escolha de $\mathbf{n} = (n_1, \dots, n_L)$: pode provar-se que se \mathbf{n} está fixo e

$$n_\ell^* = \frac{n p_\ell \sigma_{Y,\ell}}{\bar{\sigma}_Y} \quad \text{então } \text{Var}(\hat{\theta}_{n^*}) \leq \text{Var}(\hat{\theta}_{\mathbf{n}}), \quad (8.10)$$

onde $\mathbf{n}^* = (n_1^*, \dots, n_L^*)$ e para todo $\mathbf{n} = (n_1, \dots, n_L)$ tal que $n_1 + \dots + n_L = n$.

A fórmula 8.10, em outras palavras, diz que devemos escolher mais observações dos lugares onde a Y apresenta uma variabilidade maior. Os valores n_ℓ^* 's são os correspondentes ao esquema de alocação ótima da amostra (*optimal sampling allocation*) da Teoria da Amostragem.

Note-se que, com amostragem completamente aleatória o valor esperado de n_ℓ seria $n p_\ell$, o que poderia estar bem longe do ideal n_ℓ^* . Ora, na verdade, a fórmula 8.10 pode servir mais como uma guia na escolha do n_ℓ , pois as quantidades $\sigma_{Y,\ell}$ e $\bar{\sigma}_Y$ não são conhecidas. Daí que seja freqüente tomar $n_\ell = n p_\ell$.

Para ver a redução da variância usando $\hat{\theta}_{n^*}$ (ou $\hat{\theta}_{n_\ell}$ com $n_\ell \approx n_\ell^*$) sobre \bar{Y} = estimador da média amostral de θ baseado em uma amostra (Y_1, \dots, Y_n) de Y , é útil saber que

$$\text{Var}(\hat{\theta}_{n^*}) = \frac{1}{n} \left[\text{Var}(Y) - \sum_{1 \leq \ell \leq L} p_\ell \{(\mu_{Y,\ell} - \theta)^2 + (\sigma_{Y,\ell} - \bar{\sigma}_Y)^2\} \right],$$

e que $\text{Var}(\bar{Y}) = n^{-1} \text{Var}(Y)$.

Uma desvantagem desta técnica é que a estratificação mais conveniente depende da ψ , e assim deve mudar para cada ψ ; isto é, embora seja aplicável para estimar média, variância, percentis e/ou quantis da variável de resposta Y , a estratificação deve ser mudada para cada uma dessas quantidades.

Os estimadores da variância de $\hat{\theta}_{n*}$ tem que ser estimados pelo método de *batching* (veja o Apêndice B). O esquema descrito acima é chamado “amostragem pré-estratificada”; seria possível usar também uma “amostragem pós-estratificada”. Veja as referências já indicadas.

8.8 Sugestões sobre Apresentação de um Relatório de uma Experiência (ou Estudo) de Simulação

É claro que a apresentação de um relatório sobre uma pesquisa dependerá do tipo de experiência e do pesquisador. Porém, é conveniente levar em conta um certo formato já em uso habitual, ou algum proposto que se julgue adequado.

No que diz respeito ao relatório de um trabalho de simulação, a idéia principal que deve regir a sua realização é que esse trabalho nada mais é que uma experiência numérica, similar às das Ciências Aplicadas, só que em lugar de *bichinhos*, *parcelas*, etc. são usados números ou conjuntos de números.

Nessa base é que o relatório deve fornecer toda a informação necessária como para que outro experimentador possa repetir a experiência e verificar os resultados informados. É assim imprescindível informar sobre os seguintes itens:

- Computador usado (marca, modelo, etc.)
- Sistema operacional (UNIX, VM, DOS, etc.)
- “Pacotes” computacionais (SAS, S-PLUS, SOC, STATGRAPHICS, etc.)
- Linguagem de programação usada na elaboração dos programas próprios do experimentador (FORTRAN, C, PASCAL, etc.) e o compilador usado (MICROSOFT, LAHEY, TURBO-PASCAL, etc.)
- Subrotinas numéricas usadas nos cálculos dos estimadores ou de outras quantidades, mencionando de onde foram tiradas ou informando sobre o algoritmo implementado, no caso de serem elaboradas pelo mesmo experimentador, ou em caso que sejam pouco conhecidas. Neste ítem é bom informar também sobre a precisão dessas subrotinas.

Já no tocante à simulação, informar sobre:

- Gerador de uniformes: tipo, resultado de testes sobre ele realizados (pelo experimentador ou por outros); se é um dos fornecidos *built-in* em C, PASCAL ou BASIC, tratar de informar-se e transmitir a informação referente à documentação das subrotinas, etc.
- Geradores não uniformes: similar ao já dito para uniformes; principalmente informar sobre o método ou algoritmo usado pela subrotina que implementa a geração.

- Resultados dos testes que **devem** sempre ser realizados antes de usar uma certa sequência numérica pseudo-aleatória (lembra o que dissemos no Capítulo 3, sobre tudo a respeito da dependência da semente de qualquer gerador).

No referente ao relatório sobre os resultados, veremos nesta Secção um formato de apresentação que acompanha o excelente trabalho de Lewis e outros [89]. Mas, como dissemos no início desta Secção, não achamos que esta seja a melhor nem a única forma de apresentar um relatório de simulação estocástica. De qualquer maneira, achamos que é de interesse para o leitor considerá-la.

Para simplificar a exposição, usaremos certos conceitos e notações que passaremos a definir:

Definição 8.8.1 *Seja F uma fda sobre \mathbb{R}^k . Um procedimento estatístico baseado em amostras de tamanho $n \geq n_0$ de F será uma sequência $\tilde{\Theta} = (\Theta_n)_{n \geq n_0}$ de vetores (ou variáveis) aleatórios com valores em \mathbb{R}^p tal que: se para cada $n \geq n_0$, Z_1, \dots, Z_n são iid com distribuição F , então Θ_n é função de Z_1, \dots, Z_n ; digamos $\Theta_n = g_n(Z_1, \dots, Z_n)$.*

Por exemplo, o procedimento de média amostral estará definido por

$$\Theta_n = \frac{1}{n} \sum_{1 \leq i \leq n} Z_i \quad (\text{ou } \Theta_n = \bar{Z}_n).$$

Notação 8.8.1: Seja $\mathbf{a} = (a_1, \dots, a_N)$ um vetor de números reais. Por *analisar a distribuição empírica de \mathbf{a}* entenderemos calcular algumas das seguintes quantidades:

- ★ Média amostral de \mathbf{a} = Média $(\mathbf{a}) = \bar{a} = \frac{1}{N} \sum_{i=1}^N a_i$.
- ★ Variância amostral de \mathbf{a} = $\widehat{\text{Var}}(\mathbf{a}) = S^2(\mathbf{a}) = \frac{1}{N-1} \sum_{i=1}^N (a_i - \bar{a})^2$.
- ★ Desvio padrão de \mathbf{a} = $S(\mathbf{a}) = \sqrt{S^2(\mathbf{a})}$.
- ★ Coeficiente de assimetria amostral de \mathbf{a} = $\widehat{\gamma}_1(\mathbf{a}) = \frac{\widehat{\mu}_3(\mathbf{a})}{S^3(\mathbf{a})}$, onde

$$\widehat{\mu}_3(\mathbf{a}) = \frac{N}{(N-1)(N-2)} \sum_{1 \leq i \leq N} (a_i - \bar{a})^3.$$

- ★ Coeficiente de Kurtosis amostral de \mathbf{a} = $\widehat{\gamma}_2(\mathbf{a}) = \frac{\widehat{\mu}_4(\mathbf{a})}{S^4(\mathbf{a})}$, onde

$$\widehat{\mu}_4(\mathbf{a}) = \frac{N^2 - 2N + 3}{(N-1)(N-2)(N-3)} \sum_{1 \leq i \leq N} (a_i - \bar{a})^4 - \frac{3(S^2(\mathbf{a}))^2(N-1)(2N-3)}{N(N-2)(N-3)}.$$

- ★ Variância da Média Amostral de \mathbf{a} = $S^2(\bar{a}) = \frac{1}{N} S^2(\mathbf{a})$.

Seja agora $\mathbf{a}_{(.)}$ o vetor \mathbf{a} ordenado de menor a maior. Escreveremos $\mathbf{a}_{(.)} = (a_{N:1}, \dots, a_{N:N})$ (isto é $a_{N:1} \leq a_{N:2} \leq \dots \leq a_{N:N}$).

- ★ Mediana amostral de \mathbf{a} = Mediana $(\mathbf{a}) =$

$$Q_2(\mathbf{a}) = \begin{cases} a_{N:[N/2]+1} & \text{se } N \text{ é ímpar} \\ \frac{1}{2}(a_{N:N/2} + a_{N:N/2+1}) & \text{se } N \text{ é par.} \end{cases}$$

★ Quartil inferior amostral de $\mathbf{a} =$

$$Q_1(\mathbf{a}) = \begin{cases} a_{N:(\ell+1)/2} & \text{se } \ell \text{ é ímpar} \\ \frac{1}{2}(a_{N:\ell/2} + a_{N:\ell/2+1}) & \text{se } \ell \text{ é par.} \end{cases} \quad (8.11)$$

★ Quartil superior amostral de $\mathbf{a} =$

$$Q_3(\mathbf{a}) = \begin{cases} a_{N:(N+1-(\ell+1)/2)} & \text{se } \ell \text{ é ímpar} \\ \frac{1}{2}(a_{N:N+1-\ell/2} + a_{N:N-\ell/2+1}) & \text{se } \ell \text{ é par.} \end{cases} \quad (8.12)$$

Nas fórmulas 8.11 e 8.12:

$$\ell = \begin{cases} \frac{N-1}{2} & \text{se } N \text{ é ímpar} \\ \frac{N}{2} & \text{se } N \text{ é par.} \end{cases}$$

Seja $0 < \alpha < 1$, então definimos o

★ α -Quantil amostral de $\mathbf{a} = \hat{a}_\alpha =$

$$\begin{cases} a_{N:\alpha N} & \text{se } \alpha N \text{ é inteiro} \\ a_{N:[\alpha N]+1} & \text{se } \alpha N \text{ não é inteiro.} \end{cases}$$

★ Coeficiente de variação amostral de $\mathbf{a} = \tilde{C}(\mathbf{a}) = \frac{S(\mathbf{a})}{\bar{a}}$.

Uma síntese razoavelmente completa das propriedades mais relevantes das quantidades definidas quando assume-se que \mathbf{a} é um vetor aleatório com a_1, \dots, a_n variáveis independentes, está apresentada em [88].

Em todas as situações que consideraremos nesta Secção, suporemos que não estamos aplicando nenhuma técnica de redução da variância. Mas os exercícios propostos e as Secções sobre esses assuntos podem servir como guia para adaptar o esquema sugerido. Trataremos apenas duas situações:

Situação 1: Analisar o comportamento de um procedimento estatístico baseado em amostras de tamanho $n \geq n_0$ de uma fda, digamos $\hat{\Theta} = (\Theta)_{n \geq n_0}$, visando estimar uma certa característica θ da F . Por exemplo, $\theta = \mathbb{E}_F\{X\}$ ou $\theta = F^{-}(\alpha)$, etc. Para tornar mais simples as coisas, suporemos que F é uma fda sobre \mathbb{R} e que Θ_n toma valores em \mathbb{R} para todo $n \geq n_0$ (isto é, $k = p = 1$ na Definição 8.8.1).

