PL 2

Probabilidades e Variáveis Aleatórias

2.1 Probabilidade condicional, independência

Responda às seguintes questões através de simulações em Matlab e sempre que for pedido compare os resultados obtidos com os valores teóricos:

1. Considere famílias com filhos em que a probabilidade de nascimento de rapazes é igual à de nascimento de raparigas:

   $\frac{1}{2} = 0{,}5$

	M	F
M	MM	MF
F	FM	FF

   $\frac{3}{4} = 0{,}75$

   (a) Obtenha por simulação uma estimativa da probabilidade do acontecimento "ter pelo menos um filho rapaz" em famílias com 2 filhos. 0,75

   (b) Determine o valor teórico do acontecimento da alínea anterior e compare-o com a estimativa obtida por simulação. Os valores são iguais? Porquê?

   (c) Suponha que para uma família com 2 filhos escolhida ao acaso, sabemos que um dos filhos é rapaz. Qual a probabilidade do outro filho ser também rapaz? Determine o valor teórico desta probabilidade e estime a mesma probabilidade por simulação.

   $P(A|B) = \frac{P A \cap B}{P B}$

   (d) Sabendo que o primeiro filho de uma família com 2 filhos é rapaz, determine por simulação a probabilidade do segundo filho ser rapaz. O que se pode concluir do resultado obtido relativamente à independência de acontecimentos? 0,5

   A : 1º filho é rapaz

   B : 2º filho sou rapaz

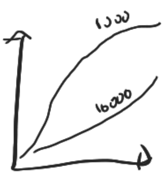
   (e) Considere uma família com 5 filhos. Sabendo que pelo menos um dos filhos é rapaz, obtenha por simulação uma estimativa para a probabilidade de um dos outros (e apenas um) ser também rapaz. 0,33

   (f) Repita a alínea (e), mas estimando a probabilidade de pelo menos um dos outros ser também rapaz.

2. Considere o seguinte "jogo": lançamento com os olhos vendados de n dardos, um de cada vez, para m alvos, garantindo-se que cada dardo atinge sempre um alvo (e apenas 1).

   (a) Estime por simulação a probabilidade de nenhum alvo ter sido atingido mais do que uma vez quando $n = 20$ dardos e $m = 100$ alvos. 0,13

   (b) Estime por simulação a probabilidade de pelo menos 1 alvo ter sido atingido 2 ou mais vezes quando $n = 20$ dardos e $m = 100$ alvos. 0,87

   (c) Considere os valores de $m = 1000$ e $m = 100000$ alvos. Para cada um destes valores, faça as simulações necessárias para desenhar um gráfico (usando a função *plot* do Matlab) da probabilidade da alinea (b) em função do número de dardos n. Considere n de 10 a 100 com incrementos de 10. Os 2 gráficos devem ser sub-gráficos de uma mesma figura (use a instrução *subplot* do Matlab). Compare os resultados dos 2 casos e retire conclusões.

   (d) Considere o valor de $n = 100$ dardos. Faça as simulações necessárias para desenhar um gráfico da probabilidade da alínea (b) em função dos valores de m = 200, 500, 1000, 2000, 5000, 10000, 20000, 50000 e 100000 alvos. O que conclui dos resultados obtidos?

3. Considere um array de tamanho T que serve de base à implementação de uma memória associativa (por exemplo em Java). Assuma que a função de *hash* devolve um valor entre 0 e $T - 1$ com todos os valores igualmente prováveis.

(a) Determine por simulação a probabilidade de haver pelo menos uma colisão (pelo menos 2 *keys* mapeadas pela função de *hash* para a mesma posição do array) se forem introduzidas 10 *keys* num array de tamanho $T = 1000$. 0,17

(b) Faça um gráfico da probabilidade da alínea (a) (estimada por simulação) em função do número de *keys* para todos os valores relevantes num array de tamanho $T = 1000$.

(c) Para um número de *keys* igual a 50, represente graficamente a variação da probabilidade (estimada por simulação) de não haver nenhuma colisão em função do tamanho T do array (assuma os tamanhos T de 100 até 1000 com incrementos de 100).

4. Considere uma festa em que está presente um determinado número n de pessoas.

   (a) Qual deve ser o menor valor de n para o qual a probabilidade de duas ou mais pessoas terem a mesma data de aniversário (mês e dia) é superior a 0,5 (assuma que um ano tem sempre 365 dias)?

   (b) Qual deve ser o valor de n para que a probabilidade da alínea anterior passe a ser superior a 0,9?

