Lista 2 - Simulação estocástica

1. Se $x_0 = 5$ e
$$x_i = (5x_{i-1} + 7) \texttt{ mod } 200,$$
determine $x_1, \ldots, x_{10}$.

2. Com $x_1 = 23$, $x_2 = 66$ e
$$x_i = 3x_{i-1} + 5x_{i-2} \texttt{ mod } 100, \quad i \geq 3,$$
determine os números pseudo-aleatórios entre 0 e 1 a partir de $u_i = x_i/100, \quad i \geq 1$. Encontre os 14 primeiros valores.

3. Gere 1000 números pseudo-aleatórios utilizando a função `runif( )` e associe o objeto ao nome `u`. Considerando a semente 19908, calcule:

   a) a média, a variância e o desvio padrão do vetor `u`.

   b) a média, a variância e o desvio padrão de uma variável aleatória $Y\ \mathcal{U}(0,1)$, ou seja, utilizando a função densidade de probabilidade da distribuição uniforme. Compare os resultados com o item a.

   c) a proporção de valores de `u` menores do que 0,6 e compare com a probabilidade de que a variável aleatória uniforme U seja menor do que 0,6.

► <u>Usar como semente</u>: **674**

4. Seja $U_i \sim \mathcal{U}(0,1), \quad i = 1, 2, \ldots$ e defina
$$N = \text{mínimo}\left\{n : \sum_{i=1}^{n} > 1\right\},$$
ou seja, N é o número de variáveis aleatórias U_i tal que a soma de $U_i's$ seja menor ou igual 1. Dessa forma, pede-se:

   a) estime $\bar{N} = \sum_{j=1}^{r} \frac{N_j}{r}$ para $r = 100$, 1000 e por último 10000.

   b) a estimativa de $\bar{N}$ se aproxima de qual valor conhecido?

5. Suponha um experimento com 100 bandejas sementeiras com 20 células. Em cada célula foi plantada sementes de uma determinada alface cuja a probabilidade de não enraizar é de 0,3. Nesse contexto, seja Y o número de células enraizadas em um total de 20 células.

   a) Qual distribuição de probabilidade pode ser associada a variável aleatória Y. Justifique a resposta.

   b) Utilizando a distribuição de probabilidade do item a simule a quantidade de células enraizadas para cada bandeja e estime as seguintes quantidades:

      - proporção de no máximo 5 células enraizadas;

- proporção de exatamente 5 células enraizadas;
- a média de células enraizadas.

c) Altere o número de bandejas para 10, 1000 e 10000. Para cada uma dessas situações, repita o item b e calcule, também, os itens do item b utilizando a definição da variável aleatória. O que se pode dizer sobre a precisão das estimativas em relação a quantidade de bandejas?

6. Simule o número de acidentes para cada ano em um período de 15 anos, quando a taxa média é de 2,8 acidentes por ano. Então, pede-se:

   a) Qual distribuição de probabilidade pode ser utilizada para simular a amostra?

   b) utilizando os dados simulados, determine a proporção de mais de 4 acidentes em um ano e a média de acidentes.

   c) Compare os resultados obtidos no item b com os resultados obtidas pela definição da variável aleatória do item a.

7. A função de probabilidade de uma variável aleatória Y é definida como

$$P(Y = y) = \left(\frac{1}{2}\right)^{y+1} + \frac{(1/2)2^{y-1}}{3^y}, \quad y = 1, 2, \ldots.$$

   Então,

   a) escreva um simples algoritmo para gerar as variáveis aleatórias;

   b) crie uma função denominada `random_fp` para gerar as variáveis aleatórias.

8. A função de probabilidade de uma variável aleatória Y com distribuição geométrica é dada por

$$P(Y = y) = (1 - p)^y \, p, \quad y = 0, 1, \ldots,$$

   em que $0 < p < 1$. Dessa forma, pede-se:

   a) escrever um simples algoritmo para gerar as variáveis aleatórias;

   b) criar uma função denominada `random_geo` para gerar variáveis aleatórias com distribuição geométrica.

   c) compare os dados gerados por meio do gráfico da distribuição empírica com a distribuição geométrica. Utilize $p = 0,5$.

9. A função densidade de probabilidade de uma variável aleatória Y é expressa como

$$f(y) = y + \frac{1}{2}, \quad 0 \leq y \leq 1.$$

   Dessa forma:

   a) escreva um simples algoritmo para gerar as variáveis aleatórias;

   b) crie uma função denominada `random_fd` para gerar as variáveis aleatórias.

   c) compare os dados gerados com a distribuição dessa variável aleatória por meio do histograma e da distribuição empírica.

