

Teoria de probabilidades

Francisco Vaz

Outubro de 2015

Contents

1	Introdução	1
1.1	Nota histórica	1
1.2	Experiências aleatórias	2
1.3	Definição frequencista de probabilidade	4
1.4	Introdução à estatística descritiva	6
1.5	Simulação de problemas aleatórios	12
1.5.1	Geração de números aleatórios	12
1.5.2	O programa Matlab	14
1.5.3	Exemplos de simulação	14
1.6	Exercícios	17
1.7	Problemas propostos	20
2	Teoria Axiomática de Probabilidades	23
2.1	Revisões da teoria de conjuntos	23
2.2	Axiomas	24
2.3	Corolários	25
2.4	Experiências com resultados igualmente possíveis	27
2.5	Probabilidade em conjuntos não numeráveis	31
2.5.1	Experiências com resultados igualmente possíveis	31
2.5.2	Lei exponencial (tempo de vida)	33
2.6	Probabilidade condicional	34
2.7	Independência de Acontecimentos	38
2.7.1	Sequências de acontecimentos independentes	40
2.7.2	Sequências de acontecimentos dependentes	41
2.8	Exercícios	42
3	Variável Aleatória	51
3.1	Definição de variável aleatória	51
3.2	Função de distribuição	53
3.2.1	Propriedades da função de distribuição	53
3.3	Função de densidade de probabilidade	55
3.3.1	Propriedades da função de densidade de probabilidade	56
3.4	Funções de probabilidade condicional	61
3.5	Esperança matemática	62

3.5.1	Variância	63
3.5.2	Momentos	63
3.6	Função característica	64
3.6.1	Propriedades	64
3.7	Função geradora de probabilidade	65
3.8	Desigualdades de Markov e Chebyshev	66
3.9	Estudo de algumas variáveis aleatórias	67
3.9.1	Variável de Bernoulli	67
3.9.2	Variável binomial	67
3.9.3	Variável binomial negativa	69
3.9.4	Variável geométrica	70
3.9.5	Variável de Poisson	71
3.9.6	Variável discreta uniforme	74
3.9.7	Variável uniforme	74
3.9.8	Variável gaussiana ou normal	75
3.9.9	Variável exponencial	79
3.9.10	Variável gama	80
3.9.11	Variável de Rayleigh	82
3.9.12	Variável t de Student	82
3.10	Função de variável aleatória	83
3.10.1	Transformação linear	86
3.10.2	Transformação quadrática	86
3.10.3	Transformação sinusoidal	88
3.10.4	Função limitadora	89
3.11	Exercícios	90
4	Variáveis aleatórias multidimensionais	101
4.1	Vetores aleatórios	101
4.1.1	Pares de variáveis aleatórias	102
4.2	Função de probabilidade conjunta	102
4.2.1	Funções de probabilidade marginal	103
4.3	Função de distribuição conjunta	104
4.3.1	Propriedades	104
4.3.2	Funções de distribuição marginal	105
4.4	Função de densidade de probabilidade conjunta	106
4.4.1	Propriedades	106
4.4.2	Funções de densidade marginal	107
4.5	Funções de probabilidade condicional	109
4.6	Independência	112
4.7	Esperança matemática	113
4.7.1	Correlação	113
4.7.2	Covariância	113
4.7.3	Média condicional	115
4.8	Extensão ao caso multivariável	116
4.8.1	Variáveis conjuntamente gaussianas	117
4.8.2	O caso bidimensional	119

4.9	Exercícios	121
5	Somas de variáveis aleatórias	129
5.1	Média e variância da soma	129
5.2	Densidade de probabilidade da soma	130
5.3	Leis dos grandes números	133
5.4	Teorema do limite central	135
5.5	Exercícios	138
A	Análise Combinatória	141
B	Geração de variáveis aleatórias	145
B.1	Métodos diretos	145
B.1.1	Variável uniforme	145
B.1.2	Variável gausseana	146
B.1.3	Variável de Bernoulli	146
B.1.4	Variável binomial	147
B.2	Método da transformação	147
B.2.1	Variável exponencial	148
B.3	Método da rejeição	148
B.3.1	Variável com uma distribuição sinusoidal	150
B.4	Geração de variáveis multidimensionais gausseanas	151
C	Formulário	153

Prefácio

Ao longo da sua formação académica e da sua vida profissional o engenheiro tem de lidar com dados que apresentam flutuações aleatórias. É pois necessário dotá-lo de conhecimentos que lhe facilitem a análise e interpretação deste tipo de informação com comportamento aparentemente errático e não inteiramente previsível. O controlo de qualidade numa linha de produção, a fiabilidade de um produto e a adopção de coeficientes de segurança apropriados a um correto dimensionamento são alguns exemplos de problemas onde uma análise baseada em modelos probabilísticos é necessária para a generalidade dos profissionais de engenharia.

Chapter 1

Introdução

1.1 Nota histórica

Desde os primórdios da história que o homem usa dispositivos físicos para introduzir a incerteza em jogos. De fato sabe-se da existência de dados de jogar bem anterior à idade de ouro da Grécia ou ao império romano. Contudo apenas no fim do século XV, princípios do século XVI se começou a usar uma metodologia rigorosa, com base no pensamento matemático de então, para caracterizar e interpretar os resultados empiricamente obtidos a partir dos jogos de azar. Deve-se ao matemático (e médico) italiano Cardano (1501-1576) uma das primeiras tentativas de caracterizar os resultados de jogos, mas o seu trabalho foi publicado postumamente e só muito mais tarde conhecido e divulgado.

É habitual atribuir o início do Cálculo das Probabilidades como disciplina da matemática a um episódio ocorrido em meados do século XVII (provavelmente em 1654): um certo Chevalier de Méré, jogador profissional para uns, escritor e importante pensador para outros, coloca a Pascal (1623-1662) uma série de problemas diretamente ligados a jogos de azar. Este, por sua vez, comunica-os a Fermat (1601-1665) e inicia-se uma troca de correspondência que constitui o ponto de partida para uma interpretação matemática dos resultados incertos dos jogos de azar. Os desenvolvimentos teóricos continuam graças aos trabalhos de, entre outros, Huyghens (1629-1695), Bernouilli (1654-1716), Leibniz (1646-1716), de Moivre (1667-1754) e Bayes (1702-1761). No início do século XIX surge um novo e importante marco no Cálculo das Probabilidades com a publicação por Laplace (1749-1827) das obras *Théorie Analytique des Probabilités* e *Essai Philosophique sur les Probabilités*. Devem também referir-se, na mesma época, os trabalhos de Gauss (1777-1855).

A partir de meados do século XIX surge a escola russa fundada por Chebyshev (1821-1894) que contou com nomes como Markov (1856-1922) e Lyapunov (1857-1918) e culminou com Kolmogorov (1903-1987) que, com a publicação em 1933 da famosa obra *Foundations of the Theory of Probability* (apenas traduzida para inglês em 1950), cria o último dos grandes marcos na construção desta ciência.

cia.

Estavam, então, estabelecidas as bases da moderna teoria de probabilidades.

1.2 Experiências aleatórias

Os resultados dos jogos de azar são da natureza dos fenómenos a que se dá o nome de aleatórios. São fenómenos tais como o lançamento ao ar de uma moeda e observação da face que fica visível, que têm duas características fundamentais: experiências conduzidas nas mesmas condições (pelo menos do ponto de vista das nossas possibilidades de realização) podem produzir resultados distintos e o resultado de uma experiência individual é imprevisível.

Para o estudo de fenómenos aleatórios vai-se considerar um quadro de referência que os permite definir e caracterizar. Para tal recorre-se ao conceito de experiência aleatória que é uma experiência que apresenta as seguintes características:

- Pode repetir-se indefinidamente nas mesmas condições ou pelo menos em condições tão próximas que, dentro dos limites disponíveis de precisão experimental, não possam ser responsáveis pela modificação dos resultados;
- É uma experiência que tem um resultado imprevisível não sendo possível, partindo da definição do procedimento experimental, fazer uma previsão credível desse resultado;
- Os resultados individuais são imprevisíveis mas, quando tomados no seu conjunto, mostram uma clara regularidade estatística.

Uma experiência aleatória fica então especificada através de um procedimento experimental e por uma série de medidas ou observações. Note-se que "experiência" tem neste contexto um significado que vai para além do atribuído nas ciências experimentais. De facto, "experiência" deve ser entendida como um conjunto de procedimentos ou de circunstâncias que conduzem a resultados observáveis.

A realização de uma experiência aleatória conduz a um dado resultado (imprevisível) - ζ e ao conjunto de todos os resultados possíveis chamaremos **conjunto ou espaço de amostragem**- S .

Quando se realiza uma experiência aleatória procura-se observar a ocorrência de um acontecimento que corresponde à observação de determinados resultados. De entre todos os resultados possíveis apenas alguns interessam para o fim em vista, isto é, apenas um subconjunto dos resultados possíveis deve ser considerado para verificar a ocorrência do acontecimento.

Define-se, então, acontecimento - A - como sendo qualquer subconjunto de S .

Dois acontecimentos particulares devem ser referidos: o acontecimento impossível que nunca pode realizar-se e, portanto, não correspondendo a qualquer elemento de S e representado pelo conjunto vazio \emptyset ; e o acontecimento que

corresponde à realização de um qualquer de todos os resultados possíveis, coincidindo portanto com o próprio conjunto de amostragem S que se designará por acontecimento certo.

Exemplo 1.1 *Selecione uma bola de uma urna contendo 49 bolas numeradas de 1 a 49 e tome nota do número da bola. O espaço de amostragem é:*

$$S_1 = \{1, 2, \dots, 49\}$$

Defina o acontecimento "Uma bola par é escolhida":

$$A_1 = \{2, 4, 6, \dots, 46\}$$

Exemplo 1.2 *Lance uma moeda três vezes consecutivas e registre a sequência de caras (C) e coroas (R). O espaço de amostragem é:*

$$S_2 = \{CCC, CCR, CRC, RCC, RRC, RCR, CRR, RRR\}$$

Defina o acontecimento "Os três lançamentos dão o mesmo resultado":

$$A_2 = \{CCC, RRR\}$$

Exemplo 1.3 *Lance uma moeda 5 vezes consecutivas e registre o número de caras. O espaço de amostragem é:*

$$S_3 = \{0, 1, 2, 3, 4, 5\}$$

Defina o acontecimento "O número de caras é igual ao de coroas":

$$A_3 = \emptyset$$

Exemplo 1.4 *Um bloco de informação é transmitido através de um canal com ruído que o pode danificar introduzindo aleatoriamente erros. Sempre que seja detetado um erro repete-se a transmissão, procedimento que se repete até se obter uma transmissão sem erros. Registe o número de transmissões requeridas. O espaço de amostragem é:*

$$S_4 = \{1, 2, \dots, \infty\}$$

Defina o acontecimento "O número de transmissões é inferior a 10":

$$A_4 = \{1, \dots, 9\}$$

Exemplo 1.5 *Meça e registre o intervalo de tempo entre duas mensagens que chegam a um centro de mensagens. O espaço de amostragem é:*

$$S_5 = \{t \in \mathbb{R} : t \geq 0\} = [0, \infty)$$

Defina o acontecimento "Menos de t_0 segundos entre a chegada de mensagens":

$$A_5 = [0, t_0)$$

Exemplo 1.6 Meça o valor de uma tensão alternada sinusoidal de valor máximo 10V num dado instante de tempo. O espaço de amostragem é:

$$S_6 = \{x \in \mathbb{R} : -10 \leq x \leq 10\} = [-10, 10]$$

Defina o acontecimento "O valor absoluto da tensão é menor que 1V":

$$A_6 = \{v \in \mathbb{R} : -1 \leq v \leq 1\} = [-1, 1]$$

Exemplo 1.7 Escolha aleatoriamente dois números reais entre 0 e 1. O espaço de amostragem é

$$S_7 = \{(x, y) \in \mathbb{R}^2 : 0 \leq x, y \leq 1\}$$

Defina o acontecimento "Escolha os pares que difiram de menos de 0.1":

$$A_7 = \{(x, y) \in \mathbb{R}^2 : |x - y| \leq 0.1\} \cap S_7$$

Uma vez que se está a associar "acontecimento" a um conjunto, dados dois acontecimentos A e B , pode-se falar de "acontecimento união" como sendo $A \cup B$ e "acontecimento interseção" como sendo $A \cap B$. Por razões que se verão mais à frente, estas designações são por vezes substituídas por "acontecimento soma" e "acontecimento produto", respetivamente.

Dois acontecimentos A e B são incompatíveis ou mutuamente exclusivos se não existe nenhum resultado $\zeta \in S$ que pertença a A e a B . Pode-se então afirmar que para dois acontecimentos incompatíveis $A \cap B = \emptyset$.

Dois acontecimentos A e B são contraditórios se a ocorrência de qualquer um deles implicar a não ocorrência do outro, ou seja se para todo $\zeta \in S$, se $\zeta \in A \implies \zeta \notin B$ e $\zeta \in B \implies \zeta \notin A$. Então $B = \bar{A}$.

Se $A \subset B$ a realização de A acarreta a realização de B , razão pela qual se afirma, por vezes, que nestas condições A implica B .

1.3 Definição frequencista de probabilidade

O número de vezes - k - que em N experiências se observa um dado acontecimento A é a frequência de ocorrência (absoluta) do acontecimento. A sua frequência relativa será

$$f_A(N) = \frac{k}{N}$$

Exemplo 1.8 Lançamento de uma moeda e registo do número de vezes que saiu caras. (este exemplo foi simulado usando o Matlab[®], como se verá adiante). Foram realizadas 100000 experiências, das quais as primeiras 20 estão registadas na seguinte tabela, onde a observação ou não de "caras" foi representada por 1 ou 0, respetivamente. Calculou-se também as frequências absoluta e relativa

do acontecimento "sair caras". Na figura 1.1 está representada graficamente a variação da frequência relativa.

Experiência n^o	Observ. "C"	Freq. abs	Freq. rel.
1	1	1	1.000
2	0	1	0.500
3	1	2	0.667
4	0	2	0.500
5	1	3	0.600
6	1	4	0.667
7	0	4	0.571
8	0	4	0.500
9	1	5	0.556
10	0	5	0.500
11	1	6	0.546
12	1	7	0.583
13	1	8	0.615
14	1	9	0.643
15	0	9	0.600
16	0	9	0.563
17	1	10	0.588
18	1	11	0.611
19	0	11	0.579
20	1	12	0.600

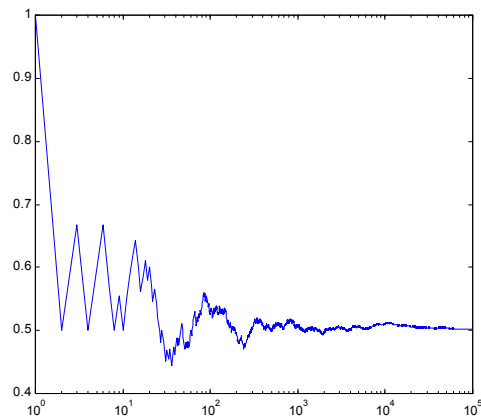


Figure 1.1: Gráfico da frequência relativa do acontecimento "sair Caras".

Da observação do gráfico pode concluir-se que frequência relativa se aproxima de 0.5 como seria de esperar intuitivamente. Na figura 1.2 apresenta-se

o resultado de 3 séries de experiências semelhantes e mais uma vez se verifica um comportamento similar de aproximação ao valor 0.5. A este comportamento regular para grandes números, que os resultados de experiências aleatórias mostram, chama-se regularidade estatística ou regularidade das séries estatísticas. Surge assim a definição frequencista de probabilidade: a probabilidade de um

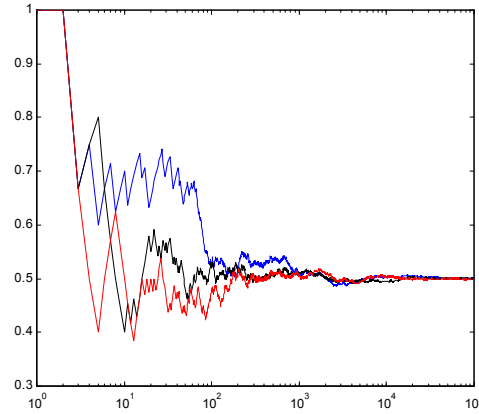


Figure 1.2: Gráficos da frequência relativa do acontecimento "sair Caras" para 3 diferentes séries de experiências.

acontecimento é o limite para que tende a frequência relativa da sua ocorrência quando o número de experiências cresce indefinidamente:

$$P[A] = \lim_{N \rightarrow \infty} f_A(N)$$

Esta definição frequencista que surgiu na primeira metade do século XIX foi a base de uma formulação axiomática de probabilidades que não será a seguida neste texto.

A frequência relativa tem algumas propriedades importantes:

- É um número real compreendido no intervalo $[0, 1]$
- A frequência relativa do acontecimento certo é 1
- Se dois acontecimentos A e B são incompatíveis é trivial provar que $f_{A \cup B}(N) = f_A(N) + f_B(N)$

1.4 Introdução à estatística descritiva

A concretização do método experimental nas ciências físico-naturais e sociais obriga à recolha de grandes quantidades de dados. Não chega, no entanto,

coleccionar dados; é fundamental organizá-los, proceder ao seu tratamento, armazenamento e apresentação. São estes os objetivos da estatística descritiva que hoje dispõe de uma ferramenta eficaz para os realizar: o computador.

Os dados organizam-se de muitas maneiras possíveis. Apresentam-se a seguir alguns exemplos simples que mostram alguns modos de organização de dados e a sua visualização.

Exemplo 1.9 *A avaliação final de uma dada disciplina teve os seguintes resultados:*

12	6	7	6	11	10	10	10	10	8	10	13
13	12	13	13	13	11	3	6	12	12	11	14
11	10	12	10	11	8	9	13	11	12	10	7
11	14	10	10	9	10	12	10	13	11	11	10
14	12	12	10	10	12	5	10	11	11	15	11
12	14	12	13	12	3	14	6	8	12	10	13
7	10	13	10	10	10	6	13	10	10	12	11
3	7	11	14	14	13	13	10	7	10	10	13
11	11	15	10	10	10	16	10	9	10	2	10
7	10	11	10	12	6	10	5	10	12	11	12
15	11	5	19	10	10	13	10	11	12	10	10
10	12	11	10	10	11	16	5	10	4	10	11
10	12	11	11	10	10	13	12	11	13	10	17

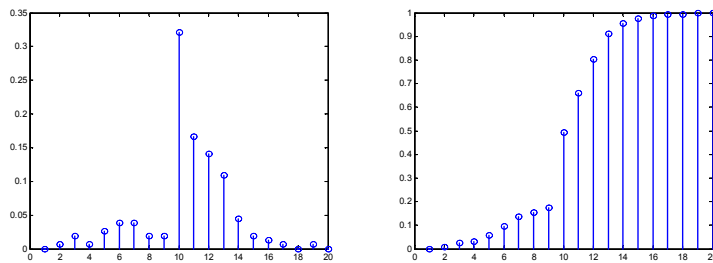
Usando a definição de frequência já apresentada pode-se determinar a frequência e a frequência relativa.

Definindo frequência cumulativa (relativa ou absoluta) para a classe de acontecimentos i como $F_i = \sum_{k=1}^i f_k$ sendo f_k a frequência relativa ou absoluta da classe k

Os dados podem então ser apresentados como tabelas de frequência:

<i>Classificação</i>	<i>Frequência absoluta</i>	<i>Frequência relativa</i>	<i>Frequência cumulativa</i>
1	0	0.0000	0.0000
2	1	0.0064	0.0064
3	3	0.0192	0.0256
4	1	0.0064	0.0321
5	4	0.0256	0.0577
6	6	0.0385	0.0962
7	6	0.0385	0.1346
8	3	0.0192	0.1538
9	3	0.0192	0.1731
10	50	0.3205	0.4936
11	26	0.1667	0.6603
12	22	0.1410	0.8013
13	17	0.1090	0.9103
14	7	0.0449	0.9551
15	3	0.0192	0.9744
16	2	0.0128	0.9872
17	1	0.0064	0.9936
18	0	0.0000	0.9936
19	1	0.0064	1.0000
20	0	0.0000	1.0000

Os mesmos dados poderão ainda ser apresentados sob a forma gráfica:

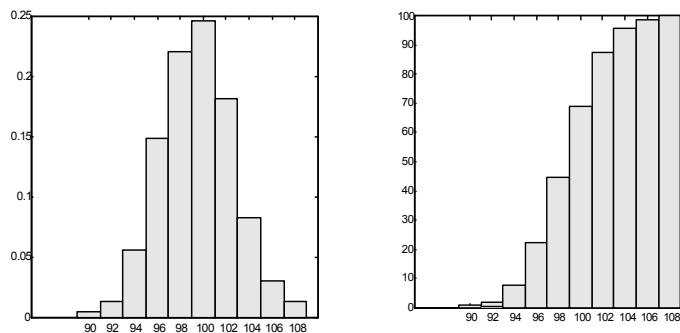


Exemplo 1.10 Num laboratório foram efetuadas as medições do valor de 1000 resistências: O resultado das medições foi agrupado em 10 classes conforme se indica na tabela (valores em Ω).

Tabela com valores medidos agrupados

Limites das classes	Ponto central	Frequência	Frequência relativa	Frequência cumulativa (%)
$h_i - h_{i+1}$	y_i	f_i	p_i	$100P_i$
89-91	90	5	0.005	0.5
91-93	92	14	0.014	1.9
93-95	94	56	0.056	7.5
95-97	96	149	0.149	22.4
97-99	98	221	0.221	44.5
99-101	100	246	0.246	69.1
101-103	102	182	0.182	87.3
103-105	104	83	0.083	95.6
105-107	106	30	0.030	98.6
107-109	108	14	0.014	100.0

Os dados de frequência podem ser apresentados graficamente através do **histograma**. O histograma é uma representação gráfica constituída por retângulos (barras) com uma área proporcional à frequência relativa (ou absoluta)¹ no intervalo. Se se representar a frequência acumulada ter-se-á um histograma cumulativo.



Na análise de dados obtidos a partir de observações ou de simulações usam-se, entre outros, os parâmetros média, mediana, variância e valor quadrático médio. Considerando uma série de observações x_1, \dots, x_N estes parâmetros definem-se como

- **Média** das observações

$$\bar{x} = \frac{1}{N} \sum_{k=1}^N x_k$$

¹Ao longo do texto será preferencialmente usado o histograma de frequências absolutas por razões de simplicidade de utilização do programa Matlab.

- **Mediana** das observações é o valor mais próximo do centro das observações: se o número de observações N for ímpar é o valor da observação a que corresponde igual número de observações inferiores e superiores; se N for par a mediana é definida pela média aritmética dos dois valores centrais

Podem também definir-se os valores máximo e mínimo, x_{max} e x_{min} , e com eles definir o **intervalo de variação** $[x_{min}, x_{max}]$ e a **gama** (de variação) das observações $x_{max} - x_{min}$.

Estes são parâmetros de localização dos dados, fornecendo informação sobre a zona dos resultados possíveis onde estão as observações. Um outro tipo de parâmetros dá informação sobre a sua dispersão, ou seja como se apresentam dentro do intervalo de variação. Os mais utilizados são:

- **Valor quadrático médio** das observações

$$x_{qm} = \sqrt{\frac{1}{N} \sum_{k=1}^N x_k^2}$$

- **Desvio padrão** das observações²

$$s_x = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (x_k - \bar{x})^2}$$

sendo s_x^2 designado por **variância** das observações

Exemplo 1.11 Foram feitas num laboratório de física as seguintes medidas em nm do comprimento de onda de uma radiação eletromagnética:

786 780 782 782 784 781 784

Determine os parâmetros característicos destas observações:

Média $\bar{x} = 782.71$; Mediana $\tilde{x} = 782$; Intervalo de variação: $[780, 786]$; Gama: 6; Valor quadrático médio $x_{qm} = 782.72$; Desvio padrão $s_x = 2.059$; Variância $s_x^2 = 4.238(\text{nm})^2$

Transformações lineares de observações

Por conveniência de cálculo e de interpretação de resultados são comumente efetuadas operações de transformação linear sobre as observações, ou seja, às observações $x_i, i = 1, \dots, N$ faz-se corresponder os valores $y_i, i = 1, \dots, N$ obtidos pela transformação $y_i = ax_i + b$.

Teorema 1.1 As médias \bar{y} e \bar{x} estão relacionadas pela relação linear da transformação

$$\bar{y} = a\bar{x} + b$$

²A razão da utilização de $N - 1$ nesta definição será explicada e justificada mais adiante, neste texto.

Demonstração.

$$\begin{aligned}\bar{y} &= \frac{1}{N} \sum_{k=1}^N y_k = \frac{1}{N} \sum_{k=1}^N (ax_k + b) = \\ &= \frac{1}{N} \left(\sum_{k=1}^N ax_k + Nb \right) = a\bar{x} + b\end{aligned}$$

■

Poder-se-ia determinar os valores para os parâmetros de dispersão, mas as fórmulas obtidas são em geral pouco interessantes, excepto no caso mais particular de se considerar $b = 0$. Neste caso a transformação será uma simples mudança de escala $y_i = ax_i$.

Teorema 1.2 *As variâncias s_y^2 e s_x^2 para uma transformação $y_k = ax_k$ estão relacionadas por*

$$s_y^2 = a^2 s_x^2$$

e os desvios padrão por

$$s_y = as_x$$

Demonstração.

$$\begin{aligned}s_y &= \sqrt{\frac{1}{N-1} \sum_{k=1}^N (y_k - \bar{y})^2} = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (ax_k - a\bar{x})^2} = \\ &= \sqrt{\frac{a^2}{N-1} \sum_{k=1}^N (x_k - \bar{x})^2} = a \sqrt{\frac{1}{N-1} \sum_{k=1}^N (x_k - \bar{x})^2} = as_x\end{aligned}$$

■

Uma transformação linear muito utilizada é a de normalização de dados definida por $y_k = \frac{x_k - \bar{x}}{s_x}$. É trivial verificar que neste caso $\bar{y} = 0$ e $s_y = 1$.

Exemplo 1.12 *Aplique a transformação de normalização aos dados do exemplo anterior e verifique os valores da média e desvio padrão dos dados transformados*

Aplicando a transformação $y_k = \frac{x_k - 782.71}{2.059}$ obtem-se os dados transformados

1.5960 -1.3185 -0.3470 -0.3470 0.6245 -0.8327 0.6245
a que corresponde $\bar{y} = 0$ e $s_y = 1$ ³.

³O leitor mais atento poderá encontrar algumas discrepâncias nos resultados indicados que resultam de erros de arredondamento efectuados na apresentação dos resultados.

1.5 Simulação de problemas aleatórios

Simular um sistema significa criar um modelo para estudar o seu comportamento, entendendo-se por modelo uma representação aproximada do sistema físico que, através de regras compreensíveis, permite compreender e imitar o comportamento do sistema. A engenharia recorre sistematicamente a modelos dos sistemas para os poder caraterizar e implementar. De facto o engenheiro ao projetar usa em geral um modelo que lhe permite, face às especificações de projeto, efetuar os cálculos das componentes do sistema.

Sempre que as caraterísticas observáveis de um sistema sejam mensuráveis recorre-se a modelos matemáticos que são conjuntos de regras reduzidas a relações matemáticas entre os parâmetros que caraterizam o sistema. A análise de circuitos é um bom exemplo de uma modelização matemática: um sistema físico onde se podem observar tensões e correntes elétricas é descrito por um conjunto de equações (leis de Kirchoff) que relacionam as correntes e parâmetros (resistências, capacitâncias, indutâncias e fontes) do circuito. Essas equações permitem uma caraterização completa do comportamento do circuito, não sendo necessária a sua implementação física.

Os modelos matemáticos são facilmente convertidos em modelos computacionais que são programas de computador que imitam o comportamento do sistema. Desta forma o computador permite um certo tipo de "experimentação" sobre modelos: fazendo correr o programa/modelo computacional do sistema e impondo diferentes valores dos parâmetros obtêm-se predições sobre o comportamento do modelo comparáveis com eventuais valores observados no sistema. Se houver concordância entre valores preditos e valores observados pode-se aceitar o modelo; caso contrário deve-se proceder a uma revisão do modelo por forma a aumentar essa concordância.

O tipo de problemas aleatórios que se pretendem simular são modeláveis por modelos matemáticos não aleatórios. A aleatoridade é introduzida na escolha dos valores dos parâmetros de um modelo.

O método que se irá seguir será:

- Criar um modelo computacional do problema
- Gerar parâmetros/entradas aleatórias do sistema e aplicá-las ao modelo e registar as predições
- Usar os métodos da estatística descritiva para analisar as predições

Para implementar esta metodologia é necessário desenvolver métodos para gerar números aleatórios e usar linguagens de programação que permitam a construção dos modelos computacionais.