Situação 2: Comparar dois ou mais procedimentos estatísticos que visam um mesmo objetivo. Por exemplo: comparar os procedimentos estatísticos “média amostral”, “mediana amostral”, “ M -estimador robusto”, etc. para estimar a posição de uma variável aleatória simétrica. Outro exemplo já desenvolvido em Secções anteriores: comparação entre t -Student aparada para testes de hipóteses.

Situação 1 - Caso 1: Seja $n \geq n_0$ fixo. Queremos analisar o comportamento de $\Theta_n = g_n(Z_1, \dots, Z_n)$. Para tanto, fazemos a seguinte experiência de simulação: sejam N inteiro e $M = [N/n]$:

1. Atribuir $m = 1$.
2. Atribuir $a(m) = (m - 1)n$.
3. Gerar $z_{a(m)+1}, \dots, z_{a(m)+n}$ amostra de Z_1, \dots, Z_n .
4. Calcular $\theta_n(m) = g_n(z_{a(m)+1}, \dots, z_{a(m)+n})$.
5. Se $m = M$, continuar em 6. Se não, fazer $m = m + 1$ e voltar ao 2.
6. Seja $\theta_n(M) = (\theta_n(1), \dots, \theta_n(M))$. Analisar a distribuição empírica de $\theta_n(M)$. Neste ponto é importante fazer uso de técnicas gráficas: "Box-Plots", "Q-Q Plots", histogramas, etc. para tirar conclusões sobre qual poderia ser a distribuição da variável aleatória Θ_n da que $(\theta_n(1), \dots, \theta_n(M))$ é uma amostra de tamanho M . Também para construir intervalos de confiança para θ , característica da F de interesse final.
7. Dispor os resultados obtidos em uma tabela como a seguinte, incluindo (eventualmente) gráficos.
 - Distribuição gerada: F .
 - Tamanho amostral de F gerada em total: N .
 - Tamanho das amostras de F nas que baseia-se o procedimento a estudar: n .
 - Número de "observações do procedimento": $M = \lfloor N/n \rfloor$.
 - Quantidades amostrais da distribuição empírica do procedimento:
 - Média amostral: $\overline{\theta_n(M)}$.
 - Variância amostral: $S^2(\theta_n(M))$.
 - Variância da Média amostral: $S^2(\theta_n(M))/M$.
 - :

Logo serão incluídos os comentários sobre os resultados das tabelas e/ou gráficos.

8. Se for conveniente aplicar a técnica de "particionamento" (*batching*, veja o Apêndice B), por exemplo porque a $\theta_n(M) = (\theta_n(1), \dots, \theta_n(M))$ não parece uma amostra de tamanho M de uma normal, ou porque estamos interessados em estimar um quantil de Θ_n a partir de uma amostra de tamanho M de Θ_n (veja o Apêndice A), continuamos a análise com as seguintes etapas: seja v o número de partições ($12 \leq v \leq 20$ segundo aconselha a experiência do pessoal da área de simulação). Seja $\ell = \lfloor M/v \rfloor$ o tamanho de cada partição:

8.1) Para cada $j = 1, \dots, v$:

- $a(j) - (j - 1)\ell$.
- $\theta_{n,j}(\ell) = (\theta_n(a(j) + 1), \dots, \theta_n(a(j) + \ell))$.
- Analisar a distribuição empírica de $\theta_{n,j}(\ell)$. Dispor os resultados em uma tabela como a Tabela 8.3 abaixo.

8.2) Analisar as distribuições empíricas de

$$\overline{\theta_{n,\cdot}(\ell)} = (\overline{\theta_{n,1}(\ell)}, \dots, \overline{\theta_{n,v}(\ell)}) \quad (8.13)$$

$$S^2(\theta_{n,\cdot}(\ell)) = (S^2(\theta_{n,1}(\ell)), \dots, S^2(\theta_{n,v}(\ell))) \quad (8.14)$$

$$\dots \quad \dots \quad (8.15)$$

Dispor os resultados em uma tabela como a Tabela 8.4 abaixo. Concluir com comentários sobre os resultados mostrados nas tabelas e/ou gráficos.

Partição	Médias Amostras	Variâncias Amostras	...
1	$\overline{\theta_{n,1}(\ell)}$	$S^2(\theta_{n,1}(\ell))$	\vdots
2	$\overline{\theta_{n,2}(\ell)}$	$S^2(\theta_{n,2}(\ell))$	\vdots
\vdots	\vdots	\vdots	\vdots
v	$\overline{\theta_{n,v}(\ell)}$	$S^2(\theta_{n,v}(\ell))$	\vdots
Sem Particionar	$\overline{\theta_n(M)}$	$S^2(\theta_n(M))$	\vdots

Tabela 8.3: Estimadores de Parâmetros Amostras por Partição

Parâmetro	Médias Amostras	Variâncias Amostras	...
Médias $\overline{\theta_{n,\cdot}(\ell)}$	$(1) = \frac{1}{v} \sum_{i=1}^v \overline{\theta_{n,i}(\ell)}$	$\frac{1}{v-1} \sum_{i=1}^v (\overline{\theta_{n,i}(\ell)} - (1))^2$...
Variâncias $S^2(\theta_{n,\cdot}(\ell))$	$(3) = \frac{1}{v} \sum_{j=1}^v S^2(\theta_{n,j}(\ell))$	$\frac{1}{v-1} \sum_{j=1}^v S^2(\theta_{n,j}(\ell) - (3))^2$...
Kurtosis $\widehat{\gamma}_2(\theta_{n,\cdot}(\ell))$
...

Tabela 8.4: Estimadores de Parâmetros Amostras

Situação 1 - Caso 2: Seja $I \geq 2$ (e não mais de 8); sejam $n_0 \leq n(1) < \dots < n(I)$. Queremos analisar o comportamento de $(\Theta_{n(1)}, \dots, \Theta_{n(I)})$ visando ter uma idéia de como varia o procedimento quando aumenta-se o tamanho das amostras de F nas quais se baseia.

Para simplificar, vamos supor que estejamos interessados em estudar o comportamento das funções

$$n \mapsto \mathbb{E}\{\Theta_n\} \text{ e} \quad (8.16)$$

$$n \mapsto \text{Var}\Theta_n, \quad (8.17)$$

extrapolando, se for possível, as funções

$$i \mapsto E\{\Theta_{n(i)}\} \quad i = 1, \dots, I \text{ e} \quad (8.18)$$

$$i \mapsto \text{Var}(\Theta_{n(i)}), \quad i = 1, \dots, I, \quad (8.19)$$

a partir de estimadores baseados em amostras obtidas por simulação das $(\Theta_{n(1)}, \dots, \Theta_{n(I)})$.

A forma conveniente de realizar o estudo Monte Carlo nesse caso é a seguinte: seja N inteiro, para cada $i = 1, \dots, I$ seja $M(i) = [N/n(i)]$.

Et. Seja $i = 1$.

E2. Seja $m = 1$.

E3. $a(m) = (m - 1)n(i)$.

E4. Gerar $z_{a(m)+1}, \dots, z_{a(m)+n(i)}$ amostra de $Z_1, \dots, Z_{n(i)}$

E5. Calcular $\theta_{n(i)} = g_{n(i)}(z_{a(m)+1}, \dots, z_{a(m)+n(i)})$.

E6. Se $m = M(i)$ então continuar em E7. Se não fazer $m = m + 1$ e voltar a E3.

E7. Se $i = I$ continuar em E8. Se não fazer $i = i + 1$ e voltar a E2.

E8. Para cada $i = 1, \dots, I$ seja $\theta_{n(i)} = \theta_{n(i)}(1), \dots, \theta_{n(i)}(M(i))$. Analisar a distribuição empírica de $\theta_{n(i)}$

E9. Dispor os resultados em uma tabela como a 8.5:

- Distribuição gerada: F .
- Tamanho amostral de F gerada em total para cada $i = 1, \dots, I$: N .
- Tamanho das amostras de F nas que baseia-se o procedimento: n_1, \dots, n_I .
- Número de observações do procedimento para cada valor de $i = 1, \dots, I$: $M(1), \dots, M(I)$. Notemos que

$$N \approx n(1)M(1) \approx n(2)M(2) \approx \dots \approx n(I)M(I). \quad (8.20)$$

n	Médias	Variâncias	Desvios Padrão	Desvios Padrão das Médias	Kurtosis	...
1	$\bar{\theta}_{n(1)}$	$S^2(\bar{\theta}_{n(1)})$	$S^2(\bar{\theta}_{n(1)})/\sqrt{n(1)}$	
2	$\bar{\theta}_{n(2)}$	$S^2(\bar{\theta}_{n(2)})$	$S^2(\bar{\theta}_{n(2)})/\sqrt{n(2)}$	
...
I	$\bar{\theta}_{n(I)}$	$S^2(\bar{\theta}_{n(I)})$	$S^2(\bar{\theta}_{n(I)})/\sqrt{n(I)}$			

Tabela 8.5: Estimadores de Parâmetros Amostrais por Tamanho de Sub-amostra

Sob certas hipóteses, (veja, por exemplo [84]) valem as seguintes expansões assintóticas:

$$\mathbb{E}\{\Theta_n\} = \theta + \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \dots \quad (8.21)$$

$$\text{Var}(\Theta_n) = \frac{b_1}{n} + \frac{b_2}{n^2} + \frac{b_3}{n^3} + \dots; \quad (8.22)$$

nesse caso poderíamos continuar o nosso estudo analisando o seguinte modelo de regressão:

$$Y_i = \theta + a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3} + \varepsilon_i, \quad i = 1, \dots, I, \quad (8.23)$$

onde $Y_i = \bar{\theta}_{n(i)}$ = estimador de $\mathbb{E}\{\Theta\}_{n(i)}$ "média amostral" sobre amostras de tamanho $M(i)$ de $\Theta_{n(i)}$; $x_{i1} = 1/n(1)$, $x_{i2} = 1/n(i)^2$, $x_{i3} = 1/n(i)^3$, $i = 1, \dots, I$.

Ora, como $\text{Var}(Y_i) = \frac{1}{M(i)} \text{Var}(\Theta_{n(i)})$ temos por 8.22, quando $n(i)$ é grande, que

$$\text{Var}(Y_i) \approx \frac{b_1}{n(i)M(i)} \approx \frac{b_1}{N} \text{ para todo } i = 1, \dots, I,$$

por 8.20. Daí que $\text{Var}(Y_1) \approx \text{Var}(Y_2) \approx \dots \approx \text{Var}(Y_I)$. Além disso, pelo esquema de geração sugerido, podemos considerar que Y_1, \dots, Y_I são independentes (observe que, para cada $i = 1, \dots, I$ a geração não volta ao início, mas continua).

Podemos então aplicar uma análise de regressão no modelo 8.23 considerando os valores $\bar{\theta}_{n(1)}, \dots, \bar{\theta}_{n(I)}$ (primeira coluna da Tabela anterior) como os valores observados de Y_1, \dots, Y_I .

