COMPLEMENTOS DE ESTATÍSTICA

Exercícios

Licenciatura em Matemática Aplicada e Computação

Instituto Superior Técnico
Universidade de Lisboa

Capítulo 1

Simulação Estocástica

Exercício 1.1 Para $\alpha > 0$, a função gama é definida por:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\,\mathrm{d}t.$$

Mostre que:

(a) $\Gamma(1) = 1$ e $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$.

(b) $\Gamma(n) = (n-1)!$ com $n \in I\!N_0$.

(c) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Exercício 1.2
Seja $X \sim U(a,b)$ e $Y = -\frac{1}{\lambda}\ln(\frac{X-a}{b-a})$, onde $Y \sim Exp(\lambda)$ e $\lambda > 0$. Mostre que $Y \sim Exp(\lambda)$.

Exercício 1.3 Considere $X \sim$W(α,λ), com $\alpha > 0$ e $\lambda > 0$.

(a) Mostre que $E[X] = \frac{1}{\lambda}\Gamma(\frac{1}{\alpha}+1)$ e $VAR[X] = \frac{1}{\lambda^2}\left[\Gamma(\frac{2}{\alpha}+1) - \Gamma^2(\frac{1}{\alpha}+1)\right]$.

(b) Mostre que $T = X^\alpha$ tem distribuição exponencial com parâmetro λ^α.

Exercício 1.4 Considere um v.a. $T \sim$Gama(α,λ), com $\alpha > 0$ e $\lambda > 0$.

(a) Mostre que $E[T] = \frac{\alpha}{\lambda}$ e $VAR[T] = \frac{\alpha}{\lambda^2}$.

(b) Verifique que $X \sim$Exp$(\lambda) \Leftrightarrow X \sim$Gama$(1,\lambda)$ e que $Z \sim \chi^2_{(n)} \Leftrightarrow Z \sim$Gama$(\frac{n}{2},\frac{1}{2})$.

(c) Admita que $\alpha = n \in I\!N$, isto é $X \sim$Gama(n,λ). Mostre que $T = 2\lambda X \sim \chi^2_{(2n)}$.

Exercício 1.5 Admita que a variável aleatória T segue uma distribuição Rayleigh com parâmetro de escala $\sigma > 0$.

(a) Sabendo que $P(T > 2) = 0.99$, identifique a distribuição de T e relacione a distribuição de T com a distribuição $W(\alpha,\lambda)$.

(b) Obtenha $E(T)$.

Exercício 1.6 Considere a variável aleatória $X \sim$N(μ,σ^2), deduza a distribuição de $T = e^X$.

Exercício 1.7 Admita que o tempo, em minutos, para a reparação de uma máquina é a variável aleatória $T \sim$LN$(4.2, 1)$.

(a) Calcule a probabilidade de ser necessário esperar pelo menos 30 minutos para reparar a máquina.

(b) Obtenha o tempo mediano de reparação da máquina e o 1^{o}quartil da distribuição de T.

Exercício 1.8 Prove que se U é uma v.a. $Unif(0, 1)$ e $\lambda > 0$, então $X = -\frac{1}{\lambda} \ln U$ tem distribuição $Exp(\lambda)$. A transformação descrita pode ser usada na geração de números pseudo-aleatórios da distribuição exponencial.

Exercício 1.9 Considere $X \sim U(0, 1)$. Deduza a distribuição de $Y = 1-X$ e calcule $Cov\,(X, Y)$.

Exercício 1.10 Suponha que $X_1, \ldots, X_n$ são v.a.'s independentes. Usando funções geradoras de momentos, determine a distribuição de $Y = \sum_{i=1}^{n} X_i$ no caso em que para $i = 1, \ldots, n$:

(a) $X_i \sim Exp(\lambda)$

(b) $X_i \sim Gama(\alpha_i, \lambda)$.

(c) $X \sim \mathcal{N}(\mu_i, \sigma_i^2)$.

(d) $X_i \sim Poi(\lambda_i)$.

(e) $X_i \sim Geom(p)$.

Exercício 1.11 Suponha que $X \sim \mathcal{N}(0, 1)$. Deduza a distribuição de $Y = e^X$.