1.5.1 Geração de números aleatórios

Um método computacional para simulação que exija aleatoridade necessita de um processo de gerar números aleatórios. Estes números deverão obedecer às

propriedades dos processos que vão permitir simular, nomeadamente as regularidades de certos comportamentos que se verificam para grandes sequências, como por exemplo, as propriedades frequenciais que foram apresentadas.

Habitualmente geram-se números "uniformemente" distribuídos no intervalo $[0, 1]$, significando que a frequência relativa com que são gerados é a mesma para todos. Pode-se em seguida através de transformações lineares, atrás referidas, obter outros limites para o intervalo de variação ou por manipulações mais complexas para comportamento não uniforme (no apêndice B este último aspeto será tratado).

Num computador, uma máquina de estados finita com capacidade limitada e finita de representação numérica, não será possível gerar números reais arbitrários em $[0, 1]$. Limitam-se pois os objetivos à geração de um conjunto finito de números inteiros $\{0, 1, \dots, M - 1\}$ que divididos por M darão $\{a_0, a_1, \dots, a_n\}$, com $0 \leq a_i < 1$. O gerador deve garantir que todos devem aparecer com a mesma frequência, e o seu número M será dependente das necessidades do problema a simular e das capacidades de representação numérica do computador. Quanto maior for M mais "denso" é o conjunto dos números gerados.

Como gerar tais números? Uma primeira resposta será realizar experiências aleatórias e registar as sequências de resultados. Por exemplo, lançar uma moeda e registar as faces como "0" ou "1" gerando assim números numa representação binária. Como as simulações devem envolver longas sequências, tais métodos revelam-se pouco práticos, não só pelos tempos envolvidos para execução de tão grande número de experiências, como também pelas necessidades de memória para os armazenar. Tem, pois, que se recorrer a outros métodos que permitam uma geração rápida e eficaz de números aleatórios. O processo mais usado atualmente é o algoritmo recursivo conhecido como **método do resíduo** e que se resume na aplicação da seguinte fórmula recursiva:

$$X_k = aX_{k-1} \bmod M$$

isto é, para gerar uma sequência de números aleatórios basta uma multiplicação, sendo o novo número o resto da sua divisão por M , o que do ponto de vista computacional pode ser implementado de maneiras muito eficazes. A necessária divisão por M também pode ser eficazmente implementada num computador.

Os problemas que se põem têm a ver com a escolha de a e de M . Demonstra-se que M deve ser primo ou potência inteira de um primo e a deve ser cuidadosamente escolhido no intervalo $[0, M]$. Um dos problemas na escolha, é que a sequência gerada é periódica e o seu período depende do valor de a . Está-se, pois, perante sequências não verdadeiramente aleatórias, mas periódicas, donde a designação de pseudo-aleatórias vulgarmente usada. Mas sendo o período da mesma ordem de M , se for muito maior que o comprimento necessário da sequência, o seu comportamento aproxima-se bastante do pretendido.

Considerando as capacidades de representação numérica dos atuais computadores pessoais, é prática generalizada o uso de $M = 2^{31} - 1$, o que dá uma sequência de comprimento 2147 483 647, número que é confortavelmente grande para grande parte das simulações que se fazem.

Para terminar, um breve comentário a X_0 , o valor inicial para a construção da sequência. Em geral os programas existentes tentam gerá-lo aleatoriamente fazendo, por exemplo, uma leitura do relógio do computador no início da execução, ou pelo menos na primeira execução do programa numa dada sessão. No entanto, em geral permitem introduzir opcionalmente o seu valor (vulgarmente designado pela palavra inglesa *seed* -semente), facilitando a execução de sucessivas realizações da mesma sequência pseudo-aleatória.

1.5.2 O programa Matlab

O programa Matlab é uma marca registrada da empresa Math Works, Inc. O Matlab é um ambiente de trabalho para computação científica, permitindo não só a concretização de cálculos complexos baseados em modelos matemáticos como também possui avançadas subrotinas para o desenho de gráficos e visualização de dados científicos. O Matlab é constituído por uma linguagem de programação desenhada para aplicações matemáticas e um conjunto de *tool-boxes*, cada uma para fins específicos.

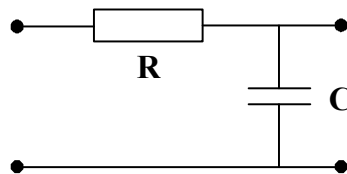
O presente texto foi desenvolvido para ser acompanhado de experimentação em computador e como ferramenta base para esse fim utilizar-se-á o Matlab, inserindo no texto programas que irão permitir verificar teoria e resultados apresentados através de simulação no computador. Aconselha-se pois o leitor a instalar a versão básica do Matlab no seu computador pessoal.

Para os fins que se pretendem a subrotina do Matlab que mais se irá utilizar é a subrotina de geração de números aleatórios, **rand()** que, conforme o argumento, gera números, vetores ou matrizes de números uniformemente distribuídos entre 0 e 1.

Usando as propriedades, anteriormente referidas, das séries de valores numéricos, e associando as subrotinas apropriadas para arredondamento poderemos gerar números dentro de gamas arbitrárias ou outros símbolos, mas sempre de distribuição uniforme (ou seja com igual probabilidade/frequência de ocorrência). Mais tarde serão introduzidos os métodos que permitem gerar números ou símbolos com probabilidade/frequência arbitrárias

1.5.3 Exemplos de simulação

Exemplo 1.13 Considere um circuito RC passa-baixo, que tem como sabe uma frequência de corte $f_h = \frac{1}{2\pi RC}$



Este problema analisado de uma forma determinística, isto é, conhecendo

exatamente o valor das componentes R e C tem uma solução única e bem determinada para o valor da frequência de corte. No entanto sabe-se que o valor das componentes não é exato, por razões de fabrico R e C não são exatas mas têm uma imprecisão (tolerância) t . Como se caracterizará a frequência de corte considerando que as componentes R e C estão nestas circunstâncias?

A solução experimental seria construir circuitos deste tipo usando diferentes componentes e efetuar medições que posteriormente podem ser analisadas segundo os princípios da estatística descritiva.

Pode-se também recorrer a um processo de simulação: geram-se N valores de R e C aleatoriamente distribuídos no intervalo da tolerância, e calcula-se a f_h para todos os casos. Em seguida determinam-se as suas propriedades estatísticas: valor médio, desvio médio quadrático, histograma, etc.

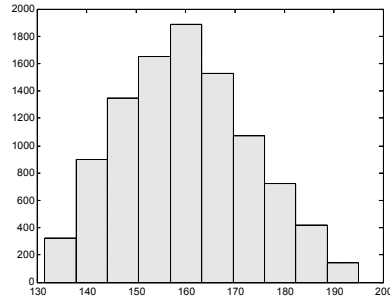
Para proceder à simulação, pode-se recorrer ao programa de simulação seguinte, exemplificando para $R = 1\text{ k}\Omega$, $C = 1\text{ }\mu\text{F}$, $t = 10\%$ e $N = 10000$,

```
%definir o número de experiências aleatórias a realizar
N=10000;
%Gerar valores de resistências de valor nominal
%1kohm uniformemente distribuídas no intervalo
%[900,1100] com Rn=1000 ohm, t=10%
t=0.1;
R=(2*(rand(1,N)-0.5)*t*1000)+1000;
%Gerar valores de capacidades de valor nominal
%1microF uniformemente distribuídas no intervalo
%[0.9,1.1] com Cn=1microF, t=10%
C=(2*(rand(1,N)-0.5)*t*1e-6)+1e-6;
%Definição algoritmica do modelo do circuito
f=1./(2*pi*R.*C);
%Análise dos resultados
fprintf(1,'valor máximo=%6.2fHz , ',max(f))
fprintf(1,'valor mínimo=%6.2fHz\n',min(f))
fprintf(1,'média=%6.2fHz , ',mean(f))
fprintf(1,'desvio padrão=%6.2fHz\n',std(f))
hist(f)
```

A execução deste programa conduz a uma solução do tipo:

```
valor máximo=195.11Hz , valor mínimo=131.58Hz
média=160.09Hz , desvio padrão= 13.05Hz
```

com um histograma dos valores de frequência



Note que a utilização do modelo determinístico da análise de circuitos permitiria determinar uma frequência de corte $f_h = 159.15$ Hz. Usando o modelo estatístico pode-se afirmar que a frequência de corte média terá um certo valor com um certo desvio padrão, valores esses que podem ser determinados para uma ou mais simulações realizadas.

Considere-se ainda um exemplo que permite calcular o valor da probabilidade de um acontecimento através do conhecimento da frequência relativa da sua ocorrência.

Exemplo 1.14 *Pretende-se determinar a probabilidade de em 5 lançamentos sucessivos de uma moeda se observar que o número de "caras" é 2.*

Este é um problema típico em que a simulação serve para verificar um modelo probabilístico: vai-se realizar um elevado número de experiências baseadas num modelo computacional e determinar a frequência relativa da ocorrência do acontecimento.

```
%definição do número de lançamentos da moeda por experiência
N=5;
%definição do número de experiências
n=4000;
%realização das experiências e determinação da frequência
%absoluta da ocorrência do acontecimento ''número de caras=T''
T=2;
for i=1:n
    f(i)=sum(moeda(N)=='C');
end
freq=sum(f==T);
fprintf(1,'frequência absoluta=%6.0f\n',freq)
prob=freq/n;
fprintf(1,'frequência relativa=%6.4f\n',prob)
O resultado de uma simulação será:
frequência absoluta= 1232
frequência relativa=0.3080
```


Conhecida a frequência relativa do acontecimento, considera-se que a probabilidade procurada é aproximadamente esse valor, dada a definição frequencista de probabilidade atrás introduzida

Note que o programa anterior recorre a um subprograma para simular o resultado da experiência aleatória "lançamento de uma moeda"

```
%Esta função gera uma sucessão aleatória de n resultados
%do lançamento de uma moeda representados
%por C para ''caras'' e R para ''coroas''
function y=moeda(n)
if nargin==0
    n=1;
end
z=round(rand(1,n));
y(1:n)='C';
y(find(z==0))='R';
```

Executando este programa para 10 lançamentos podem obter-se os resultados:

```
moeda(10) CRRRCRRRC
moeda(10) RCRRCCRCR
```

1.6 Exercícios

Exercício 1.1 *Programa para simular o resultado da experiência aleatória "lançamento de um dado"*

```
%Esta função gera uma sucessão aleatória de n resultados
%do lançamento de um dado
function y=dado(n)
if nargin==0
    n=1;
end
y=floor(rand(1,n)*6)+1;
Executando este programa podem obter-se os resultados:
dado      5
dado(10) 3 1 4 5 4 6 3 6 3 4
```

Exercício 1.2 *Programa para simular o resultado da experiência aleatória "extração de bolas numeradas de uma urna"*

```
%subrotina que simula a extração de n bolas
%de uma urna contendo m, numeradas de 1 a m,
%nao havendo reposição
function y=extract(n,m)
x=(1:m);
for i=1:n
    ind=floor((m-i+1)*rand)+1;
```

```

y(i)=x(ind);
x=[x(1:ind-1) x(ind+1:m-i+1)];
end
Resultados da simulação para m=49 e n=6 (totoloto)
extract(6,49)    47 12 30 24 44 37
extract(6,49)    23 1 41 22 31 39

```

Exercício 1.3 *Faça um programa que permita simular a extração de cartas de um baralho de 52 cartas*

```

%Esta função gera uma sucessão aleatória de n resultados
%da extração de uma carta de um baralho de 52 cartas
%com reposição (n>0) ou sem reposição (n<0)
function y=carta(n)
if nargin==0
    n=1;
end
if n>0
    for i=1:n
        y(i)=conv(floor(rand*52)+1);
    end
else
    n=-n;x=extract(n,52);
    for i=1:n
        y(i)=conv(floor(rand*52)+1);
    end
end
%converte um número entre 1 e 52 num string
%com o nome de uma das cartas de um baralho
function s=conv(x)
if nargin~=1|x<1|x>52
    error('Entrada fora dos limites 1,...52');return
end
naipe={'Paus','Ouros','Copas','Espadas'};
ordem={'2','3','4','5','6','7','8','9','10',...
    'Valete','Dama','Rei','Ás'};
ind1=mod(x,13);
ind2=floor(x/13)+1;
if ind1==0
    ind1=13;
    ind2=ind2-1;
end
s=strcat(ordem(ind1),'-',naipe(ind2));
Resultados de simulação
carta    'Dama-Espadas'
carta(3) 'Ás-Paus' '7-Copas' 'Ás-Ouros'

```

Exercício 1.4 Considere que para duas séries de observações x_1, \dots, x_N e y_1, \dots, y_M se conhecem os parâmetros \bar{x}, \bar{y}, s_x^2 e s_y^2 . Se se considerarem as duas séries concatenadas determine a sua média e variância em função das médias e variâncias das partes.

Seja z_1, \dots, z_{N+M} a série resultante com $z_i = x_i$, $i = 1, \dots, N$ e $z_{i+N} = y_i$, $i = 1, \dots, M$

$$\begin{aligned}\bar{z} &= \frac{1}{N+M} \sum_{i=1}^{N+M} z_i = \frac{1}{N+M} \left(\sum_{i=1}^N z_i + \sum_{i=N+1}^{N+M} z_i \right) = \\ &= \frac{1}{N+M} \left(\sum_{i=1}^N x_i + \sum_{i=1}^M y_i \right) = \frac{N\bar{x} + M\bar{y}}{N+M}\end{aligned}$$

Uma vez que é trivial provar que

$$\sum_{i=1}^N (x_i - \bar{x})^2 = \sum_{i=1}^N x_i^2 - N\bar{x}^2$$

que conduz a

$$(N-1)s_x^2 = \sum_{i=1}^N x_i^2 - N\bar{x}^2$$

virá:

$$\begin{aligned}s_z^2 &= \frac{1}{N+M-1} \sum_{i=1}^{N+M} (z_i - \bar{z})^2 = \frac{1}{N+M-1} \left(\sum_{i=1}^N z_i^2 - (N+M)\bar{z}^2 \right) = \\ &= \frac{1}{N+M-1} \left(\sum_{i=1}^N x_i^2 + \sum_{i=1}^M y_i^2 - \frac{(N\bar{x} + M\bar{y})^2}{N+M} \right) = \\ &= \frac{1}{N+M-1} \left((N-1)s_x^2 + N\bar{x}^2 + (M-1)s_y^2 + M\bar{y}^2 - \frac{(N\bar{x} + M\bar{y})^2}{N+M} \right)\end{aligned}$$

Casos particulares de interesse:

1. Acrescentar um valor à série de observações

Neste caso basta fazer $M = 1$ nas fórmulas anteriores e substituir

$$\begin{aligned}z_1, \dots, z_{N+1} &\rightarrow x_1, \dots, x_{N+1}, \\ \bar{x} &\rightarrow \bar{x}_N \\ y &\rightarrow x_{N+1}\end{aligned}$$

para se obter

$$\bar{x}_{N+1} = \frac{N\bar{x}_N + x_{N+1}}{N+1}$$

$$s_{x_{N+1}}^2 = \frac{1}{N} \left((N-1)s_{x_N}^2 + N\bar{x}_N^2 + x_{N+1}^2 - \frac{(N\bar{x}_N + x_{N+1})^2}{N+1} \right)$$

2. Retirar um valor à série de observações

Resolvendo as fórmulas anteriores em ordem a \bar{x}_N e a $s_{x_N}^2$

$$\bar{x}_N = \frac{(N+1)\bar{x}_{N+1} - x_{N+1}}{N}$$

$$s_{x_N}^2 = \frac{1}{N-1} \left(Ns_{x_{N+1}}^2 - N\bar{x}_N^2 - x_{N+1}^2 + \frac{(N\bar{x}_N + x_{N+1})^2}{N+1} \right)$$

e substituindo $N = K - 1$

$$\bar{x}_{K-1} = \frac{K\bar{x}_K - x_K}{K-1}$$

$$s_{x_{K-1}}^2 = \frac{1}{K-2} \left((K-1)s_{x_K}^2 - (K-1)\bar{x}_{K-1}^2 - x_K^2 + \frac{((K-1)\bar{x}_{K-1} + x_K)^2}{K} \right)$$

1.7 Problemas propostos

Exercício 1.5 Demonstre a validade da seguinte fórmula recursiva para o cálculo da média de observações: $\bar{x}_N = \bar{x}_{N-1} + \frac{x_N - \bar{x}_N}{N}$, com $\bar{x}_0 = 0$.

Exercício 1.6 Foram observadas duas séries de valores de uma dada grandeza física com os seguintes resultados:

n° de obs.	média	desvio padrão
10	0.6122	0.4285
15	0.5034	0.2579

Determine a média e o desvio padrão do conjunto

Exercício 1.7 A média e o desvio padrão de um dado conjunto de 500 dados é $\bar{x} = 7.467$ e $s_x = 0.234$. Mais tarde verificou-se que um dos valores fora incorretamente introduzido: o verdadeiro valor 7.3 fora substituído por 4.3. Faça a correção da média e do desvio padrão.

Exercício 1.8 *Faça um programa para simular o resultado da experiência aleatória "extração de bolas numeradas de uma urna" supondo que não há reposição.*

Com base neste programa determine por simulação a probabilidade de na extração de duas bolas de uma urna contendo bolas numeradas de 1 a 20, se obter uma bola que tenha um número par múltiplo de 7.

Chapter 2

Teoria Axiomática de Probabilidades

2.1 Revisões da teoria de conjuntos

Como se viu, um acontecimento foi definido a partir das noções da teoria de conjuntos e já se utilizaram as operações sobre conjuntos para definir formas de se comporem acontecimentos. Recordemos então as definições e principais propriedades das operações união, intercepção e complemento.

União¹ dos conjuntos A e B é o conjunto constituído pelos elementos de A e de B .

$$A \cup B = \{x : x \in A \vee x \in B\}$$

Interseção dos conjuntos A e B é o conjunto constituído pelos elementos comuns a A e a B .

$$A \cap B = \{x : x \in A \wedge x \in B\}$$

Complemento do conjunto A em relação a S é o conjunto constituído pelos elementos de S que não pertencem a A .

$$\bar{A} = \{x : x \notin A \wedge x \in S\}$$

Apresentam-se em seguida as principais propriedades destas operações:

1. Propriedade comutativa

$$A \cup B = B \cup A, A \cap B = B \cap A$$

2. Propriedade associativa

$$A \cup (B \cup C) = (A \cup B) \cup C, A \cap (B \cap C) = (A \cap B) \cap C$$

¹A designação "reunião" é também usada em português.

3. Propriedade distributiva

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad , \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

4. Leis de De Morgan

$$\overline{A \cup B} = \bar{A} \cap \bar{B} \quad , \quad \overline{A \cap B} = \bar{A} \cup \bar{B}$$

Definição 2.1 Os conjuntos B_1, \dots, B_n constituem uma partição de S se para $\forall_{i \neq j} B_i \cap B_j = \emptyset$ se verificar $\bigcup_{k=1}^n B_k = S$ com n finito ou infinito.

Um acontecimento é um qualquer subconjunto de S . Considere-se então o conjunto \mathcal{A} de todos os subconjuntos de S . Então qualquer acontecimento $A \in \mathcal{A}$

De acordo com as definições apresentadas de acontecimento, é trivial demonstrar, para conjuntos finitos, que:

1. Se $A, B \in \mathcal{A}$ então $A \cup B \in \mathcal{A}$
2. Se $A \in \mathcal{A}$ então $\bar{A} \in \mathcal{A}$
3. $S \in \mathcal{A}$

Nestas condições \mathcal{A} é uma algebra de subconjuntos de S , e as propriedades que acabamos de referir mostram que é fechada para as operações união e complemento. Desta forma pode-se afirmar que partindo de um número finito de elementos dessa algebra e efetuando um número finito de operações de união, complemento e (portanto também) de interseção o resultado continua a pertencer a esta algebra, ou seja partindo de acontecimentos e efetuando este tipo de operações chega-se a acontecimentos válidos.

O número de elementos de \mathcal{A} pode ser calculado do seguinte modo: seja n o número de elementos de S ; pode-se apenas construir um subconjunto com 0 elementos, o conjunto vazio $\emptyset \subset S$; pode-se construir $n = C_1^n$ subconjuntos com 1 elemento e é trivial verificar que se poderão construir C_k^n subconjuntos com k elementos. Sendo assim n_A o número total de elementos de \mathcal{A} será $n_A = \sum_{k=0}^n C_k^n = 2^n$ (Anexo 1)

Quando se passa de conjuntos finitos para conjuntos infinitos surgem obviamente dificuldades que são ultrapassadas com a definição de σ -algebra: \mathcal{F} é uma σ -algebra se \mathcal{F} é uma algebra e se para uma sequência infinita de acontecimentos $\{A_i\}, i \in \mathbb{N}$ em \mathcal{F} se verificar que $\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$. Nestas condições se se efetuarem um número finito ou infinito numerável de operações de conjuntos sobre acontecimentos continuar-se-á a obter um acontecimento válido.

2.2 Axiomas

Axioma 2.1 A probabilidade de um acontecimento A é um número real não negativo

$$P[A] \geq 0 \tag{2.1}$$

Axioma 2.2 *A probabilidade do acontecimento certo S é igual a 1*

$$P[S] = 1 \quad (2.2)$$

Axioma 2.3 *Se dois acontecimentos forem mutuamente exclusivos a probabilidade da sua união é a soma das suas probabilidades*

$$A \cap B = \emptyset \Rightarrow P[A \cup B] = P[A] + P[B] \quad (2.3)$$

Se o espaço de amostragem não for finito mas infinito numerável, este axioma deve ser reformulado para

Axioma 2.4 *Se A_1, A_2, \dots for uma sequência infinita de acontecimentos mutuamente exclusivos, isto é, $\forall_{i \neq j} A_i \cap A_j = \emptyset$ então*

$$P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{i=1}^{\infty} P[A_k] \quad (2.4)$$

Como se vê os axiomas introduzidos correspondem às propriedades apresentadas das frequências relativas e, embora uma teoria axiomática não exija justificação dos axiomas donde parte, é confortável mostrar a plausibilidade dos axiomas, esperando assim que as teorias desenvolvidas sobre eles conduzam a resultados compatíveis com a experiência.²

2.3 Corolários

Corolário 2.1 $P[A] = 1 - P[\bar{A}]$

Demonstração. Como $A \cap \bar{A} = \emptyset$ e $A \cup \bar{A} = S$, pelos axiomas 2 e 3 vem $P[A] + P[\bar{A}] = 1$ ■

Corolário 2.2 $P[A] \leq 1$

Demonstração. Como $P[A] = 1 - P[\bar{A}]$ e $P[\bar{A}] \geq 0 \Rightarrow P[A] \leq 1$ ■

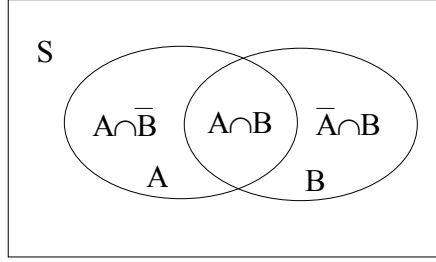
Corolário 2.3 $P[\emptyset] = 0$

Demonstração. Basta considerar no corolário 1 $A = S$ e $\bar{A} = \emptyset$ ■

Corolário 2.4 $P[A \cup B] = P[A] + P[B] - P[A \cap B]$

²Uma introdução mais formal à teoria de probabilidade é a seguinte:

Um espaço de probabilidades é um triplete (S, \mathcal{F}, P) em que \mathcal{F} é uma σ -álgebra não vazia de subconjuntos de S e P é uma aplicação de \mathcal{F} em \mathbb{R} satisfazendo aos 3 axiomas apresentados



Demonstração.

$$A = (A \cap B) \cup (A \cap \bar{B}) \implies P[A] = P[A \cap B] \cup P[A \cap \bar{B}]$$

$$\Leftrightarrow P[A \cap \bar{B}] = P[A] - P[A \cap B]$$

$$B = (A \cap B) \cup (\bar{A} \cap B) \implies P[B] = P[A \cap B] \cup P[\bar{A} \cap B]$$

$$\Leftrightarrow P[\bar{A} \cap B] = P[B] - P[A \cap B]$$

Da figura pode facilmente verificar-se que

$$A \cup B = (A \cap \bar{B}) \cup (\bar{A} \cap B) \cup (A \cap B)$$

e que $(A \cap \bar{B})$, $(\bar{A} \cap B)$ e $(A \cap B)$ são disjuntos. Então

$$P[A \cup B] = P[A \cap \bar{B}] + P[\bar{A} \cap B] + P[A \cap B]$$

e substituindo os valores de $P[A \cap \bar{B}]$ e de $P[\bar{A} \cap B]$ conduz ao resultado pretendido ■

Este resultado generaliza-se para um número n de acontecimentos A_1, \dots, A_n :

$$\begin{aligned} P[A_1 \cup \dots \cup A_n] &= \sum_{i=1}^n P[A_i] - \sum_{1 \leq i_1 < i_2 \leq n} P[A_{i_1} \cap A_{i_2}] \\ &\quad + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} P[A_{i_1} \cap A_{i_2} \cap A_{i_3}] \\ &\quad - \dots \pm P[A_1 \cap \dots \cap A_n] \\ &= \sum_{r=1}^n (-1)^{r-1} \sum_{1 \leq i_1 < \dots < i_r \leq n} P[A_{i_1} \cap \dots \cap A_{i_r}] \end{aligned} \quad (2.5)$$

Este resultado pode ser demonstrado por indução.

Corolário 2.5 $P[A \cup B] \leq P[A] + P[B]$

Demonstração. Basta considerar no anterior $P[A \cap B] \geq 0$ ■

Este resultado é generalizável para um número m finito ou infinito de acontecimentos

$$P\left[\bigcup_{i=1}^m A_i\right] \leq \sum_{i=1}^m P[A_i] \quad (2.6)$$

resultado conhecido como desigualdade de Boole.

Corolário 2.6 Se $A \subset B \implies P[A] \leq P[B]$

Demonstração.

$$\begin{aligned} B &= B \cap S = B \cap (A \cup \bar{A}) = (B \cap A) \cup (B \cap \bar{A}) = \\ &= A \cup (B \cap \bar{A}) \implies P[B] = P[A] + P[B \cap \bar{A}] \end{aligned}$$

e como

$$P[B \cap \bar{A}] \geq 0 \implies P[A] \leq P[B]$$

■

2.4 Experiências com resultados igualmente possíveis

Considere-se um espaço de amostragem finito $S = \{a_i, i = 1, \dots, n\}$ que corresponde a uma experiência aleatória que pode dar n resultados igualmente possíveis.

Considerem-se os acontecimentos $A_i = \{a_i\}, i = 1, \dots, n$, isto é, os acontecimentos para que contribui um e só um dos resultados da experiência. É trivial verificar que os $A_i, i = 1, \dots, n$ constituem uma partição de S e portanto:

$$\begin{aligned} S &= A_1 \cup \dots \cup A_n \\ P[S] &= P[A_1] + \dots + P[A_n] = 1. \end{aligned}$$

Como os acontecimentos são igualmente possíveis deve ser atribuída a mesma probabilidade a todos eles e, portanto,

$$P[A_i] = \frac{1}{n}, i = 1, \dots, n$$

Considere-se agora um outro qualquer acontecimento $B = \{a_j, j = 1, \dots, k\}$, ou seja, um acontecimento para o qual podem contribuir k resultados da experiência aleatória. A sua probabilidade será então

$$P[B] = \sum_{j=1}^k P[A_j] = \sum_{j=1}^k \frac{1}{n} = \frac{k}{n} \quad (2.7)$$

Este resultado é conhecido como a **Regra de Laplace** para experiências com resultados igualmente possíveis, e é apresentado habitualmente como:

Num espaço de resultados igualmente possíveis a probabilidade de um acontecimento é o quociente entre o número de resultados favoráveis e o número de resultados possíveis.

Nesta situação o problema do cálculo de probabilidades passa a ser um problema de contagem (ver Anexo A).

Esta foi a definição usada por Laplace para apresentar todo o seu trabalho teórico sobre o Cálculo das Probabilidades. Note que esta forma de atribuir probabilidades cumpre os axiomas gerais.

Exemplo 2.1 Considere a experiência aleatória que consiste retirar uma bola de uma urna contendo 10 bolas numeradas de 0 a 9 e anotar o número. Defina os seguintes acontecimentos e determine as respectivas probabilidades:

1. A - O número da bola é par;
2. B - O número da bola é múltiplo de três;
3. C - O número da bola é menor que 3;
4. D - $A \cup B$.