Obteremos assim os seguintes estimadores dos coeficientes de 8.23 (por exemplo, por mínimos quadrados):

$$\hat{\theta}_{\text{reg}}, \hat{a}_{1,\text{reg}}, \hat{a}_{2,\text{reg}}, \hat{a}_{3,\text{reg}}. \quad (8.24)$$

Para estimar a precisão dos estimadores de 8.24 poderíamos **continuar** a simulação repetindo todo o esquema E1, E2, ..., E8 obtendo, digamos R valores como 8.24:

$$\begin{aligned} &\hat{\theta}_{\text{reg}}(1), \dots, \hat{\theta}_{\text{reg}}(R) \\ &\hat{a}_{1,\text{reg}}(1), \dots, \hat{a}_{1,\text{reg}}(R) \\ &\vdots \end{aligned} \quad (8.25)$$

Então, analisariamos as distribuições empíricas de

$$\begin{aligned} &(\hat{\theta}_{\text{reg}}(1), \dots, \hat{\theta}_{\text{reg}}(R)) \\ &(\hat{a}_{1,\text{reg}}(1), \dots, \hat{a}_{1,\text{reg}}(R)), \text{ etc.}, \end{aligned}$$

e daí obteríamos estimadores da precisão dos estimadores por mínimos quadrados dos coeficientes do modelo 8.23. Lembremos que o nosso principal interesse será, na maioria dos casos, o θ mas também a_1 poderia sé-lo pois indica o vício assintótico do procedimento.

Observemos que, depois de fazer essas R repetições disporemos, para analisar a variável aleatória $\Theta_{n(i)}$, de $RM(i)$ amostras: $(\theta_{n(i)}(1), \dots, \theta_{n(i)}(M(i)), \dots, \theta_{n(i)}(RM(i)))$.

Finalmente, seja no Caso 1 ou no Caso 2, poderíamos também estudar o comportamento do procedimento $\hat{\Theta}$ sob outras distribuições além da F ; digamos G , H , etc. Nesse caso, teremos que repetir a experiência completa gerando amostras da G , da H , etc. As tabelas a serem apresentadas poderiam ser do tipo já descrito, uma para cada distribuição, ou também dividindo as colunas dessas tabelas em sub-divisões para alocar em cada sub-divisão os resultados correspondentes a cada distribuição.

Sobre a Situação 2, é pouco o que veremos aqui. Recomendamos ao leitor rever neste ponto os exemplos e exercícios ao longo deste trabalho, visando a comparação entre diversos procedimentos estatísticos. Por exemplo: Exercício 7.3 (continuado no Exercício 8.12); Exemplo 8.3.1 e Exercício 8.5; Exemplo 8.4.1 e Exercício 8.6. Já um estudo de proporções gigantescas de comparação de estimadores pode ser visto no excelente trabalho [11]. Uma boa apresentação de trabalhos de simulação também é [93].

8.9 Exercícios

Exercício 8.1: Sejam Z e Z^* variáveis não necessariamente independentes tais que $\text{IPr}(Z = 1) = \text{IPr}(Z^* = 1) = p$ e $\text{IPr}(Z = 0) = \text{IPr}(Z^* = 0) = 1 - p$ com $0 < p < 1$. Prove que (a)

Correlação $(Z, Z^*) \geq -\frac{p}{1-p}$; (b) vale a igualdade em (a) se $\text{IPr}(Z+Z^* = 1) = 2p$ e $\text{IPr}(Z+Z^* = 0) = 1 - 2p$.

Exercício 8.2: Analisar a aplicação da RVAV a cada um dos métodos do Exemplo 7.3.1.

Exercício 8.3: (a) Faça uma lista de funções de distribuição F para as quais pareça simples definir $X = F^{-1}(U)$ e $X^a = F(1 - U)$ com $U \sim \mathcal{U}(0, 1)$. (b) Nos casos em que a relação entre X e X^a não seja trivial, faça um gráfico (X, X^a) por dois métodos: (i) gerando n pares (x_i, x_i^a) , $i = 1, \dots, n$, (ii) usando uma grade fixa de valores (x_i, x_i^a) .

Exercício 8.4: Dedução das condições para redução da variância na técnica RVCV, no caso $\sigma^2(C) \neq \sigma^2(Y)$: (a) prove que $\sigma^2(\bar{C} - \mu_C) = 2\text{Cov}(\bar{Y}, \bar{C} - \mu_C) < 0 \iff \rho_{C,Y} > \frac{\sigma(C)}{2\sigma(Y)}$; (b) tente dar um exemplo de C e de Y tais que $\sigma^2(\bar{Y}') > \sigma^2(\bar{Y})$.

Exercício 8.5: Fazer a experiência de simulação sugerida no Exemplo 8.3.1 com $n = 30$ e $M = 20$.

Exercício 8.6: Fazer a experiência de simulação sugerida no Exemplo 8.4.1, Método 3, com $n = 30$, $K = 20$ e $R = 100$ (logo $M = KR = 2000$).

Exercício 8.7: Método da Partição das amostras para estimação do β na técnica RVCVR: sejam $m = 2L$, $(Y_1, C_1), \dots, (Y_M, C_M)$ vetores aleatórios iid com distribuição (Y, C) . Sejam

$$\begin{aligned}\hat{\sigma}_1(Y) &= \sqrt{\frac{1}{L-1} \sum_{1 \leq \ell \leq L} (Y_\ell - \bar{Y}_1)^2}; & \bar{Y}_1 &= \frac{1}{L} \sum_{1 \leq \ell \leq L} Y_\ell \\ \hat{\sigma}_1(C) &= \sqrt{\frac{1}{L-1} \sum_{1 \leq \ell \leq L} (C_\ell - \bar{C}_1)^2}; & \bar{C}_1 &= \frac{1}{L} \sum_{1 \leq \ell \leq L} C_\ell \\ \rho_{1,CY} &= \frac{\frac{1}{L} \sum_{1 \leq \ell \leq L} (C_\ell - \bar{C}_1)(Y_\ell - \bar{Y}_1)}{\hat{\sigma}_1(Y)\hat{\sigma}_1(C)}; \\ \hat{\beta}_1 &= -\rho_{1,CY} \frac{\hat{\sigma}_1(Y)}{\hat{\sigma}_1(C)}; \\ \hat{\sigma}_2(Y) &= \sqrt{\frac{1}{L-1} \sum_{L+1 \leq \ell \leq M} (Y_\ell - \bar{Y}_2)^2}; & \bar{Y}_2 &= \frac{1}{L} \sum_{L+1 \leq \ell \leq M} Y_\ell \\ \hat{\sigma}_2(C) &= \sqrt{\frac{1}{L-1} \sum_{L+1 \leq \ell \leq M} (C_\ell - \bar{C}_2)^2}; & \bar{C}_2 &= \frac{1}{L} \sum_{L+1 \leq \ell \leq M} C_\ell \\ \rho_{2,CY} &= \frac{\frac{1}{L} \sum_{L+1 \leq \ell \leq M} (C_\ell - \bar{C}_2)(Y_\ell - \bar{Y}_2)}{\hat{\sigma}_2(Y)\hat{\sigma}_2(C)}; \\ \hat{\beta}_2 &= -\rho_{2,CY} \frac{\hat{\sigma}_2(Y)}{\hat{\sigma}_2(C)}; \\ \bar{Y}_1^* &= \bar{Y}_1 + \beta_2(\bar{C}_1 - \mathbb{E}\{\bar{C}_1\}) \\ \bar{Y}_2^* &= \bar{Y}_2 + \beta_1(\bar{C}_2 - \mathbb{E}\{\bar{C}_2\})\end{aligned}$$

$$\overline{Y^*} = \frac{\overline{Y_1^*} + \overline{Y_2^*}}{2}.$$

(a) Por que $\mathbb{E}\{\overline{Y^*}\} = \mathbb{E}\{\overline{Y}\}$. (b) Qual é $\text{Var}(\overline{Y^*})$? Compará-la com $\text{Var}(\overline{Y})$, onde $\overline{Y} = (\overline{Y_1} + \overline{Y_2})/2 = M^{-1} \sum_{m=1}^M Y_m$.

Exercício 8.8: Aplicação da técnica RVCVR com controle ajustado por regressão polinômica aos Métodos 3 e 4 do Exemplo 7.3.1. Ripley ([110, pág. 125]), depois de uma experiência de simulação prévia para encontrar os coeficientes do ajuste por regressão polinômica:

$$Y_3^* = \phi_3(X) + \beta_1(X^2 - \mathbb{E}\{X^2\}) + \beta_2(X^4 - \mathbb{E}\{X^4\}),$$

com $\phi_3(t) = 2f(t) = \frac{2}{\pi(1+t^2)}$ e $X \sim \mathcal{U}(0,1)$, e

$$Y_4^* = \phi_4(X) + \gamma_1(X^2 - \mathbb{E}\{X^2\}) + \gamma_2(X^4 - \mathbb{E}\{X^4\}),$$

com $\phi_4(t) = f(t)/2 = \frac{1}{2\pi(1+t^2)}$ e $X \sim \mathcal{U}(0,1)$, sugere usar como estimadores da quantidade $\theta = \int_2^{+\infty} \frac{1}{(1+x^2)} dx = \mathbb{P}(\text{Cauchy} > 2)$:

$$Y_3^* = \frac{1}{2} - \left[\phi_3(X) - 0.15(X^2 - \frac{4}{3}) - 0.025(X^4 - \frac{16}{5}) \right],$$

$$Y_4^* = \phi_4(X) + 0.312(X^2 - \frac{1}{12}) - 0.233(X^4 - \frac{1}{80}).$$

Prove que

$$\text{Var}(\overline{Y_3^*}) \approx 6.3 \frac{10^{-4}}{n} \text{ e } \text{Var}(\overline{Y_4^*}) \approx 1.1 \frac{10^{-9}}{n}$$

trabalhando com amostras de tamanho n das X 's correspondentes. Lembre que, no Exemplo 7.3.1 tínhamos obtido

$$\text{Var}(T_3) \approx \frac{0.028}{n} \text{ quando } T_3 = \frac{1}{2} - \frac{1}{n} \sum_{1 \leq i \leq n} \phi_3(X_i) \text{ e}$$

$$\text{Var}(T_4) \approx 9.3 \frac{10^{-5}}{n} \text{ quando } T_4 = \frac{1}{n} \sum_{1 \leq i \leq n} \phi_4(X_i).$$

Exercício 8.9: (Novamente a estatística $T_{n,1}$) Sugerir e implementar um método para estimar os quantis da $T_{n,1}$ usando como controle $\overline{T}_{n,1}$ (a média amostral de várias observações da $T_{n,1}$).

Exercício 8.10: (A estatística $T_{n,1}$ mais uma vez) Seja K par; implementar e discutir um método para estimar o $(K-k)$ -ésimo quantil da distribuição $T_{n,1} = t_{n,1}(X_{n,1}, \dots, X_{n,n})$ (com $X_{n,1}, \dots, X_{n,n}$ i.i.d. $\mathcal{N}(0,1)$) usando como controle a soma desse mesmo $(K-k)$ -ésimo quantil com o k -ésimo quantil da mesma $T_{n,1}$ (os estimadores desses quantis estão, obviamente, baseados em “amostras” de tamanho K da $T_{n,1}$).

Exercício 8.11: Uma aplicação de técnica RVAC.

(a) Sejam V e W independentes com densidades ($0 < k < 1$):

$$f_V(x) = \frac{1}{\Gamma(k)\Gamma(1-k)} x^{k-1} (1-x)^{(1-k)-1} \mathbb{I}_{(0,1)}(x) \text{ e } f_W(x) = \beta \exp\{-\beta x\} \mathbb{I}_{(0,+\infty)}(x),$$

isto é $V \sim \text{Beta}(k, 1-k)$ e $W \sim \mathcal{E}(\beta)$. Provar que $X = VW$ tem distribuição $\text{Gamma}(\beta, k)$, isto é X tem densidade $f_X(x) = \frac{\beta^k}{\Gamma(k)} x^{k-1} \exp\{-\beta x\} \mathbb{I}_{(0,+\infty)}(x)$. Sejam agora $n \geq 2$ e X_1, \dots, X_n iid com distribuição $\mathcal{G}(\beta, k)$, e $M_n = \min\{X_1, \dots, X_n\}$. Problema: estimar $\theta = \mathbb{E}\{M_n\}$

Método 1: (Monte Carlo à força bruta) Seja M inteiro

1. Para cada $m = 1, \dots, M$:

(a) Gerar $X_{m,1}, \dots, X_{m,n}$ amostra da $\mathcal{G}(\beta, k)$.