Exercício 1.12 A transformação seguinte (onde U_1 e U_2 são variáveis uniformes em $(0, 1)$ e independentes) é muito utilizada na prática:

$$X_1 = (-2 \ln U_1)^{1/2} \cos{(2\pi U_2)}$$
$$X_2 = (-2 \ln U_1)^{1/2} \operatorname{sen}{(2\pi U_2)}\,,$$

onde $X_i \sim \mathcal{N}\,(0, 1)$, com $X_1 \amalg X_2$. Usando o R, com a semente em geração fixa, gere uma amostra de dimensão 1000 para cada variável aleatória $X_i, i = 1, 2$. Com base no pacote ggplot2 do R, construa dois gráficos da função de distribuição empírica associada a cada amostra. Em cada gráfico apresente (com cor diferente) a função de distribuição da $\mathcal{N}\,(0, 1)$.

Exercício 1.13 Sejam X_i v.a.'s i.i.d. com $X_i \sim Exp(\lambda)$, $i = 1 \ldots, n$ e seja $Y = \sum_{i=1}^{n} X_i$. Suponha que tem disponível um simulador de números pseudo-aleatórios uniformemente distribuídos em $(0, 1)$.

(a) Usando uma função geradora conveniente deduza a distribuição de Y assim como o seu valor esperado e variância.

(b) Diga como proceder para simular números pseudo-aleatórios da distribuição de X_i.

(c) Diga como proceder para simular números pseudo-aleatórios da distribuição de Y.

Exercício 1.14 Sejam X_i v.a.'s i.i.d a X com $X \sim N(0, 1)$, $i = 1 \ldots, n$. Suponha que tem disponível um simulador de números pseudo-aleatórios uniformemente distribuídos em $(0, 1)$.

(a) Usando uma função geradora conveniente, que deve calcular, deduza a distribuição de $Y = \sum_{i=1}^{n} X_i$ assim como o seu valor esperado e variância.

(b) Diga como proceder para simular números pseudo-aleatórios oriundo de uma distribuição aproximada à da v.a. X.

(c) Considere a v.a. $W = e^X$, deduza a distribuição de W e diga como proceder para simular números pseudo-aleatórios da distribuição de W.

Exercício 1.15 Um par aleatório (X, Y) em $\mathbb{R}^2$ tem distribuição normal bivariada com parâmetros $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ se a sua função densidade de probabilidade conjunta é:

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}$$

(a) Mostre que $X \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$ e $Y \sim \mathcal{N}\left(\mu_2, \sigma_2^2\right)$.

(b) Mostre que X e Y são independentes se e só se $\rho = 0$.

Exercício 1.16 (X, Y) tem distribuição normal bivariada com $\mu_X = 5$, $\mu_Y = 10$, $\sigma_X^2 = 1$ e $\sigma_Y^2 = 25$. Determine ρ sabendo que $P\left(4 < Y < 16 | X = 5\right) = 0.954$.

(Sol: $\rho = \pm 0.7990$.)

Exercício 1.17 Sejam X_1 e X_2 v.a.'s i.i.d. a $X \sim N(0, 1)$ e considere $Y_1 = 2X_1 + X_2$ e $Y_2 = X_1 - X_2$. Identifique a distribuição conjunta do par (Y_1, Y_2) e os respetivos parâmetros.

Exercício 1.18 Considere o vetor aleatório $\mathbf{X} = (X_1, X_2) \sim N_2(\mathbf{0}, \sigma^2\mathbf{I_2})$ e o quadrado da distância de Mahalanobis $\Delta^2 = (\mathbf{X} - \mu)'\mathbf{\Sigma}^{-1}(\mathbf{X} - \mu)$. Deduza a distribuição de Δ^2 e indique como simular um valor dessa distribuição.

Exercício 1.19 Suponha que $X_1, \ldots, X_n$ são v.a. independentes. Determine, usando a função geradora de probabilidades, a distribuição de $Y = X_1 + \ldots + X_n$ nos seguintes casos em que para $i = 1, 2, \ldots, n$:

(a) $X_i \sim Bin(n, p)$, $n \in \mathbb{N}$ $0 < p < 1$.

(b) $X_i \sim Poi(\lambda)$, $\lambda > 0$.

(c) $X_i \sim Geom(p)$, $0 < p < 1$.

Exercício 1.20 Se $X \sim BinNeg\left(5, \frac{1}{4}\right)$, calcule $P\left\{X = 12\right\}$ usando a tabela da distribuição binomial.

Exercício 1.21 Para levar a cabo uma investigação relativa à eficácia de um novo tratamento para uma doença rara, cuja incidência na população em geral é de 0.5%, é preciso usar 30 pessoas com a doença para realizar um ensaio clínico. Escreva a função de probabilidade do número de pessoas que é preciso observar até encontrar as 30 pessoas necessárias para realizar o ensaio. Qual é o número médio de pessoas que é preciso entrevistar?