Respostas:

1. $A = \{0, 2, 4, 6, 8\}$, com probabilidades $P[A] = 0.5$
2. $B = \{3, 6, 9\}$, com probabilidades, $P[B] = 0.3$
3. $C = \{0, 1, 2\}$, com probabilidades $P[C] = 0.3$
4. $D = \{0, 2, 3, 4, 6, 8, 9\}$ com probabilidades $P[D] = 0.7$.

Esta probabilidade também pode ser calculada por $P(D) = P(A) + P(B) - P(A \cap B) = 5/10 + 3/10 - 1/10 = 0.7$.

Exemplo 2.2 Considere a experiência aleatória: lançamento de 3 vezes consecutivas de uma moeda e registo da sequência caras-coroas. Calcule a probabilidade do acontecimento A "saírem duas vezes caras".

O espaço de amostragem é como já se viu

$$S_2 = \{CCC, CCR, CRC, RCC, RRC, RCR, CRR, RRR\}$$

e aplicando a regra de Laplace 2.7 obtem-se $P[A] = 3/8$.

Poder-se-ia ter seguido o seguinte raciocínio (errado): se se contar o número de vezes que sai caras o espaço de amostragem será $S = \{0, 1, 2, 3\}$ e, e aplicando a regra de Laplace 2.7 obtem-se $P[A] = 1/4$. O erro cometido está em considerar-se que no novo espaço de amostragem os 4 resultados são igualmente possíveis, o que uma análise mais cuidada contradiz.³

Exemplo 2.3 Num jogo de dados, em que se faz o lançamento de 3 dados, a soma obtida pode variar entre 3 e 18. Se se considerar as somas 9 e 10 pode parecer numa primeira análise que estes dois resultados são igualmente

³Sobre o problema da atribuição de igual probabilidade a diferentes resultados é costume referir-se a seguinte polémica mantida por Pascal com Roberval e bastante semelhante ao problema acabado de apresentar. Trata-se da seguinte situação: uma moeda é lançada sucessivamente até se obter o resultado "caras" ou o máximo de 2 lançamentos. Os resultados possíveis são C, RC e RR. Segundo Roberval dever-se-ia atribuir as probabilidades de 1/3 a cada um dos casos mas Pascal defendeu, e bem, que as probabilidades a atribuir seriam 1/2, 1/4 e 1/4, argumentando que quando o 2º lançamento é necessário o seu resultado é independente do resultado do 1º.

2.4. EXPERIÊNCIAS COM RESULTADOS IGUALMENTE POSSÍVEIS 29

possíveis. Com efeito as hipóteses de se obterem são as 6 para cada caso que se apresentam na tabela:

Soma =9	Soma =10
1 2 6	1 3 6
1 3 5	1 4 5
1 4 4	2 2 6
2 2 5	2 3 5
2 3 4	2 4 4
3 3 3	3 3 4

e portanto, qualquer que seja o número total de casos possíveis, ter-se-á a mesma probabilidade. No entanto, a experiência mostrava a jogadores que a soma 10 era mais frequente, como aliás a simulação que em seguida se apresenta confirma. A razão desta discrepância deve estar na hipótese de se considerar igualmente possíveis todos os resultados da tabela. Com efeito tal não se verifica porque cada dado individualmente é responsável por um resultado e portanto os casos possíveis são $A_3^6 = 6^3 = 216$. O resultado 1 2 6 não é um único mas um dos $P_3 = 3! = 6$ que obtêm permutando os valores pelos 3 dados. O resultado 2 2 5 também não o é, pois pode ser obtido de $P_{2,1}^3 = \frac{3!}{2!1!} = 3$ maneiras, permutando um grupo de 2 dados e o terceiro de todas as maneiras possíveis. Apenas o resultado 3 3 3 corresponde a um único agrupamento dos 3 dados. Então a tabela dos casos igualmente possíveis e favoráveis às somas 9 e 10 deve ser alterada para

Soma =9	n° casos favor.	Soma =10	n° casos favor.
1 2 6	6	1 3 6	6
1 3 5	6	1 4 5	6
1 4 4	3	2 2 6	3
2 2 5	3	2 3 5	6
2 3 4	6	2 4 4	3
3 3 3	1	3 3 4	3
Total	25	Total	27

e as probabilidades dos acontecimentos serão

$$P[\text{soma} = 9] = \frac{25}{216} = 0.116, \quad P[\text{soma} = 10] = \frac{27}{216} = 0.125$$

Este resultado pode ser avaliado pela seguinte simulação:

```
%Simulação do lançamento de 3 dados N vezes para determinação
%da frequência relativa da soma de resultado ser igual
%a x1 e a x2
N=100000;
x1=9;
x2=10;
resultado=dado(N)+dado(N)+dado(N);
somax1=sum(resultado==x1);
```

```

somap2=sum(resultado==x2);
fx1=somap1/N;
fx2=somap2/N;
fprintf(1,'P[soma=%2.0f]=%5.3f, P[soma=%2.0f]=%5.3f\n',x1,fx1,x2,fx2)
com o resultado
P[soma= 9]=0.116, P[soma=10]=0.126

```

Vejamos ainda mais um exemplo onde, mais uma vez, se chama a atenção para o problema da contagem dos casos igualmente possíveis.

Exemplo 2.4 *No bridge usa-se um baralho de 52 cartas que são igualmente distribuídas pelos 4 jogadores, cabendo portanto a cada um 13 cartas, a que vulgarmente se chama uma "mão". Determine a probabilidade de uma mão com pelo menos um naipe em falta (acontecimento A)*

Uma maneira (apressada e incorreta) de resolver este problema seria:

1. Escolha o naipe em falta: há 4 possibilidades
2. Escolha as 13 cartas entre as restantes 39 cartas: há C_{13}^{39} maneiras possíveis

Como os casos possíveis são C_{13}^{52} , a probabilidade procurada será

$$P[A] = \frac{4 \cdot C_{13}^{39}}{C_{13}^{52}} = \frac{17063919}{333515525} = 5.1164 \times 10^{-2}$$

Esta solução está de facto errada pois estamos a contar como diferentes casos que são iguais. Por exemplo, considere a mão com copas sendo o naipe em falta e escolha 13 paus. Segundo o raciocínio seguido a mão com espadas em falta e 13 paus seria diferente e obviamente não é, com o resultado de contarmos duas vezes a mesma mão.

Seja $\bar{E}, \bar{C}, \bar{O}$ e \bar{P} os acontecimentos mão sem espadas, copas, ouros e paus, respetivamente. O acontecimento poderá então ser definido por $A = \bar{E} \cup \bar{C} \cup \bar{O} \cup \bar{P}$. Então, tendo em conta o corolário 2.5:

$$\begin{aligned}
P[A] &= P[\bar{E}] + P[\bar{C}] + P[\bar{O}] + P[\bar{P}] - \underbrace{(P[\bar{E} \cap \bar{C}] + P[\bar{E} \cap \bar{O}] + \dots)}_{C_2^4 \text{ termos do tipo não haver dois naipes}} \\
&\quad + \underbrace{(P[\bar{E} \cap \bar{C} \cap \bar{O}] + P[\bar{E} \cap \bar{C} \cap \bar{P}] + \dots)}_{C_3^4 \text{ termos do tipo não haver três naipes}} - \underbrace{P[\bar{E} \cap \bar{C} \cap \bar{O} \cap \bar{P}]}_{=0}
\end{aligned}$$

Então os casos possíveis serão $4 \cdot C_{13}^{39} - C_2^4 \cdot C_{13}^{26} + C_3^4 \cdot C_{13}^{13}$ e a probabilidade correta será

$$P[A] = \frac{4 \cdot C_{13}^{39} - C_2^4 \cdot C_{13}^{26} + C_3^4 \cdot 1}{C_{13}^{52}} = \frac{1621364909}{31750677980} = 5.1066 \times 10^{-2}$$

2.5 Probabilidade em conjuntos não numeráveis

Em muitas experiências os resultados possíveis ocorrem no domínio dos números reais, portanto os respectivos conjuntos de amostragem são infinitos e não numeráveis. Normalmente os acontecimentos que interessam nestes casos são subconjuntos compactos definidos dentro dos respectivos espaços de amostragem e portanto ter-se-á que definir regras de definição de probabilidade para conjuntos deste tipo. Se se estiver a trabalhar em \mathbb{R} os subconjuntos compactos de interesse são intervalos abertos ou fechados em \mathbb{R} .

2.5.1 Experiências com resultados igualmente possíveis

Uma das situações mais simples é a de considerar que sendo $S = [a, b]$ o conjunto de amostragem, qualquer valor é igualmente possível, generalizando a Regra de Laplace. Nestas condições a probabilidade de um qualquer resultado $x \in S$ é $P[x] = 0$.

Podemos no entanto definir a probabilidade para acontecimentos do tipo $A = [c, d]$ a partir de uma medida definida para intervalos compactos, por exemplo a sua dimensão $d - c$. Define-se então

$$P[A] = \frac{d - c}{b - a} \quad (2.8)$$

Para acontecimentos deste tipo é trivial verificar que esta lei de probabilidade cumpre os axiomas.

Note que como

$$\begin{aligned} \{c\} \cup (c, d) \cup \{d\} &= [c, d] \\ P[\{c\}] + P[(c, d)] + P[\{d\}] &= P[(c, d)] = P[[c, d]] \end{aligned}$$

pois para $\forall x \in S P[\{x\}] = 0$, isto é, a probabilidade de um intervalo aberto ou fechado é igual, já que a probabilidade de um valor singular é sempre nula. Este resultado é também válido para intervalos abertos à esquerda ou à direita.

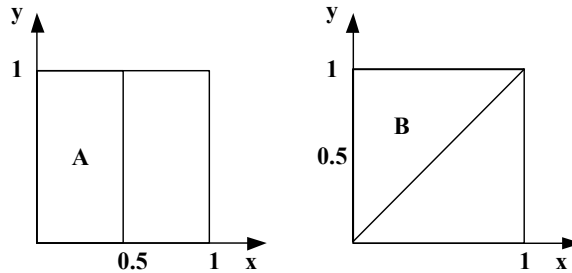
Uma lei semelhante se pode definir em \mathbb{R}^2 para acontecimentos que sejam subconjuntos compactos de \mathbb{R}^2 . Neste caso para um conjunto compacto $S \subset \mathbb{R}^2$ pode definir-se uma lei de probabilidade para qualquer $A \subset S$ como a relação entre uma medida de A e de S , que pode ser a sua área:

$$P[A] = \frac{\text{Área de } A}{\text{Área de } S} \quad (2.9)$$

Mais uma vez é trivial verificar que se cumprem os axiomas gerais e que as probabilidades de acontecimentos que difiram por incluir ou não as linhas limites, são iguais.

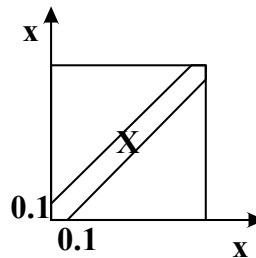
Poder-se-á generalizar para \mathbb{R}^n esta tipo de lei de probabilidade substituindo "área" por "volume" em \mathbb{R}^3 ou "hipervolume" em \mathbb{R}^n .

Exemplo 2.5 Determine as probabilidades dos acontecimentos A e B considerando que o espaço de amostragem é $S = \{(x, y) : 0 \leq x, y \leq 1\}$. Por aplicação direta de 2.9 obtém-se $P[A] = P[B] = \frac{1}{2}$



Note que se Área de $S = 1$, vem $P[A] = \text{Área de } A$. Desta forma pode-se determinar a área através do cálculo de uma probabilidade, o que talvez não seja muito interessante. Mas este resultado demonstra-se muito útil em termos de simulação pois simulando uma experiência aleatória que conduza à realização de acontecimento A , podemos determinar um valor aproximado da área a partir da determinação da frequência relativa.

Exemplo 2.6 Determine a área da figura geométrica definida por $X \cap S$, com $X = \{(x, y) : |x - y| \leq 0.1\}$ e $S = \{(x, y) : 0 \leq x, y \leq 1\}$

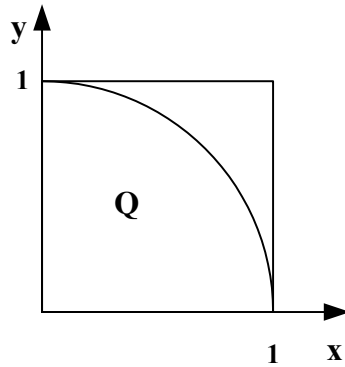


O seguinte programa permite determinar a frequência relativa do acontecimento X em S

```
N=100000;
x=rand(1,N);
y=rand(1,N);
difabs=abs(x-y);
area=sum(difabs<0.1)/N
que conduz aos seguintes possíveis resultados:
0.1897    0.1909    0.1908    0.1864
```

Exercício 2.1 Determine o valor de π , a partir da probabilidade do acontecimento Q da figura seguinte:

Exemplo 2.7 Atendendo a 2.9 $P[Q] = \frac{\pi \cdot 1^2}{1^2} = \frac{\pi}{4}$; então $\pi \simeq 4 \cdot f_Q$



O seguinte programa permite então obter um valor aproximado de π :

```
%Esta função permite estimar o valor de pi calculando
%estatisticamente a área de um setor circular de raio
%unitário
function y=api(n)
x1=rand(1,n);
x2=rand(1,n);
z=sqrt(x1.^2+x2.^2);
freq=sum(z<=1);
y=4*freq/n;
Resultados para diversas simulações com 200000 valores:
3.1418    3.1428    3.1415
```

2.5.2 Lei exponencial (tempo de vida)

A medição do tempo de vida de uma componente num sistema sugere a seguinte lei: o número de componentes que funciona para além de um tempo t decai exponencialmente a um ritmo constante α . Qual será a lei de probabilidade que se pode definir para satisfazer esta condição? O espaço de amostragem será a semireta real não negativa $[0, \infty)$ considerando que uma componente inicia a sua vida no instante $t = 0$ e que poderá manter-se em funcionamento por um período infinito.

A condição inicial pode ser interpretada como "a probabilidade da vida da componente exceder um tempo t decresce exponencialmente ao ritmo α " que se pode exprimir definindo a probabilidade de acontecimentos do tipo $[t, \infty)$ como sendo dados por $P[[t, \infty)) = e^{-\alpha t}$ para $t \geq 0$ e $\alpha > 0$. Nestas condições é fácil verificar que os axiomas 2.1 e 2.2 se verificam, uma vez que

$$P[[0, \infty)) = P[S] = 1$$

e

$$0 \leq e^{-\alpha t} \leq 1$$

Considerando que $(r, s] \cup [s, \infty) = [r, \infty)$, e usando o axioma 2.3

$$P[(r, s]] + P[[s, \infty)) = P[[r, \infty))$$

ou seja

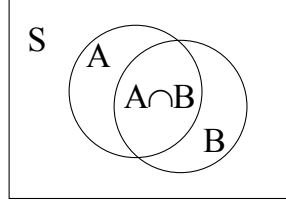
$$P[(r, s]] = P[[r, \infty)) - P[[s, \infty)) = e^{-\alpha r} - e^{-\alpha s}$$

Desta forma pode-se determinar a probabilidade de qualquer intervalo pertencendo à semireta real não negativa.

2.6 Probabilidade condicional

Dados dois acontecimentos A e B e sendo $P[B] > 0$, a probabilidade de A se B se realizou define-se como sendo a probabilidade condicional de A se B , $P[A|B]$ como sendo

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \quad (2.10)$$



Como se pode ver da figura, esta definição corresponde a uma renormalização dos valores de probabilidade a um conjunto mais restrito.

Atendendo a que $P[A|B] \geq 0$, $P[B|B] = 1$ e se $A_1 \cap A_2 = \emptyset$,

$$\begin{aligned} P[(A_1 \cup A_2) | B] &= \frac{P[(A_1 \cup A_2) \cap B]}{P[B]} = \frac{P[(A_1 \cap B) \cup (A_2 \cap B)]}{P[B]} = \\ &= \frac{P[A_1 \cap B] + P[A_2 \cap B]}{P[B]} = P[A_1|B] + P[A_2|B] \end{aligned}$$

a probabilidade condicional cumpre os axiomas gerais das probabilidades e, portanto podemos aplicar os corolários e teoremas obtidos para o caso geral, como por exemplo $P[A|B] = 1 - P[\bar{A}|B]$

Da equação 2.10 é imediato obter

$$P[A \cap B] = P[A|B] P[B] \quad (2.11)$$

resultado conhecido com **lei ou teorema das probabilidades compostas**.

Exemplo 2.8 Um dado é lançado duas vezes e regista-se o número de pontos que saem em cada lançamento. Considere o acontecimento A "o número total de pontos é par" e o acontecimento B "o número de pontos em cada lançamento é ímpar". Determine $P[A|B]$ e $P[A|B]$.

Como o número de hipóteses é reduzido construamos a tabela de resultados possíveis para A e B

	1	2	3	4	5	6		1	2	3	4	5	6
1	×		×		×		1	×		×		×	
2		×		×		×	2						
3	×		×		×		3	×		×		×	
4		×		×		×	4						
5	×		×		×		5	×		×		×	
6		×		×		×	6						

Como o número de casos possíveis é 36 e como $B \subset A$ e portanto $A \cap B = B$ ter-se-á: $P[A] = \frac{18}{36} = \frac{1}{2}$, $P[B] = \frac{9}{36} = \frac{1}{4}$, $P[A \cap B] = \frac{9}{36} = \frac{1}{4}$ e aplicando 2.10 $P[A|B] = \frac{1/4}{1/4} = 1$ e $P[B|A] = \frac{1/4}{1/2} = \frac{1}{2}$

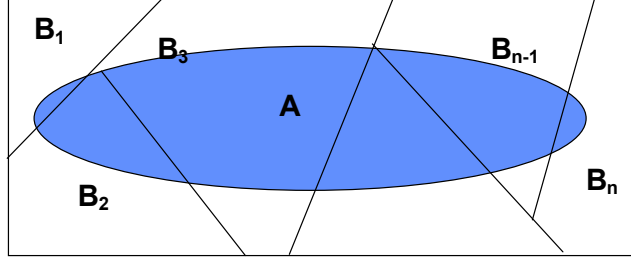
Exemplo 2.9 Num baralho de 52 cartas retiram-se duas cartas sem reposição. Determine a probabilidade de terem saído 2 Reis (2R) sabendo que pelo menos uma das cartas escolhidas é uma figura (consideram-se figuras 4 cartas por naipe: ás, rei, dama, valete).

$$P[2R | \text{pelo menos uma figura}] = \frac{P[(2R) \cap (\text{pelo menos uma figura})]}{P[\text{pelo menos uma figura}]}$$

Como o acontecimento "pelo menos uma figura" está contido no acontecimento 2R e usando o 1º corolário

$$\begin{aligned}
 P[2R | \text{pelo menos uma figura}] &= \frac{P[2R]}{1 - P[\text{nenhuma figura}]} = \\
 &= \frac{\frac{C_2^4}{C_2^{52}}}{1 - \frac{C_2^{36}}{C_2^{52}}} = 8.6207 \times 10^{-3}
 \end{aligned}$$

Seja $B_i, i = 1, \dots, n$ uma partição de S e seja A um acontecimento qualquer neste espaço de amostragem.



A pode ser expresso por:

$$A = A \cap S = A \cap (B_1 \cup \dots \cup B_n) = (A \cap B_1) \cup \dots \cup (A \cap B_n)$$

Atendendo a que

$$(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) = A \cap \emptyset = \emptyset$$

pode-se escrever que

$$P[A] = P[A \cap B_1] + \dots + P[A \cap B_n]$$

e aplicando a definição de probabilidade condicional

$$P[A] = P[A|B_1]P[B_1] + \dots + P[A|B_n]P[B_n] = \sum_{i=1}^n P[A|B_i]P[B_i] \quad (2.12)$$

resultado que é conhecido como o **teorema das probabilidades totais**

Exemplo 2.10 Uma doença (D) tem uma taxa de incidência de 0.001 nas mulheres (F) e de 0.005 nos homens (M). Qual a taxa de doentes numa população constituída por 52% de mulheres e 48% de homens.

Os subconjuntos F e M constituem uma partição da população total e de acordo com o enunciado estão definidas as seguintes probabilidades: $P[F] = 0.52$, $P[M] = 0.48$, $P[D|F] = 0.001$, $P[D|M] = 0.005$. Aplicando o teorema 2.12 obtém-se

$$\begin{aligned} P[D] &= P[D|F]P[F] + P[D|M]P[M] = \\ &= 0.001 \times 0.52 + 0.005 \times 0.48 = \\ &= 0.00292 \end{aligned}$$

Seja $B_i, i = 1, \dots, n$ uma partição de S com probabilidades $P[B_i], i = 1, \dots, n$. Supondo que ocorre um acontecimento A quais serão as novas probabilidades a atribuir aos acontecimentos B_j , isto é, as $P[B_j|A]$? Aplicando a definição de probabilidade condicional e o teorema das probabilidades totais obtém-se

$$P[B_j|A] = \frac{P[B_j \cap A]}{P[A]} = \frac{P[A|B_j]P[B_j]}{\sum_{i=1}^n P[A|B_i]P[B_i]} \quad (2.13)$$

resultado conhecido como **Teorema ou Regra de Bayes**.

Neste tipo de problemas as probabilidades iniciais $P[A|B_j]$ são designadas probabilidades *a priori* e os resultados $P[B_j|A]$ por probabilidades *a posteriori*. Este importante resultado mostra como o conhecimento sobre a evolução de um sistema - neste caso o conhecimento da ocorrência de A - permite modificar a forma de atribuir probabilidades.

Exemplo 2.11 *Dispõe-se de 2 urnas contendo bolas brancas e bolas pretas. A urna 1 tem 3 bolas brancas e 1 preta e a urna 2 tem 2 bolas brancas e 2 pretas. Faz-se a seguinte experiência: escolhe-se uma urna através do lançamento de uma moeda e extraem-se dela sucessivamente duas bolas havendo reposição. Se se observar que as bolas extraídas são de cor diferente, qual a probabilidade da urna escolhida ter sido a nº 1?*

Começemos por simbolizar por B e P a extração de bola branca e preta, respetivamente e por CD o acontecimento "observação de cores diferentes". As probabilidades de escolher as urnas são iguais, $P[1] = P[2]$; as probabilidades na urna 1 são

$$\begin{aligned} P[BP|1] &= P[PB|1] = \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{16} \\ P[CD|1] &= P[BP|1] + P[PB|1] = \frac{6}{16} \end{aligned}$$

; analogamente tem-se para na urna 2

$$P[BP|2] = P[PB|2] = \frac{2}{4} \cdot \frac{2}{4} = \frac{4}{16}$$

e

$$P[CD|2] = P[BP|2] + P[PB|2] = \frac{8}{16}$$

De acordo com a eq. 2.12

$$P[CD] = P[CD|1]P[1] + P[CD|2]P[2] = \frac{6}{16} \cdot \frac{1}{2} + \frac{8}{16} \cdot \frac{1}{2} = \frac{7}{16}$$

e pela regra de Bayes 2.13

$$P[1|CD] = \frac{P[CD|1]P[1]}{P[CD]} = \frac{\frac{6}{16} \cdot \frac{1}{2}}{\frac{7}{16}} = \frac{3}{7}$$

As probabilidades *a priori* das duas urnas eram iguais mas a observação do acontecimento CD modificou as probabilidades que se devem atribuir a cada urna.

Exemplo 2.12 *Num sistema de comunicação binário o utilizador introduz na entrada sequências de símbolos 0 ou 1 que são em seguida transmitidos e sujeitos a erros no canal de transmissão. No receptor uma decisão é tomada quando o*

sinal é recebido atribuindo-se de novo os valores 0 ou 1. Se α for a probabilidade de erro, calcule qual entrada é mais provável se a saída temos um 1, admitindo que as entradas 0 ou 1 são equiprováveis.

Seja A_k o acontecimento "a entrada é k " e B_k o acontecimento "a saída é k ", $k = 0, 1$. A_0 e A_1 constituem uma partição do espaço de amostragem do espaço de amostragem das entradas e pode-se aplicar o teorema das probabilidades totais

$$P[B_1] = P[B_1|A_0]P[A_0] + P[B_1|A_1]P[A_1] = \alpha \cdot \frac{1}{2} + (1 - \alpha) \cdot \frac{1}{2} = \frac{1}{2}$$

As probabilidades a posteriori serão

$$P[A_0|B_1] = \frac{P[B_1|A_0]P[A_0]}{P[B_1]} = \frac{\frac{\alpha}{2}}{\frac{1}{2}} = \alpha$$

$$P[A_1|B_1] = \frac{P[B_1|A_1]P[A_1]}{P[B_1]} = \frac{\frac{1 - \alpha}{2}}{\frac{1}{2}} = 1 - \alpha$$

Se $\alpha < 1/2$ então a entrada mais provável será 1 se a saída for 1.

2.7 Independência de Acontecimentos

Quando o conhecimento sobre a ocorrência de B não altera a probabilidade de um outro acontecimento A , diz-se que os dois acontecimentos são independentes. Portanto quando dois acontecimentos são independentes $P[A|B] = P[A]$, ou seja atendendo a 2.10

$$P[A] = P[A|B] = \frac{P[A \cap B]}{P[B]}$$

e se $P[B] \neq 0$, $P[A \cap B] = P[A]P[B]$.

Definição 2.2 Dois acontecimentos A e B são independentes se e só se

$$P[A \cap B] = P[A]P[B] \quad (2.14)$$

quaisquer que sejam as $P[A]$ e $P[B]$.

Nestas condições $P[A|B] = P[A]$ e $P[B|A] = P[B]$.

Em geral dois acontecimentos que sejam mutuamente exclusivos e tenham probabilidade não nula, não podem ser independentes. De facto se $A \cap B = \emptyset \Rightarrow P[A \cap B] = 0$ e portanto ou $P[A] = 0$ ou $P[B] = 0$.

Um conjunto de n acontecimentos A_1, \dots, A_n são independentes quando o são 2 a 2, 3 a 3, ...e n a n , isto é, quando se verificarem as condições

$$P[A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}] = P[A_{i_1}]P[A_{i_2}] \dots P[A_{i_k}]$$

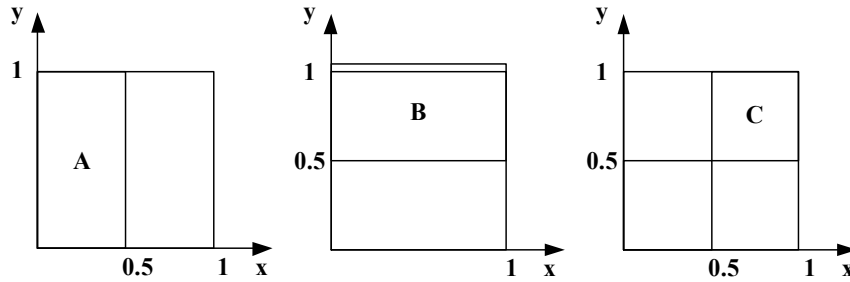
para $2, 3, \dots, n$ e $1 \leq i_1 < i_2 < \dots < i_k \leq n$.

Para tal é necessário definir um número N_T de condições que é igual a:

$$\begin{aligned} N_T &= C_2^n + C_3^n + \dots + C_n^n = \\ &= C_0^n + C_1^n + C_2^n + C_3^n + \dots + C_n^n - C_0^n - C_1^n = \\ &= \sum_{k=0}^n C_k^n - C_0^n - C_1^n = 2^n - 1 - n \end{aligned}$$

Note que se se verificar a independência 2 a 2, isso não implica independência global. Para tal verificar basta o seguinte exemplo que existe independência para os pares de acontecimento mas não existe independência quando se consideram os 3 acontecimentos A, B e C .

Exemplo 2.13 Considerem-se os 3 acontecimentos definidos geometricamente na seguinte figura, num espaço de amostragem $S = \{(x, y) : 0 \leq x \leq 1 \wedge 0 \leq y \leq 1\}$



$$\begin{aligned} A &= \{(x, y) : 0 < x < 1/2 \wedge 0 < y < 1\}, \\ B &= \{(x, y) : 0 < x < 1 \wedge 1/2 < y < 1\}, \\ C &= \{(x, y) : (0 < x < 1/2 \wedge 0 < y < 1/2) \vee (1/2 < x < 1 \wedge 1/2 < y < 1)\} \end{aligned}$$

Exemplo 2.14 As probabilidades dos acontecimentos são:

$$P[A] = P[B] = P[C] = 1/2$$

São independentes 2 a 2 pois

$$P[A \cap B] = 1/4 = P[A]P[B], P[AC] = \frac{1}{4} = P[A]P[C], P[CB] = \frac{1}{4} = P[C]P[B]$$

Mas não são independentes os três quando tomados em conjunto já que

$$P[A \cap B \cap C] = 0 \text{ e } P[A]P[B]P[C] = \frac{1}{8}$$

A definição geral de independência de acontecimentos não é um conceito fácil de utilizar dada a complexidade de condições a que obriga. A aplicação mais comum do conceito de independência é na análise de problemas associados a sucessivas experiências. A hipótese comum que se aceita é a de que acontecimentos resultantes de diferentes experiências são independentes.