5. Considere um dado de seis faces numeradas de 1 a 6 lançado 2 vezes. Assuma que o dado é equilibrado (mesma probabilidade para todas as faces ficarem para cima). Considere os acontecimentos seguintes: "A – a soma dos dois valores é igual a 9", "B – o segundo valor é par", "C – pelo menos um dos valores é igual a 5" e "D – nenhum dos valores é igual a 1". $P_A = 0,11$ $P_B = 0,5$ $P_C = 0,3$ $P_D = 0,69$

   (a) Estime por simulação a probabilidade da cada um dos 4 acontecimentos.

   (b) Determine teoricamente se os acontecimentos A e B são independentes. $P_{A \cap B} = \frac{1}{18} = P_A \times P_B = \frac{1}{18} = 0,06$
   são Independentes

   (c) Determine teoricamente se os acontecimentos C e D são independentes. $P_{C \cap B} = \frac{9}{36} = 0,25 = P_C \times P_D =$
   Não são Independentes

6. Considere uma linguagem com apenas 3 palavras {"um", "dois", "três"} e que permite sequências de 2 palavras. Considere que todas as sequências são equiprováveis e que as duas palavras de uma sequências podem ser iguais. As respostas às questões seguintes devem ser baseadas nos valores teóricos.

   (a) Qual a probabilidade da sequência "um dois"? $\frac{1}{9}$

   (b) Qual a probabilidade de "um" aparecer pelo menos uma vez numa sequência? $\frac{5}{9}$

   (c) Qual a probabilidade de ocorrer "um" ou "dois" numa sequência? $\frac{8}{9}$

   (d) Qual o valor de P["sequência incluir a palavra um" | "sequência inclui palavra dois"]? $\frac{P(A \cap B)}{P_B} = \frac{\frac{2}{9}}{\frac{5}{9}} = \frac{2}{5}$
   A B

7. Considere que uma empresa tem 3 programadores (André, Bruno e Carlos) e que a probabilidade de um programa de cada um deles ter problemas ("bugs") e o número de programas desenvolvidos assumem os valores apresentados na tabela seguinte.

Programador	Prob("erro num programa")	programas
André	0.01	20
Bruno	0.05	30
Carlos	0.001	50

0,0002

O Diretor da empresa seleciona de forma aleatória um dos 100 programas produzidos pelos seus 3 programadores e descobre que este contém um erro sério.

   (a) Qual é a probabilidade de o programa ser do Carlos? 0,0286

   (b) De quem é mais provável ser o programa? Bruno

Carlos Andre Bruno

$$\frac{0,001 \times 50}{100} + \frac{0,01 \times 20}{100} + \frac{0,05 \times 30}{100} = 0,0175 = et^{\text{erro total}}$$

/et

=

0,0286

0,05

2.2 Variáveis e distribuições aleatórias

1. Considere a variável aleatória X correspondente à face que fica para cima no lançamento de 1 dado. Usando os valores teóricos:

   (a) produza um gráfico, em Matlab, que represente a função massa de probabilidade[1] de X;

   (b) num segundo gráfico da mesma figura, desenhe o gráfico da função de distribuição acumulada (use a função `stairs` do Matlab).

2. Considere uma caixa contendo 90 notas de 5 Euros, 9 notas de 50 e 1 de 100.

   S tem tamanho 100 cada espaço é uma nota

   $S = \{5, 50, 100\}$ — $p = 0{,}9$, $p = 0{,}09$, $p = 0{,}01$

   (a) Descreva o espaço de amostragem da experiência aleatória, retirar uma nota da caixa, e as probabilidades dos acontecimentos elementares. $p(5) = \frac{90}{100} = 0{,}9$ $p(50) = \frac{9}{100} = 0{,}09$ $p(100) = \frac{1}{100} = 0{,}01$

   (b) Considere agora a variável aleatória X como sendo o valor de uma nota retirada à sorte da caixa acima descrita. Descreva o espaço de amostragem e a função massa de probabilidade de X.

   (c) Determine a função distribuição acumulada de X e efectue a sua representação gráfica em Matlab.

3. Considere 4 lançamentos de uma moeda equilibrada. Seja X a variável aleatória representativa do número de coroas observados nos 4 lançamentos.

   (a) Estime por simulação a função massa de probabilidade $p_X(x)$ da variável aleatória X.

   valores ?

   (b) Estime o valor esperado, a variância e o desvio padrão de X com base em $p_X(x)$.