(b) Calcular $X_{(1)}(m) = \min\{X_{m,1}, \dots, X_{m,n}\}$.

2. Calcular $\bar{\theta} = M^{-1} \sum_{m=1}^M x_{(1)}(m)$ e $S^2(\bar{\theta}) = (M(M-1))^{-1} \sum_{m=1}^M (X_{(1)}(m) - \bar{\theta})^2$.

Assim, $\bar{\theta}$ é um valor observado do estimador Monte Carlo à força bruta de θ , e $S^2(\bar{\theta})$ é um estimador da sua variância. Ora, este método pode acelerar-se muito levando em conta o resultado a seguir, que o leitor pode provar:

(b) Sejam Y_1, \dots, Y_n variáveis com distribuição comum F contínua e estritamente crescente sobre $\{y: F(y) > 0\}$; sejam $0 < p < 1$, $Y_{(1)} = \min\{Y_1, \dots, Y_n\}$ e $y = F_{Y_{(1)}}^{-1}(p)$. Então $y = F^{-1}(1 - (1-p)^{1/n})$. Daí que $Y_{(1)}$ tem a mesma distribuição que a v.a. $F^{-1}(1 - V^{1/n})$, com $V \sim \mathcal{U}(0,1)$.

Método 1 modificado:

1. Para cada $m = 1, \dots, M$:

(a) Gerar v_m da $U \sim \mathcal{U}(0,1)$.

(b) Calcular $x_{(1)}(m) = F_X^{-1}(1 - v^{1/n})$.

2. Como no Método 1.

A desvantagem deste método é que F_X^{-1} , neste caso, é difícil de calcular (ver, por exemplo [20])

Método 2: (Usando RVAC) Pela parte (a) $X_i = V_i W_i$ com $V_1, \dots, V_n, W_1, \dots, W_n$ independentes; V_1, \dots, V_n igualmente distribuídas com f_V e W_1, \dots, W_n igualmente distribuídas com f_W . Seja M inteiro:

1. Para cada $m = 1, \dots, M$:

(a) Gerar $b_{m,1}, \dots, b_{m,n}$ amostra de tamanho n da $\text{Beta}(k, 1-k)$, isto é, f_V .

(b) Calcular

$$h(b_{m,1}, \dots, b_{m,n}) = \mathbb{E}\{M_n | V_1 = b_{m,1}, \dots, V_n = b_{m,n}\} = \left(\beta \sum_{i=1}^n \frac{1}{b_{m,i}} \right)^{-1}$$

(por que?)

2. Calcular

$$\begin{aligned} \theta^* &= \frac{1}{M} \sum_{1 \leq m \leq M} h(b_{m,1}, \dots, b_{m,n}) \text{ e} \\ S^2(\theta^*) &= \frac{1}{(M(M-1))} \sum_{1 \leq m \leq M} (h(b_{m,1}, \dots, b_{m,n}) - \theta^*)^2. \end{aligned}$$

Assim, θ^* é um “valor observado” do estimador de θ usando RVAC, e $S^2(\theta^*)$ é um estimador da variância desse estimador.

(c) Faça uma experiência Monte Carlo comparando os Métodos 1 e 2 para $n = 5, 10, 15, 20, 25, 30, 35$ e 40 . Mais precisamente, sejam $N = 2000$, $n(1) = 5$, $n(2) = 10$, $n(3) = 15$, $n(4) = 20$, $n(5) = 25$, $n(6) = 30$, $n(7) = 35$ e $n(8) = 40$. Sejam $\beta = 1$ e $k = 0.96$; para cada $I = 1, \dots, 8$ sejam $A(I) = 2000(I - 1)$ e $M(I) = [2000/n(I)]$, e para cada $m = 1, \dots, M(I)$ seja $b(I, m) = (m - 1)n(I)$.

Simulação 1: (Método 1)

1. Gerar uma amostra de tamanho $8N = 16000$, digamos X_1, \dots, X_{16000} da distribuição $\mathcal{G}(1, 0.96)$.
2. Para cada $I = 1, \dots, 8$:
 - (a) Para cada $m = 1, \dots, M(I)$ calcular

$$x_{(1)}(I, m) = \min\{x_{a(I)+b(I,m)+1}, \dots, x_{a(I)+b(I,m)+n(I)}\}.$$

- (b) Calcular $\bar{\theta}(I) = (M(I))^{-1} \sum_{m=1}^{M(I)} x_{(1)}(I, m)$ e $S^2(\bar{\theta}(I)) = (M(I)(M(I) - 1))^{-1} \sum_{m=1}^{M(I)} (x_{(1)}(I, m) - \bar{\theta}(I))^2$.

Simulação 2: (Método 2)

1. Gerar uma amostra de tamanho $8N = 16000$, digamos X_1, \dots, X_{16000} da distribuição $\mathcal{G}(1, 0.96)$.
2. Para cada $I = 1, \dots, 8$:
 - (a) Para cada $m = 1, \dots, M(I)$ calcular $h(I, m) = (\sum_{i=1}^{n(I)} (b_{a(I)+b(I,m)+1})^{-1})^{-1}$
 - (b) Calcular $\theta^*(I) = (M(I))^{-1} \sum_{m=1}^{M(I)} h(I, m)$ e $S^2(\theta^*(I)) = (M(I)(M(I) - 1))^{-1} \sum_{m=1}^{M(I)} (h(I, m) - \theta^*(I))^2$

Acabadas as duas simulações, mostrar os resultados construindo uma tabela como a 8.6. Nesta tabela temos sugerido multiplicar os estimadores de $\mathbb{E}\{M_n\}$ por n pois $\mathbb{E}\{nM_n\} = 1$

$n(i)$	$n(I)\bar{\theta}(I)$	$n(I)S(\bar{\theta}(I))$	$n(I)\theta^*(I)$	$n(I)S^2(\theta^*(I))$	VR
5
10
15
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
40

Tabela 8.6: Exercício 8.11

quando $\beta = k = 1$ (distribuição $\mathcal{E}(1)$). A última coluna da tabela, VR , é para comparar as

variâncias e assim medir uma percentagem da redução obtida usando o Método 1 sobre o Método 2. Esses valores definem-se como

$$VR(I) = \frac{|S^2(\bar{\theta}(I)) - S^2(\theta^*(I))|}{S^2(\bar{\theta}(I))}.$$

Exercício 8.12: Continuação do Exercício 7.3: O verdadeiro valor aproximado de $\sigma(\hat{\rho}_{14})$ é, pelo visto na Secção 7.1.2, dado por

$$\sigma(\hat{\rho}_{14}) \approx \frac{1 - \rho^2}{\sqrt{14 - 3}} = \frac{1 - (0.5)^2}{\sqrt{11}} \approx 0.226. \quad (8.26)$$

Tendo realizado o Exercício 7.3 teremos dois arquivos, digamos NORMAL e NONORMAL, cada um com os valores ($n = 14$)

$$\hat{\rho}_n(r), \quad \hat{\sigma}(\hat{\rho}_n)(r), \quad \hat{\sigma}_B(\hat{\rho}_n)(r), \quad r = 1, \dots, 1000,$$

o NORMAL com F de a) do Exercício 7.3 (normal bivariada) e NAONORMAL obtido com F de b) do mesmo Exercício.

1. (Apenas para controle da experiência) Analisar, para ambos os arquivos, as distribuições empíricas de $(\hat{\rho}_n(1), \dots, \hat{\rho}_n(1000))$. Dispor os resultados segundo o indicado na Secção 8.8 ou na forma que o leitor considerar melhor.
2. Para ambos os arquivos, analisar as distribuições empíricas de

$$(\hat{\sigma}(\hat{\rho}_n)(1), \dots, \hat{\sigma}(\hat{\rho}_n)(1000)) \text{ e } (\hat{\sigma}_B(\hat{\rho}_n)(1), \dots, \hat{\sigma}_B(\hat{\rho}_n)(1000)),$$

e dispor os resultados numa tabela como a 8.7: onde $F(a)$ e $F(b)$ são as distribuições

Parâmetros Amostrais	$\hat{\sigma}(\hat{\rho}_n)$		$\hat{\sigma}_B(\hat{\rho}_n)$	
	$F(a)$	$F(b)$	$F(a)$	$F(b)$
Médias
Variância
Desvio Padrão da Média
Assimetria
Kurtosis
⋮	⋮	⋮	⋮	⋮

Tabela 8.7: Exercício 8.12

de a) e b), respectivamente, do Exercício 7.3 Os resultados mostrados na linha das “Médias” têm que ser comparados com o 0.226 de 8.26.

Exercício 8.13: Continuação do Exercício 7.4. A partir da parte b) do Exercício 7.4 é fácil provar que

$$\text{Var}(T) = \frac{\mu^2}{n-2}. \quad (8.27)$$

Em [26] propõe-se comparar os intervalos de confiança construídos no Exercício 7.4 com os que resultam da 8.27 aplicando aproximação normal, isto é comparar para $n = 10$ e 100 os intervalos

$$\left[\hat{\mu}(r) - z_{(1-\alpha/2)} \frac{\hat{\mu}(r)}{\sqrt{n-2}}, \hat{\mu}(r) + z_{(1-\alpha/2)} \frac{\hat{\mu}(r)}{\sqrt{n-2}} \right], \quad (8.28)$$

onde $z_{(1-\alpha/2)}$ satisfaz $\text{IPr}(N \leq z_{(1-\alpha/2)}) = 1 - \alpha/2$, com $N \sim \mathcal{N}(0, 1)$; e os intervalos

$$\left(\frac{\hat{\mu}(r)^2}{U(r)}, \frac{\hat{\mu}(r)^2}{L(r)} \right), \quad (8.29)$$

para $r = 1, \dots, 30000/n$ e $\alpha = 0.05$.