2.7.1 Sequências de acontecimentos independentes

Considere-se uma experiência aleatória composta por uma sucessão E_1, E_2, \dots, E_n de subexperiências. Em geral o seu resultado será um n -tuplo $r = (r_1, r_2, \dots, r_n)$ em que r_k é o resultado da experiência parcelar E_k . Este resultado pertencerá ao espaço de amostragem que se obtém pelo produto cartesiano dos espaços de amostragem para cada experiência parcelar. De acordo com a hipótese anteriormente posta, pode-se admitir que os resultados parcelares são independentes. Se se considerar os acontecimentos $A_k, k = 1, \dots, n$, em que A_k é um acontecimento que diz respeito apenas à experiência de ordem k , é razoável admitir que os acontecimentos $A_k, k = 1, \dots, n$ sejam independentes e

$$P[A_1 \cap A_2 \cap \dots \cap A_n] = P[A_1]P[A_2] \dots P[A_n] \quad (2.15)$$

Define-se experiência de Bernouilli como uma experiência em que se pretende observar se ocorre ou não um dado acontecimento A . Haverá, então, apenas dois resultados possíveis mutuamente exclusivos: a ocorrência de A a que se pode associar "sucesso" e a não ocorrência ou "insucesso". Pode-se também usar outra simbologia do tipo binário, como "sim"/"não" ou "1/0". O espaço de amostragem será pois $S_B = \{\text{sucesso}, \text{insucesso}\}$, $S_B = \{\text{sim}, \text{não}\}$ ou $S_B = \{1, 0\}$.

Suponhamos que o acontecimento A tem uma probabilidade de sucesso de $p = P[A]$ e portanto uma de insucesso de $1 - p = P[\bar{A}]$.

Determinemos a probabilidade de k sucessos em n experiências sucessivas, que se representará por $p_n(k)$. Procura-se portanto a probabilidade de uma sequência de acontecimento independentes do tipo

$$\underbrace{A, \dots, A}_{k \text{ vezes}}, \underbrace{\bar{A}, \dots, \bar{A}}_{n-k \text{ vezes}}$$

e portanto

$$P[A \cap A \cap \dots \cap A \cap \bar{A} \cap \dots \cap \bar{A}] = p^k(1-p)^{n-k}$$

Como os k sucessos em n podem ocorrer não apenas por esta ordem, mas segundo uma ordem arbitrária das C_k^n possíveis, a probabilidade que se procura determinar obtém-se somando as probabilidades para todas as C_k^n hipóteses de se obterem k sucessos em n experiências, ou seja

$$p_n(k) = C_k^n p^k (1-p)^{n-k} \quad (2.16)$$

Esta lei de probabilidade é conhecida por **Lei Binomial**.

Exemplo 2.15 Qual a probabilidade de em 10 lançamentos de uma moeda saírem 5 vezes caras?

$$p_{10}(5) = C_5^{10} p^5 (1-p)^5 \text{ e se } p = 1/2, p_{10}(5) = \frac{10!}{5!5!} \left(\frac{1}{2}\right)^{10} = \frac{63}{256}$$

Exemplo 2.16 Um canal de comunicação introduz erros com uma probabilidade $\epsilon = 10^{-3}$. O transmissor envia cada bit 3 vezes e no receptor existe um decodificador que decide qual o bit recebido depois de aplicar uma regra de maioria. Esta regra consiste em decidir pelo bit mais representado nos 3 recebidos, ou seja, se 001 decide 0, se 101 decide 1, etc. Qual a probabilidade de tomar a decisão incorreta? Se houver apenas um erro a decisão será correta, o erro é detetado e corrigido. A resposta será pois a probabilidade de haver 2 ou mais erros em 3 tentativas. então aplicando a lei de probabilidade 2.16 obtem-se

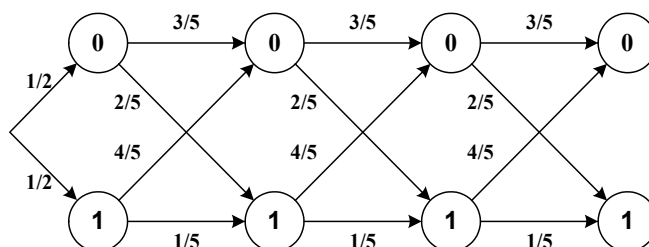
$$\begin{aligned} P[n^\circ \text{ de erros} > 1] &= \sum_{k=2}^3 C_k^3 \epsilon^k (1 - \epsilon)^{3-k} = \sum_{k=2}^3 C_k^3 0.001^k (1 - 0.001)^{3-k} = \\ &= 2.998 \times 10^{-6} \end{aligned}$$

Note que este exemplo apresenta uma técnica de corrente uso nas modernas telecomunicações: o melhoramento do comportamento do canal de transmissão usando redundância, isto é, usando mais informação (neste caso 3 vezes mais) do que a necessária para compor a mensagem.

2.7.2 Sequências de acontecimentos dependentes

Considere-se uma cadeia de experiências tal que o resultado de uma experiência determina a maneira de se realizar a seguinte.

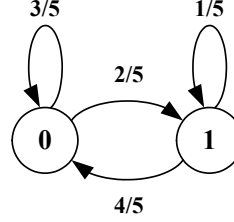
Exemplo 2.17 Considere a experiência aleatória que consiste em retirar uma bola numerada de uma de duas urnas (havendo reposição): uma urna (identificada pelo número 0) tem 3 bolas com o número 0 e 2 com o número 1; a outra urna (identificada pelo número 1) tem 4 bolas com o número 0 e 1 com o 1. A urna inicial é determinada lançando uma moeda ao ar e em seguida utiliza-se a primeira se o resultado for 0 e a segunda se o resultado for 1.



A observação que se faz é o registo da sequência das urnas utilizadas na extração. Determine a probabilidade de se observar a sequência inicial 0011. Atendendo ao enunciado a probabilidade desta sequência será

$$P[0] P[0|0] P[1|0] P[1|1] = \frac{1}{2} \cdot \frac{3}{5} \cdot \frac{2}{5} \cdot \frac{1}{5} = \frac{3}{125}$$

Note que em alternativa ao esquema gráfico anterior, diagrama em árvore em que se mostram as possíveis transições entre urnas para sucessivas experiências, se pode indicar as transições e respetivas probabilidades através do seguinte diagrama, conhecido por diagrama de (transição de) estados:



Suponhamos que queremos calcular a probabilidade de uma sequência de observações o_0, o_1, \dots, o_n : atendendo a 2.11

$$P[o_0 \cap o_1 \cap \dots \cap o_n] = P[o_n | o_0 \cap o_1 \cap \dots \cap o_{n-1}] P[o_0 \cap o_1 \cap \dots \cap o_{n-1}]$$

Aplicando sucessivamente 2.11 e reordenando, obtém-se:

$$P[o_0 \cap o_1 \cap \dots \cap o_n] = P[o_0] P[o_1 | o_0] P[o_2 | o_0 \cap o_1] \dots P[o_n | o_0 \cap o_1 \cap \dots \cap o_{n-1}]$$

Há um caso de especial importância: se para qualquer $j \in \{0, \dots, n-1\}$

$$P[o_{j+1} | o_0 \cap o_1 \cap \dots \cap o_j] = P[o_{j+1} | o_j]$$

a sequência assume o nome de **Cadeia de Markov** e a probabilidade da observação passa a ser

$$P[o_0 \cap o_1 \cap \dots \cap o_n] = P[o_0] P[o_1 | o_0] P[o_2 | o_1] \dots P[o_n | o_{n-1}]$$

Associando o índice da observação a um tempo, pode-se afirmar que nestas condições a observação futura $(n+1)$ depende apenas do presente (n) e não do passado $(n-1)$.

2.8 Exercícios

Exercício 2.2 Quais das seguintes afirmações são verdadeiras:

1. $b \in \{a, b, c\}$
2. $b \subset \{a, b, c\}$
3. $\{b\} \in \{a, b, c\}$
4. $\{b\} \subset \{a, b, c\}$

Respostas:

1. verdadeiro
2. falso
3. falso
4. verdadeiro

Exercício 2.3 Quantos subconjuntos pode definir a partir do conjunto $S = \{1, 2, 3\}$ e liste-os.

$2^3 = 8$ subconjuntos. $\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, S$

Exercício 2.4 Se $A_n = [0, 2^{1/n}]$, $n \geq 1$ determine $\bigcap_{i=1}^{\infty} A_n$
Atendendo a

$$2^{1/n} > 1 \Rightarrow \forall_{n \geq 1} [0, 1] \subset A_n$$

e

$$[0, 1] \subset \bigcap_{i=1}^{\infty} A_n$$

Como

$$\lim_{n \rightarrow \infty} 2^{1/n} = 1, \forall_{x > 1} x \notin \bigcap_{i=1}^{\infty} A_n$$

Então

$$\bigcap_{i=1}^{\infty} A_n = [0, 1]$$

Exercício 2.5 Demonstre que se A e B são independentes, os pares de acontecimentos A e \bar{B} , \bar{A} e B e \bar{A} e \bar{B} também o são.

$$A = A \cap S = A \cap (B \cup \bar{B}) = (A \cap B) \cup (A \cap \bar{B})$$

Pelo axioma 2.3

$$\begin{aligned} P[A] &= P[A \cap B] + P[A \cap \bar{B}] \\ P[A \cap \bar{B}] &= P[A] - P[A \cap B] = P[A](1 - P[B]) = P[A]P[\bar{B}] \end{aligned}$$

Analogamente para A e \bar{B}

$$\bar{A} = \bar{A} \cap S = \bar{A} \cap (B \cup \bar{B}) = (\bar{A} \cap B) \cup (\bar{A} \cap \bar{B})$$

Pelo axioma 2.3

$$\begin{aligned} P[\bar{A}] &= P[\bar{A} \cap B] + P[\bar{A} \cap \bar{B}] \\ P[\bar{A} \cap \bar{B}] &= P[\bar{A}] - P[\bar{A} \cap B] = P[\bar{A}](1 - P[B]) = P[\bar{A}]P[\bar{B}] \end{aligned}$$

Exercício 2.6 Demonstre que se $A \subset B$ e $C \subset D$ então $(A \cap C) \subset (B \cap D)$
 Se

$$A \subset B : x \in A \Rightarrow x \in B$$

$$C \subset D : x \in C \Rightarrow x \in D$$

$$A \cap C = \{x : x \in A \wedge x \in C\} \subset \{x : x \in B \wedge x \in D\} = B \cap D$$

Exercício 2.7 Num certo país estão registados 50000 automóveis numerados de 1 a 50000. Escolhendo-se um automóvel ao acaso qual a probabilidade do número de matrícula começar por 4?

Número de casos possíveis= 50000

Número de casos favoráveis:

intervalo	n_f
[1, 9]	1
[10, 99]	10
[100, 999]	100
[1000, 9999]	10000
total	11111

$$p = \frac{11111}{50000} = 0.22$$

Exercício 2.8 Numa urna com M bolas vermelhas e N bolas pretas extraem-se com reposição t bolas. Determine a probabilidade de serem m vermelhas e n pretas?

Número de casos possíveis= $C_t^{M+N} = C_{m+n}^{M+N}$

Número de casos favoráveis=número de grupos possíveis de m em M bolas vermelhas \times número de grupos possíveis de n em N bolas pretas= $C_m^M \cdot C_n^N$

$$p = \frac{C_{m+n}^{M+N}}{C_m^M C_n^N}$$

Exercício 2.9 Considere a seguinte experiência aleatória: Lançamento de um dado 12 vezes e registe a sequência numérica. Qual a probabilidade de cada número aparecer duas vezes?

Exercício 2.10 Número de casos possíveis = $A_{12}^6 = 6^{12}$

$$\text{Número de casos favoráveis} = P_{2,2,2,2,2,2}^{12} = \frac{12!}{2!2!2!2!2!2!}$$

$$p = \frac{1925}{559872} = 3.4383 \times 10^{-3}$$

Exercício 2.11 Uma série de 50 peças fabricadas contém 10 defeituosas. Suponha que escolhe aleatoriamente 10 peças e as ensaia. Qual a probabilidade de encontrar 5 defeituosas?

Exercício 2.12 Número de casos possíveis $= C_{10}^{50} = 10\,272\,278\,170$

Número de casos favoráveis = número maneiras de escolher 5 peças defeituosas em $10 \times$

\times o número de maneiras de escolher 5 peças boas em $40 =$

$$= C_5^{10} \cdot C_5^{40} = \frac{10!}{5!5!} \cdot \frac{40!}{5!35!} = 165\,818\,016$$

$$p = \frac{165\,818\,016}{10\,272\,278\,170} = 1.6142 \times 10^{-2}$$

Exercício 2.13 Calcule a probabilidade de dois ou mais indivíduos de um grupo de n tenham o aniversário no mesmo dia do ano (considere o ano como tendo sempre 365 dias)

$P[2 \text{ ou mais aniversários no mesmo dia}] = 1 - P[\text{todos os aniversários em dias diferentes}]$

Vamos, pois, determinar a probabilidade deste segundo acontecimento.

1- Usando técnicas de contagem e a regra de Laplace 2.7 :

Número de casos possíveis $= A_k^{365} = 365^k$

Número de casos favoráveis $= A_k^{365} = 365 \dots (365 - k + 1)$

2- Encarando a questão como uma cadeia de acontecimentos independentes:

probabilidade do 1º fazer anos num dia do ano $= \frac{365}{365};$

probabilidade do 2º fazer anos num dia diferente $= \frac{364}{365}; \dots;$

probabilidade do de ordem k fazer anos num dia diferente de todos os anteriores $= \frac{365 - k + 1}{365}$

Em qualquer dos casos

$$P[2 \text{ ou mais aniversários no mesmo dia}] = 1 - \frac{365 \dots (365 - k + 1)}{365^k}$$

Para $k = 30$ esta probabilidade é de 0.71 e para $k = 40$ esta probabilidade é de 0.89

Exercício 2.14 Num teste de avaliação com 50 perguntas de escolha entre 4 hipóteses suponha que responde aleatoriamente.

1. Qual a probabilidade de acertar em todas as respostas?
2. Qual a probabilidade de não acertar em nenhuma?
3. Qual a probabilidade de acertar em metade das respostas?

Como as respostas são experiências independentes, o número de vezes que se realiza o acontecimento "acertar a resposta" segue uma lei binomial: $p_{50}(k) = C_k^{50} (1/4)^k (3/4)^{50-k}$

$$1. p_{50}(50) = (1/4)^{50} = 7.9 \times 10^{-31}$$

$$2. p_{50}(0) = (3/4)^{50} = 5.6632 \times 10^{-7}$$

$$3. p_{50}(25) = \frac{50!}{25!25!} (1/4)^{25} (3/4)^{25} = 8.5 \times 10^{-5}$$

Exercício 2.15 Um fabricante de material eletrônico recebe transistores de 3 diferentes fabricantes A, B e C que têm uma probabilidade de defeito de 0.001, 0.005 e 0.010 respectivamente e considere o stock constituído por transistores vindos dos três fabricantes na proporção p_A, p_B e p_C . Se um transistor escolhido ao acaso é defeituoso, qual a probabilidade de ter sido pelo fornecedor A, B ou C ?

Exercício 2.16 Pela regra de Bayes 2.13 :

$$P[A|def] = \frac{P[def|A] \cdot P[A]}{P[def]} = \frac{10^{-3}p_A}{10^{-3}p_A + 5 \cdot 10^{-3}p_B + 10 \cdot 10^{-3}p_C}$$

$$P[A|def] = \frac{p_A}{p_A + 5p_B + 10p_C}$$

e analogamente

$$P[B|def] = \frac{5p_B}{p_A + 5p_B + 10p_C}$$

$$P[C|def] = \frac{10p_C}{p_A + 5p_B + 10p_C}$$

Exercício 2.17 Um processo de fabrico produz uma mistura de chips de memória bons e maus. Um chip bom tem um tempo de vida que segue uma lei exponencial com uma taxa de avarias de a . Os chips maus seguem também uma lei semelhante mas com uma taxa de avarias $1000a$. Suponha que a fração de chips bons é $1 - p$ e a de maus é p .

1. Calcule a probabilidade de um chip arbitrariamente escolhido ainda estar a funcionar ao fim de t
2. Suponha que para diminuir as falhas do produto colocado no mercado, todos os chips são testados em fabrica durante t_1 horas e os que falham são destruídos. Determine o valor de t_1 para o qual 99% dos chips expedidos são bons, admitindo que $p = 0.1$ e $1/a = 20000$ horas.

A lei exponencial significa que a probabilidade do chip exceder t é

$$P[\text{vida} > t] = P[[t, \infty)) = e^{-at}$$

Seja C o acontecimento "chip a funcionar ao fim de t ", B o acontecimento "chip bom" e M o acontecimento "chip mau".

Do enunciado sabe-se que: $P[B] = 1 - p, P[M] = p, P[C|B] = e^{-at}$ e $P[C|M] = e^{-1000at}$.

1. Pelo teorema das probabilidades totais:

$$P[C] = P[C|B] P[B] + P[C|M] P[M] = (1 - p)e^{-at} + pe^{-1000at}$$

2. Pede-se a probabilidade de estar bom depois de se ter verificado o acontecimento C , ou seja, $P[B|C]$.

Aplicando a Regra de Bayes 2.13:

$$\begin{aligned} P[B|C] &= \frac{P[C|B] P[B]}{P[C]} = \frac{(1-p)e^{-at}}{(1-p)e^{-at} + pe^{-1000at}} = \\ &= \frac{1}{1 + \frac{pe^{-1000at}}{(1-p)e^{-at}}} = \frac{1}{1 + \frac{(1-p)e^{-999at}}{p}} \end{aligned}$$

Procura-se o valor de t para que $P[B|C] = 0.99$; resolvendo a equação em ordem a t obtém-se

$$t_1 = \frac{1}{999a} \ln\left(99 \frac{p}{1-p}\right)$$

que para $a = 20000^{-1}/\text{hora}$ e $p = 0.1$ conduz ao resultado $t_1 \simeq 48\text{horas}$.

Exercício 2.18 Problema da roleta russa

A roleta russa é jogada por dois jogadores que utilizam um revólver que tem uma peça cilíndrica rotativa (o tambor) com lugar para seis balas e onde apenas é colocada uma. O primeiro jogador faz rodar o cilindro e aponta à sua cabeça e dispara. Se não morrer passa ao outro jogador que repete os mesmos passos. O jogo continua até à morte de um dos jogadores. Determine a probabilidade de morte para cada jogador.

Considere o acontecimento X_i como "o disparo ocorreu na experiência de ordem i " que tem a probabilidade $P[X_i] = \left(\frac{5}{6}\right)^{i-1} \frac{1}{6}$. Então sendo $P[A]$ a probabilidade do primeiro jogador morrer, ela será:

$$P[A] = \frac{1}{6} \left(1 + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^4 + \dots \right) = \frac{1}{6} \frac{1}{1 - (5/6)^2} = \frac{6}{11}$$

A probabilidade de morte do segundo será $P[B] = 1 - P[A] = \frac{5}{11}$

Exercício 2.19 Problema da coincidência⁴

Numa urna existem N bolas numeradas de 1 a N . As bolas vão sendo retiradas da urna até esta ficar vazia. Se a bola número i for a obtida na extração de ordem i , diz-se que houve uma coincidência. Determine a probabilidade de haver pelo menos uma coincidência.

Se $A_i =$ "coincidência na extração i " a probabilidade procurada é $p = P\left[\bigcup_{i=1}^N A_i\right]$ que de acordo com 2.5 vale

$$p = \sum_{r=1}^N (-1)^{r-1} \sum_{1 \leq i_1 < \dots < i_r \leq N} P[A_{i_1} \cap \dots \cap A_{i_r}]$$

⁴Conhecido pela sua designação em francês, problema do *rencontre*.

e como

$$\sum_{1 \leq i_1 < \dots < i_r \leq r} P[A_{i_1} \cap \dots \cap A_{i_r}] = C_r^N \frac{(N-r)!}{N!} = \frac{1}{r!}$$

coduz a

$$p = \sum_{r=1}^N \frac{(-1)^{r-1}}{r!} = 1 - \frac{1}{2!} + \dots + (-1)^{N-1} \frac{1}{N!}$$

que tende para $1 - e^{-1} \simeq 0.632$ quando $N \rightarrow \infty$

Exercício 2.20 O prémio na embalagem

Em cada embalagem de comida é distribuído uma senha que pode ser de N diferentes tipos, e que foram distribuídas aleatoriamente nas caixas com igual probabilidade. Um prémio será entregue se se apresentar um conjunto completo das senhas. Se uma pessoa comprar n embalagens, qual a probabilidade de obter um conjunto completo e ganhar o prémio?

Designando por A_i o acontecimento "falta a senha nº i nas n embalagens", a probabilidade de faltar pelo menos uma senha será $P\left[\bigcup_{i=1}^N A_i\right]$ e a probabilidade procurada será $p = 1 - P\left[\bigcup_{i=1}^N A_i\right]$. Como a probabilidade de uma embalagem não conter a senha de tipo i_1, \dots, i_r é $(1 - r/N)$ a $P[A_{i_1} \cap \dots \cap A_{i_r}] = (1 - r/N)^n$, a probabilidade p será, atendendo a 2.5:

$$\begin{aligned} p &= 1 - \sum_{r=1}^N (-1)^{r-1} \sum_{1 \leq i_1 < \dots < i_r \leq r} P[A_{i_1} \cap \dots \cap A_{i_r}] = \\ &= 1 - \sum_{r=1}^N (-1)^{r-1} C_r^N (1 - \frac{r}{N})^n. \end{aligned}$$

Para $N = 5$ e $n = 10$ obtem-se $p = 1 - \sum_{r=1}^{10} (-1)^{r-1} C_r^{10} (1 - \frac{r}{5})^{10} = 0.37$.

Exercício 2.21 Considere a experiência aleatória 2 escolha de um número no intervalo $[-1, 1]$ e os dois acontecimentos $A = \{|x - 1/2| < 1\}$ e $B = \{x > 3/4\}$.

Determine as probabilidades de $A, B, A|B$ e $B|A$ e faça um programa que permita calcular uma aproximação a estas probabilidades.

Como

$$\begin{aligned} A &= \{|x - 1/2| < 1\} \cap S = (-1/2, 3/2) \cap [-1, 1] = (-1/2, 1] \\ P[A] &= \frac{1 - (-1/2)}{1 - (-1)} = \frac{3}{4} \end{aligned}$$

$$\begin{aligned}
P[B] &= \frac{1 - 3/4}{1 - (-1)} = \frac{1}{8} \\
P[A|B] &= \frac{P[A \cap B]}{P[B]} = \frac{P[B]}{P[B]} = 1 \\
P[B|A] &= \frac{P[A \cap B]}{P[A]} = \frac{P[B]}{P[A]} = \frac{1/8}{3/4} = \frac{1}{6}
\end{aligned}$$

```

n=100000;
x=2*(rand(1,n)-.5);
fa=sum(abs(x-.5)<1)/n;
fb=sum(x>.75)/n;
fa_e_b=sum(abs(x-.5)<1&x>.75)/n;
fa_se_b=fa_e_b/fb;
fb_se_a=fa_e_b/fa;
fprintf(1,'P[A]=%5.3f,P[B]=%5.3f\n',fa,fb)
fprintf(1,'P[A|B]=%5.3f,P[B|A]=%5.3f\n',fa_se_b,fb_se_a)
que conduz aos resultados
P[A]=0.750,P[B]=0.125                P[A]=0.748,P[B]=0.127
P[A|B]=1.000,P[B|A]=0.166            P[A|B]=1.000,P[B|A]=0.170

```

Exercício 2.22 Um circuito contém 10 chips. Admitindo que cada chip tem um tempo de vida que segue uma lei exponencial com uma constante a , determine a probabilidade de pelo menos metade dos chips estarem a funcionar ao fim de um tempo igual a $1/a$

Para cada chip $P[\text{vida} > 1/a] = e^{-a/a} = \frac{1}{e}$.

Admitindo que o estado de funcionamento de cada chip é uma experiência de Bernouilli, pois só pode assumir dois estados - a funcionar com probabilidade $\frac{1}{e}$ ou avariado com a probabilidade complementar - e como é independente dos restantes estamos perante um problema que segue a lei binomial 2.16 e a probabilidade procurada será dada por

$$P[k \geq 5] = \sum_{k=5}^{10} C_k^{10} \left(\frac{1}{e}\right)^k \left(1 - \frac{1}{e}\right)^{10-k} = 0.28897$$

Exercício 2.23 Um bloco de informação contendo 100 bits é transmitido através de um canal de comunicação que tem uma probabilidade de erro de 10^{-3} . Determine a probabilidade de na recepção um bloco tenha 3 ou mais erros.

Faça um programa que permita calcular uma aproximação a esta probabilidade

Considerando que a transmissão de cada bit é uma experiência isolada e independente da transmissão dos outros bits, o número de erros segue uma lei

binomial de parâmetros $n = 100$ e $p = 0.001$.

$$\begin{aligned} P[k > 3] &= 1 - P[k \leq 2] = 1 - \sum_{k=0}^2 C_k^{100} \cdot 0.001^k \cdot 0.999^{100-k} = \\ &= 1.5 \times 10^{-4} \end{aligned}$$

```
%Defina o número de experiências N
%e o número de séries de experiências n
N=10000;n=20;
%Gere valores N valores de uma variável
%de Poisson com os parâmetros 100,0.001
% e repita n vezes
for i=1:n
x=binomial(N,100,.001);
freq(i)=sum(x\>=3);
end
terros=mean(freq)/N;
fprintf(1,'~taxa de erros=%8.6f\n',terros)
que conduz ao resultado
taxa de erros=0.000145
```

Exercício 2.24 *Determine o número mínimo de lançamentos de um dado que deve fazer para ter uma probabilidade de 90% de pelo menos uma vez ter obtido 6.*

Como os lançamentos são independentes, o número de vezes que se observa 6 segue uma lei de probabilidade binomial e a problema pode ser reduzido à seguinte inequação:

$$P[k \geq 1] = \sum_{k=1}^n C_k^n (1/6)^k (5/6)^{n-k} > 0.9$$

Para resolver esta inequação pode-se recorrer ao seguinte programa de Matlab

```
n=1;p=0;
while p<0.9
p=0;
for k=1:n
p=p+prod(n-k+1:n)/prod(1:k)*5^(n-k)/6^n;
end
n=n+1;
end
n=n-1;
fprintf(1,'nº de lança.=%3.0f, P[k>1]=%5.3f\n',n,p)
com o resultado
nº de lança.= 13, P[k>1]=0.907
```

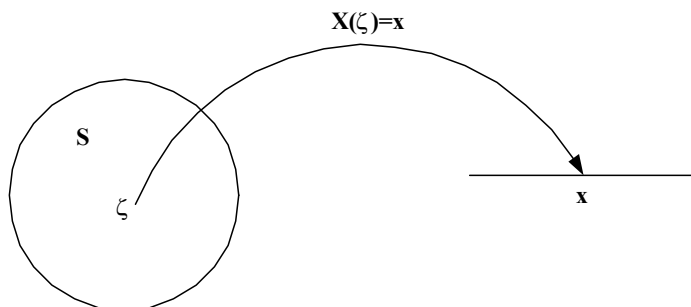
Chapter 3

Variável Aleatória

3.1 Definição de variável aleatória

Variável aleatória é uma aplicação de S em \mathbb{R} , isto é, para cada elemento de S associa-se um número real $x = X(\zeta)$.

$$X : \zeta \in S \rightarrow X(\zeta) = x \in \mathbb{R}$$



O espaço de amostragem S é o domínio da variável aleatória e o conjunto $S_X \subset \mathbb{R}$ de todos os valores possíveis para X é o seu contradomínio ¹ em \mathbb{R} .

Exemplo 3.1 Considere a experiência aleatória "lançamento de uma moeda 4 vezes" e defina a variável aleatória X como sendo a o número de vezes que sai "caras" nos 4 lançamentos.

Caraterize o conjunto de amostragem e o contradomínio de X .

$$S = \left\{ \begin{array}{l} CCCC, CCCR, CCRC, CRCC, RCCC, CRRR, CRRC, RCCR, \\ CRRC, RCRC, RRCC, CRRR, CCRC, CRCC, RCCC, RRRR \end{array} \right\}$$

$$S_X = \{0, 1, 2, 3, 4\}$$

¹Na literatura de língua inglesa designado por *range*, que se poderá traduzir por "gama".