10. A função densidade de probabilidade de uma variável aleatória Y com distribuição logística é dada por
$$f(y) = \frac{\exp(-y)}{(1+\exp(-y))^2}, \quad y \in \mathbb{R}$$
Dessa forma:
    a) escreva um simples algoritmo para gerar as variáveis aleatórias;
    b) crie uma função denominada `random_lg` para gerar as variáveis aleatórias.
    c) compare os dados gerados com a distribuição logística por meio do histograma e da distribuição empírica.

11. Suponha que um certo tipo de bateria tem o tempo de vida distribuído exponencialmente com média igual 55 horas, ou seja, para Y = tempo de vida, sua função densidade de probabilidade é dada por
$$f(y) = \frac{1}{55}\exp(-\frac{1}{55}y), \quad y > 0.$$
Utilize a simulação para gerar os tempo de vida de 1000 baterias e determine:
    a) a proporção de baterias com tempo de vida menor do que 100 horas.
    b) a proporção de baterias com tempo de vida entre 100 e 150 horas.
    c) compare os resultados com os valores teóricos utilizando a função densidade da exponencial.

12. Considere uma variável aleatória Y com função densidade de probabilidade dada por
$$f(y) = \frac{2}{\pi}\sqrt{1-y^2}, \quad -1 \leq y \leq 1.$$
    a) Escreva um simples algoritmo para gerar as variáveis aleatórias pelo método da aceitação-rejeição;
    b) Crie uma função parar gerar os números aleatórios;
    c) Compare os valores gerados com a função densidade da variável aleatória por meio do histograma e da distribuição empírica.

13. Suponha que queremos gerar uma variável aleatória Y com função densidade de probabilidade
$$f(y) = 2y, \quad 0 < y < 1.$$
    a) Escreva um simples algoritmo para gerar as variáveis aleatórias pelo método da aceitação-rejeição e pelo método da transformação inversa;
    b) Crie uma função parar gerar os números aleatórios para cada método;
    c) Utilize um gráfico histograma e da distribuição empírica para verificar os dados gerados.

14. Gerar números aleatórios de uma distribuição normal padrão, ou seja $Z \sim N(0,1)$, utilizando o método da aceitação-rejeição. Com esse objetivo, considere as seguintes funções densidade:
    - *half*-normal
$$f(y) = \sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{2}y^2\right), \quad y > 0.$$

- Exponencial como o envelope

$$g(x) = \exp(-x), \quad x > 0,$$

para gerar Z.

Observação: as funções tem suporte nos reais positivos, porém a distribuição normal tem suporte em todos os reais. Então usar a seguinte condição:

$$z = \begin{cases} x & \text{se } u_1 \leq 1/2 \\ -x & \text{se } u_1 > 1/2, \end{cases}$$

em que $U_1 \sim \mathcal{U}(0,1)$.

Dessa forma, pede-se:

a) Escreva um simples algoritmo para gerar as variáveis aleatórias;

b) Crie uma função parar gerar os números aleatórios;

c) Utilize um gráfico histograma e da distribuição empírica para verificar os dados gerados.

15. Uma versão do teorema central do limite diz que se Y é uma variável aleatória binomial com parâmetros n e p, então

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

quando n é suficientemente grande. Mostre numericamente, como a distribuição de Z muda quando n aumenta.

16. Suponha que uma determinada fazenda de soja tem uma produção média diária de 500 toneladas com um desvio padrão de 100 toneladas e que seja razoável imaginar que a produção de soja da fazenda segue uma distribuição normal. Então, simule a produção de soja para 200 dias e calcule:

a) a proporção de proporção de dias com produção de soja menor do que 300 toneladas.

b) a proporção de dias com a produção de soja entre 400 e 600 toneladas.

c) a proporção de dias com a produção de soja maior do que 600 toneladas.

d) compare os itens anteriores com as probabilidades teóricas obtidas pela função densidade de probabilidade da distribuição normal.

17. Suponha que a variabilidade presente no índice de desenvolvimento da educação básica (IDEB) pode ser explicada pelo índice de desenvolvimento de analfabetismo (IA) de municípios do estado de São Paulo. Suponha, também, que 100 municípios do estado de São Paulo foram escolhidos aleatoriamente e que IDEB segue uma distribuição normal. Nesse contexto, pede-se:

a) qual o intervalo de valores prováveis para IDEB e para IA?

b) qual a suposição que deve ser feita a respeito da correlação entre IA e IDEB?

c) para IA fixo, gerar os dados de IDEB levando em conta o item b e de modo que IDEB e IA podem ser representados por um modelo de regressão linear simples. Esboce também, um algoritmo simples.

d) Faça um gráfico de dispersão com os dados de gerados.