ESTATÍSTICA COMPUTACIONAL E SIMULAÇÃO

Folha Prática 1: Geração de Números Pseudo Aleatórios

Ano letivo 2021/22

1. Gere valores de uma distribuição Normal de valor médio 2 e variância 4.

   (a) Determine as características amostrais média, mediana e desvio padrão e quantis.

   (b) Calcule valores da função densidade de probabilidade e da função distribuição nesses pontos gerados e represente-as graficamente.

   (c) faça o histograma da amostra gerada.

   (d) Considere um sequencia de valores apropriados á amostra gerada.

      i. Calcule valores da fdp da distribuição N(2,4) nessa sequencia e represente graficamente.

      ii. Num único gráfico represente o histograma de amostras geradas com $n = 100$ a fdp calculada nos valores da sequencia criada; refaça o grafico usando uma amostra de dimensão 1000.

2. Gere valores de uma distribuição Binomial (100,0.25).

   (a) Determine características amostrais dos valores gerados e calcule valores da fp e fd em vários pontos.

   (b) Faça o gráfico de barras de 5000 valores gerados usando B(10, 0.25), B(50, 0.25) e B(100,0.25) e considere no mesmo gráfico os valores da fmp calculada numa sequência de valores (usando as binomiais anteriormente referidas).

   (c) Considere a distribuição B(100, 0.25). Gere 5000 valores de uma distribuição N(25,var=18.75) e compare as duas distribuições graficamente.

3. Diz-se que uma variável aleatória X tem distribuição de Pareto de tipo I, com parâmetros α e β, $(\alpha > 0, \beta > 0)$ - escreve-se $X \sim Pa(\alpha, \beta)$- se a sua função densidade de probabilidade tiver a seguinte forma:

$$f(x|\alpha, \beta) = \alpha\beta^{\alpha}x^{-(\alpha+1)}, x \geq \beta.$$

   com $E(X) = \frac{\alpha\beta}{\alpha-1}$ e $Var(X) = \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)}$; suponha $\alpha = 3, \beta = 1.5$.

   (a) Sem usar as funções específicas implementadas no R para a distribuição Pareto, escreva um conjunto de funções em R que permita obter, para esta distribuição de Pareto:
   - O valor da função densidade de probabilidade (f.d.p.) para um determinado x (x pode ser um vector de valores onde se quer calcular a f.d.p.)
   - O valor da função de distribuição (f.d.) para um determinado x (x pode ser um vector de valores onde se quer calcular a f.d.)
   - os quantis de probabilidade p (p pode ser um vector de probabilidades)
   - n números pseudo-aleatórios dessa distribuição. Sugestão: Resolva o exercício anterior usando os comandos do R para a distribuição Pareto (pareto), instalando package: extremefit.

   (b) Fazendo variar os parâmetros α e β, represente graficamente a função densidade de probabilidade de uma $Pa(\alpha, \beta)$. Comente.

   (c) Mostre, por simulação, que se
   - $\alpha \leq 1$ não existe valor médio; se $\alpha > 1$, determine estimativa para o valor médio.
   - $\alpha \leq 2$ não existe variância; se $\alpha > 2$,determine estimativa para a variância.

4. Diz-se que uma variável aleatória X tem distribuição logística de parâmetros α, β $(\alpha \in R, \beta > 0)$ - escreve-se $X \sim Lo(x|\alpha, \beta)$- se a sua função densidade de probabilidade for dada por

$$f(x|\alpha, \beta) = \beta^{-1}\frac{exp\{-(\frac{x-\alpha}{\beta})\}}{[1 + exp\{-(\frac{x-\alpha}{\beta})\}]^2}, x \in R$$

Sem usar os comandos do R para a distribuição logística e tomando $\alpha = 1, \beta = 2$, escreva um conjunto de funções em R que permita obter:

(a) o valor da função densidade de probabilidade (f.d.p.) para um determinado x (x pode ser um vector de valores onde se quer calcular a f.d.p.)

(b) o valor da função de distribuição (f.d.) para um determinado x (x pode ser um vector de valores onde se quer calcular a f.d.)

(c) os quantis de probabilidade p (p pode ser um vector de probabilidades)

(d) n números pseudo-aleatórios dessa distribuição.

Sugestão: Resolva o exercício anterior usando os comandos do R para a distribuição logística (logis).

5. Dada uma sucessão $(X_1, \cdots, X_n)$ NPA´s U(0,1)podemos implementar a nuvem de pontos (X_i, X_{i+1}) para $i = 1, \cdots, n-1$ de forma avaliar acerca da independência dos X_i.

   (a) Crie um gráfico da nuvem de pontos usando o gerador de NPA´s do R, represente a função de autocorrelação $acf(X)$. Como interpreta os gráficos resultantes?

   (b) Crie agora gráficos semelhantes usando a função LCG - Gerador Congruencial Linear, das aulas para os valores $m = 8, a = 5, c = 1$ e seed=0. Comente o gráfico de nuvens de pontos resultante em paralelo com o gráfico de $acf(X)$.

   (c) Considere o novo conjunto de valores: $m = 1024, a = 401, c = 201$ e seed=0. Discuta os resultados.

6. Considere para X, o modelo Beta de parâmetros (2,2) com função densidade de probabilidade dada por $f(x) = 6x(1-x), 0 < x < 1$.Pretende-se gerar valores desta variável usando o Método da Rejeição e escolhendo para função proponente o modelo uniforme (0,1).

   (a) Em média, quantos NPA´s U(0,1) serão necessários para gerar 1000 valores dessa variável aleatória?

   (b) Gere 1000 valores, inserindo um contador para o número de iterações que efectivamente são necessárias para obter as 1000 observações e compare com o valor esperado de iterações.

   (c) Compare pontualmente os percentis empíricos, média e variância amostrais com os correspondentes teóricos da distribuição Be(2,2).

   (d) Represente graficamente o histograma associado para os n valores gerados e sobreponha a curva da fdp da Be(2,2).

   (e) Aumente o número de valores gerados para $n = 10000$ e comente os resultados obtidos anteriormente.

7. Considere a seguinte mistura de Gamas:

   (a) $0.3Ga(1.5, 3) + 0.7Ga(5, 3)$

   (b) $\sum_{i=1}^{5} \theta_i f_{Ga(3,\theta_i)}$, onde $\theta_1 = 0.1$,$\theta_2 = 0.2$, $\theta_3 = 0.2$,$\theta_4 = 0.3$, $\theta_5 = 0.2$

   Gere valores de cada mistura e represente graficamente.

8. Diz-se que uma variável aleatória discreta tem uma distribuição Beta-Binomial de parâmetros α, β e n e escreve-se $X \sim Bb(\alpha, \beta, n)$ se a sua função massa de probabilidade for da forma

$$f(x|\alpha, \beta, n) = c\binom{n}{x}\Gamma(\alpha + x)\Gamma(\beta + n - x), x = 0, \cdots, n; \quad c = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + n)}.$$

   (a) Mostre que esta distribuição é gerada pela mistura de uma distribuição binomial com uma distribuição beta, isto é

$$f(x|\alpha, \beta, n) = \int_0^1 \binom{x}{n} y^x(1-y)^{n-x} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}dy, x = 0, \cdots, n.$$

   (b) Considerando o vector parâmetro $(\alpha, \beta, n) = (22, 4, 10)$, gere valores desta distribuição.