1. Contar o número de vezes em que o verdadeiro valor ($\mu = 2$) é coberto em cada caso.
2. Para cada $n = 10$ e 100 e cada $r = 1, \dots, 30000/n$ definir

$$d_{2,n}(r) = 2z_{(1-\alpha/2)} \frac{\hat{\mu}(r)}{\sqrt{n-2}} = \text{comprimento dos intervalos definidos em 8.28}$$

$$d_{3,n}(r) = \hat{\mu}(r)^2 \left(\frac{1}{L(r)} - \frac{1}{U(r)} \right) = \text{comprimento dos intervalos definidos em 8.29.}$$

Analisar as distribuições empíricas de

$$d_{2,n} = (d_{2,n}(1), \dots, d_{2,n}(30000/n)) \text{ e } d_{3,n} = (d_{3,n}(1), \dots, d_{3,n}(30000/n)).$$

3. Dispor os resultados de 1) e 2) numa tabela como a 8.8, tendo como guia (talvez o sugerido na Secção 8.8; onde 8.28 indica o método dos intervalos definidos em 8.28, 8.29 indica o método dos intervalos definidos em 8.29, e “Proporção de Cobrimentos” refere-se às quantidades calculadas na parte 1)

Exercício 8.14: Técnica de Particionamento para vários procedimentos amostrais sobre amostras da $\text{Gamma}(\alpha, \beta)$. Seja f a densidade da $\mathcal{G}(\alpha, \beta)$, isto é

$$f(x) = (\Gamma(\alpha)\beta^\alpha)^{-1} x^{\alpha-1} \exp\{-x/\beta\} \mathbb{I}_{(0,+\infty)}(x),$$

com $\alpha > 0$ e $\beta > 0$. Seja Z com densidade f . Pode ver-se que

$$\theta_1 = \mathbb{E}\{Z\} = \alpha\beta$$

$$\theta_2 = \sigma(Z) = \sqrt{\alpha\beta}$$

$$\theta_3 = \text{coef. de variação de } Z = \frac{\sigma(Z)}{\mathbb{E}\{Z\}} = \frac{1}{\sqrt{\alpha}}$$

$$\theta_4 = \text{coef. de assimetria de } Z = \gamma_1(Z) = \frac{(\mathbb{E}\{Z - \mathbb{E}\{Z\}\})^3}{(\sigma(Z))^3} = \frac{2}{\sqrt{\alpha}}$$

$$\theta_5 = \text{coef. de kurtosis de } Z = \gamma_2(Z) = \frac{(\mathbb{E}\{Z - \mathbb{E}\{Z\}\})^4}{(\sigma(Z))^4} - 3 = \frac{6}{\alpha}. \quad (8.30)$$

Parâmetros Amostrais	n = 10		n = 100	
	8.28	8.29	8.28	8.29
Médias
Variâncias
Desvios Padrão
Desvios Padrão da Média
Assimetria
Kurtosis
⋮	⋮	⋮	⋮	⋮
Proporção de Cobrimento

Tabela 8.8: Exercício 8.13

O objetivo deste Exercício é aplicar a técnica de particionamento vista no Apêndice B aos seguintes procedimentos:

$${}_k\tilde{\Theta} = ({}_k\Theta_n)_{n \geq 2} \text{ para estimar } \theta_k; \quad k = 1, 2, 3, 4, 5$$

definidas para cada n por: seja $\mathbf{Z} = (Z_1, \dots, Z_n)$ variad com densidade f ,

$${}_1\Theta_n = \text{média amostral de } \mathbf{Z} = \frac{1}{n} \sum_{i \leq i \leq n} Z_i$$

$${}_2\Theta_n = \text{desvio padrão amostral de } \mathbf{Z} = \sqrt{\frac{1}{n-1} \sum_{1 \leq i \leq n} (Z_i - {}_1\Theta_n)^2}$$

$${}_3\Theta_n = \frac{{}_2\Theta_n}{{}_1\Theta_n}$$

$${}_4\Theta_n = \text{coef. de assimetria amostral de } \mathbf{Z}$$

$${}_5\Theta_n = \text{coef. de kurtosis amostral de } \mathbf{Z}$$

Para as duas últimas definições veja a Notação 8.8.1.

Faça a seguinte experiência Monte Carlo: sejam $M = 1000$, $v = 20$ (logo $\ell = M/v = 50$), $\alpha = 0.2$ e $\beta = 1$

1. Gerar z_1, \dots, z_M da $\mathcal{G}(\alpha, \beta)$.
2. Para cada $j = 1, \dots, 20$ sejam

$$\begin{aligned} a(j) &= (j-1)\ell = (j-1)50 \\ \mathbf{z}_j(\ell) &= (z_{a(j)+1}, \dots, z_{a(j)+\ell}). \end{aligned}$$

Calcular

$${}_1\theta_{M,j}(\ell) = \text{média amostral de } \mathbf{z}_j(\ell)$$

$$\begin{aligned}
{}_2\theta_{M,j}(\ell) &= \text{desvio padrão amostral de } z_j(\ell) \\
{}_3\theta_{M,j}(\ell) &= \frac{{}_2\theta_{M,j}(\ell)}{{}_1\theta_{M,j}(\ell)} \\
{}_4\theta_{M,j}(\ell) &= \text{coef. de assimetria amostral de } z_j(\ell) \\
{}_5\theta_{M,j}(\ell) &= \text{coef. de kurtosis amostral de } z_j(\ell)
\end{aligned}$$

3. Também calcular ${}_k\theta_M$ para $k = 1, 2, 3, 4, 5$ que são as correspondentes à amostra completa $\mathbf{z} = (z_1, \dots, z_M)$.
4. Dispor os resultados obtidos numa tabela como a 8.9

j	${}_1\theta_{M,j}(\ell)$	${}_2\theta_{M,j}(\ell)$	${}_3\theta_{M,j}(\ell)$	${}_4\theta_{M,j}(\ell)$	${}_5\theta_{M,j}(\ell)$	${}_2\theta_{M,j}(\ell)v^{-1/2}$
1
2
\vdots
20
\mathbf{z}	${}_1\theta_M$	${}_2\theta_M$	${}_3\theta_M$	${}_4\theta_M$	${}_5\theta_M$	$\frac{{}_2\theta_M}{\sqrt{1000}}$

Tabela 8.9: Exercício 8.14

Nota: A coluna ${}_2\theta_{M,j}(\ell)v^{-1/2}$, que é obtida dividindo cada elemento da ${}_2\theta_{M,j}(\ell)$ da linha correspondente por $\sqrt{20}$, é um estimador do desvio padrão do ${}_1\theta_{M,j}(\ell)$. Também, $\frac{{}_2\theta_M}{\sqrt{1000}}$ é um estimador do desvio padrão de ${}_1\theta_M = (1000)^{-1} \sum_{m=1}^{1000} z_m$.

5. Para cada $k = 1, \dots, 5$ calcular

$$\begin{aligned}
{}_k\theta_{M,\bullet}(\ell) &= \frac{1}{20} \sum_{1 \leq j \leq 20} {}_k\theta_{M,j}(\ell) \\
S({}_k\theta_{M,\bullet}(\ell)) &= \sqrt{\frac{1}{20} \frac{1}{19} \sum_{1 \leq j \leq 20} ({}_k\theta_{M,j}(\ell) - {}_k\theta_{M,\bullet}(\ell))^2}.
\end{aligned}$$

6. Comparar os intervalos de confiança para $\theta_1 (= 0.2)$ com nível de confiança $1 - \gamma = 0.95$ dados por

$$\left[{}_1\theta_M \pm z_{1-\gamma/2} \frac{{}_2\theta_M}{\sqrt{1000}} \right], \text{ e } \left[{}_1\theta_{M,\bullet}(\ell) \pm t_{19,1-\gamma/2} S({}_1\theta_{M,\bullet}(\ell)) \right].$$

É o verdadeiro valor de θ_1 coberto por eles?

7. Para $k = 2, 3, 4$ e 5 o θ_k é coberto pelo intervalo

$$\left[{}_k\theta_{M,\bullet}(\ell) \pm t_{19,1-\gamma/2} S({}_k\theta_{M,\bullet}(\ell)) \right]?$$

Seria interessante repetir a experiência com outro v , por exemplo $v = 10$.

Apêndice A

Estimação dos Quantis

Sejam $F: \mathbb{R} \rightarrow [0, 1]$ uma fda; X uma va com fda igual à F . Dado $0 < \alpha < 1$ chama-se “ α -quantil de F ” a $x_\alpha = F^{-1}(\alpha) = \inf\{x: F(x) \geq \alpha\}$. Evidentemente, se conhecêssemos x_α para todo α , então conheceríamos toda a função F ; no entanto, do ponto de vista prático, o estatístico interessa-se apenas por certos valores de α — $\alpha = 0.5$ (mediana), $\alpha = 0.25$ (quartil inferior), $\alpha = 0.75$ (quartil superior), $\alpha = 0.025$ e $\alpha = 0.975$ (para formar intervalos de confiança para $\mathbb{E}_F\{X\}$), etc. Alguns desses valores, e outros mais, ajudam a formar-se uma idéia da posição, dispersão e forma da distribuição F .

Neste Apêndice veremos muito brevemente o mais essencial (e simples) do problema de: dado α estimar x_α . Seja $K \geq 2$ inteiro e sejam X_1, \dots, X_K valid com distribuição comum F .

O estimador natural de x_α baseado em X_1, \dots, X_K é

$$\hat{x}_{\alpha, K} = \begin{cases} X_{K:\alpha K} & \text{se } \alpha K \text{ é inteiro} \\ X_{K:[\alpha K]+1} & \text{se não,} \end{cases} \quad (\text{A.1})$$

onde $(X_{K:1}, \dots, X_{K:K})$ é o vetor “estatísticas de ordem” do vetor (X_1, \dots, X_K) .

Pode ver-se que para K pequeno o estimador $\hat{x}_{\alpha, K}$ pode ser muito ruim, especialmente para α 's perto de 0 ou de 1. Isto não seria tão problemático no caso de simulações onde o K está sob controle do experimentador (lembramos que um estudo de simulação é uma experiência numérica). Mas é problema para dados obtidos em experiências das Ciências Aplicadas, onde o tamanho das amostras freqüentemente não é *suficientemente grande*. Daí que têm sido sugeridos estimadores alternativos ao definido na equação A.1. O leitor interessado pode consultar a referência [103].

Como já dissemos, certos quantis têm recebido uma consideração especial na Estatística, e tem-se sugerido outros estimadores alternativos ao A.1 (veja, por exemplo, [105]): estimador do 0.50-quantil (mediana)

$$\widehat{\text{Med}}_K = \begin{cases} X_{K:(K+1)/2} & \text{se } K \text{ é ímpar} \\ \frac{1}{2}(X_{K:K/2} + X_{K:K/2+1}) & \text{se não.} \end{cases}$$

Estimador do 0.25-quantil (quartil inferior)

$$\widehat{\text{Q1}}_K = \begin{cases} X_{K:(\ell+1)/2} & \text{se } \ell \text{ é ímpar} \\ \frac{1}{2}(X_{K:\ell/2} + X_{K:\ell/2+1}) & \text{se não,} \end{cases}$$

$$\text{onde } \ell = \begin{cases} (K-1)/2 & \text{se } K \text{ é ímpar} \\ K/2 & \text{se não} \end{cases}.$$

Estimador do 0.75-quantil (quartil superior)

$$\widehat{Q}_{3K} = \begin{cases} X_{K:K=1-(\ell+1)/2} & \text{se } \ell \text{ é ímpar} \\ \frac{1}{2}(X_{K:K+1-\ell/2} + X_{K:K-\ell/2+1}) & \text{se não,} \end{cases}$$

onde ℓ é como acima.

A.0.1 Algumas Propriedades da Distribuição do Estimador \hat{x}_α

Do ponto de vista teórico, a distribuição do \hat{x}_α não apresenta problema nenhum, desde que é uma estatística de ordem. Mas na prática, a sua distribuição depende fortemente da F e do K . Particularmente, se $\hat{x}_{\alpha,K} = X_{\alpha,K}$ para algum $1 \leq k \leq K$, em [103] pode ver-se que

$$\text{Pr}(\hat{X}_{\alpha,K} \leq x) = F_{\hat{x}_\alpha}(x) = \sum_{k \leq j \leq K} \binom{K}{j} (F(x))^j (1 - F(x))^{K-j},$$

que pode ser computacionalmente horrível.

Intervalos de confiança quando K é pequeno, não paramétricos, podem ver-se em [33,88]. Se K é grande (o que é habitual nos trabalhos de simulação), os intervalos não paramétricos referidos acima podem ser usados, mas o cálculo deles pode ser muito pesado. Daí que é melhor usar a teoria assintótica para contruir intervalos de confiança aproximados. Essa teoria pode ser vista em vários textos de Estatística Matemática, por exemplo no já citado [103].