A variável aleatória $X(\zeta)$ fica então definida por:

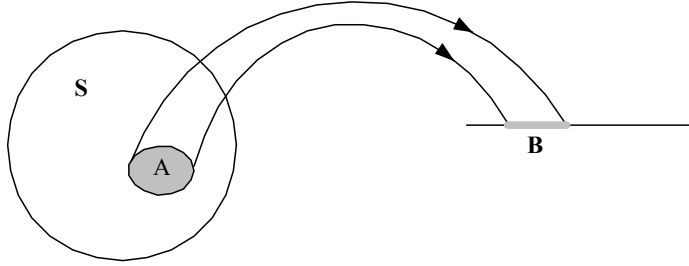
ζ	$CCCC$	$CCCR$	$CCRC$	$CRCC$
$X(\zeta)$	4	3	3	3
ζ	$RCCC$	$CCRR$	$CRCR$	$RCCR$
$X(\zeta)$	3	2	2	2
ζ	$CRRC$	$RCRC$	$RRCC$	$CRRR$
$X(\zeta)$	2	2	2	1
ζ	$CCRC$	$CRCC$	$RCCC$	$RRRR$
$X(\zeta)$	1	1	1	0

Para se definirem as probabilidades de acontecimentos em S_X é necessário recorrer à atribuição de probabilidades em S . Seja A um acontecimento em S e $B \subset \mathbb{R}$ o conjunto imagem de A através de X

$$A = \{\zeta : X(\zeta) \in B\}$$

Então como B ocorre sempre que A ocorre, deve-se atribuir a mesma probabilidade a ambos. Nesse sentido designar-se-ão A e B como acontecimentos equivalentes:

$$P[B] = P[A] = P[\{\zeta : X(\zeta) \in B\}]$$



Exemplo 3.2 Definam-se , então, as probabilidades para o exemplo anterior, supondo $P[C] = p$, $P[R] = 1 - p$

Usando a definição de acontecimentos equivalentes pode-se atribuir as probabilidades em S_X da seguinte forma:

O acontecimento $\{0\}$ é equivalente a $\{RRRR\}$ e terão a mesma probabilidade $p_0 = (1 - p)^4$

O acontecimento $\{1\}$ é equivalente a $\{CRRR, CCRC, CRCC, RCCC\}$ e terão a mesma probabilidade $p_1 = C_1^4 p(1 - p)^3$

Em geral, ter-se-á:

$$p_k = P[X(\zeta) = k] = C_k^4 p^k (1 - p)^{4-k}$$

Note que se o resultado de uma experiência aleatória é um valor numérico, pode-se considerar esse resultado como uma variável aleatória definida pela função identidade $X(\zeta) = \zeta$. Sendo assim a maior parte dos casos apresentados no capítulo 2 podem ser reinterpretados como variáveis aleatórias.

A partir daqui, por razões de simplificação da notação, vai usar-se apenas o símbolo X (ou qualquer outra maiúscula) em vez de $X(\zeta)$ para representar uma variável aleatória.

3.2 Função de distribuição

Definição 3.1 *Função de distribuição de uma variável aleatória X é a função real de variável real*

$$F_X : x \in \mathbb{R} \rightarrow F_X(x) = P[X(\zeta) \leq x] \in \mathbb{R}$$

3.2.1 Propriedades da função de distribuição

$$1. 0 \leq F_X(x) \leq 1$$

Demonstração. $F_X(x)$ é uma probabilidade e pelos axioma 2.1 e corolário 2 apresentados qualquer probabilidade pertence a este intervalo ■

$$2. \lim_{x \rightarrow \infty} F_X(x) = 1$$

Demonstração. $\{X < \infty\}$ é o acontecimento certo e a sua probabilidade é igual a 1 ■

$$3. \lim_{x \rightarrow -\infty} F_X(x) = 0$$

Demonstração. $\{X < -\infty\} = \emptyset$ é o acontecimento impossível e a sua probabilidade é igual a 0 ■

$$4. F_X(x) \text{ é uma função não decrescente de } x, \text{ isto é, } a < b \Rightarrow F_X(a) \leq F_X(b)$$

Demonstração. Se $a < b \Rightarrow \{X \leq a\} \subset \{X \leq b\}$ e pelo corolário ?? $P[X \leq a] \leq P[X \leq b]$ ■

$$5. F_X(x) \text{ é uma função contínua à direita, isto é, para } h > 0,$$

$$F_X(b) = \lim_{h \rightarrow 0} F_X(b+h) = F_X(b^+)$$

Demonstração. Quando $h \rightarrow 0$ por valores positivos

$$\begin{aligned} \{X \leq b+h\} &\rightarrow \{X \leq b\} \\ P[X \leq b+h] &\rightarrow P[X \leq b] \end{aligned}$$

■

$$6. P[a < X \leq b] = F_X(b) - F_X(a)$$

Demonstração.

$$\{X \leq a\} \cup \{a < X \leq b\} = \{X \leq b\}$$

e como os conjuntos no primeiro membro são disjuntos pelo 3º axioma

$$\begin{aligned} P[X \leq a] + P[a < X \leq b] &= P[X \leq b] \\ F_X(a) + P[a < X \leq b] &= F_X(b) \end{aligned}$$

■

$$7. P[X = b] = F_X(b) - F_X(b^-)$$

Demonstração. Para $h > 0$, da propriedade anterior

$$P[b - h < X \leq b] = F_X(b) - F_X(b - h)$$

Como quando $h \rightarrow 0$

$$\begin{aligned} \{b - h < X \leq b\} &\rightarrow \{b\} \\ P[b - h < X \leq b] &\rightarrow P[\{b\}] = P[X = b] \end{aligned}$$

■

Se a função $F_X(x)$ for contínua em b , a $P[X = b] = 0$ e se for descontínua o valor da $P[X = b]$ será o do salto - diferença entre os valores à direita e à esquerda - no ponto b .

Usando estas últimas propriedades podem ainda deduzir-se as probabilidades para outros tipos de intervalo

$$\begin{aligned} P[a \leq X \leq b] &= F_X(b) - F_X(a^-) \\ P[a < X < b] &= F_X(b^-) - F_X(a) \\ P[a \leq X < b] &= F_X(b^-) - F_X(a^-) \end{aligned}$$

$$8. P[X > x] = 1 - F_X(x)$$

Demonstração. Como os acontecimentos $\{X > x\}$ e $\{X \leq x\}$ são complementares, o corolário 1 prova esta propriedade ■

Exemplo 3.3 Defina a função de distribuição para o exemplo anterior

Recorrendo novamente ao conceito de equivalência de acontecimentos em S e S_X :

Para qualquer $x < 0$ o conjunto de resultados em S que conduzem a $\{X \leq x\}$ é \emptyset e $F_X(x) = P[X \leq x] = P[\emptyset] = 0$.

Mas se $x = 0$ o conjunto de resultados em S que conduzem a $\{X \leq x\}$ é $\{RRRR\}$ e $F_X(0) = P[X \leq 0] = p_0$.

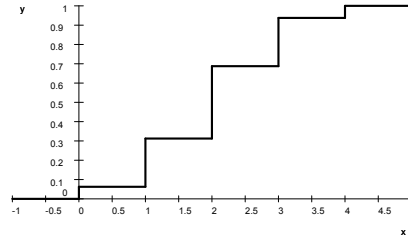
Para qualquer $0 \leq x < 1$ o conjunto de resultados em S que conduzem a $\{X \leq x\}$ é $\{RRRR\}$ e $F_X(x) = P[X \leq x] = P[X \leq 0] = p_0$.

Mas se $x = 1$ o conjunto de resultados em S que conduzem a $\{X \leq x\}$ é $\{RRRR, CRRR, CCRC, CRCC, RCCC\}$ e $F_X(1) = P[X \leq 1] = p_0 + p_1$.

Continuando com este tipo de raciocínio conclui-se que

$$F_X(x) = \begin{cases} 0 & , x < 0 \\ p_0 & , 0 \leq x < 1 \\ p_0 + p_1 & , 1 \leq x < 2 \\ p_0 + p_1 + p_2 & , 2 \leq x < 3 \\ p_0 + p_1 + p_2 + p_3 & , 3 \leq x < 4 \\ p_0 + p_1 + p_2 + p_3 + p_4 & , x \geq 4 \end{cases}$$

Para $p = 1/2$ esta função de distribuição terá o seguinte aspeto gráfico:



Note que os traços verticais estão a ser usados como indicadores de descontinuidade à esquerda da função.

Se se introduzir a função degrau unitário $u(x)$:

Definição 3.2

$$u(x) = \begin{cases} 0 & , x < 0 \\ 1 & , x \geq 0 \end{cases}$$

Pode-se então reescrever a anterior função de distribuição de uma forma mais compacta:

$$\begin{aligned} F_X(x) &= p_0 u(x) + p_1 u(x-1) + p_2 u(x-2) + p_3 u(x-3) + p_4 u(x-4) = \\ &= \sum_{k=0}^4 p_k u(x-k) \end{aligned}$$

A função $F_X(x)$ que apresenta descontinuidades em $x = 0, 1, 2, 3, 4$ pode ser representada por uma soma de degraus com a transição nesses pontos.

3.3 Função de densidade de probabilidade

Definição 3.3 Se F_X for diferenciável, a função densidade de probabilidade é a função real de variável real

$$f_X : x \in \mathbb{R} \rightarrow f_X(x) = \frac{dF_X(x)}{dx} \in \mathbb{R} \quad (3.1)$$

3.3.1 Propriedades da função de densidade de probabilidade

1. $f_X(x) \geq 0$

Demonstração. $F_X(x)$ é uma função não decrescente logo a sua derivada será não negativa ■

2. $P[a \leq X \leq b] = \int_a^b f_X(x)dx$

Demonstração. Como $F_X(x)$ é uma primitiva de $f_X(x)$ então

$$\int_a^b f_X(x)dx = F_X(b) - F_X(a)$$

■

Esta propriedade justifica o nome de densidade de probabilidade para esta função. Se $f_X(x)$ for contínua e se se considerar um intervalo Δx centrado em x a probabilidade desse intervalo será

$$P[x - \Delta x/2 < x < x + \Delta x/2] = \int_{-\Delta x/2}^{+\Delta x/2} f_X(x)dx \simeq f_X(x)\Delta x$$

com um erro da ordem $(\Delta x)^2$, portanto a probabilidade de X pertencer a um intervalo pequeno é dado pelo produto de um valor da densidade de probabilidade no interior do intervalo pela dimensão do intervalo.

3. $F_X(x) = \int_{-\infty}^x f_X(x)dx$

Demonstração. Resulta imediatamente da propriedade anterior substituindo o intervalo $[a, b]$ por $(-\infty, x)$. ■

4. $\int_{-\infty}^{+\infty} f_X(x)dx = 1$

Demonstração. O intervalo $(-\infty, +\infty) = \mathbb{R}$ é a imagem por X do acontecimento certo pelo que terá de ter a probabilidade = 1 ■

Desta propriedade resulta que para qualquer função real da variável real $g(x)$ definida não negativa, isto é, $\forall x \in \mathbb{R} g(x) \geq 0$ e que verifique $\int_{-\infty}^{+\infty} g(x)dx = c < \infty$ se pode construir uma função densidade de probabilidade $f_X(x) = g(x)/c$.

Como se sabe do estudo das funções reais de variável real, a derivada não existe em pontos em que a função é descontínua. Retomando o exemplo que se tem vindo a apresentar a função densidade de probabilidade é sempre nula excepto para $x = 0, 1, 2, 3, 4$, onde não está definida atendendo às discontinuidades existente. No entanto é nesses pontos que está a informação interessante sobre o comportamento da variável X . Este problema é geral para as variáveis discretas que apresentam valores de probabilidade concentrados em pontos isolados da reta real.

Uma maneira de contornar o problema é definir-se para variáveis discretas uma outra função caracterizadora do seu comportamento, a função de probabilidade:

Definição 3.4 Para uma variável aleatória X que tem um contradomínio no conjunto numerável $S_X = \{x_1, x_2, \dots\}$ a função de probabilidade é

$$f_X(x_k) = P[X = x_k], k = 1, 2, \dots$$

Ir-se-á usar a notação $p_{x_k}(x_k) = P[X = x_k], k = 1, 2, \dots$. No caso particular de $x_k = k$ a notação será evidentemente $p_k(k) = P[X = k], k = 1, 2, \dots$.

Vai-se, no entanto, seguir uma outra via que pretende englobar num único quadro de referência variáveis discretas, contínuas ou com comportamento misto. Para tal recorre-se à definição de uma nova função, a função impulso unitário ou função de Dirac, que se irá representar por $\delta(x)$:

Definição 3.5

$$\delta(x) = \begin{cases} 0, & x \neq 0 \\ 1, & x = 0 \end{cases}$$

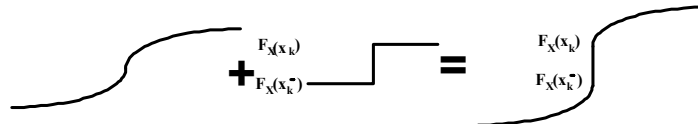
e que tem, por definição, as seguintes relações com $u(x)$:

$$\int_{-\infty}^x \delta(x) dx = u(x) \text{ e } \frac{du(x)}{dx} = \delta(x)$$

Uma descontinuidade no ponto $x = x_k$ da função de distribuição pode ser sempre representada somando a uma função contínua um degrau de amplitude $P[X = x_k] = F_X(x_k) - F_X(x_k^-)$, ou seja, uma função

$$P[X = x_k] \cdot u(x - x_k)$$

tal como se explicita na seguinte figura:



Usando a definição geral apresentada de densidade de probabilidade e a regra de derivação de $\delta(x)$, a função de densidade será então

$$f_X(x) = P[X = x_k] \delta(x - x_k) + F_X'(x)$$

Está-se a atribuir à função densidade de probabilidade nos pontos de descontinuidade o valor da função de probabilidade somado ao valor da parte contínua.²

Note que da definição apresentada $g(x)\delta(x - a) = g(a)\delta(x - a)$

²As propriedades de $\delta(x)$ aqui introduzidas por definição estão plenamente justificadas na Teoria de Distribuições. O leitor mais curioso deve consultar, por exemplo, J. Sebastião e Silva.....

Exemplo 3.4 Defina a função de densidade de probabilidade para o exemplo anterior

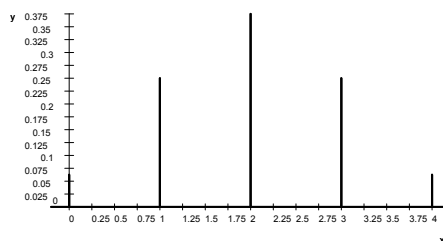
Derivando

$$F_x(x) \frac{d}{dx} F_x(x) = \frac{d}{dx} \sum_{k=0}^4 p_k u(x-k)$$

obtem-se

$$f_x(x) = \sum_{k=0}^4 p_k \delta(x-k)$$

com a seguinte representação gráfica:

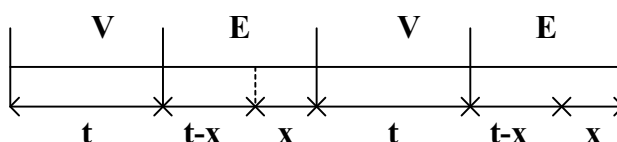


Exemplo 3.5 Num cruzamento há um semáforo que mostra alternadamente a luz verde (V) e vermelha (E) durante t segundos. Considere que um condutor chega aleatoriamente ao cruzamento durante o ciclo de luzes e é obrigado a esperar um tempo x . Associando a esta observação uma variável aleatória X , determine a sua função de distribuição.

Se o condutor chega durante o período da luz verde, o tempo de espera é nulo e portanto $P[X=0] = 1/2$

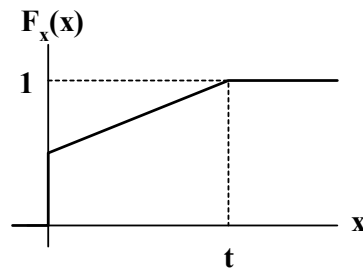
Se o condutor chegar $t-x$ segundos após o início do período da luz vermelha, o seu tempo de espera é maior que x segundos. Então, para $0 < x < t$,

$$P[X > x] = \frac{t-x}{2t} \Rightarrow P[X \leq x] = 1 - \frac{t-x}{2t} = \frac{1}{2} + \frac{x}{2t}$$

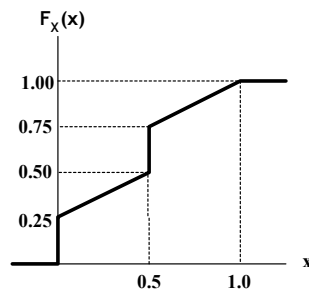


A função de distribuição será então:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2} & x = 0 \\ \frac{1}{2} + \frac{x}{2t} & 0 < x < t \\ 1 & x \geq t \end{cases}$$



Exemplo 3.6 Considere uma variável aleatória que tem a função de distribuição da seguinte figura



Determine:

1. $P[0 < x < 0.5]$, $P[0 < x \leq 0.5]$, $P[0 \leq x < 0.5]$, $P[0 \leq x \leq 0.5]$, $P[x > 0.5]$, $P[x \geq 0.5]$
2. A função densidade de probabilidade

Respostas:

1. $P[0 < x < 0.5] = F_X(5^-) - F_X(0) = 0.5 - 0.25 = 0.25$
 $P[0 < x \leq 0.5] = F_X(5) - F_X(0) = 0.75 - 0.25 = 0.5$
 $P[0 \leq x < 0.5] = F_X(5^-) - F_X(0^-) = 0.5 - 0 = 0.5$
 $P[0 \leq x \leq 0.5] = F_X(5) - F_X(0^-) = 0.75 - 0 = 0.75$
 $P[x > 0.5] = 1 - P[x \leq 0.5] = 1 - F_X(5) = 1 - 0.75 = 0.25$
 $P[x \geq 0.5] = 1 - P[x < 0.5] = 1 - F_X(5^-) = 1 - 0.5 = 0.5$

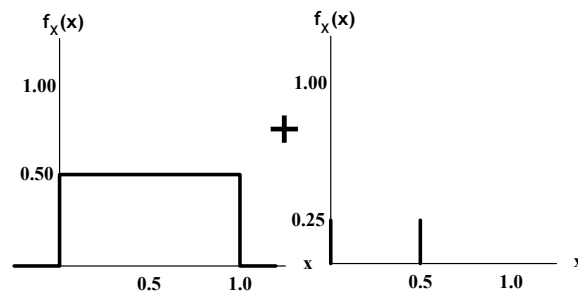
2. A função de distribuição pode expressar-se analiticamente por

$$F_X(x) = \frac{1}{2} (xu(x) - (x-1)u(x-1)) + \frac{1}{4} (u(x) + u(x-0.5))$$

Derivando em ordem a x :

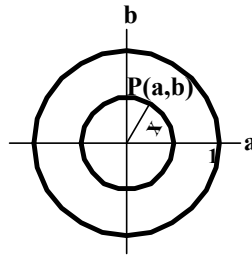
$$\begin{aligned} f_X(x) &= \frac{1}{2}(u(x) - u(x-1)) + \frac{1}{2} \left(\underbrace{x\delta(x)}_{=0} - \underbrace{(x-1)\delta(x-1)}_{=0} \right) + \\ &\quad + \frac{1}{4}(\delta(x) + \delta(x-0.5)) \\ f_X(x) &= \frac{1}{2}(u(x) - 0.5u(x-1)) + \frac{1}{4}(\delta(x) + \delta(x-0.5)) \end{aligned}$$

com a seguinte representação gráfica



Considere-se agora o seguinte exemplo de uma variável aleatória contínua

Exemplo 3.7 Considere a experiência aleatória "escolha de um ponto no interior do círculo unitário". Defina a variável aleatória X como sendo a distância do ponto escolhido à origem



O espaço de amostragem será o círculo unitário

$$\{(a, b) \in \mathbb{R}^2 : a^2 + b^2 < 1\}$$

Para qualquer $\zeta = (a, b) \in S$ define-se $x = X(\zeta) = \sqrt{a^2 + b^2}$.

Como $0 \leq a^2 + b^2 < 1 \Rightarrow 0 \leq x < 1$ e $S_X = [0, 1)$

O acontecimento $\{X \leq x\}$ é equivalente em S ao interior de um círculo de raio x cuja probabilidade será dada pela relação das áreas dos dois círculos.

Atendendo à equivalência de acontecimentos, para $x \in S_X$,

$$P[X \leq x] = \frac{P[\{(a, b) : a^2 + b^2 < x\}]}{P[\{(a, b) : a^2 + b^2 < 1\}]} = \frac{\pi \cdot x^2}{\pi \cdot 1^2} = x^2.$$

A função de distribuição será então:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

Utilizando a definição, a função de densidade de probabilidade obtém-se por derivação da expressão anterior em ordem a x :

$$f_X(x) = \begin{cases} 0, & x < 0 \vee x \geq 1 \\ 2x, & 0 \leq x < 1 \end{cases}$$

3.4 Funções de probabilidade condicional

A introdução de condicionamentos nas funções que caracterizam uma variável aleatória é simples e direta bastando substituir nas definições o conceito de probabilidade pelo de probabilidade condicional. Assim, se se pretender caracterizar o comportamento da variável X conhecida a ocorrência de A pode-se definir a função de distribuição condicional de X dada a ocorrência de A :

Definição 3.6

$$F_X(x|A) = P[\{X \leq x\} | A] = \frac{P[\{X \leq x\} \cap A]}{P[A]}$$

Se $F_X(x|A)$ for diferenciável, a função de distribuição densidade de probabilidade condicional de X dada a ocorrência de A :

$$f_X(x|A) = \frac{d}{dx} F_X(x|A)$$

Exemplo 3.8 O tempo de vida de um tipo de máquina é modelado por uma variável aleatória X . Se uma máquina estiver a funcionar no instante t qual a nova caracterização da variável que representa o tempo de vida.

Pretende-se caracterizar X dado o acontecimento $X > t$. A sua função de distribuição será

$$F_X(x|X > t) = \frac{P[\{X \leq x\} \cap \{X > t\}]}{P[\{X > t\}]}$$

Como

$$\begin{aligned} \{X \leq x\} \cap \{X > t\} &= \begin{cases} \emptyset, & x < t \\ t < X \leq x & x \geq t \end{cases} \\ P[\{X > t\}] &= 1 - P\{X \leq t\} \end{aligned}$$

e portanto

$$F_X(x|X > t) = \begin{cases} 0, & x < t \\ \frac{F_X(x) - F_X(t)}{1 - F_X(t)} & x \geq t \end{cases}$$

e se $F_X(x|X > t)$ for diferenciável

$$f_X(x|X > t) = \begin{cases} 0, & x < t \\ \frac{f_X(x)}{1 - F_X(t)} & x \geq t \end{cases}$$

Considere por exemplo que a variável X é uma variável exponencial com $\lambda = 1$ e $t = 0.9$. As funções de distribuição serão:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \geq 0 \end{cases}$$

e

$$F_X(x|X > t) = \begin{cases} 0 & x < 0.9 \\ 1 - e^{-(x-0.9)} & x \geq 0.9 \end{cases}$$

havendo como se vê uma simples translação.

3.5 Esperança matemática

Definição 3.7 Para uma variável aleatória X define-se esperança matemática $E[X]$, valor médio esperado ou média como

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx$$

Para variáveis aleatórias discretas

$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} \sum_k x_k p_{x_k}(x_k) \delta(x - x_k) dx = \\ &= \sum_k p_{x_k}(x_k) \int_{-\infty}^{+\infty} x \delta(x - x_k) dx = \sum_k x_k p_{x_k}(x_k) \end{aligned}$$

A média de uma variável aleatória pode não existir. Por exemplo considere-se a variável aleatória com a seguinte função de distribuição:

$$F_X(x) = \begin{cases} 1 - 1/x, & x \leq 1 \\ 0, & x < 1 \end{cases},$$

cuja densidade é

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

Como $\int_1^{+\infty} x \cdot \frac{1}{x^2} dx$ não converge, não existirá média para esta variável.

Um outro exemplo de não existência de valor médio pode ser dado para a variável discreta X que tenha a função de probabilidade

$$p_X(k) = \frac{1}{ak^2}, k = 1, 2, \dots \text{ e } a = \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

uma vez que

$$E[X] = \sum_{k=1}^{\infty} k \cdot \frac{1}{ak^2} = \sum_{k=1}^{\infty} \frac{1}{ak}$$

diverge.

3.5.1 Variância

Definição 3.8 A variância de uma variável aleatória é

$$Var(X) = E[(X - E[X])^2] = \int_{-\infty}^{+\infty} (x - E[X])^2 f_X(x) dx$$

Definição 3.9 O desvio padrão de uma variável aleatória é

$$\sigma_X = \sqrt{Var(X)}$$

Definição 3.10 O valor médio quadrático de uma variável aleatória é³

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x) dx$$

Partindo da definição de variância pode-se determinar a seguinte relação:

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] = \\ &= E[X^2 - 2XE[X] + E[X]^2] = \\ &= E[X^2] - 2E[XE[X]] + E[X]^2 = \\ &= E[X^2] - E[X]^2 \end{aligned}$$

3.5.2 Momentos

Definição 3.11 O momento de ordem n de uma variável aleatória é

$$E[X^n] = \int_{-\infty}^{+\infty} x^n f_X(x) dx$$

³Se uma variável aleatória representa um valor com um determinado significado físico, será sempre medida através de uma unidade dimensional. Neste caso, a média e o desvio padrão serão grandezas físicas medidas pela mesma unidade e a variância e valor médio quadrático pelo quadrado dessa unidade.

Definição 3.12 O momento central de ordem n de uma variável aleatória é

$$E[(X - E[X])^n] = \int_{-\infty}^{+\infty} (x - E[X])^n f_X(x) dx$$

Das definições é imediato ver que a média e o valor quadrático médio são os momentos de 1ª e 2ª ordem, repetivamente, e a variância é o momento central de 2ª ordem. Também é trivial verificar que o momento central de 1ª ordem é nulo.

3.6 Função caraterística

Definição 3.13 Função caraterística de uma variável aleatória X é ⁴

$$\Phi_X(w) = E[e^{jwX}] = \int_{-\infty}^{+\infty} f_X(x) e^{jwx} dx$$

Da definição pode ver-se que a função caraterística é o valor médio da função e^{jwX} ou a transformada de fourier da densidade de probabilidade de X . Nesta última hipótese e se existir transformada inversa, o par $(f_X(x), \Phi_X(w))$ caracteriza totalmente a variável aleatória X .

3.6.1 Propriedades

1. $\Phi_X(0) = 1$

Demonstração. Como $e^{jwX}|_{w=0} = 1 \Rightarrow E[1] = 1$ ■

2. $|\Phi_X(w)| \leq 1$

Demonstração. Como

$$\begin{aligned} |e^{jwX}| &= 1 \\ \left| \int_{-\infty}^{+\infty} f_X(x) e^{jwx} dx \right| &= \int_{-\infty}^{+\infty} |f_X(x) e^{jwx}| dx \leq \int_{-\infty}^{+\infty} f_X(x) dx = 1 \end{aligned}$$

■

3. $\Phi_X(-w) = \Phi_X^*(w)$ ⁵

Demonstração. Como $e^{-jwX} = (e^{jwX})^*$ e $f_X(x)$ é real $\Rightarrow \int_{-\infty}^{+\infty} f_X(x) e^{-jwx} dx = \left(\int_{-\infty}^{+\infty} f_X(x) e^{jwx} dx \right)^*$ ■

Teorema 3.1 Teorema dos momentos

$$E[X^n] = \frac{1}{j^n} \frac{d^n}{dw^n} \Phi_X(w) \Big|_{w=0}$$

⁴Note que a menos do sinal do expoente, se trata da transformada de Fourier da função de densidade de probabilidade.

⁵* significa complexo conjugado.

Demonstração.

$$\begin{aligned}\frac{d^n}{dw^n} \int_{-\infty}^{+\infty} f_X(x) e^{iwx} dx &= \int_{-\infty}^{+\infty} f_X(x) \frac{d^n}{dw^n} e^{iwx} dx = \\ &= \int_{-\infty}^{+\infty} f_X(x) (ix)^n e^{iwx} dx = \\ &= i^n E [x^n e^{iwx}]\end{aligned}$$

. Para $w = 0$ o resultado fica demonstrado. ■

Este resultado pode também ser deduzido desenvolvendo a $\Phi_X(w)$ em série de Mac Laurin na vizinhança de $w = 0$:

$$\Phi_X(w) = 1 + E[X] iw + E[X^2] \frac{(iw)^2}{2!} + \dots = 1 + \sum_{k=1}^m E[X^k] \frac{(iw)^k}{k!} + R(w^m) \quad (3.2)$$

3.7 Função geradora de probabilidade

Em problemas relacionados com variáveis discretas é preferível usar uma outra função, a função geradora de probabilidades, $G_X(z)$:

Definição 3.14 Para uma variável aleatória X que assume valores em $x_k = k, k = 0, 1, 2, \dots$ a função geradora de probabilidades é ⁶

$$G_X(z) = E[z^X] = \sum_{k=0}^{\infty} p_k(k) z^k$$

Se a variável for finita, a função geradora de probabilidade é um polinómio em z .