Exercício 1.22 Um dispositivo eletrónico faz parar uma máquina automática de encher consecutivamente pacotes de um determinado produto assim que é detectado o terceiro pacote com peso inferior ao nominal. Sendo $p = 0.2$ a probabilidade de um pacote ter peso inferior ao peso nominal, calcule a probabilidade de a máquina ser parada antes de ter conseguido encher 7 pacotes.

Exercício 1.23 Uma v.a. Λ tem distribuição $Gama\,(n, \alpha)$, com n inteiro. Outra v.a. X tem, para cada λ fixo, distribuição $Poisson\,(\lambda)$. Mostre que X tem distribuição binomial negativa* e identifique os seus parâmetros.

Exercício 1.24 Suponha que uma experiência tem r resultados possíveis e que o i-ésimo resultado ocorre com probabilidade p_i, $\left(0 < p_i < 1, \sum_{i=1}^{r} p_i = 1\right)$. São efectuadas n repetições independentes da experiência. Define-se X_i como sendo o número de observações que são iguais ao i-ésimo resultado $(i = 1, 2, \ldots, r)$.

(a) Justifique que $\mathbf{X} = (X_1, \ldots, X_r)$ tem função massa de probabilidade

$$P\left\{\mathbf{X} = (n_1, \ldots, n_r)\right\} = \begin{cases} \frac{n!}{n_1! \ldots n_r!} p_1^{n_1} \ldots p_r^{n_r}, & n_i \in \{0, 1, \ldots, n\}, \sum_{i=1}^{r} n_i = n \\ 0, & c.c. \end{cases}$$

Diz-se que $\mathbf{X} \sim Multinomial\,(n, p_1, \ldots, p_r)$.

(b) Verifique que se $r = 2$ então a multinomial reduz-se à binomial.

(c) Calcule $E\{X_i\}$ e $Var\{X_i\}$.

(d) Calcule, para $i \neq j$, $Cov\,(X_i, X_j)$.

(e) Serão X_i e X_j v.a. independentes $(i \neq j)$?

Exercício 1.25 Numa transmissão digital de mensagens (de comprimento fixo) sabe-se que as probabilidades de que uma mensagem sofra distorção elevada, média ou baixa são 0.01, 0.04 e 0.95, respectivamente. Admite-se que as distorções ocorrem de forma independente. Suponha que são transmitidas 20 mensagens. Considere as seguintes variáveis aleatórias:

· X_1: nº de mensagens (nas 20) que chegaram com distorção elevada;

· X_2: nº de mensagens (nas 20) que chegaram com distorção média;

· X_3: nº de mensagens (nas 20) que chegaram com distorção baixa.

(a) Indique a distribuição do vector aleatório $\mathbf{X} = (X_1, X_2, X_3)$ e calcule o respectivo valor esperado e matriz de covariâncias.

(b) Calcule $P(X_1 = 2)$, $P(X_1 = 0, X_2 = 1, X_3 = 19)$ e $P(X_3 = 18|X_2 = 0)$.

(c) Construa um algoritmo para simular valores do vector aleatório (X_1, X_2, X_3).

Exercício 1.26 Construa um algoritmo eficiente para simular valores da v.a. X:
$P(X = 1) = 0.3$, $P(X = 2) = 0.2$, $P(X = 3) = 0.35$ e $P(X = 4) = 0.15$.

Exercício 1.27 Considere uma v.a. $X \sim BN(r, p)$.

(a) Usando a relação entre as v.a's binomial negativa e geométrica construa um algoritmo para simular valores da v.a. X.

(b) Usando uma relação recursiva entre as probabilidades de X nos pontos $j + 1$ e j construa um algoritmo que permite simular valores da v.a. X.

Exercício 1.28 Considere uma v.a. $R \sim Unif\{1, 2, 3, 4\}$ e $X_1, X_2, \ldots$ uma sequência de v.a.'s i.i.d. a $X \sim Exp(1)$. Admita ainda que, outra v.a. Y é tal que $(Y|X = x) \sim Poi(x)$.

(a) Descreva algoritmos, com N passos, que permitem:

  i) Estimar $p = P(S > 8)$, onde $S = \sum_{i=1}^{r} X_i$.

  ii) Estimar o valor médio de Y, utilizando simulação condicional.

(b) Mostre que a função de probabilidade marginal de Y é $P(Y = y) = (\frac{1}{2})(\frac{1}{2})^y, y \in I\!N_0$. Construa um algoritmo para simular um valor desta v.a..