Serão de interesse os seguintes resultados, válidos para F estritamente crescente com densidade $f(x) > 0$:

$$\begin{aligned} \lim_{K \rightarrow \infty} E_F\{\hat{x}_{\alpha,K}\} &= x_\alpha \\ \lim_{K \rightarrow \infty} \text{Var}_F(\hat{x}_{\alpha,K}) &= \lim_{K \rightarrow \infty} \frac{\alpha(1-\alpha)}{K(f(x_\alpha))^2} \\ \frac{\hat{x}_{\alpha,K} - x_\alpha}{\frac{\sqrt{\alpha(1-\alpha)}}{\sqrt{K}f(x_\alpha)}} &\xrightarrow{D} \mathcal{N}(0,1) \quad \text{quando } K \rightarrow \infty. \end{aligned} \quad (\text{A.2})$$

Infelizmente, o K pode ser muito grande para aplicar as aproximações sugeridas por essas fórmulas. Contudo, levando em conta a última fórmula, podemos tomar como intervalo de confiança aproximado para x_α de nível $100(1 - \gamma/2)\%$ ao intervalo

$$\left[\hat{x}_{\alpha,K} - \sqrt{\frac{\alpha(1-\alpha)}{K\hat{f}^2(\hat{x}_{\alpha,K})}} z_{1-\gamma/2}, \hat{x}_{\alpha,K} + \sqrt{\frac{\alpha(1-\alpha)}{K\hat{f}^2(\hat{x}_{\alpha,K})}} z_{1-\gamma/2} \right] \quad \text{onde}$$

$$\text{Pr}(Z \leq z_{1-\gamma/2}) = 1 - \frac{\gamma}{2}, \quad Z \sim \mathcal{N}(0,1),$$

e \hat{f} é um estimador da densidade F .

Ora, esse intervalo pode ser muito ruim caso K seja pequeno ou caso α esteja perto de 0 ou de 1, ou caso \hat{f} não seja um estimador estável de F . Daí que talvez seja útil usar outros intervalos de confiança, por exemplo os sugeridos em [33].

Da fórmula A.2 vemos que para ter uma precisão razoável com o estimador $\hat{x}_{\alpha,K}$ precisaremos de K bem grande. Isto, juto com o fato de que se α está ou perto de 1 ou de 0, pode fazer com que o processo de calcular um estimador da variância do $\hat{x}_{\alpha,K}$ possa levar muito tempo de computador. Existem formas de estimar os quantis que são mais eficientes e realistas que a aplicação da equação A.2. O leitor interessado no tema pode ver [59] para uma resenha sobre técnicas em uso e propostas, e [111] para uma certa técnica um tanto sofisticada.

Apêndice B

Técnica de Particionamento para Cálculo de Variâncias

Consideremos a seguinte situação: seja F uma fda sobre \mathbb{R}^k ; θ uma certa característica da F (por exemplo $\theta = \mathbb{E}_F\{g(X)\}$, com $g: \mathbb{R}^k \rightarrow \mathbb{R}^p$); Z um vetor aleatório com distribuição F ; e seja $\tilde{\Theta} = (\Theta_n)_{n \geq n_0}$ um procedimento estatístico baseado em amostras de tamanho $n \geq n_0$ de F , considerado como um método para estimar θ . Isto é (Definição 8.11): para cada $n \geq n_0$ existe uma função mensurável g_n definida sobre um subconjunto mensurável de $(\mathbb{R}^k)^n$ com valores em \mathbb{R}^p tal que $\Theta_n = g_n(Z_1, \dots, Z_n)$, onde Z_1, \dots, Z_n são i.i.d com distribuição F . Problema: dado $M \geq n_0$ estimar $\text{Var}(\Theta_M)$.

Exemplo B.1: Suponhamos que $k = p = 1$, que $M > n_0$, que $\theta = \mathbb{E}_F\{Z\}$, e que $\tilde{\Theta}$ é o procedimento “média amostral”; isto é

$$\Theta_M = \frac{1}{M} \sum_{1 \leq m \leq M} Z_m = \bar{Z}_M.$$

Como já sabemos, $\text{Var}(\Theta_M) = M^{-1} \text{Var}_F(Z)$.

Em geral, $\text{Var}_F(\Theta)$ não será conhecida e poderia ser estimada por

$$S_M^2(Z) = \frac{1}{M-1} \sum_{1 \leq m \leq M} (Z_m - \bar{Z}_M)^2.$$

Daí que um estimador da $\text{Var}_F(\Theta_M)$ será $S^2(\Theta_M) = S^2(\bar{Z}_M) = M^{-1} S_M^2(Z)$, e um estimador do desvio padrão de Θ_M será $S(\Theta_M) = S(\bar{Z}_M) = M^{-1/2} S_M(Z)$.

Poderíamos usar esse valor para ter uma idéia da precisão de \bar{Z}_M e construir um intervalo de confiança para θ de nível $1 - \alpha$, digamos (L, U) , definido por

$$L = \bar{Z}_M - t_{M-1, 1-\alpha/2} S(\bar{Z}_M) \text{ e } U = \bar{Z}_M + t_{M-1, 1-\alpha/2} S(\bar{Z}_M),$$

onde $\text{Pr}(T \leq t_{M-1, 1-\alpha/2}) = 1 - \alpha/2$ com $T \sim t$ -Student com $M - 1$ graus de liberdade.

Mas isso funciona se Z_1, \dots, Z_M é uma amostra de uma $\mathcal{N}(\theta, \sigma^2)$ ou, então, “muito perto” dela. Claro, poderíamos usar o Teorema Central do Limite e pensar que

$$\frac{\bar{Z}_M - \theta}{S(\bar{Z}_M)} \simeq \mathcal{N}(0, 1), \quad (\text{B.1})$$

entendendo “ \simeq ” como “distribuída aproximadamente como”, e daí deduzir o intervalo de confiança para θ de nível $1 - \alpha$ definido por (L, U) com

$$L = \bar{Z}_M - z_{1-\alpha/2} S(\bar{Z}_M) \text{ e } U = \bar{Z}_M + z_{1-\alpha/2} S(\bar{Z}_M),$$

onde $\text{IPr}(N \leq z_{1-\alpha/2}) = 1 - \alpha/2$ com $N \sim \mathcal{N}(0, 1)$.

Ora, em geral $\Theta_M = g_M(Z_1, \dots, Z_M)$ não será uma média amostral como \bar{Z}_M e aí como estimar $\text{Var}(\Theta_M)$? Na Secção 7.1.2 já vimos uma forma *bootstrap*; outra é a chamada *jackknifing* (veja [39]). Neste Apêndice veremos uma outra técnica, chamada *sectioning*, *batching* e que traduziremos por *particionamento*, que é particularmente conveniente e direta para o problema colocado acima quando estamos analisando o comportamento de Θ_M usando simulação.

A presente técnica consiste em substituir Θ_M como estimador de θ por um outro estimador Θ_M^* de θ , também baseado em amostras de tamanho M de Z do qual possamos estimar $\text{Var}(\Theta_M^*)$. Se for possível provar (ou se for razoável esperar) que $\text{Var}(\Theta_M) \approx \text{Var}(\Theta_M^*)$, então poderíamos ficar satisfeitos.

A idéia básica está fundamentada no Exemplo anterior. Isto é, suponhamos que Θ_M^* pode ser escrito como

$$\Theta_M^* = \frac{1}{v} \sum_{1 \leq \ell \leq v} Y_\ell^*,$$

onde Y_1^*, \dots, Y_v^* são *valid*. Então um estimador da variância de Θ_M^* será

$$S^2(\Theta_M^*) = \frac{1}{v} S^2(Y_1^*) \text{ com } S^2(Y_1^*) = \frac{1}{v-1} \sum_{1 \leq \ell \leq v} (Y_\ell^* - \Theta_M^*)^2$$

(que é um estimador da $\text{Var}(Y_1^*)$).

Ora, dada a amostra Z_1, \dots, Z_M como poderíamos estimar θ , usando toda ela, por médio de um estimador da forma B.1? Vejamos. Sejam v e ℓ inteiros convenientemente escolhidos. Para cada $j = 1, \dots, v$ sejam

$$a(j) = (j-1)\ell \text{ e } \Theta_{M,j}(\ell) = g_\ell(Z_{a(j)+1}, \dots, Z_{a(j)+\ell}).$$

Agora observemos que $\Theta_{M,1}(\ell), \dots, \Theta_{M,v}(\ell)$ na verdade formam uma amostra de tamanho v de Θ_ℓ . Daí que

$$\Theta_{M,\bullet}(\ell) = \frac{1}{v} \sum_{1 \leq j \leq v} \Theta_{M,j}(\ell)$$

será também um estimador de θ que chamaremos “estimador de θ baseado no procedimento $\tilde{\Theta}$ obtido particionando amostras de tamanho M em subconjuntos de tamanho ℓ ”.

Pelo observado recentemente, temos que um estimador da $\text{Var}(\Theta_\ell)$ será

$$S^2(\Theta_\ell) = \frac{1}{v-1} \sum_{1 \leq j \leq v} (\Theta_{M,j}(\ell) - \Theta_{M,\bullet}(\ell))^2,$$

e daí que um estimador da $\text{Var}(\Theta_{M,\bullet}(\ell))$ será

$$S^2(\Theta_{M,\bullet}(\ell)) = \frac{1}{v} S^2(\Theta_\ell).$$

É necessário destacar que, em geral, teremos $\Theta_M \neq \Theta_{M,\bullet}(\ell)$ mas, em certas circunstâncias, pode esperar-se que $\text{Var}(\Theta_M) \approx \text{Var}(\Theta_{M,\bullet}(\ell))$. Por exemplo, se $\Theta_M = \bar{Z}_M$; quer dizer, o procedimento é a “média amostral”, então $\Theta_M = \Theta_{M,\bullet}(\ell)$, pois nesse caso

$$\Theta_M = \frac{1}{M} \sum_{1 \leq j \leq M} Z_j \text{ e } \Theta_{M,\bullet}(\ell) = \frac{1}{v\ell} \sum_{1 \leq j \leq v} \sum_{1 \leq i \leq \ell} Z_{a(j)+i},$$

mas $S(\bar{Z}_M) \neq S(\Theta_{M,\bullet}(\ell))$.

Mais geralmente, sob certas condições, temos que o procedimento $\tilde{\Theta} = (\Theta_n)_{n \geq n_0}$ satisfaz

$$\sqrt{n}(\Theta_n - \theta) \xrightarrow{D} \mathcal{N}(0, \sigma^2) \text{ quando } n \rightarrow \infty. \quad (\text{B.2})$$

Daí que, se ℓ é “grande”, (e assim M também o é) temos

$$\sqrt{\ell}(\Theta_\ell - \theta) \simeq \mathcal{N}(0, \sigma^2) \quad (\text{B.3})$$

(lembre B.1 e $\sqrt{M}(\Theta_M - \theta) \simeq \mathcal{N}(0, \sigma^2)$), donde deduz-se que

$$\text{Var}(\sqrt{\ell}\Theta_\ell) \approx \sigma^2 \approx \text{Var}(\sqrt{M}\Theta_M),$$

então

$$\text{Var}(\Theta_M) \approx \frac{\ell}{M} \approx \text{Var}(\Theta_\ell) \frac{1}{v} \text{Var}(\Theta_\ell) \approx \frac{1}{v} S^2(\Theta_\ell) = S^2(\Theta_{M,\bullet}(\ell)).$$

Logo, sob B.2 escolhendo ℓ “grande” (logo v “pequeno”) poderíamos esperar que B.3 seja válida.

