Exercícios

Carlos Marcelo Pedroso, Universidade Federal do Paraná

Lista de exercícios sobre ajuste de distribuições de probabilidade para disciplina EELT7029 do curso de Pós Graduação em Engenharia Elétrica da UFPR. Esta parte da disciplina é uma revisão de modelos de probabilidade, com a adição de ajuste de distribuições de probabilidade. Estes fundamentos serão necessários para compreender a sequência da disciplina.

1 Fundamentos

1.1 Distribuições 1

Considere uma variável aleatória contínua X que possui distribuição exponencial com média 6,0. Determine:

1. $P(X \leq 5)$.
2. $P(3 < X \leq 7)$.
3. $P(X > 9)$.

1.2 Distribuições 2

Considere uma variável aleatória contínua X que possui distribuição normal com média 5,0 e desvio padrão 4,0. Determine:

1. $P(X \leq 3)$.
2. $P(3 < X \leq 6)$.
3. $P(X > 10)$.

1.3 Distribuições 3

Considere uma variável aleatória contínua X que possui distribuição de Pareto com $\alpha = 2,1$ e $\beta = 1,0$. Determine:

1. $E[X]$ e $V[X]$.
2. $P(X \leq 2)$.
3. $P(2 < X \leq 10)$.
4. $P(X > 10)$.

1.4 Distribuições 4

Considere uma variável aleatória X com a densidade de probabilidade dada por $f(x) = Rx^{R-1}$ para $0 < x \leq 1$, onde $R > 0$ é um inteiro.

1. Demonstre que esta função pode ser utilizada como uma distribuição de probabilidade.
2. Determine $F(x)$ (apresente todo o desenvolvimento matemático).
3. Determine $E(x)$ (apresente todo o desenvolvimento matemático).
4. Determine $V(x)$ (apresente todo o desenvolvimento matemático).

1.5 Distribuições 5

Suponha uma variável aleatória X que descreve o peso de um grupo de pessoas. Faça considerações sobre qual deve ser o comportamento da distribuição de probabilidade correspondente e escreva um gráfico com o esboço de $P(X = x)$.

2 Ajuste de distribuições

Nestes exercícios você deve esboçar os gráficos exploratórios, como histograma e distribuição acumulada, e comentar/discutir estes resultados no texto. Lembre-se que imprimir o gráfico não é o suficiente.

2.1 Ajuste 1 (nível fácil)

Realize o teste de aderência para os dados a seguir.

15.6	1.7	12.3	11.9	7.6	14.2	9.0	14.1	9.4	10.1
15.6	12.5	9.9	17.2	-3.0	10.8	13.5	2.2	16.5	7.1
11.8	4.8	7.5	10.4	14.1	9.7	14.0	13.2	12.4	8.0
16.2	11.2	13.1	9.4	12.5	10.0	11.1	9.7	14.1	17.6
8.5	11.5	10.2	3.5	9.4	8.4	3.4	9.7	8.8	6.8
18.4	5.7	5.9	15.8	12.4	7.1	5.5	12.3	1.5	4.7
13.1	10.2	7.5	1.3	14.7	13.9	11.0	11.0	14.3	9.4
8.3	19.9	3.0	6.7	9.4	12.5	18.3	19.1	7.5	13.6
15.9	10.0	14.8	15.0	21.2	14.9	17.8	6.9	2.0	15.7
14.3	13.9	3.1	-0.2	9.7	7.6	18.1	11.8	7.3	15.8

Você deve realizar a caracterização utilizando distribuições de probabilidade: (a) identificar a distribuição teórica que melhor se adapta a VA observada (use histograma, distribuição acumulada, esboce gráficos comparativos) e (b) realizar os testes de aderência de Chi-Quadrado e o QPlot para confirmar sua hipótese.

3 Ajuste 2 (nível médio)

Considere as notas da disciplina TE144 Eletricidade Aplicada, turma 2011: (<http://www.eletrica.ufpr.br/pedroso/2011/TE144/NotasTE144.html>).

Procure por uma distribuição de probabilidade que melhor se ajusta aos dados. Realize os testes de aderência de Chi-Quadrado e QPlot para confirmar a sua hipótese. Faça considerações sobre a distribuição encontrada. Este era o resultado esperado? Porque?

4 Ajuste 3 (nível médio)

Considere a população das maiores cidades do Brasil, conforme os dados disponíveis em https://en.wikipedia.org/wiki/List_of_largest_cities_in_Brazil (use a estimativa para 2020).

Procure por uma distribuição de probabilidade que melhor se ajusta aos dados, realize os testes de aderência de Chi-Quadrado e QQPlot para confirmar a sua hipótese.

4.1 Lei de Zipf (nível difícil)

Para o problema anterior, realize a modelagem utilizando a Lei de Zipf.

Considere que os usuários acessam o servidor de acordo com a Lei de Zipf, que argumenta que para uma população de N elementos a frequência normalizada do rank k , $f(k, \alpha, N)$ é dado por:

$$\sum_{i=1}^N P_i = 1, P_k = \frac{\lambda}{k^{1-\alpha}}, \lambda = \frac{1}{\sum_{i=1}^N \frac{1}{i^{1-\alpha}}} \quad (1)$$

onde N é o número de elementos, k é o rank do elemento e α é o parâmetro que caracteriza a distribuição.