   Distribuição Binomial

   (c) Identifique o tipo da distribuição da variável aleatória X e escreva a expressão teórica da respectiva função de probabilidade. $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$

   1 4 6 4 1 — k — 1, 2, 3, 4 — len(S) = 16

   (d) Calcule os valores teóricos da função massa de probabilidade de X e compare-os com os valores estimados por simulação obtidos em (a). $p(0) = \frac{1}{16}$ $p(1) = \frac{4}{16}$ ...

   (e) Calcule os valores teóricos de $E[x]$ e de $Var(X)$ e compare-os com os valores obtidos em (b).

   $E[x] = p(0)\times 0 + p(1)\times 1 + p(2)\times 2 + p(3)\times 3 + p(4)\times 4 = 2$ $Var(x) = p(0)\times 0^2 + p(1)\times 1^2 + p(2)\times 2^2 + p(3)\times 3^2 + p(4)\times 4^2 - E(x)^2 = 1$

   (f) Com base nos valores teóricos da função massa de probabilidade desta distribuição, calcule:

      i. a probabilidade de obter pelo menos 2 coroas; $p(2) + p(3) + p(4) = 0{,}69$

      ii. a probabilidade de obter até 1 coroa; $p(0) + p(1) = 0{,}31$

      iii. a probabilidade de obter entre 1 e 3 coroas. $p(1) + p(2) + p(3) = 0{,}875$

4. Sabendo que um processo de fabrico produz 30% de peças defeituosas e considerando a variável aleatória X, representativa do número de peças defeituosas numa amostra de 5 peças tomadas aleatoriamente, obtenha:

   (a) Por simulação:

      i. estimativa para a função massa de probabilidade de X;

      ii. o gráfico representativo da função distribuição acumulada de probabilidades de X;

      iii. estimativa para probabilidade de, no máximo, 2 das peças de uma amostra serem defeituosas.

   (b) Analiticamente:

      ?

      i. a função distribuição acumulada de X; $\sum 0{,}3^k \times 0{,}7^{5-k}$ $k\ (0 \to 5)$

      ii. a probabilidade de, no máximo, 2 das peças de uma amostra serem defeituosas.

      $p(0) + p(1) + p(2)$

5. Suponha que o(s) motor(es) de um avião pode(m) falhar com probabilidade p e que as falhas são independentes entre motores. Suponha ainda que o avião se despenha se mais de metade dos motores falharem. Nestas condições, prefere voar num avião com 2 ou 4 motores? Utilize a distribuição que considerar mais adequada.

[1] A função massa de probabilidade é muitas vezes designada simplesmente por função de probabilidade

Sugestão: Tem pelo menos 2 alternativas: (1) obter expressões para a probabilidade de cada tipo de avião se despenhar em função de p e usar o quociente entre ambas para responder à questão, (2) efectuar os cálculos para um conjunto de valores concretos[2] de p (ex: `p= logspace(-3,log10(1/2),100))` e usar um gráfico mostrando simultaneamente as probabilidades de cada tipo de avião se despenhar.

6. A distribuição de Poisson é uma forma limite da distribuição binomial (quando $n \to \infty$, $p \to 0$ e np permanece constante) e portanto pode ser usada para aproximar e simplificar os cálculos envolvidos com a binomial em situações em que as condições anteriores se verifiquem.

   Num processo industrial de fabrico de chips, alguns aparecem defeituosos tornando-os inapropriados para comercialização. É sabido que em média por cada 1000 chips há um defeituoso.

   → n

   (a) Usando a distribuição binomial, determine a probabilidade de numa amostra de 8000 chips aparecerem 7 defeituosos. → k

   (b) Determine a mesma probabilidade usando a aproximação de Poisson e compare o resultado com o da alínea anterior.
   Lei de Poisson: $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$

7. Suponha que o número de mensagens que chega a um servidor de *email* segue uma lei de Poisson com média de 15 (mensagens por segundo). Calcule a probabilidade de num intervalo de um segundo:

   ↳ λ

   (a) o servidor não receber nenhuma mensagem;

   (b) o servidor receber mais de 10 mensagens.

   → n

8. Assumindo que o número de erros tipográficos numa página de um livro tem uma distribuição de Poisson com $\lambda = 0.02$, calcule a probabilidade de que exista no máximo 1 erro num livro de 100 páginas. → n
   ↳ k
   Considere que o número de erros em cada página é independente do número de erros nas outras páginas.