Mas também poderíamos verificar isso analisando a distribuição empírica do

$$(\Theta_{M,1}(\ell), \dots, \Theta_{M,v}(\ell)) \quad (\text{B.4})$$

obtido de uma realização de (Z_1, \dots, Z_M) . Para fazer essa verificação é necessário que v não seja pequeno demais.

Conclusão: se não podemos rejeitar que a realização obtida do vetor B.4 é uma amostra de tamanho v de uma normal, então podemos tomar como intervalo de confiança para θ a (L, U) com

$$L = \Theta_{M,\bullet}(\ell) - t_{v-1, 1-\alpha/2} S(\Theta_{M,\bullet}(\ell)) \text{ e } U = \Theta_{M,\bullet}(\ell) + t_{v-1, 1-\alpha/2} S(\Theta_{M,\bullet}(\ell)).$$

Uma questão fica por ser esclarecida: qual é o v conveniente? Não consideraremos as respostas para essa questão; o leitor interessado pode vê-las em [88, Capítulo 9]. Esses autores sugerem $12 \leq v \leq 20$; assim, se $M \geq 1000$ então as chances de que $\Theta_{M,1}(\ell), \dots, \Theta_{M,v}(\ell)$ podam ser consideradas como amostras normais são boas. Também, nesse caso, $\sigma(\Theta_M) \approx S(\Theta_{M,\bullet}(\ell))$.

Exemplo B.2: Estimação dos quantis da estatística t -Student aparada (Exemplo 8.4.1 e Exercício 8.6). O método 3 sem uso da variável de controle consiste da análise do seguinte procedimento estatístico: dado n fixo, estimar o α -quantil amostral sobre amostras de tamanho K .

A notação usada no Exemplo 8.4.1 em termos da usada neste Apêndice é: Z = variável t -Student aparada com $n-1$ graus de liberdade; $\theta = F_Z^{-1}(\alpha)$; $v = R$; $\ell = K$ (logo $M = KR$). As variáveis Z_1, \dots, Z_M com distribuição igual à Z nesse Método tomam valores $z(1), z(2), \dots, z(M)$ dados por

$$\begin{aligned} z(1) &= t_{n,1}(1,1), \dots, z(K) = t_{n,1}(K,1), \\ z(K+1) &= t_{n,1}(1,2), \dots, z(2K) = t_{n,1}(K,2), \dots, \\ z((K-1)R+1) &= t_{n,1}(1,R), \dots, z(KR) = t_{n,1}(K,R). \end{aligned}$$

Também

$$\Theta_{M,j}(\ell) = \Theta_{M,j}(K) = t_{n,1}(K : k_0, j) \text{ onde } k_0 = \begin{cases} \alpha K & \text{se } \alpha K \text{ inteiro} \\ [\alpha K] & \text{se não,} \end{cases}$$

para $j = 1, \dots, v$ ($v = R$). Finalmente,

$$\begin{aligned} \Theta_{M,\bullet}(\ell) &= \overline{t_{n,1}(K : k_0)} \quad (\ell = K) \\ S^2(\Theta_\ell) &= \widehat{\sigma^2}(Y(k_0)) \text{ e} \\ S^2(\Theta_{M,\bullet}(\ell)) &= \widehat{\sigma^2}(\bar{Y}(k_0)) = \frac{1}{R} \widehat{\sigma^2}(Y(k_0)). \end{aligned}$$

Para um outro exemplo desta técnica, o leitor é convidado a fazer o Exercício 8.14.

Bibliografia

- [1] *Library IMSL*. IMSL Inc., Houston, 1984.
- [2] *A Million Random Digits with 100.000 Normal Deviates*. Rand Corporation, Glenco, USA, 1955.
- [3] *The NAG FORTRAN Library Manual, Mark-14*. NAG, Oxford, 1990.
- [4] *SAS User's Guide*. SAS Institute, Inc., Cary, New York, 1982.
- [5] *SPSSX Statistical Algorithms*. SPSS Inc., Chicago, 1983.
- [6] *Stat. Library IMSL*. IMSL Inc., Houston, 1987.
- [7] E. Aarts and J. Korst. *Simulated Annealing and Boltzman Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*. Wiley, 1989.
- [8] E. Aarts and P. Laarhoven. Simulated annealing: an introduction. *Statistica Neerlandica*, 43, 1988.
- [9] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1965.
- [10] D. F. Andrews. Contribuição à discussão de A. C. Atkinson and M. C. Pearce (1976), em "The computer generation of beta, gamma and normal random variables". *J. Roy. Statist. Soc. A*, 139:452-453, 1976.
- [11] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers and J. W. Tuckey. *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, New Jersey, 1972.
- [12] A. C. Atkinson. Tests of pseudo-random numbers. *Appl. Statist.*, 29:164-171, 1980.
- [13] A. C. Atkinson and J. Whittaker. A switching algorithm for the generation of beta random variables with at least one parameter less than 1. *J. Roy. Statist. Soc. A*, 139:464-467, 1976.
- [14] R. J. Baker and J. A. Nelder. *GLIM Manual Release 3*. NAG, Oxford, 1978.
- [15] D. L. Banks. Improving the bayesian bootstrap. 1988. Em andamento (citado em [88]).
- [16] V. C. Barbosa. Redes neuronais e "simulated annealing" como ferramentas para otimização combinatória. *Investigación Operativa*, 1:125-142, 1989.
- [17] J. D. Beasley and S. G. Springer. Algorithm AS111: the percentage points of the normal distribution. *Appl. Statist.*, 26:118-121, 1977.

- [18] D. J. Best. Letter to the editor. *Appl. Statist.*, 27:181, 1978.
- [19] D. J. Best. A simple algorithm for the computer generation of random samples from a Student's t or symmetric beta distribution. In L. C. Corsten and J. Hermans, editors, *COMPSTAT 1978: Proc. in Computational Statistics*, pages 341–347, Physica Verlag, Wien, 1978.
- [20] D. J. Best and D. E. Roberts. AS91 the percentage points of the χ^2 distribution. *Appl. Statist. C*, 24:385–388, 1975.
- [21] K. Binder, editor. *Applications of the Monte Carlo Method in Statistical Physics*. Springer-Verlag, Berlin, 2nd. edition, 1987.
- [22] K. Binder, editor. *Monte Carlo Methods in Statistical Physics*. Springer-Verlag, Berlin, 2nd. edition, 1986.
- [23] Z. W. Birnbaum. Computers and unconventional test statistics. In *Reliability and Biometry: Statistical Analysis of Lifelenght*, pages 441–458, SIAM, Philadelphia, 1974.
- [24] G. E. Box, W. G. Hunter and J. S. Hunter. *Statistics for Experimenters*. Wiley, New York, 1978.
- [25] M. Brown and H. Solomon. On combining pseudorandom number generators. *Ann. Statist.*, 7:691–695, 1979.
- [26] S. T. Buckland. Monte Carlo confidence intervals. *Biometrics*, 40:811–817, 1984.
- [27] O. H. Bustos. *A TURBO C Implementation of a GFSR[0,1] Pseudorandom Number Algorithm*. Informes de Matemática B-62, IMPA, Rio de Janeiro, 1990.
- [28] P. Carnevali, L. Coletti and S. Patarnello. Image processing by simulated annealing. *IBM J. Res. Dev.*, 29:569–579, 1985.
- [29] J. M. Chambers, C. L. Mallows and B. W. Stuck. A method for simulation of stable random variables. *J. Amer. Statist. Assoc.*, 71:340–344, 1976.
- [30] R. C. Cheng. Antithetic variate methods for simulation of processes with peaks and troughs. *European J. of Operations Research*, 15:227–236, 1984.
- [31] R. C. Cheng. The generation of gamma variables with non-integral shape parameters. *Appl. Statist.*, 26:71–75, 1977.
- [32] K. L. Chung. *A Course in Probability Theory*. Academic Press, New York, 2nd. edition, 1974.
- [33] W. J. Conover. *Practical Nonparametric Statistics*. Wiley, New York, 2nd. edition, 1980.
- [34] R. R. Coveyou and R. D. MacPherson. Fourier analysis of uniform random number generators. *J. ACM*, 14:100–119, 1967.
- [35] J. Dagpunar. *Principles of Random Variate Generation*. Claredon Press, Oxford, 1988.
- [36] L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, Berlin, 1986.
- [37] P. J. Diggle. *Statistical Analysis of Spatial Point Patterns*. Academic Press, London, 1983.

- [38] R. Durrett. *Probability Theory and Examples*. Wadsworth & Brooks/Cole, Pacific Grove, California, 1991.
- [39] B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia, 1982.
- [40] B. Efron. Nonparametric estimates of standard error: the jackknife, the bootstrap and other resampling methods. *Biometrika*, 68:589-599, 1981.
- [41] B. Efron and G. Gong. A leisurely look at the bootstrap, the jackknife and cross-validation. *American Statistician*, 37:36-48, 1983.
- [42] B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*, 1:54-77, 1986.
- [43] G. Fishman. *Principles of Discrete Event Simulation*. Wiley, 1978.
- [44] G. Fishman and L. Moore. An exhaustive analysis of multiplicative congruential random number generators with modulus $2^{31} - 1$. *SIAM J. of Scientific and Statist. Computing*, 7:24-45, 1985.
- [45] A. C. Frery. Monte Carlo determination of the stopping time of two cluster-flip algorithms in the Ising model. *Revista Brasileira de Física*, 21(4):482-491, 1991.
- [46] J. García. Simulación de sucesiones de uniformes en $(0,1)$ usando micro y mini computadores. Trabajo especial para Lic. en Matemática, 1992.
- [47] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, 1979.
- [48] D. Geman. Random fields and inverse problems in imaging. In P. L. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XVIII-1988*, Springer-Verlag, Berlin, 1988. Lecture Notes in Mathematics 1427.
- [49] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. PAMI*, 6:721-741, 1984.
- [50] J. E. Gentle. Portability considerations for random number generators. In W. F. Eddy, editor, *Computer Science and Statistics: Proc. 13th Symp. on the Interface*, pages 158-164, Springer-Verlag, New York, 1981.
- [51] S. W. Golomb. *Shift Register Sequences*. Holden-Day, San Francisco, 1967.
- [52] T. González, S. Sahni and W. R. Franta. An efficient algorithm for the Kolmogorov-Smirnov and Lilliefors tests. *ACM Trans. Math. Soft.*, 3:60-64, 1977.
- [53] I. J. Good. On the serial test for random sequences. *Ann. Math. Statist.*, 28:262-264, 1957.
- [54] I. J. Good. The serial test for sampling numbers and other tests for randomness. *Proc. Camb.*, 1953.
- [55] B. F. Green Jr., J. E. Smith and L. Klem. Empirical tests of an additive random number generator. *J. ACM*, 6:527-537, 1959.