É fácil de provar que $p_X(k) = \frac{1}{k!} \frac{d^k}{dz^k} G_X(z) \Big|_{z=0}$ razão pela qual se chama função geradora de probabilidades.

Determinando as duas primeiras derivadas em ordem a z e avaliando-as em $z = 1$ obtém-se:

$$\begin{aligned}\frac{d}{dz} G_X(z) \Big|_{z=1} &= \sum_{k=0}^{\infty} k p_k(k) z^{k-1} \Big|_{z=1} = \sum_{k=0}^{\infty} k p_k(k) = E[X] \\ \frac{d^2}{dz^2} G_X(z) \Big|_{z=1} &= \sum_{k=0}^{\infty} k(k-1) p_k(k) z^{k-2} \Big|_{z=1} = \\ &= \sum_{k=0}^{\infty} k(k-1) p_k(k) = E[X^2] - E[X]\end{aligned}$$

⁶Note que a menos do sinal do expoente, é a transformada- z da função de probabilidade.

Então:

$$E[X] = G'_X(1) \text{ e } Var(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2 \quad (3.3)$$

3.8 Desigualdades de Markov e Chebyshev

Os seguintes teoremas conhecidos respetivamente por desigualdades de Markov e Chebyshev permitem estabelecer facilmente majorantes para as probabilidades de certas classes de acontecimentos, partindo apenas do conhecimento da média e variância de uma variável aleatória

Teorema 3.2 *Desigualdade de Markov*

Para qualquer variável aleatória X não negativa

$$P[X \geq a] \leq \frac{E[X]}{a} \quad (3.4)$$

Demonstração.

$$\begin{aligned} E[X] &= \int_0^a x f_X(x) dx + \int_a^\infty x f_X(x) dx \geq \\ &\geq \int_a^\infty x f_X(x) dx \geq \int_a^\infty a f_X(x) dx = a P[X \geq a] \end{aligned}$$

■

Exemplo 3.9 A média de altura de uma população é de 1.65m. Qual o limite superior de probabilidade de um indivíduo ultrapassar os 2m?

$$P[X \geq 2] < \frac{1.65}{2} = 0.825$$

O limite obtido não é, porventura, muito útil nem significativo, No entanto para o seu cálculo apenas se considerou o valor médio da variável em questão.

Teorema 3.3 *Desigualdade de Chebyshev*

Seja $m = E[X]$,

$$P[|X - m| \geq a] \leq \frac{\sigma_X^2}{a^2} \quad (3.5)$$

Demonstração. Defina-se a função de variável aleatória $D^2 = (X - m)^2$. Sendo óbvio que $D^2 \geq 0$

e $D^2 \geq a^2 \Leftrightarrow |D| \geq a$ pode-se aplicar o teorema anterior

$$P[|D| \geq a] = P[D^2 \geq a^2] \leq \frac{E[(X - m)^2]}{a^2} = \frac{\sigma_X^2}{a^2}$$

■

Se se fizer $a = t\sigma_X$ o resultado anterior pode ser reescrito como

$$P[|X - m| \geq t\sigma_X] \leq \frac{1}{t^2}$$

ou ainda

$$P[|X - m| < t\sigma_X] \geq 1 - \frac{1}{t^2}$$

Exemplo 3.10 Considere o exemplo anterior e admita um desvio padrão de 0.1m. Determine o novo limite para a probabilidade do desvio em relação à média ser superior a 0.35m.

$$P[|X - 1.65| \geq 0.35] \leq \frac{0.1^2}{0.35^2} = 0.08$$

Embora esta probabilidade não se refira ao mesmo acontecimento que o do problema anterior, pois inclui também o acontecimento "indivíduo com uma altura inferior a $1.65 - 0.35 = 1.30$ m", o limite de probabilidade obtido já é bastante mais restritivo. A utilização da informação sobre a dispersão dos valores em torno do seu valor médio, dada pelo desvio padrão, assim o permite.

3.9 Estudo de algumas variáveis aleatórias

3.9.1 Variável de Bernoulli

Seja A um acontecimento que possa resultar de uma dada experiência aleatória. A variável de Bernoulli ou função indicadora de A , é:

$$I_A(\zeta) = \begin{cases} 0, & \zeta \notin A \\ 1, & \zeta \in A \end{cases}$$

O seu contradomínio é $S_I = \{0, 1\}$ e $p_I(1) = p$ e $p_I(0) = 1 - p$ sendo $p = P[A]$.

Valor médio:

$$E[I_A] = 0 \cdot p_I(0) + 1 \cdot p_I(1) = 0 \cdot (1 - p) + 1 \cdot p = p$$

Variância:

$$Var(I_A) = E[I_A^2] - E[I_A]^2 = 0^2 \cdot (1 - p) + 1^2 \cdot p - p^2 = p(1 - p)$$

3.9.2 Variável binomial

Se uma experiência aleatória é repetida n vezes e seja X o número de vezes, k , que ocorre um acontecimento A . Então esta variável conhecida por binomial, pode ser considerada como a soma de n variáveis de Bernoulli $X = I_1 + \dots + I_n$ e será simbolizada por $Bin(n, p)$

O contradomínio de X é $S_X = \{0, 1, \dots, n\}$ e as probabilidades para $X = k$ já foram calculadas no capítulo 1 em 2.16

$$P[X = k] = C_k^n p^k (1 - p)^{n-k}, k = 0, \dots, n$$

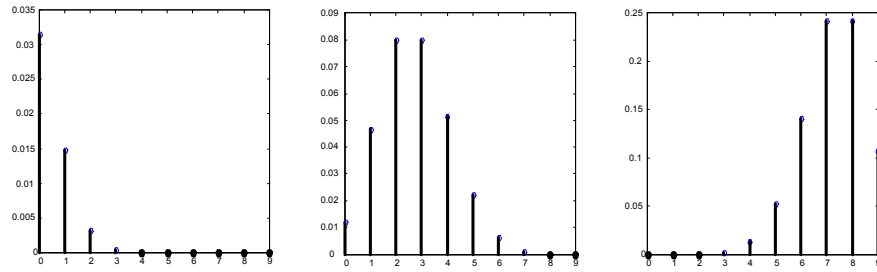
Valor médio:

$$E[X] = E\left[\sum_{k=1}^n I_k\right] = \sum_{k=1}^n E[I_k] = np$$

Variância ⁷:

$$\text{Var}(X) = \sum_{k=1}^n \text{Var}(I) = np(1-p)$$

Alguns exemplos de funções de densidade de probabilidade para a lei binomial para $n = 9$ e $p = 0.05, 0.3$ e 0.8 , respetivamente:



Exemplo 3.11 Num armazém existem 500 peças das quais 10% não estão em boas condições. É efetuada uma inspeção de controlo de qualidade sobre uma amostra de 10 unidades escolhidas ao acaso. Admitindo como critério de rejeição a existência de não mais de 3 peças defeituosas determine:

1. A probabilidade de rejeição do lote
2. Qual a dimensão da amostra para que a probabilidade de rejeição seja superior a 5%

Está-se perante um problema modelado por uma variável binomial com $n = 10$ e $p = 0.1$. portanto:

1. $P[k > 3] = \sum_{i=4}^{10} C_i^{10} 0.1^i 0.9^{10-i} = 0.0129$
2. $P[k > 3] = \sum_{i=4}^N C_i^N 0.1^i 0.9^{N-i} > 0.05$

Exemplo 3.12 Recorrendo a um programa de cálculo numérico pode construir-se a tabela

N	11	12	13	14	15
$P[k > 3]$	0.019	0.026	0.034	0.044	0.056

e a solução é uma amostra com 15 unidades

⁷Esta igualdade é facilmente justificada a partir da independência de variáveis, assunto que é tratado um pouco mais à frente, neste capítulo.

3.9.3 Variável binomial negativa

Na variável binomial procura-se caracterizar o número de sucessos k que se verificam em n experiências de Bernoulli. Pode-se também caracterizar um outro problema: qual o número n de experiências de Bernoulli que se devem realizar para obter r sucessos. Esta variável aleatória que se designa por binomial negativa tem um contradomínio $S_X = \{r+1, r+2, \dots\}$ e se p for a probabilidade de sucesso a função de probabilidade será

$$P[X = n] = P[r-1 \text{ sucessos em } n-1 \text{ experiências}] \times \\ \times P[\text{sucesso na última experiência}]$$

$$P[X = n] = C_{r-1}^{n-1} p^{r-1} (1-p)^{n-r} p = C_{r-1}^{n-1} p^r (1-p)^{n-r}$$

Valor médio

$$E[X] = \frac{r}{p}$$

Variância:

$$Var(X) = \frac{r}{p^2}$$

Alternativamente pode definir-se uma variável Y que represente o número de insucessos m até se obterem r sucessos nas $n = m + r$ experiências.

Se p for a probabilidade de sucesso a função de probabilidade será $P[X = m] = P[m \text{ insucessos em } m + r - 1 \text{ experiências}] \times P[\text{sucesso na última experiência}]$

$$P[X = m] = C_m^{m+r-1} p^{r-1} (1-p)^m p = C_m^{m+r-1} p^r (1-p)^m$$

com $S_X = \{0, 1, \dots\}$

Valor médio:

$$E[X] = r \frac{1-p}{p}$$

Variância:

$$Var(X) = r \frac{1-p}{p^2}$$

Exemplo 3.13 A probabilidade de avaria de uma máquina quando efetua uma série de fabrico entre períodos de manutenção é de $p = 0.05$. Se a máquina avariar a série em curso não pode ser aproveitada.

As avarias podem ser facilmente reparadas pela substituição de uma peça de que existem 3 unidades em stock. Supõe-se que as avarias são independentes e que no início de cada série, após a manutenção, a máquina está sempre nas mesmas condições (isto é, não há envelhecimento).

Considere a variável aleatória X que é o número de séries executadas até não haver mais peças de substituição. X é então o número de insucessos (séries sem avaria) até se verificarem 4 sucessos (séries com avaria), portanto, uma variável aleatória binomial negativa com parâmetros $r = 4$ e $p = 0.05$.

Determine:

1. O número esperado de séries completas antes de se esgotarem as peças de substituição em stock
2. A probabilidade de se completarem 100 séries
3. O número de peças em stock que, com probabilidade não inferior a 0.95, assegurem o fabrico de mais de 100 séries

Respostas

1. O valor médio esperado de X é: $E[X] = r \frac{1-p}{p} = 4 \frac{1-0.05}{0.05} = 76$
- 2.

$$\begin{aligned} P[X > 100] &= \sum_{k=101}^{\infty} C_k^{k+4-1} 0.05^4 0.95^k = \\ &= 1 - \sum_{k=0}^{100} C_k^{k+3} 0.05^4 0.95^k = 0.2308 \end{aligned}$$

- 3.

$$\begin{aligned} P[X > 100] &= \sum_{k=101}^{\infty} C_k^{k+r-1} 0.05^r 0.95^k = \\ &= 1 - \sum_{k=0}^{100} C_k^{k+r-1} 0.05^7 0.95^k > 0.95 \end{aligned}$$

que, recorrendo a um programa de cálculo numérico conduz a $r = 10$, ou seja, 9 peças em stock.

3.9.4 Variável geométrica

Um caso particular é para $r = 1$, isto é, procura-se caracterizar a probabilidade do número de experiências até se obter um sucesso.

Alternativamente pode-se considerar a variável aleatória que seja igual ao número de insucessos até se obter um sucesso, usualmente designada por variável aleatória geométrica e que apresenta as seguintes características:

O contradomínio de X é $S_X = \{0, 1, \dots\}$ e as probabilidades para $X = k$

$$P[X = k] = p(1-p)^k, k = 0, 1, \dots$$

Valor médio:

$$E[X] = E\left[\sum_{k=0}^{\infty} kp(1-p)^k\right] = p \frac{1-p}{p^2} = \frac{1-p}{p}$$

Variância e valor médio quadrático:

Atendendo (ver apêndice C) a que $\frac{2}{(1-x)^3} = \sum_{k=0}^{\infty} k(k-1)x^{k-2}$ fazendo $x = 1-p$ e multiplicando ambos os membros por $p(1-p)^2$ obtem-se:

$$\begin{aligned} \frac{2p(1-p)^2}{p^3} &= \sum_{k=0}^{\infty} k(k-1)(1-p)^{k-2}p(1-p)^2 = \\ &= \sum_{k=0}^{\infty} k(k-1)p(1-p)^k = E[X^2] - E[X] \end{aligned}$$

donde se pode concluir, atendendo ao valor de $E[X]$ já determinado

$$E[X^2] = \frac{(2-p)(1-p)}{p^2}$$

e

$$Var(X) = \frac{1-p}{p^2}$$

Exemplo 3.14 Uma máquina é usada com uma probabilidade de avaria de 0.01 por semana. Qual a probabilidade de funcionar pelo menos 25 semanas?

Está se perante uma variável geométrica de parâmetro $P = 0.01$ e a probabilidade procurada será:

$$P[X > 25] = \sum_{k=26}^{\infty} 0.01 \cdot 0.99^k = 1 - \sum_{k=0}^{25} 0.01 \cdot 0.99^k = 0.77$$

3.9.5 Variável de Poisson

Esta variável foi introduzida por Poisson como forma limite de estudar as probabilidades binomiais.

Considere-se que, para uma variável binomial, n cresce e p decresce por forma a que $np \rightarrow \alpha > 0$. Para n grande pode-se então fazer a aproximação $p \simeq \frac{\alpha}{n}$ e $1-p \simeq 1 - \frac{\alpha}{n}$;

$$\begin{aligned} P[X = k] &= C_k^n p^k (1-p)^{n-k} \simeq \\ &\simeq \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\alpha}{n}\right)^k \left(1 - \frac{\alpha}{n}\right)^{n-k} = \\ &= \frac{\alpha^k}{k!} \frac{n(n-1)\dots(n-k+1)}{n^k} \left(1 - \frac{\alpha}{n}\right)^n \left(1 - \frac{\alpha}{n}\right)^{-k} = \\ &= \frac{\alpha^k}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\alpha}{n}\right)^n \left(1 - \frac{\alpha}{n}\right)^{-k} \end{aligned}$$

Como $\lim_{n \rightarrow \infty} \left(1 - \frac{j}{n}\right) = 1, j = 1, \dots, k-1$ e $\lim_{n \rightarrow \infty} \left(1 - \frac{\alpha}{n}\right)^n = e^{-\alpha}$ e o limite do produto é o produto dos limites dos fatores, vem

$$\lim_{n \rightarrow \infty} \frac{\alpha^k}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\alpha}{n}\right)^n \left(1 - \frac{\alpha}{n}\right)^{-k} = \frac{\alpha^k}{k!} e^{-\alpha}$$

Para que esta expressão seja uma lei válida de probabilidade é necessário verificar que $\sum_{k=0}^{\infty} \frac{\alpha^k}{k!} e^{-\alpha} = 1$. Isto pode ser facilmente provado recordando que o desenvolvimento em série de Maclaurin $e^{\alpha} = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!}$.

Pode-se então definir a variável de Poisson como o limite para que tende a distribuição binomial quando $n \rightarrow \infty$.

O contradomínio de X é $S_X = \{0, 1, \dots\}$ e a função de probabilidade

$$p_X(k) = P[X = k] = \frac{\alpha^k}{k!} e^{-\alpha}$$

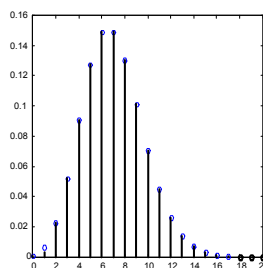
Valor médio:

$$E[X] = \alpha$$

Variância ⁸:

$$Var(X) = \alpha$$

Exemplo de função de densidade de probabilidade para a lei de Poisson, com $\alpha = 7$



Está demonstrado que o erro cometido ao tomar a lei de Poisson em vez da lei binomial está limitado da seguinte maneira: dados dois inteiros r, s e $\alpha = np$ verifica-se

$$\left| \sum_{k=r}^s C_k^n p^k (1-p)^{n-k} - \sum_{k=r}^s \frac{\alpha^k}{k!} e^{-\alpha} \right| < \frac{\alpha^2}{n}$$

Desta desigualdade pode-se concluir que:

- Se se fixar α a aproximação melhora aumentando o número de experiências n
- Se se fixar o número de experiências n , a aproximação melhora com $p \rightarrow 0$

⁸Veja o exercício nº ??? para o cálculo da média e variância.

- Se se fixar p por menor que seja, a aproximação piora se se aumentar o número de experiências n .

A variável de Poisson modela muitos sistemas físicos, por exemplo, o número de desintegrações verificadas numa massa de material radioativo. O material radioativo é constituído por um número n de átomos têm uma probabilidade p de se desintegrar no período considerado e sendo a desintegração de cada um um acontecimento independente. Está-se perante uma situação de determinar a probabilidade de k sucessos (desintegrações) em n acontecimentos independentes. Atendendo a que n é muito elevado (da ordem de $10^{16}/\mu\text{g}$ para certas substâncias radioativas) e que a probabilidade de desintegração é muito pequena para períodos curtos em substâncias naturalmente radiativas, a aproximação pela lei de Poisson justifica-se plenamente, fornecendo muito bons resultados.

A variável de Poisson também se mostra muito útil noutra classe de problemas em que se pode considerar que uma sequência de experiências de Bernoulli independentes ocorrem aleatoriamente no tempo, num período T , sendo a taxa média de ocorrência igual a λ . Pode-se dividir o intervalo T em n subintervalos suficientemente pequenos para que a probabilidade de ocorrência de mais do que um acontecimento seja desprezável face à probabilidade de ocorrência de um acontecimento que valerá $p = \lambda T/n$. Mais uma vez se está perante uma situação modelável por uma lei binomial pois procura-se caracterizar λT sucessos em n possíveis. Se n for grande e p pequeno pode-se aproximar por uma lei de Poisson com $\alpha = np = \lambda T$.

Exemplo 3.15 *É conhecido que o número de mensagens que chega a uma central telefónica por segundo se comporta de acordo com a lei de Poisson. Suponha que uma central telefónica recebe em média 15 mensagens por segundo.*

1. Calcule a probabilidade de que num segundo não se receba nenhuma mensagem
2. Calcule a probabilidade de que mais de 10 mensagens cheguem á central telefónica no período de um segundo.

Respostas:

1. $\alpha = 15; P[k = 0] = \frac{15^0}{0!} e^{-15} = 3.059 \times 10^{-7}$
2. $P[k > 10] = \sum_{k=11}^{\infty} \frac{15^k}{k!} e^{-15} = 1 - \sum_{k=1}^{10} \frac{15^k}{k!} e^{-15} = 0.88$

Exercício 3.1 *Numa comunidade com 15000 pessoas, a probabilidade de uma pessoa necessitar de ser internada no hospital local é de $1/3000$. Determine o número mínimo de camas C que o hospital deve ter para que todas as pessoas que num determinado dia recorram ao hospital tenham cama, com uma probabilidade superior a 99%.*

Utilizando a lei binomial ter-se-á

$$P[X \leq C] = \sum_{k=0}^C C_k^{15000} \left(\frac{1}{3000}\right)^k \left(\frac{2999}{3000}\right)^{10000-k} > 0.95$$

problema de difícil resolução numérica como o leitor pode verificar recorrendo ao Matlab.

Recorrendo à lei de Poisson com parâmetro $\alpha = 15000 \frac{1}{3000} = 5$ o problema simplifica-se significativamente :

$$P[X \leq C] = \sum_{k=0}^C \frac{5^k}{k!} e^{-5} > 0.95$$

que, resolvido numericamente, permite determinar $C = 9$ com $P[X \leq C] = 0.97$

3.9.6 Variável discreta uniforme

Uma variável é uniforme e discreta em n pontos quando a sua função de probabilidade é

$$P[U = x_k] = \frac{1}{n}, x_k \in \mathbb{Z}, k = 1, \dots, n.$$

Um caso de especial interesse é quando $x_k = 1, 2, \dots, n$ e então ter-se-á:

$$S_U = \{1, 2, \dots, n\}$$

$$P[U = k] = \frac{1}{n}$$

Valor médio:

$$E[U] = \sum_{k=1}^n \frac{1}{n} \cdot k = \frac{n+1}{2}$$

Valor médio quadrático:

$$E[U_k^2] = \sum_{k=1}^n \frac{1}{n} \cdot k^2 = \frac{1}{n} \left(\frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} \right) = \frac{n^2}{3} + \frac{n}{2} + \frac{1}{6}$$

Variância:

$$Var(U) = \frac{n^2}{3} + \frac{n}{2} + \frac{1}{6} - \left(\frac{n+1}{2} \right)^2 = \frac{n^2-1}{12}$$

3.9.7 Variável uniforme

Uma variável aleatória X diz-se uniforme no intervalo $[a, b]$ e representa-se por $U(a, b)$ se a sua densidade de probabilidade for:

$$f_X(x) = \begin{cases} 0, & x < a \vee x > b \\ \frac{1}{b-a}, & a \leq x \leq b \end{cases}$$

a que corresponde uma função de distribuição

$$F_F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

Valor médio:

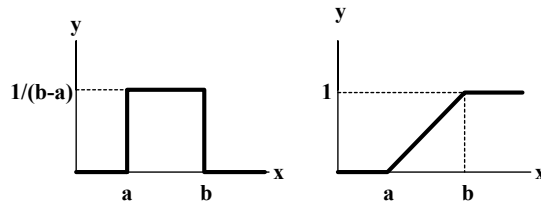
$$E[X] = \int_{-\infty}^{+\infty} x f_x(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$$

Valor médio quadrático:

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{+\infty} x^2 f_x(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \\ &= \frac{1}{3}(b^2 + ab + a^2) \end{aligned}$$

Variância:

$$\begin{aligned} Var(X) &= \int_{-\infty}^{+\infty} (x - E[x])^2 f_x(x) dx = \\ &= \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} dx = \frac{(b-a)^2}{12} \end{aligned}$$



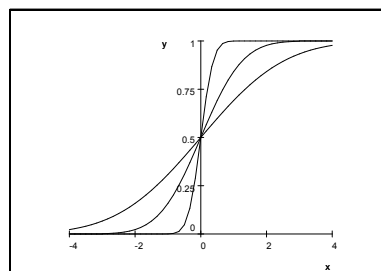
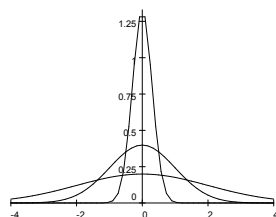
3.9.8 Variável gausseana ou normal

A distribuição gausseana ou normal é, talvez, a mais importante das distribuições de probabilidade contínua.

Uma variável aleatória X tem uma distribuição gausseana ou normal com parâmetros m e σ , e representa-se por $N(m, \sigma)$ se a sua densidade de probabilidade for:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Na figura mostram-se as funções de densidade de probabilidade e as funções de distribuição para $m = 0$ e $\sigma = 0.3, 1$ e 2



Valor médio:

$$E[X] = m$$

Valor médio quadrático:

$$E[X^2] = \sigma^2 + m^2$$

Variância :

$$\text{Var}(X) = \sigma^2$$

Definição 3.15 Define-se a variável gausseana normalizada como sendo a variável $N(0, 1)$, tendo portanto uma densidade de probabilidade

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

a que corresponde uma função de distribuição normalizada $\Phi(x) = \int_{-\infty}^x \phi(x) dx$

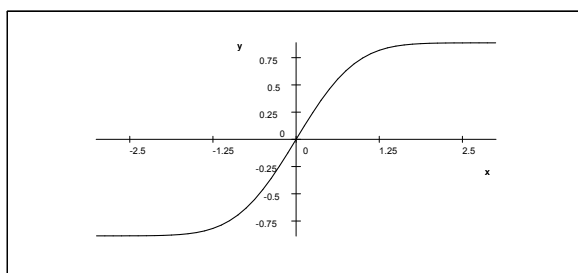
A função normalizada obtem-se através da transformação $x \rightarrow \frac{x-m}{\sigma}$ que para $\sigma \neq 0$ é sempre possível e facilmente realizável. É trivial provar que

$$F_X(x) = \int_{-\infty}^{\frac{x-m}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi\left(\frac{x-m}{\sigma}\right)$$

Associada a esta é habitual definir-se ainda a função $Q(x) = 1 - \Phi(x)$, função muito utilizada para o cálculo de probabilidades envolvendo a variável gausseana por se encontrar comunemente tabelada. Os modernos meios de cálculo dispensam estas tabelas, no entanto nos exercícios propostos encontram-se alguns ligados ao cálculo e utilização de $Q(x)$.

Define-se ainda a função de erro, $\text{err}(x)$ ou $\text{er}(x)$ como sendo

$$\text{err}(x) = \int_0^x e^{-t^2} dt$$



Teorema 3.4 A densidade de probabilidade de $N(m, \sigma)$ é simétrica em relação à reta $x = m$

Demonstração. É imediato verificar que $\forall a \in \mathbb{R} f_X(x - a - m) = f_X(x + a - m)$

■

É, então, trivial provar que $\phi(x) = \phi(-x)$ e, em consequência, $\int_x^\infty \phi(x) dx = \int_{-\infty}^{-x} \phi(x) dx$ o que é equivalente a $1 - \Phi(x) = \Phi(-x) \Leftrightarrow Q(x) = 1 - Q(-x)$ que implica a seguinte relação:

$$F_X(x) = 1 - Q\left(\frac{x - m}{\sigma}\right)$$

Exemplo 3.16 Determine $P[|X - m| \leq k\sigma]$ e compare com os valores que se obtêm usando a desigualdade de Chebyshev

$$\begin{aligned} P[|X - m| \leq k\sigma] &= P[-k\sigma \leq X - m \leq k\sigma] = \\ &= P[-k\sigma + m \leq X \leq k\sigma + m] = \\ &= F_X(k\sigma + m) - F_X(-k\sigma + m) = \\ &= \Phi(k) - \Phi(-k) = 1 - 2Q(k) \end{aligned}$$

Usando a desigualdade de Chebyshev

$$P[|X - m| \leq k\sigma] \geq 1 - \frac{1}{k^2}$$

k	$1 - 2Q(k)$	$1 - \frac{1}{k^2}$
1	0.6827	0.000
2	0.9545	0.750
3	0.9973	0.889
4	0.9999	0.938

Como se pode verificar os limites dados pela desigualdade de Chebyshev são bastante grosseiros, mas apenas recorrem aos valores da média e da variância.

Exemplo 3.17 Uma empresa, monopolista no mercado de um certo produto, tem uma produção mensal de 80 toneladas(t)/mês. e sabe que a procura mensal segue uma lei normal $N(75, 10)$. Determine:

1. A probabilidade da procura se situar entre 78 e 80t
2. A probabilidade da procura ser excedentária
3. Qual o stock que deve constituir para fazer face a uma procura excedentária por forma que a probabilidade da procura excedentária seja de 0.025.

Respostas

1.

$$\begin{aligned} P[78 \leq X \leq 80] &= P\left[\frac{78-75}{10} \leq U_X \leq \frac{80-75}{10}\right] = \\ &= P[0.3 \leq U_X \leq 0.5] = Q(0.3) - Q(0.5) = 0.074 \end{aligned}$$

2.

$$P[X > 80] = P\left[U_X > \frac{80-75}{10}\right] = P[U_X > 0.5] = Q(0.5) = 0.309$$

3.

$$\begin{aligned} P[X > 80 + S] &= P\left[U_X > \frac{80+S-75}{10}\right] = Q\left(\frac{5+S}{10}\right) = 0.025 \\ \frac{5+S}{10} &= 1.96 \Leftrightarrow S = 14.6 \end{aligned}$$

Exemplo 3.18 Numa fábrica de rolamentos sabe-se que os diâmetros das esferas metálicas fabricadas para os rolamentos segue uma lei normal. Para rapidamente se determinar os parâmetros da lei normal é usado o seguinte procedimento: As esferas são colocadas em caixas que têm orifícios de diferentes dimensões e contam-se as que os atravessaram. Suponha que numa dada experiência se verificou que 21% das esferas passavam por orifícios de 7.50 mm e 65% passavam por orifícios de 8.00mm. Determine a média e o desvio padrão do diâmetro das esferas.