9. Considerando a variável aleatória X, representativa das classificações dos alunos de um determinado curso, contínua [3] e com distribuição normal [4] (média 14 e desvio padrão 2), obtenha através de simulação uma estimativa para as probabilidades de:

   ↓ ↳ σ
   $E[X] = m$

   (a) um aluno do curso ter uma classificação entre 12 e 16;

   (b) os alunos terem classificações entre 10 e 18;

   (c) um aluno passar (ter classificação maior ou igual a 10);

   (d) verifique a correção dos resultados anteriores usando a função Matlab `normcdf()`.

normrnd

[2]Correr `help logspace` no Matlab para perceber os argumentos do `logspace` usados no exemplo.
[3]Equivale a considerar que as classificações são números reais.
[4]Utilize a função Matlab `randn()`.

2.3 Exercícios suplementares

Considere uma empresa fabricante de brinquedos que produz um determinado brinquedo. O brinquedo é composto por dois componentes (1 e 2) que são produzidos separadamente e posteriormente montados. No final, os brinquedos são embalados para comercialização em caixas com n brinquedos cada.
O processo de fabrico do Componente 1 produz $p_1 = 0,2\%$ de componentes com defeito. O processo de fabrico do Componente 2 produz $p_2 = 0,5\%$ de componentes com defeito. Um brinquedo está com defeito se pelo menos um de seus componentes estiver com defeito. O processo de montagem produz $p_a = 1\%$ de brinquedos com defeito (mesmo quando nenhum dos 2 componentes está com defeito).

1. Considere o evento "A - uma caixa de brinquedos tem pelo menos 1 brinquedo com defeito".

   (a) Estime por simulação a probabilidade do evento A quando $n = 8$ brinquedos.

   (b) Estime por simulação o número médio de brinquedos defeituosos apenas devido ao processo de montagem quando ocorre o evento A e $n = 8$ brinquedos.

2. Considere o evento "B - uma caixa de brinquedos não tem brinquedos com defeito".

   (a) Estime por simulação a probabilidade do evento B quando $n = 8$ brinquedos. Verifique a consistência deste resultado com o obtido na questão 1(a).

   (b) Determine o valor teórico da probabilidade do evento B e compare-o com o valor estimado por simulação na questão 2(a). O que conclui?

   (c) Faça as simulações necessárias para desenhar um gráfico $plot$ da probabilidade do evento B em função da capacidade da caixa n. Considere todos os valores de n de 2 a 20. Descreva e justifique os resultados obtidos.

   (d) Analisando o gráfico traçado na questão anterior, 2(c), qual deve ser a capacidade máxima da caixa se a empresa quiser garantir que a probabilidade de cada caixa não ter brinquedos com defeito seja de pelo menos 90%?

3. Considere a variável aleatória X que representa o número de brinquedos defeituosos numa caixa.

   (a) Estime por simulação a função massa de probabilidade $p_X(x)$ de X quando $n = 8$ brinquedos e desenhe-a num gráfico $stem$. Descreva os resultados obtidos e verifique a sua consistência com o resultado obtido na questão 2(a).

   (b) Com base em $p_X(x)$, calcule a probabilidade de $X >= 2$. O que conclui?

   (c) Com base em $p_X(x)$, estime o valor esperado, a variância e o desvio padrão de X.

   (d) Repita as questões 3(a), 3(b) e 3(c), mas agora considerando $n = 16$ brinquedos. Compare todos os resultados com os obtidos anteriormente (com $n = 8$ brinquedos) e justifique as diferenças.

4. Suponha agora que a empresa pretende comercializar os brinquedos em caixas de $n = 20$ brinquedos garantindo que a probabilidade de uma caixa comercializada não ter brinquedos com defeito seja de pelo menos 90%.

   Para atingir este objetivo, o processo de montagem foi melhorado reduzindo p_a para 0,1% e foi implementado um processo de garantia de qualidade da seguinte forma: uma amostra de m brinquedos (com $1 \leq m < 20$) é selecionada de cada caixa para teste; a caixa não é comercializada se pelo menos um dos brinquedos selecionados estiver com defeito, ou é comercializada caso contrário.

   (a) Estime por simulação a probabilidade de uma caixa ser comercializada quando o processo de garantia de qualidade é implementado com $m = 1$ (verifique a utilidade da função Matlab $randperm$ na implementação da simulação). O que conclui?

   (b) Estime por simulação o menor valor de m necessário para atingir o objectivo desejado.