Exercício 1.29 Considere uma sucessão de variáveis v.a.'s i.i.d. a $X \sim Geo(p)$ e $Y = \sum_{i=1}^{r} X_i$. Admita ainda que outra v.a. W é tal que $Z = (W|Y = y) \sim Uni(0, y)$.

(a) Deduza a função geradora de probabilidades de X e utilize-a para identificar a distribuição da v.a. Y.

(b) Descreva um algoritmo para simular K valores da v.a. X.

(c) Descreva algoritmos com N passos que permitem:

  i) Estimar $p = P(Y > 8)$.

  ii) Estimar o valor médio de W.

(d) Considere a v.a. Z e seja $I = I(Z \leq ay)$, onde I é a função indicadora da realização de $Z \leq ay$, com $0 < a < 1$. Suponha que se pretende estimar, recorrendo a simulação de Monte-Carlo, a probabilidade seguinte: $p = P(Z \leq ay)$. Considere para esse efeito o estimador $T = I + c(Z - E(Z))$. Determine o valor de c que minimiza a variância de T. T é um estimador de que tipo?

Exercício 1.30 Considere o vector aleatório $\mathbf{X} = (X_1, X_2, X_3) \sim$ Multinomial$(20; 0.3, 0.5, 0.2)$. Construa um algoritmo para simular N valores deste vector aleatório. Diga como diminuir a variabilidade dos resultados à custa de variáveis antitéticas.

Exercício 1.31 Suponha que X e W são obtidos por simulação e $E(X) = E(W) = \theta$. Considere o estimador combinado de θ, $T = \alpha X + (1 - \alpha)W$. Encontre o valor de α que minimiza a variância de T.

Exercício 1.32 Suponha que pretende estimar a $P_f(X > a)$, onde X é uma v.a. positiva com f.d.p. $f_X(x)$ e a um valor na cauda sua distribuição. Usando o método de simulação com amostragem por importância, e gerando X_i a partir da densidade de uma v.a. exponencial com parâmetro λ, escreva o algoritmo para simular a probabilidade pretendida.

Exercício 1.33 Construa um algoritmo para a geração de uma valor da variável aleatória X, cuja função densidade de probabilidade é:

$$f_X(x) = \begin{cases} x/2, & 0 \leq x < 1 \\ 1/2, & 1 \leq x < 2 \\ (3-x)/2, & 2 \leq x < 3 \\ 0, & c.c. \end{cases}$$

Exercício 1.34 Determine um algoritmo que usa o método da transformação inversa para a geração de um valor da variável aleatória cuja função massa de probabilidade é:

$$P(X = x) = \binom{3}{x}(1/3)^x(2/3)^{3-x}, \;\; x = 0, 1, 2, 3.$$

Exercício 1.35 Determine um algoritmo que usa o método da transformação inversa para a geração de um valor da variável aleatória X, com a função de distribuição:

$$F_X(x) = \left(1 - \frac{2cos^{-1}(x)}{\pi}\right), \;\; 0 \leq x \leq 1.$$

Exercício 1.36 Construa um algoritmo de aceitação-rejeição para gerar um valor de uma variável aleatória $X \sim G(\frac{3}{2}, 1)$.

Exercício 1.37 Construa um algoritmo de aceitação-rejeição para gerar um valor da variável aleatória X com função densidade de probabilidade $f_X(x) = 12x(1-x)^2, \;\; 0 < x < 1$.

Exercício 1.38 Construa um algoritmo de simulação de Monte Carlo, recorrendo à geração de números pseudo-aleatórios da distribuição $U(0, 1)$, para estimar a função de distribuição $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{t^2/2} dt$ de uma variável aleatória normal padrão:

(a) Sem redução de variância.

(b) Com redução de variância usando variáveis antitéticas.

Exercício 1.39 Considere o parâmetro $\theta = E(e^U)$ onde $U \sim U(0, 1)$ e sejam $T = e^U$ e $T_1 = T + c(U - E(U))$.

(a) Escreva, com base em T, um algoritmo de simulação de Monte Carlo com N passos para estimar θ.

(b) Deduza a expressão da constante c que torna mínimo o erro quadrático médio de T_1. T_1 é um estimador de que tipo?

(c) Calcule o valor de c e a eficiência de T_1 relativamente a T.

(d) Escreva, com base em T_1, um algoritmo de simulação de Monte Carlo com N passos para estimar θ.

Exercício 1.40 Considere a variável aleatória $X \sim Exp(2)$ e outra variável aleatória Y da qual se sabe que $(Y|X = x) \sim N(x, 9)$. Escrevendo $\theta = P(Y > 2)$ como o valor esperado de uma função de X, indique um algoritmo com N passos que permita estimar θ.