Sejam x_1, x_2 dois valores da variável X a que correspondem as probabilidades

$$P[X \leq x_1] = a_1, P[X \leq x_2] = a_2$$

Fazendo a normalização da variável

$$\begin{aligned} P\left[\frac{X-m}{\sigma} \leq \frac{x_1-m}{\sigma}\right] &= a_1 \\ P\left[\frac{X-m}{\sigma} \leq \frac{x_2-m}{\sigma}\right] &= a_2 \end{aligned}$$

o que conduzirá às duas equações

$$\frac{x_i - m}{\sigma} = z_{a_i}$$

com $z_{a_i} = \Phi^{-1}(a_i)$, $i = 1, 2$. Reescrevendo sob a forma de um sistema ter-se-á

$$\begin{cases} m + z_{a_1}\sigma = x_1 \\ m + z_{a_2}\sigma = x_2 \end{cases}$$

com a solução

$$m = \frac{x_2 z_{a_1} - x_1 z_{a_2}}{z_{a_1} - z_{a_2}} \text{ e } \sigma = \frac{x_2 - x_1}{z_{a_1} - z_{a_2}}$$

Recorrendo a um programa de cálculo numérico a solução procurada será $m = 7.8383$ e $\sigma = 0.4196$

3.9.9 Variável exponencial

A variável exponencial aparece no modelamento de intervalos de tempo entre acontecimentos ou no modelamento do tempo de vida de dispositivos ou sistemas.

Uma variável aleatória X tem uma distribuição exponencial com parâmetro λ se a sua densidade de probabilidade for:

$$f_X(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

e, consequentemente, tiver uma função de distribuição

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

Valor médio:

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

Variância:

$$\begin{aligned} Var(X) &= \int_{-\infty}^{+\infty} (x - E[X])^2 f_X(x) dx = \\ &= \int_0^{\infty} \left(x - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2} \end{aligned}$$

Teorema 3.5 $P[X > t + a | X > a] = P[X > t]$

Demonstração.

$$\begin{aligned} P[X > t + a | X > a] &= \frac{P[X > t + a \cap X > a]}{P[X > a]} = \\ &= \frac{P[X > t + a]}{P[X > a]} = \frac{e^{-\lambda(t+a)}}{e^{-\lambda a}} = \\ &= e^{-\lambda t} = P[X > t] \end{aligned}$$

■

Este teorema traduz uma propriedade de falta de memória desta variável que é sua característica entre as variáveis contínuas.

Exemplo 3.19 *A duração da vida útil, em milhares de horas, de uma componente de uma máquina é uma variável exponencial de valor médio 10 mil horas. Determine a probabilidade de uma componente seleccionada ao acaso:*

1. Dure pelo menos 4000 horas
2. Dure entre 3000 e 10000 horas
3. Dure mais de 12000 horas

Exemplo 3.20 *Sendo a média de vida 10k hora $\Rightarrow \lambda = 0.1/k$ hora, as probabilidades pretendidas serão:*

1. $P[X < 4] = \int_0^4 0.1e^{-0.1x} dx = F_X(4) = 0.33$
2. $P[3 < X < 10] = \int_3^{10} 0.1e^{-0.1x} dx = F_X(10) - F_X(3) = 0.37$
3. $P[X > 12] = \int_{12}^{\infty} 0.1e^{-0.1x} dx = 1 - F_X(12) = 0.30$

3.9.10 Variável gama

A variável gama é muito flexível e permite modelar muitos e diversos sistemas. Como exemplos pode-se citar os tempos de espera numa fila, o tempo de vida de sistemas ou dispositivos em estudos de fiabilidade e o comportamento de variáveis aleatórias em testes de hipóteses (um caso particular desta variável come será amplamente referido).

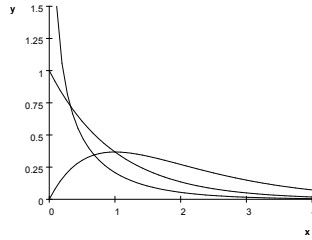
Uma variável aleatória tem uma distribuição gama com parâmetros α e λ se a sua densidade de probabilidade for

$$f_X(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$$

com $x \in (0, +\infty)$ e sendo $\Gamma(\cdot)$ a função gama definida por

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt, x > 0$$

Apresentam-se na figura seguinte 3 casos correspondentes a $\lambda = 1$ e $\alpha = 1/2, 1$ e 2 :



Valor médio:

$$E[X] = \frac{\alpha}{\lambda}$$

Variância:

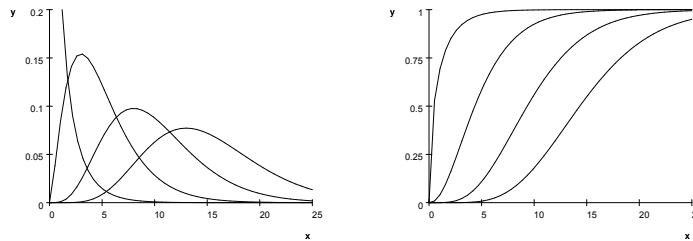
$$Var(X) = \frac{\alpha}{\lambda^2}$$

Um caso particular de larga importância é o definido por $\lambda = 1/2$ e $\alpha = k/2$. A variável aleatória assume a designação de variável de χ^2 com k graus de liberdade e a sua densidade de probabilidade é

$$f_X(x) = \frac{x^{(k-2)/2} e^{-x/2}}{2^{k/2} \Gamma(k/2)}$$

com $x \in (0, +\infty)$

Apresentam-se na figura seguinte 3 casos correspondentes $k = 1, 5, 10, 15$ e $0 \leq x \leq 25$.



Valor médio:

$$E[X] = k$$

Variância:

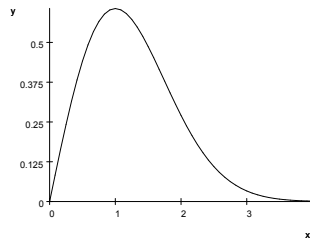
$$Var(X) = 2k$$

Esta variável está relacionada com a soma quadrática de variáveis gaussianas como adiante se verá.

3.9.11 Variável de Rayleigh

Uma variável aleatória tem uma distribuição de Rayleigh com parâmetro α se a sua densidade de probabilidade for

$$f_X(x) = \frac{x}{\alpha^2} e^{-\frac{x^2}{2\alpha^2}}$$



Valor médio:

$$E[X] = \alpha\sqrt{\pi/2}$$

Variância:

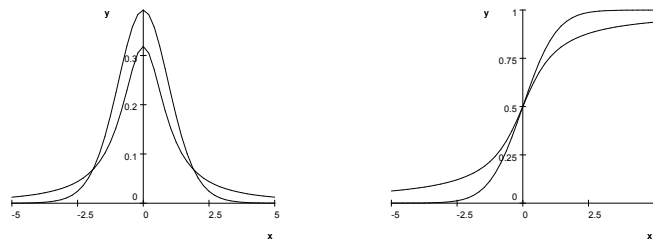
$$Var(X) = 2 - \pi/2 \alpha^2$$

3.9.12 Variável t de Student

Uma variável aleatória tem uma distribuição t de Student⁹ com n graus de liberdade se a sua densidade de probabilidade for

$$f_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \left(1 + \frac{1}{n}t^2\right)^{-\frac{n+1}{2}}$$

As figuras seguintes são os gráficos das funções de densidade e de distribuição da variável t de Student para $n = 1$ e $n = 15$ com $-5 \leq x \leq 5$.



⁹O nome deve-se ao pseudônimo A.Student usado pelo seu autor W.S.Gosset na publicação original.

Valor médio:

$$E[T] = 0$$

Variância:

$$Var(T) = \frac{n}{n-2}$$

Demonstra-se que o limite quando $n \rightarrow \infty$ desta distribuição é a distribuição normal.

3.10 Função de variável aleatória

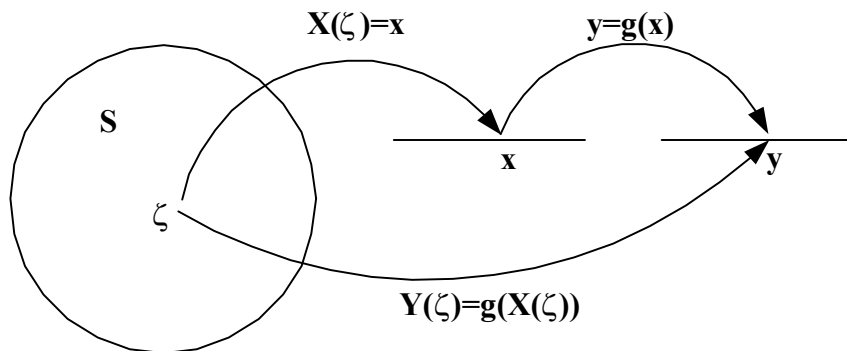
Seja X uma variável aleatória e seja $g(x)$ uma função real de variável real definidas por:

$$\begin{aligned} X &: \zeta \in S \rightarrow X(\zeta) = x \in \mathbb{R} \\ g &: x \in \mathbb{R} \rightarrow g(x) \in \mathbb{R} \end{aligned}$$

Pode-se então definir uma aplicação de S em \mathbb{R}

$$Y : \zeta \in S \rightarrow g(X(\zeta)) = y \in \mathbb{R}$$

ou seja, $Y = g(X)$ é também uma variável aleatória.



Definição 3.16 Define-se média ou valor médio esperado de uma função de variável aleatória como sendo

$$E[Y] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

Exemplo 3.21 Seja $X = a \cos(\omega t + \Theta)$ com $\Theta = U(0, 2\pi)$. Determine a média e variância de X

$$f_{\theta}(\theta) = \begin{cases} \frac{1}{2\pi} & 0 \leq \theta \leq 2\pi \\ 0 & 0 < \theta < 2\pi \end{cases}$$

$$\begin{aligned}
E[X] &= E[a \cos(wt + \Theta)] = \int_0^{2\pi} \frac{1}{2\pi} a \cos(wt + \theta) d\theta = \\
&= \frac{a}{2\pi} [\sin(wt + \theta)]_0^{2\pi} = 0
\end{aligned}$$

$$\begin{aligned}
Var(X) &= E[X^2] = E[a^2 \cos^2(wt + \Theta)] = \\
&= \int_0^{2\pi} \frac{1}{2\pi} a^2 \cos^2(wt + \theta) d\theta = \\
&= \frac{a^2}{2\pi} \int_0^{2\pi} \left(\frac{1}{2} \cos 2(wt + \theta) + \frac{1}{2} \right) d\theta = \\
&= \frac{1}{2} a^2
\end{aligned}$$

Exemplo 3.22 Seja $X = A \cos(wt + \theta)$ com $A = U(-A_m, A_m)$. Determine a média e variância de X

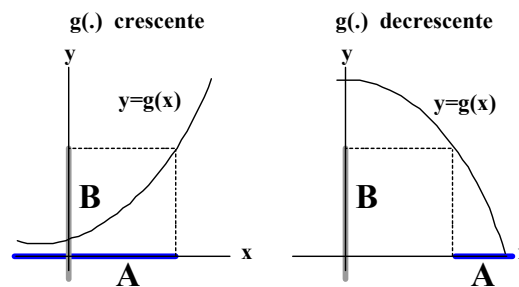
$$f_A(a) = \begin{cases} \frac{1}{2A_m} & |a| \leq A_m \\ 0 & |a| > A_m \end{cases}$$

$$E[X] = E[A \cos(wt + \theta)] = \cos(wt + \theta) E[A] = 0$$

$$\begin{aligned}
Var(X) &= E[X^2] = E[A^2 \cos^2(wt + \theta)] = \\
&= \cos^2(wt + \theta) E[A^2] = \\
&= \frac{A_m^2}{3} \cos^2(wt + \theta)
\end{aligned}$$

Um dos problemas que se põe em funções de variável aleatória, é o de determinar as funções de distribuição e de densidade, $F_Y(y)$ e $f_Y(y)$ conhecidas $g(x)$, $F_X(x)$ e $f_X(x)$.

Seja $g(x)$ uma função real de variável real contínua, com contradomínio em $(a, b) = (g(-\infty), g(+\infty))$ e com inversa $x = g^{-1}(y)$ derivável em todo o seu domínio (a, b)



1. Seja $g(x)$ crescente

Para determinar as relações procuradas vai-se recorrer de novo ao conceito de acontecimentos equivalentes. Atendendo à definição de função de variável aleatória, os acontecimentos A definido sobre X e B definido sobre Y são a imagem de um mesmo acontecimento em S logo são equivalentes e terão a mesma probabilidade. Como

$$A = \{X \leq x\} = \{X \leq g^{-1}(y)\}$$

e

$$B = \{Y \leq y\}$$

são equivalentes, as suas probabilidades são iguais

$$P[X \leq g^{-1}(y)] = P[Y \leq y]$$

ou seja

$$F_Y(y) = F_X(g^{-1}(y)), y \in (a, b)$$

Atendendo a que $\frac{dF}{dy} = \frac{dF}{du} \frac{du}{dy}$ com $u = g(x)$ e $x = g^{-1}(y)$

$$f_Y(y) = f_X(g^{-1}(y)) \frac{dx}{dy}, y \in (a, b)$$

2. Seja $g(x)$ decrescente

Analogamente, como

$$A = \{X \geq x\} = \{X > g^{-1}(y)\}$$

e

$$B = \{Y \leq y\}$$

são equivalentes, as suas probabilidades são iguais

$$P[X > g^{-1}(y)] = 1 - P[X \leq g^{-1}(y)] = P[Y \leq y]$$

ou seja

$$F_Y(y) = 1 - F_X(g^{-1}(y)), y \in (a, b)$$

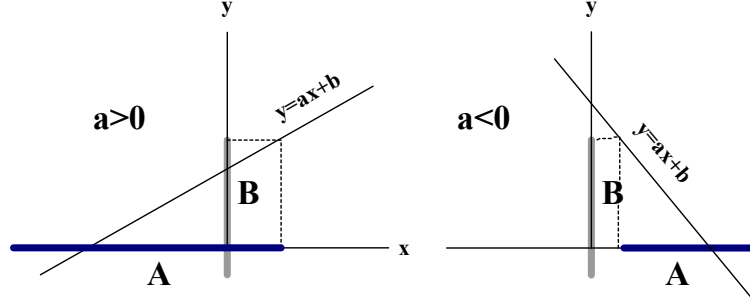
e derivando em ordem a y :

$$f_Y(y) = -f_X(g^{-1}(y)) \frac{dx}{dy}, y \in (a, b)$$

Como $\frac{dx}{dy} < 0$ quando $g(x)$ é decrescente, a densidade de probabilidade pode representar-se por

$$f_Y(y) = -f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right|, y \in (a, b)$$

3.10.1 Transformação linear



Considere-se a transformação linear $Y = aX + b$.

Neste caso tem-se $g^{-1}(y) : x = \frac{y-b}{a}$ e $\frac{dx}{dy} = \frac{1}{a}$ e portanto:

$$F_Y(y) = \begin{cases} F_X\left(\frac{y-b}{a}\right), & a > 0 \\ 1 - F_X\left(\frac{y-b}{a}\right), & a < 0 \end{cases}$$

e

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right), a \neq 0 \quad (3.6)$$

Exemplo 3.23 Se X for uma variável aleatória gausseana $N(\mu, \sigma)$. Caraterize a função $Y = aX + b$

Aplicando 3.6, obtem-se

$$f_Y(y) = \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-\frac{(y - (a\mu + b))^2}{2a^2\sigma^2}}$$

ou seja a transformada linear de uma variável aleatória gausseana é uma variável gausseana $N(a\mu + b, |a|\sigma)$

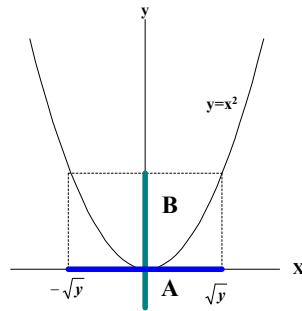
3.10.2 Transformação quadrática

Uma outra transformação é a transformação quadrática $Y = X^2$. O acontecimento $B = \{Y \leq y\}$ ocorre sempre que ocorre $A = \{X^2 \leq y\} = \{-\sqrt{y} \leq X \leq \sqrt{y}\}$ e sendo A e B equivalentes $P[A] = P[B]$ e $P[Y \leq y] = P[-\sqrt{y} \leq X \leq \sqrt{y}]$ e:

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

Derivando em ordem a y e atendendo a que $\frac{dF_X}{dy} = \frac{dF_X}{du} \frac{du}{dx}$ com $u = \sqrt{y}$,

$$f_Y(y) = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}} \quad (3.7)$$



Exemplo 3.24 Se X for uma variável aleatória uniforme $U(-2, 2)$ caracterize a função $Y = X^2$

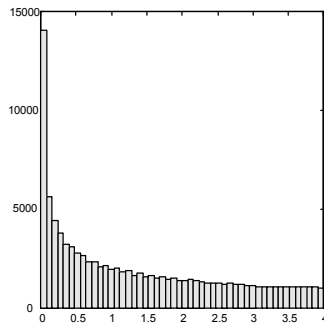
Se $y < 0 : F_Y(y) = 0;$

$$\text{Se } 0 \leq y \leq 4 : F_Y(y) = \frac{\frac{1}{4} + \frac{1}{4}}{2\sqrt{y}} = \frac{1}{4\sqrt{y}}$$

Se $y > 4 : F_Y(y) = 0$ porque $F_X(x) = 0$ se $x > 2$

Simulação:

```
x=4*(rand(1,100000)-0.5);
y=x.^2;
hist(y,50)
```



Exemplo 3.25 Se X for uma variável aleatória gaussiana de valor médio nulo e variância unitária caracterize a função $Y = X^2$

Aplicando 3.7, obtem-se:

$$f_X(\sqrt{y}) = f_X(-\sqrt{y}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}}$$

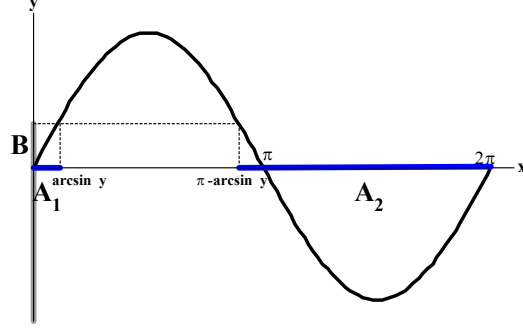
e portanto

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}$$

com $y \geq 0$ que é uma densidade de probabilidade de uma variável aleatória χ^2 com um grau de liberdade.

3.10.3 Transformação sinusoidal

Exemplo 3.26 Transformação $Y = \sin X$



Utilizando mais uma vez o conceito de equivalência de acontecimentos $P[A] = P[B]$, com $A = A_1 \cup A_2$:

Sendo $B = \{Y \leq y\}$:

Se $|y| \leq 1$: $A_1 \cup A_2 = \{0 \leq X \leq \arcsin y\} \cup \{\pi - \arcsin y \leq X \leq 2\pi\}$

$$P[Y \leq y] = P[A_1 \cup A_2] = P[[0, \arcsin y] \cup [\pi - \arcsin y, 2\pi]]$$

Se $y > 1$ $B = \{Y \leq y\}$ é o acontecimento certo pois $\sin \alpha \leq 1$

Se $y < -1$ $B = \{Y \leq y\} = \emptyset$ pois $\sin \alpha \geq -1$

Considerando que X é uniformemente distribuída em $[0, 2\pi]$, ou seja, $X = U(0, 2\pi)$:

$$\begin{aligned} P[[0, \arcsin y] \cup [\pi - \arcsin y, 2\pi]] &= \frac{\arcsin y + 2\pi - (\pi - \arcsin y)}{2\pi} = \\ &= \frac{1}{2} + \frac{\arcsin y}{\pi} \end{aligned}$$

e, portanto:

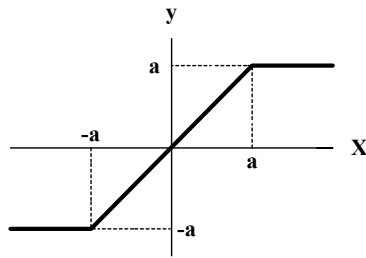
$$F_Y(y) = \begin{cases} 1, & y > 1 \\ \frac{1}{2} + \frac{\arcsin y}{\pi}, & |y| \leq 1 \\ 0, & y < -1 \end{cases}$$

e derivando em ordem a y :

$$f_Y(y) = \begin{cases} \frac{1}{\pi\sqrt{1-y^2}}, & |y| \leq 1 \\ 0, & |y| > 1 \end{cases}$$

3.10.4 Função limitadora

Exemplo 3.27 Considere a função $g(x)$ representada na figura e determine as relações entre $F_Y(y)$ e $F_X(x)$, e $f_Y(y)$ e $f_X(x)$.



Se $|Y| < a$ a transformação é a identidade e $F_Y(y) = F_X(y)$;

Se $Y > a$ $F_Y(y) = 1$ e se $Y < -a$ $F_Y(y) = 0$.

Os pontos críticos serão precisamente $Y = a$ e $Y = -a$. Nestes pontos surgem eventualmente descontinuidades se $F_X(a) \neq 1$ e $F_X(-a) \neq 0$ que terão os seguintes valores:

$$F_Y(a) - F_Y(a^-) = 1 - F_X(a) = F_Y(-a) - F_Y(-a^-) = F_X(-a) - 0.$$

Usando a função $u(x)$ pode-se representar a função de distribuição de y por:

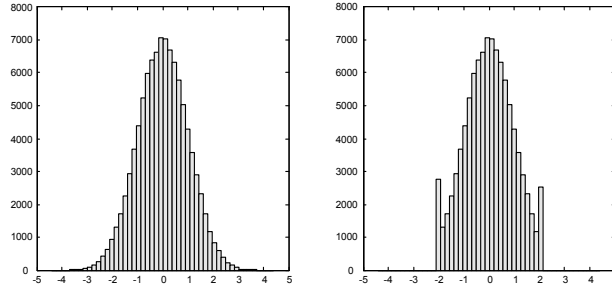
$$F_Y(y) = F_X(y) (u(y + a) - u(y - a)) + u(y - a)$$

Derivando em ordem a y

$$f_Y(y) = f_X(y) (u(y + a) - u(y - a)) + F_X(-a)\delta(y + a) + (1 - F_X(a))\delta(y - a)$$

Com o Matlab pode-se fazer a seguinte simulação: uma variável aleatória gaussiana $N(0,1)$ é transformada por uma função deste tipo com $a = 2$. Os resultados são apresentados sob a forma de histogramas para X e Y

```
x=randn(1,100000);y=x;
y(find(x<=-2))=-2;
y(find(x>=2))=2;
[N,centros]=hist(x,50); hist(x,50)
hist(y,centros)
```



Este exemplo baseia-se no comportamento de um circuito eletrônico vulgarmente usado, o circuito limitador. Como se viu um circuito destes pode alterar significativamente as propriedades estatísticas de um sinal, transformando um sinal monomodal (um único máximo) num sinal multimodal (com vários máximos relativos).

3.11 Exercícios

Exercício 3.2 *Demonstre que*

$$\begin{aligned} P[a \leq X \leq b] &= F_X(b) - F_X(a^-) \\ P[a < X < b] &= F_X(b^-) - F_X(a) \\ P[a \leq X < b] &= F_X(b^-) - F_X(a^-) \end{aligned}$$

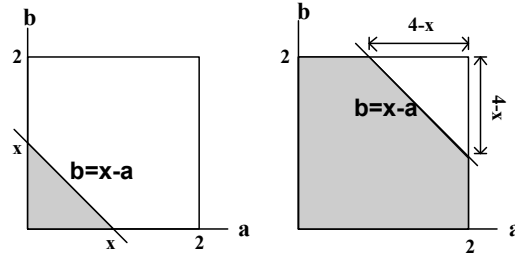
Exercício 3.3 $\{a \leq x \leq b\} = \{a\} \cup \{a < x \leq b\} \Rightarrow$

$$\begin{aligned} P[a \leq X \leq b] &= F_X(a) - F_X(a^-) + F_X(b) - F_X(a) = F_X(b) - F_X(a^-) \\ \{a < x < b\} \cup \{b\} &= \{a < x \leq b\} \Rightarrow \\ P[a < X < b] + F_X(b) - F_X(b^-) &= F_X(b) - F_X(a) \\ P[a < X < b] &= F_X(b^-) - F_X(a) \\ \{a \leq x < b\} \cup \{b\} &= \{a \leq x \leq b\} \Rightarrow \\ P[a \leq X < b] + F_X(b) - F_X(b^-) &= F_X(a) - F_X(a^-) + F_X(b) - F_X(a) \\ P[a \leq X < b] &= F_X(b^-) - F_X(a^-) \end{aligned}$$

Exercício 3.4 *Considere a experiência aleatória "escolha de um ponto (a, b) no interior de um quadrado de lado 2 com vértices opostos em $(0, 0)$ e $(2, 2)$. Defina a variável aleatória como sendo a soma das coordenadas do ponto, $X = a + b$. Determine:*

1. O contradomínio S_X da variável X
2. A função de distribuição da variável X
3. A função de densidade de probabilidade da variável X

Para qualquer $\zeta = (a, b) \in S$ define-se $x = X(\zeta) = a + b$



1. Como $0 \leq a \leq 2$ e $0 \leq b \leq 2 \Rightarrow 0 \leq a + b \leq 4 \Rightarrow 0 \leq x \leq 2 \Rightarrow S_X = [0, 4]$
2. Se $x < 0$ acontecimento $\{X \leq x\} = \{X \leq a + b\}$ é equivalente em S a \emptyset e $P[X \leq x] = 0$;

Se $0 \leq x < 2$ acontecimento $\{X \leq x\} = \{X \leq a + b\}$ é equivalente em S

ao triângulo da figura e $P[X \leq x] = \frac{\frac{x^2}{2}}{4} = \frac{x^2}{8}$.

Se $x < 0$ acontecimento $\{X \leq x\} = \{X \leq a + b\}$ é equivalente em S a \emptyset e $P[X \leq x] = 0$;

Se $2 \leq x < 4$ acontecimento $\{X \leq x\} = \{X \leq a + b\}$ é equivalente em S

ao pentágono da figura e $P[X \leq x] = \frac{4 - \frac{(4-x)^2}{2}}{4} = 1 - \frac{1}{8}(4-x)^2$.

Se $x \geq 4$ o acontecimento $\{X \leq x\} = \{X \leq a + b\}$ é equivalente em S ao acontecimento certo e $P[X \leq x] = 1$;

A função de distribuição será então:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{8}x^2, & 0 \leq x < 2 \\ 1 - \frac{1}{8}(4-x)^2, & 2 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$

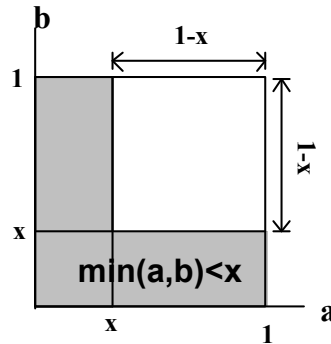
3. Utilizando a definição a função de densidade de probabilidade será

$$f_X(x) = \begin{cases} 0, & x < 0 \vee x \geq 4 \\ \frac{1}{4}x, & 0 \leq x < 2 \\ \frac{1}{4}(4-x) & 2 \leq x < 4 \end{cases}$$

Exercício 3.5 Considere a experiência aleatória "escolha de um ponto (a, b) no interior de um quadrado de lado 1 com vértices opostos em $(0, 0)$ e $(1, 1)$. Defina sobre esta experiência aleatória a variável aleatória X como sendo a maior das coordenadas do ponto, $X = \min(a, b)$. Determine:

1. O contradomínio S_X da variável X
2. A função de distribuição da variável X
3. A função de densidade de probabilidade da variável X

Para qualquer $\zeta = (a, b) \in S$ define-se $x = X(\zeta) = \min(a, b)$



1. Como $0 \leq a \leq 1$ e $0 \leq b \leq 1 \Rightarrow 0 \leq \min(a, b) \leq 1 \Rightarrow 0 \leq x \leq 1 \Rightarrow S_X = [0, 1]$
2. Se $x < 0$ o acontecimento $\{X \leq x\} = \{X \leq \min(a, b)\}$ é equivalente em S a \emptyset e $P[X \leq x] = 0$;
 Se $0 \leq x \leq 1$ o acontecimento $\{X \leq x\}$ é equivalente em S à área sombreada da figura e $P[X \leq x] = \frac{1 - (1-x)^2}{1} = 2x - x^2$.
 Se $x > 1$ o acontecimento $\{X \leq x\} = \{X \leq \min(a, b)\}$ é equivalente em S ao acontecimento certo e $P[X \leq x] = 1$;
 A função de distribuição será então:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 2x - x^2, & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

3. Utilizando a definição a função de densidade de probabilidade será

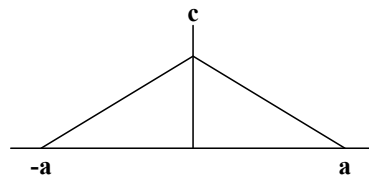
$$f_X(x) = \begin{cases} 0, & x < 0 \vee x > 1 \\ 2 - 2x, & 0 \leq x \leq 1 \end{cases}$$

Exercício 3.6 Uma variável aleatória tem uma densidade de probabilidade

$$f_X(x) = \begin{cases} 0, & |x| > a \\ c(1 - \frac{|x|}{a}), & |x| \leq a \end{cases}, a > 0$$

1. Determine a constante c
2. Determine a função de distribuição
3. Determine b de modo a que $P[|X| < b] = \frac{1}{2}$
4. Elabore um programa em Matlab que permita gerar valores de uma variável aleatória que tenha esta densidade de probabilidade.