- [56] J. Gruber and K. Freimann. Combined response surface regressions in Monte Carlo studies of small sample properties of estimators: theory and application. Discussion Paper, 1986.
- [57] F. G. Gustavson and W. Liniger. A fast random number generator with good statistical properties. *Computing*, 6:221–226, 1970.
- [58] J. M. Hammersley and D. C. Handscomb. *Monte Carlo Methods*. Methuen & Co. Ltd., London, 1964.
- [59] P. Heidelberg and P. A. Lewis. Quantile estimation in dependent sequences. *Operations Research*, 32:185–209, 1984.
- [60] P. G. Hoel, S. C. Port and C. J. Stone. *Introdução à Teoria da Probabilidade*. Editora Interciência, Rio de Janeiro, 1978. Tradução de Chiyoski, F. Y.
- [61] A. C. Hope. A simplified Monte Carlo significance test procedure. *J. Roy. Statist. Soc. B*, 30:582–598, 1968.
- [62] T. R. Hopkins. A revised algorithm for the spectral test. *Appl. Statist.*, 328–335, 1983.
- [63] F. Hsuan. Generating uniform polygonal random pairs. *Appl. Statist.*, 28:170–172, 1979.
- [64] D. W. Hutchinson. A new uniform pseudo-random number generator. *Comm. ACM*, 9:432–433, 1966.
- [65] B. R. James. *Probabilidade: um curso em nível intermediário*. IMPA, Rio de Janeiro, 1981. Projeto Euclides.
- [66] K. H. Jöckel. Computational aspects of Monte Carlo tests. In *COMPSTAT 1984 Proc. Physica*, pages 183–188, Springer-Verlag, Viena, 1984.
- [67] M. D. Johnk. Erzeugung von betaverteilten und gammaverteilten zufallszahlen. *Metrika*, 8:5–15, 1964.
- [68] M. E. Johnson. *Multivariate Statistical Generation*. Wiley, New York, 1987.
- [69] N. L. Johnson and S. Kotz. *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley, New York, 1972.
- [70] H. Kahn. Use of different Monte Carlo sampling techniques. In H. A. Meyer, editor, *Symposium on Monte Carlo Methods*, pages 146–190, Wiley, New York, 1956.
- [71] H. Katzen Jr. *APL Programming and Computer Techniques*. Van Nostrand, New York, 1970.
- [72] M. G. Kendall and B. Babington-Smith. Tables of random sampling numbers. *Tracts for Computers*, XXIV, 1939. Cambridge University Press, London.
- [73] A. J. Kinderman and J. F. Monahan. Computer generation of random variables using the ratio of uniform deviates. *ACM Trans. Math. Soft.*, 3:257–260, 1977.
- [74] A. J. Kinderman and J. G. Ramage. *FORTTRAN programs for generating normal random variables*. Technical Report 23, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, P. A., 1977.

- [75] S. Kirkpatrick, C. D. Gelatt Jr. and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671-680, 1983.
- [76] J. P. Kleijnen. *Statistical Techniques in Simulation, Part I*. Dekker, New York, 1974.
- [77] J. P. Kleijnen. *Statistical Techniques in Simulation, Part II*. Dekker, New York, 1975.
- [78] D. E. Knuth. *The Art of Computer Programming. Vol. 2: Seminumerical Algorithms*. Addison Wesley, Reading, Massachussetts, 2nd edition, 1981.
- [79] S. S. Lavenberg, T. L. Moeller and P. D. Welch. Statistical results on control variables with application to queueing network simulation. *Operations Research*, 30:182-202, 1982.
- [80] S. S. Lavenberg and P. D. Welch. A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Management Science*, 27:322-335, 1981.
- [81] A. J. Lawrance and P. A. Lewis. Properties of the bivariate delayed Poisson process. *J. Appl. Prob.*, 12:257-268, 1975.
- [82] G. P. Learmonth. Empirical tests of multipliers for the prime modulus random number generator $x_{i+1} = ax_i \pmod{2^{31} - 1}$. In D. C. Hoaglin and R. E. Welsch, editors, *Proc. Ninth Interface Symp. on Computer Science and Statist.*, pages 178-183, Prindle, Weber & Schmidt, Boston, 1976.
- [83] G. P. Learmonth and P. W. Lewis. Statistical tests of some widely used and recently proposed uniform random number generators. In W. J. Kennedy, editor, *Proc. Seventh Conf. Computer Science and Statist. Interface*, Iowa State University Press, Ames, 1974.
- [84] E. L. Lehmann. *Theory of Point Estimation*. Wiley, New York, 1983.
- [85] D. H. Lehmer. Mathematical methods in large scale computing units. In *Proc. Second Symp. Large Scale Digital Calculating Machinery*, pages 141-146, Harvard University Press, Cambridge, Massachussetts, 1951.
- [86] H. Levene and J. Wolfowitz. The covariance matrix of runs up and down. *Ann. Math. Statist.*, 15:58-69, 1944.
- [87] P. A. Lewis, A. S. Goodman and J. M. Miller. A pseudo-random number generator for the system 360. *IBM Sys. J.*, 8:136-145, 1969.
- [88] P. A. Lewis and E. J. Orav. *Simulation Methodology for Statisticians, Operation Analysts and Engineers*. Volume 1, Wadsworth & Brooks, Pacific Grove, California, 1989. Cole Advanced Books & Software.
- [89] P. A. Lewis, E. J. Orav and L. C. Uribe. *Enhanced Simulation and Statistical Package*. Wadsworth & Books/Cole, Pacific Grove, California, 1988.
- [90] P. A. Lewis and W. H. Payne. Generalized feedback shift register pseudorandom number algorithm. *J. ACM*, 20:456-468, 1973.
- [91] P. A. Lewis and L. C. Uribe. *The New Naval Postgraduate School Random Number Generator Package LLRAN DOMII*. Technical Report PS-55-81-005, Naval Postgraduate School, Monterrey, California, 1981.

- [92] T. G. Lewis and W. H. Payne. Generalized feedback shift register pseudorandom number algorithm. *J. ACM*, 20:456–458, 1973.
- [93] S. Lira Filho. Um estudo comparativo de estimadores adaptáveis do parâmetro de localização. 1991. Dissertação de Mestrado em Estatística, UNICAMP, São Paulo.
- [94] M. D. MacLaren and G. Marsaglia. Uniform random number generators. *J. ACM*, 12:83–89, 1965.
- [95] K. V. Mardia. Statistics of directional data. *J. R. Statist. Soc. B*, 37:349–393, 1975.
- [96] F. H. Marriott. Barnard's Monte Carlo tests: how many simulations? *Appl. Statist.*, 28:75–77, 1979.
- [97] G. Marsaglia. A current view of random number generators. In L. Billiard, editor, *Computer Science and Statistics: Proc. 16th Symp. Interface*, pages 3–10, North-Holland, Amsterdam, 1985.
- [98] G. Marsaglia. Random number generation. In A. Ralston, editor, *Encyclopedia of Computer Science*, Van Nostrans Reinhold, New York, 1976.
- [99] G. Marsaglia. The structure of linear congruential sequences. In S. K. Zaremba, editor, *Applications of Number Theory to Numerical Analysis*, Academic Press, New York, 1972.
- [100] F. Martinelli, E. Olivieri and E. Scoppola. On the Swendsen-Wang dynamics I: exponential convergence to equilibrium. *J. Statist. Phys.*, 62:117–133, 1991.
- [101] F. Martinelli, E. Olivieri and E. Scoppola. On the Swendsen-Wang dynamics II: critical droplets and homogeneous nucleation at low temperature for the 2 dimensional Ising model. *J. Statist. Phys.*, 62:135–159, 1991.
- [102] C. A. Mauro. Efficient identification of important factors in large scale simulations. In J. Wilson, J. Hendriksen and S. Roberts, editors, *Proc. 1986 Winter Simulation Conference*, pages 296–305, Institute of Electrical and Electronics Engineers, Piscataway, N. J., 1986.
- [103] A. M. Mood, F. A. Graybill and D. C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, New York, 1974.
- [104] P. A. Moran. The Monte Carlo evaluation of orthant probabilities for multivariate normal distributions. *Australian J. Statist.*, 26:39–44, 1984.
- [105] F. Mosteller and J. W. Tuckey. *Data Analysis and Regression*. Addison-Wesley, Reading, Massachusetts, 1977.
- [106] B. L. Nelson and B. W. Schmeiser. Decomposition of some well-known variance reduction techniques. *J. Statist. Computation and Simulation*, 23:183–209, 1986.
- [107] E. S. Pearson and H. O. Hartley. *Biometrika Tables for Statisticians*. Volume I, The University Press, Cambridge, 1966.
- [108] W. H. Press, B. P. Flannery, S. A. Teulosky and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 1989.

- [109] B. D. Ripley. Computer generation of random variables: a tutorial. *Int. Statist. Rev.*, 51:301-319, 1983.
- [110] B. D. Ripley. *Stochastic Simulation*. Wiley, New York, 1987.
- [111] D. W. Robinson. *Nonparametric Quantile Estimation Through Stochastic Approximation*. PhD thesis, Naval Postgraduate School, Monterrey, California, 1975.
- [112] M. Rosenblatt. Multiply schemes and shuffling. *Mathematics of Computation*, 29:929-934, 1975.
- [113] Y. Sawada and H. Hyosu. Growth velocity of electrochemical deposition and its concentration dependence. *Physica D*, 38(1-3):299-303, 1989.
- [114] R. H. Schonmann. Finite size scaling behaviour of a biased majority rule cellular automaton. *Physica A*, 167:619-627, 1990.
- [115] L. Schrage. A more portable fortran random number generator. *ACM Trans. Math. Soft.*, 5:132-139, 1979.
- [116] J. Spanier and E. M. Gelbard. *Monte Carlo Principles and Neutron Transport Problems*. Addison-Wesley, 1969.
- [117] D. Sparks. RND is fatally flawed. *CALL APPLE*, 29-32, January 1983.
- [118] H. E. Stanley. Learning concepts of fractals and probability by "Doing Science". *Physica D*, 38(1-3):330-340, 1989.
- [119] R. Swendsen and J. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86-88, 1987.
- [120] J. D. Sylwestrowicz. *Parallel Processing in Statistics*, pages 131-136. Springer-Verlag, Vienna, 1982.
- [121] L. H. Tippett. Random sampling numbers. *Tracts for Computers*, XV, 1927. Cambridge University Press, London.
- [122] J. W. Tuckey and D. M. McLaughlin. Less vulnerable confidence and significance procedures for location based on a single sample: trimming/winsorization 1. *Sankhyā A*, 25:331-352, 1963.
- [123] J. von Neumann. Various techniques in connection with random digits. *NBS Appl. Math. Ser.*, 12:36-38, 1951.
- [124] A. J. Walker. An efficient method for generating discrete random variables with general distributions. *ACM Trans. Math. Soft.*, 3:253-256, 1974.
- [125] A. J. Walker. New fast method for generating discrete random numbers with arbitrary frequency distribution. *Electronic Letters*, 10:127-128, 1974.
- [126] B. A. Whichmann and I. D. Hill. Correction to AS183: an efficient and portable pseudo-random number generator. *Appl. Statist.*, 33:123, 1984.
- [127] B. A. Whichmann and I. D. Hill. An efficient and portable pseudo-random number generator. *Appl. Statist.*, 31:188-190, 1982.

- [128] J. R. Wilson. Variance reduction techniques for digital simulation. *Amer. J. of Math. and Management Sciences*, 4:277-312, 1984.
- [129] J. R. Wilson and A. A. Pritsker. Experimental evaluation of variance reduction techniques for queueing simulation using generalized concomitant variables. *Management Sciences*, 30:1459-1472, 1984.
- [130] U. Wolff. Collective Monte Carlo updating for spin systems. *Phys. Rev. Lett.*, 62:361-364, 1989.

Impresso na Gráfica do



pelo Sistema Xerox/1065