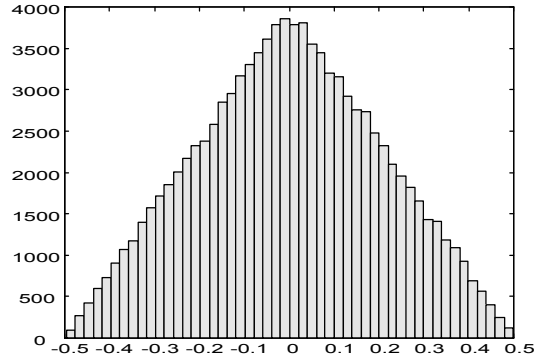
A função $f_X(x)$ tem a seguinte representação gráfica:



1. Como $\int_{-\infty}^{+\infty} f_X(x)dx = 1 \Leftrightarrow \int_{-a}^a f_X(x)dx = 1 \Leftrightarrow \text{Área do triângulo} = 1$ e portanto $\frac{c \cdot 2a}{2} = 1 \Rightarrow c = \frac{1}{a}$
2. Se $x < -a$, $F_X(x) = \int_{-\infty}^x f_X(x)dx = 0$
 Se $-a \leq x < 0$, $F_X(x) = \int_{-a}^x \frac{1}{a}(1 + \frac{x}{a})dx = \frac{1}{2} + \frac{x}{a} + \frac{x^2}{2a^2}$
 Se $0 \leq x < a$, $F_X(x) = \int_{-a}^0 \frac{1}{a}(1 + \frac{x}{a})dx + \int_0^x \frac{1}{a}(1 - \frac{x}{a})dx = \frac{1}{2} + \frac{x}{a} - \frac{x^2}{2a^2}$
 Se $x \geq a$, $F_X(x) = 1$
3. $P[|X| < b] = P[-b < x < b] = F_X(b) - F_X(-b) = \frac{1}{2}$
 $\frac{1}{2} + \frac{b}{a} - \frac{b^2}{2a^2} - (\frac{1}{2} + \frac{-b}{a} + \frac{b^2}{2a^2}) = \frac{1}{2} \Leftrightarrow 2\frac{b}{a} - \frac{b^2}{a^2} = \frac{1}{2} \Rightarrow \frac{b}{a} = 1 - \frac{1}{2}\sqrt{2}$
4. Determinação da função inversa de $F_X(x)$
 $x = a(-1 + \sqrt{2u})$, para $0 \leq u < \frac{1}{2}$
 $x = a(1 - \sqrt{2 - 2u})$, para $\frac{1}{2} \leq u < 1$

O programa seguinte gera 100000 amostras, com $a = 0.5$ e traça em seguida o histograma das amostras.

```
n=100000;a=0.5;
x=[(-1+sqrt(rand(1,n/2))) (1-sqrt(2-(1+rand(1,n))))]*a;
x=x(randperm(n));
hist(x,50)
```



Exercício 3.7 *Problema da agulha ou problema de Buffon*

Uma agulha de comprimento l é lançada sobre um plano coberto por linhas retas paralelas separadas de uma distância $d, d > l$. Determine a probabilidade da agulha cair sobre uma das linhas retas do plano.

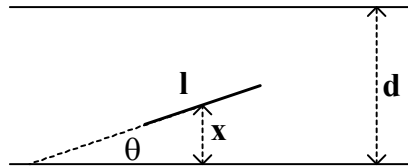
Considere x a distância do ponto médio da agulha à reta mais próxima e θ o ângulo que a agulha forma com as retas do plano. Os seus valores são variáveis aleatórias X e Θ independentes e uniformemente distribuídas nos intervalos $[0, d/2]$ e $[0, \pi]$. Dada a independência a função de densidade conjunta será

$$f_{X\Theta}(x, \theta) = \frac{2}{d\pi}, 0 \leq x \leq d/2 \wedge 0 \leq \theta \leq \pi$$

e uniformemente nula no resto do plano.

O acontecimento "agulha cai interceptando uma das linhas retas do plano" pode ser facilmente caracterizado por $x < \frac{l}{2} \sin \theta$ e portanto a sua probabilidade dada por

$$P \left[X < \frac{l}{2} \sin \Theta \right] = \int_0^\pi \int_0^{\frac{l}{2} \sin \theta} \frac{2}{d\pi} dx d\theta = \frac{2l}{d\pi}$$



Exercício 3.8 *Considere a seguinte função real de variável real*

$$f(x) = \begin{cases} |ax|, & |x| \leq c \\ 0, & |x| > c \end{cases}, a > 0$$

Indique as condições que deve impor a c para que a função $f(x)$ possa representar uma função de densidade de probabilidade e determine a respectiva função de distribuição.

$$\begin{aligned}\int_{-\infty}^{+\infty} f(x)dx &= 1 \Leftrightarrow \int_{-\infty}^{+\infty} |ax| dx = \int_{-c}^0 -axdx + \int_0^c axdx \\ &= c^2a = 1 \Rightarrow c = \frac{1}{\sqrt{a}}\end{aligned}$$

Função de distribuição:

Se $x < -c$, $F_X(x) = 0$;

Se $-c \leq x < 0$, $F_X(x) = \int_{-c}^x -axdx = -\frac{1}{2}x^2a + \frac{1}{2}c^2a = \frac{1}{2} - \frac{1}{2}ax^2$;

Se $0 \leq x < c$, $F_X(x) = \int_{-c}^0 -axdx + \int_0^x axdx = \frac{1}{2}c^2a + \frac{1}{2}x^2a = \frac{1}{2} + \frac{1}{2}ax^2$;

Se $x \geq c$, $F_X(x) = 1$

Exercício 3.9 Considere a seguinte função real de variável real

$$f(x) = ce^{-a|x|}, a > 0$$

Indique as condições que deve impor a c para que a função $f(x)$ possa representar uma função de densidade de probabilidade e determine a respectiva função de distribuição.

$$\begin{aligned}\int_{-\infty}^{+\infty} f(x)dx &= 1 \Leftrightarrow \int_{-\infty}^{+\infty} ce^{-a|x|}dx = \int_{-\infty}^0 ce^{ax}dx + \int_0^{\infty} ce^{-ax}dx = 2\frac{c}{a} = \\ 1 &\Rightarrow c = \frac{a}{2}\end{aligned}$$

Função de distribuição:

Se $x < 0$, $F_X(x) = \int_{-\infty}^x \frac{1}{2}ae^{ax}dx = \frac{1}{2}e^{ax}$

Se $0 \leq x < c$, $F_X(x) = \int_{-\infty}^0 \frac{1}{2}ae^{ax}dx + \int_0^x \frac{1}{2}ae^{-ax}dx = 1 - \frac{1}{2}e^{-ax}$

Exercício 3.10 Sendo c um número real, demonstre as seguintes igualdades

$$1. E[c] = 0; E[X + c] = E[X] + c; E[cX] = cE[X]$$

$$2. Var(c) = 0; Var(X + c) = Var(X); Var(kX) = c^2Var(X)$$

$$1. E[c] = \int_{-\infty}^{+\infty} cf_X(x)dx = c \int_{-\infty}^{+\infty} f_X(x)dx = c$$

$$\begin{aligned}E[X + c] &= \int_{-\infty}^{+\infty} (x + c)f_X(x)dx = c \int_{-\infty}^{+\infty} f_X(x)dx + \int_{-\infty}^{+\infty} xf_X(x)dx = \\ &= c + E[x]\end{aligned}$$

$$E[cX] = \int_{-\infty}^{+\infty} cx f_X(x)dx = c \int_{-\infty}^{+\infty} x f_X(x)dx = cE[x]$$

$$\begin{aligned}
2. \text{Var}(c) &= \int_{-\infty}^{+\infty} (c - E[c])^2 f_X(x) dx = \int_{-\infty}^{+\infty} 0 \cdot f_X(x) dx = 0 \\
\text{Var}(X + c) &= \int_{-\infty}^{+\infty} (x + c - E[x + c])^2 f_X(x) dx = \\
&= \int_{-\infty}^{+\infty} (x + c - E[x] - c)^2 f_X(x) dx = \text{Var}(X) \\
\text{Var}(cX) &= \int_{-\infty}^{+\infty} (cx - E[cx])^2 f_X(x) dx = \int_{-\infty}^{+\infty} (cx - cE[x])^2 f_X(x) dx = \\
&= \int_{-\infty}^{+\infty} c^2 (x - E[x])^2 f_X(x) dx = c^2 \text{Var}(X)
\end{aligned}$$

Exercício 3.11 Determine a média e a variância de uma variável obtida por transformação linear de uma variável uniforme discreta.

Seja U uma variável discreta uniforme, a aplicação de uma transformação linear conduz a $V = a + bU$

$$E[V] = E[a + bU] = a + bE[U] = a + \frac{b}{2}n(n+1) \text{ pois } E[U] = \frac{1}{2}n(n+1)$$

$$\text{Var}(V) = \text{Var}(a + bU) = \text{Var}(bU) = b^2 \text{Var}(U) = b^2 \frac{n^2-1}{12}$$

Exercício 3.12 Usando a função geradora de probabilidade determine a média e a variância de uma variável aleatória de Poisson

A função geradora de probabilidade é, de acordo com a definição:

$$\begin{aligned}
G_X(z) &= \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} e^{-\alpha} z^k = e^{-\alpha} \sum_{k=0}^{\infty} \frac{(\alpha z)^k}{k!} = \\
&= e^{-\alpha} e^{\alpha z} = e^{\alpha(z-1)}
\end{aligned}$$

$$\frac{d}{dz} e^{\alpha(z-1)} = \alpha e^{\alpha(z-1)}$$

$$\frac{d^2}{dz^2} e^{\alpha(z-1)} = \alpha^2 e^{\alpha(z-1)}$$

De acordo com as fórmulas em 3.3 :

$$E[X] = G'_X(1) = \alpha$$

$$\text{Var}(X) = G''_X(1) + G'_X(1) - \left(G'_X(1)\right)^2 = \alpha^2 + \alpha - \alpha^2 = \alpha$$

Exercício 3.13 Demonstre que para a variável normal $N(m, \sigma)$ a média é m e a variância σ^2

Exercício 3.14 Considere-se a variável normalizada $X_N = N(0, 1)$.

$$E[X_N] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}x^2} dx = \left[-e^{-\frac{1}{2}x^2} \right]_{-\infty}^{+\infty} = 0 - 0 = 0$$

$$\begin{aligned}
\text{Var}(X_N) &= E[X_N^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} xxe^{-\frac{1}{2}x^2} dx = \\
&= \frac{1}{\sqrt{2\pi}} \left[-xe^{-\frac{1}{2}x^2} \right]_{-\infty}^{+\infty} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (-e^{-\frac{1}{2}x^2}) dx = \\
&= 0 - 0 + 1 = 1 \\
E[\sigma X_N + m] &= \sigma E[X_N] + m = m \\
\text{Var}(\sigma X_N + m) &= \text{Var}(\sigma X_N) = \sigma^2
\end{aligned}$$

Exercício 3.15 O desvio médio definido como o valor médio absoluto em relação à média é $E[|X - m|]$. Calcule o desvio médio da variável gaussiana

$$\begin{aligned}
E[|X - m|] &= \int_{-\infty}^0 \frac{(-x - m)}{\sqrt{2\pi}\sigma} e^{-\frac{(-x - m)^2}{2\sigma^2}} dx + \int_0^{\infty} \frac{(x - m)}{\sqrt{2\pi}\sigma} e^{-\frac{(x - m)^2}{2\sigma^2}} dx = \\
&= \sigma \sqrt{\frac{2}{\pi}} \simeq 0.8\sigma
\end{aligned}$$

Exercício 3.16 Para uma variável X gaussiana, determine k por forma que $p = P[|X - m| \leq k\sigma]$ para $p = 0.5, 0.9, 0.95$ e 0.99

$$p = P[|X - m| \leq k\sigma] = 1 - 2Q(k)$$

Recorrendo a um programa de cálculo numérico:

p	k
0.50	0.6745
0.90	1.6448
0.95	1.9598
0.99	2.5752

Exercício 3.17 Dois chips estão a ser considerados para serem usados num dado circuito. O tempo de vida em horas dos chips é modelado por uma variável $N(20000, 4000)$ para o chip 1 e $N(22000, 1000)$ para o nº 2. (Nota: despreze as probabilidades de vida negativa)

Que chip deve escolher se o tempo de vida de projeto para o circuito for de 20000 horas? e se for de 24000?

$$\begin{aligned}
P[X_1 > 20000] &= P\left[U_X > \frac{20000 - 20000}{4000}\right] = \\
&= P[U_X > 0] = Q(0) = 0.5
\end{aligned}$$

$$\begin{aligned}
P[X_1 > 20000] &= P\left[U_X > \frac{20000 - 22000}{1000}\right] = \\
&= P[U_X > -2] = Q(-2) = \\
&= 1 - Q(2) = 0.972
\end{aligned}$$

Nesta situação deve escolher o chip 2.

$$\begin{aligned} P[X_1 > 24000] &= P\left[U_X > \frac{24000 - 20000}{4000}\right] = \\ &= P[U_X > 1] = Q(0) = 0.158 \end{aligned}$$

$$\begin{aligned} P[X_1 > 24000] &= P\left[U_X > \frac{24000 - 22000}{1000}\right] = \\ &= P[U_X > 2] = Q(2) = 0.028 \end{aligned}$$

Nesta situação deve escolhe o chip 1.

Exercício 3.18 0.1% de um determinado tipo de chips apresentam um defeito que impede de serem usados numa dada aplicação. Supondo que precisa de 50 a funcionar, quantos deve comprar para que tenha uma probabilidade de pelo menos 99% de ter 50 a funcionar em boas condições.

1. Resolva o problema usando uma lei de probabilidade binomial e refaça os cálculos admitindo uma lei de Poisson.
2. Repita os cálculos anteriores se pretender ter 500 chips em boas condições.

Respostas:

1. Procura-se um número n de chips dos quais pelo menos 50 têm as boas características, sabendo que a probabilidade de estar nestas condições é $p = 0.999$.

(a) Usando a lei binomial

$$P[k \geq 50] = \sum_{k=50}^n C_k^n \cdot 0.999^k \cdot 0.001^{n-k} \geq 0.99$$

A resolução numérica desta inequação (ver exercício nº do capítulo 2) conduz a $n \geq 51$

- (b) A lei de Poisson pode ser usada como aproximação da binomial fazendo $\alpha = np$

Pode-se por o problema em termos do número de avarias, isto é, procure-se n por forma a que só possa haver $n - 50$ em más condições. Então como a probabilidade de estar em más condições é $p = 0.001$

$$P[k \leq n - 50] = \sum_{k=0}^{n-50} \frac{(0.001n)^k}{k!} e^{-0.001n} \geq 0.99$$

A resolução numérica conduz também a $n \geq 51$ e pois

$$\sum_{k=0}^{51-50} \frac{(0.001 \cdot 51)^k}{k!} e^{-0.001 \cdot 51} = 0.999$$

2. Se se pretende 500 chips em boas condições:

(a)

$$P[k \geq 500] = \sum_{k=500}^n C_k^n \cdot 0.999^k \cdot 0.001^{n-k} \geq 0.99$$

com uma solução numérica $n \geq 503$

ou

(b)

$$P[k \leq n - 500] = \sum_{k=0}^{n-500} \frac{(0.001n)^k}{k!} e^{-0.001n} \geq 0.99$$

com a mesma solução numérica $n \geq 503$

Exercício 3.19 *Uma companhia de aviação utiliza aviões de 200 lugares e pratica overbooking, ou seja, vende mais bilhetes do que a capacidade do avião sabendo que 3% dos seus clientes que reservam uma viagem não comparecem à partida.*

1. *Sabendo que para uma determinada viagem vendeu 210 bilhetes, qual a probabilidade de todos os passageiros que comparecerem à partida terem lugar no avião?*
2. *Quantos bilhetes pode vender por forma a que todos os passageiros que comparecerem à partida terem lugar no avião com uma probabilidade não inferior a 95%?*

O número de passageiros k que comparece e tem lugar, é escolhido entre n que adquiriram bilhete, seguindo portanto uma lei binomial com parâmetro $p = 1 - 0.03 = 0.97$

$$1. \sum_{k=0}^{200} C_k^{210} \cdot 0.97^k \cdot 0.03^{210-k} = 0.103$$

$$2. P = \sum_{k=0}^{200} C_k^n \cdot 0.97^k \cdot 0.03^{n-k} > 0.95$$

A solução numérica para $n = 203, P = 0.9445$ e $n = 202, P = 0.985$, portanto $n = 202$.

Exercício 3.20 *Escreva um programa em Matlab que lhe permita calcular o valor da função $Q(x)$ com uma dada precisão.*

```
function y=q1(x,prec)
%calcula a função Q(x) com uma dada precisão
%por defeito a precisão será 1.e-6
if nargin==1
    prec=.000001;
end
a=x;
```

```
b=x+1;
inic=quad('exp2',a,b);
final=-1;
while abs(inic-final)>prec
    b=b+1;
    final=inic;
    inic=quad('exp2',a,b);
end
y=inic;
function y=exp2(x)
y=1/sqrt(2*pi)*exp(-x.^2/2);
q1(2)      0.0228
q1(0)      0.5000
q1(2,.0001) 0.0227
```

Chapter 4

Variáveis aleatórias multidimensionais

4.1 Vetores aleatórios

Em muitas situações uma experiência aleatória não corresponde apenas a um valor observado mas sim a um conjunto finito de observações. Sendo S o conjunto de amostragem e \mathcal{F} uma σ -álgebra de S na qual se encontra definida uma lei de probabilidade, no caso multidimensional ter-se-á que, para qualquer $\zeta \in S$, se faz corresponder um vetor $\mathbf{X} = (X_1, \dots, X_n) \in \mathbb{R}^n$.

Para qualquer $B \subset \mathbb{R}^n$ pode-se atribuir uma probabilidade recorrendo à noção de equivalência de acontecimentos já apresentada para o caso unidimensional: $P[B] = P[\mathbf{X}(\zeta) \in B] = P[\zeta \in \mathbf{X}^{-1}(B)]$ em que $\mathbf{X}^{-1}(B) = A$, a imagem inversa de B por \mathbf{X} pertence a \mathcal{F} e, por hipótese, tem uma probabilidade definida.

Definição 4.1 Quando para qualquer ponto $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, o acontecimento

$$A = \{\zeta : X_1(\zeta) \leq x_1, X_n(\zeta) \leq x_n\}$$

pertencer a \mathcal{F} diz-se que $\mathbf{X}(\zeta) = (X_1(\zeta), \dots, X_n(\zeta))$ é um vetor aleatório ou uma variável aleatória de dimensão n .

Tal como se fez para o caso unidimensional e para simplificar a notação, será omitida geralmente a variável ζ .

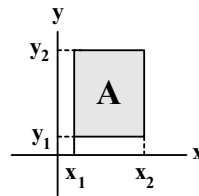
Exemplo 4.1 Numa população humana, para cada indivíduo escolhido ao acaso, definem-se os seguintes parâmetros biométricos: X_1 = altura, X_2 = peso e X_3 = idade. O vetor (X_1, X_2, X_3) é uma variável aleatória tridimensional.

Exemplo 4.2 Um sinal elétrico é amostrado e medido a partir de um instante de tempo ζ arbitrário. São recolhidas e registadas 100 amostras $V_i, i = 1, \dots, 100$. O vetor $\mathbf{V} = (V_1, \dots, V_{100})$ é uma variável aleatória de dimensão 100.

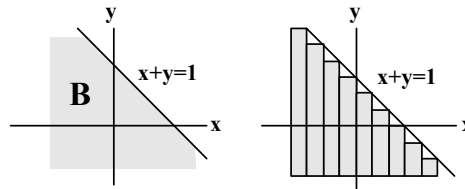
Um acontecimento A em \mathbb{R}^n será definido, em geral, por uma ou mais relações que envolvem as diferentes variáveis X_i , mas pode ser definido exclusivamente a partir de uma única variável. Nesta situação será referido como o acontecimento $\{X_i \text{ em } A_i\}$ significando "a propriedade de X_i que caracteriza o acontecimento A_i ". Uma classe importante de acontecimentos será a que se pode expressar pela intercepção de acontecimentos deste tipo, isto é, acontecimentos $A = \{X_1 \text{ em } A_1\} \cap \dots \cap \{X_n \text{ em } A_n\}$. Esta forma de definir acontecimentos designa-se por forma produto ou formulação por produto que será largamente utilizada no texto que se segue.

Exemplo 4.3 O acontecimento $A = \{(x, y); x_1 \leq X \leq x_2 \wedge y_1 \leq Y \leq y_2\}$ é um acontecimento descrito na forma de produto pois

$$A = \{x_1 \leq X \leq x_2\} \cap \{y_1 \leq Y \leq y_2\}$$



enquanto que $B = \{(x, y) : x + y < 1\}$ não o é. No entanto este acontecimento pode ser considerado como o limite de uma sucessão de acontecimentos constituídos por uniões de acontecimentos produto, como se exemplifica na figura.



4.1.1 Pares de variáveis aleatórias

Como o caso bidimensional é bastante mais simples de descrever e de visualizar que o caso geral, justifica-se um estudo detalhado. Como notação para um vetor bidimensional será usado (X, Y) em vez de (X_1, X_2) e será também designado por par de variáveis aleatórias.

4.2 Função de probabilidade conjunta

Considere-se o par de variáveis (X, Y) que tem valores em

$$S_{XY} = \{(x_j, y_k), j = 1, 2, \dots, k = 1, 2, \dots\}$$

Definição 4.2 A função de probabilidade conjunta de X e Y é

$$\begin{aligned} p_{XY}(x_j, y_k) &= P[\{X = x_j\} \cap \{Y = y_k\}] = \\ &= P[X = x_j, Y = y_k], \quad j = 1, 2, \dots, k = 1, 2, \dots \end{aligned}$$

E pelo axioma 2.1

$$\sum_{j,k} p_{XY}(x_j, y_k) = 1$$

Um acontecimento A definido sobre o par (X, Y) terá uma probabilidade

$$P[A] = \sum_{(x_j, y_k) \in A} p_{XY}(x_j, y_k)$$

4.2.1 Funções de probabilidade marginal

Definição 4.3 A função de probabilidade marginal de X é

$$p_X(x_j) = P[\{X = x_j\} \text{ e qualquer } Y]$$

Como

$$\begin{aligned} P\left[\{X = x_j\} \cap \bigcup_k \{Y = y_k\}\right] &= P\left[\bigcup_k \{X = x_j\} \cap \{Y = y_k\}\right] = \\ &= \sum_k P[\{X = x_j\} \cap \{Y = y_k\}] \end{aligned}$$

vem

$$p_X(x_j) = \sum_k p_{XY}(x_j, y_k), \quad j = 1, 2, \dots$$

e analogamente

$$p_Y(y_k) = \sum_j p_{XY}(x_j, y_k), \quad k = 1, 2, \dots$$

Exemplo 4.4 O comportamento conjunto não pode ser inferido do comportamento marginal, veja-se por exemplo, o caso de duas moedas cujo lançamento conduz a resultados com a probabilidade conjunta da tabela:

	C	R
C	$\frac{3}{8}$	$\frac{1}{8}$
R	$\frac{1}{8}$	$\frac{3}{8}$

Cada moeda é normal com $P[C] = P[R] = \frac{1}{2}$, mas o seu comportamento conjunto não é o de duas moedas normais.

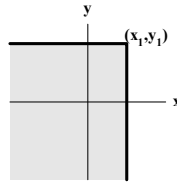
4.3 Função de distribuição conjunta

Para o par de variáveis aleatórias (X, Y) a função de distribuição conjunta define-se como:

Definição 4.4

$$F_{XY}(x, y) = P[\{X \leq x\} \cap \{Y \leq y\}] = P[X \leq x, Y \leq y]$$

Para um ponto (x_1, y_1) a função de distribuição conjunta $F_{XY}(x_1, y_1)$ é a probabilidade do acontecimento que se indica na figura:



4.3.1 Propriedades

1. $x_1 \leq x_2 \wedge y_1 \leq y_2 \Rightarrow F_{XY}(x_1, y_1) \leq F_{XY}(x_2, y_2)$

Demonstração.

$$(x_1 \leq x_2 \Rightarrow \{X \leq x_1\} \subset \{X \leq x_2\})$$

$$(y_1 \leq y_2 \Rightarrow \{Y \leq y_1\} \subset \{Y \leq y_2\})$$

$$(\{X \leq x_1\} \cap \{Y \leq y_1\}) \subset (\{X \leq x_2\} \cap \{Y \leq y_2\})$$

$$P[X \leq x_1, Y \leq y_1] \leq P[X \leq x_2, Y \leq y_2]$$

■

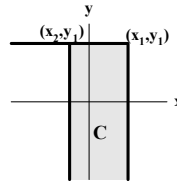
2. $\lim_{x \rightarrow -\infty} F_{XY}(x, y) = 0$ e $\lim_{y \rightarrow -\infty} F_{XY}(x, y) = 0$

Demonstração. Quando $x \rightarrow -\infty$ o conjunto $\{X \leq x\} \rightarrow \{X \leq -\infty\} = \emptyset$. Analogamente para y ■

3. $\lim_{x, y \rightarrow \infty} F_{XY}(x, y) = 1$

Demonstração. Quando $x, y \rightarrow \infty$ o conjunto $\{X \leq x\} \cap \{Y \leq y\} \rightarrow \mathbb{R}^2$, ou seja, para o acontecimento certo. ■

4. $P[x_2 < X \leq x_1, Y \leq y_1] = F_{XY}(x_1, y_1) - F_{XY}(x_2, y_1)$

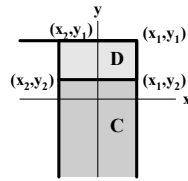


Demonstração.

$$\begin{aligned}
 \{X \leq x_1, Y \leq y_1\} &= C \cup \{X \leq x_2, Y \leq y_1\} \\
 P[C] &= P[x_2 < X \leq x_1, Y \leq y_1] = \\
 &= P[X \leq x_1, Y \leq y_1] - P[X \leq x_2, Y \leq y_1] = \\
 &= F_{XY}(x_1, y_1) - F_{XY}(x_2, y_1)
 \end{aligned}$$

■

$$5. P[x_2 < X \leq x_1, y_2 < Y \leq y_1] = F_{XY}(x_1, y_1) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_2, y_2)$$



Demonstração.

$$P[x_2 < X \leq x_1, y_2 < Y \leq y_1] = P[D \cup C] - P[C]$$

$$\begin{aligned}
 P[D \cup C] &= F_{XY}(x_1, y_1) - F_{XY}(x_2, y_1) \\
 P[C] &= (F_{XY}(x_1, y_2) - F_{XY}(x_2, y_2))
 \end{aligned}$$

■

$$6. \lim_{x \rightarrow a^+} F_{XY}(x, y) = F_{XY}(a, y) \text{ e } \lim_{y \rightarrow b^+} F_{XY}(x, y) = F_{XY}(x, b)$$

Demonstração. É uma consequência direta da continuidade à direita da função de distribuição para uma variável

Para $h > 0$, quando $h \rightarrow 0$

$$\{X \leq b + h, Y \leq y\} \rightarrow \{X \leq b, Y \leq y\}$$

e

$$P[X \leq b + h, Y \leq y] \rightarrow P[X \leq b, Y \leq y]$$

■

4.3.2 Funções de distribuição marginal

Definição 4.5 A função de distribuição marginal de X é

$$F_X(x) = P[\{X \leq x\} \text{ e qualquer } Y] = P[\{X \leq x\} \cap \{Y \leq \infty\}] = \lim_{y \rightarrow \infty} F_{XY}(x, y)$$

Analogamente

$$F_Y(y) = P[\{Y \leq y\} \text{ e qualquer } X] = P[\{X \leq \infty\} \cap \{Y \leq y\}] = \lim_{x \rightarrow \infty} F_{XY}(x, y)$$

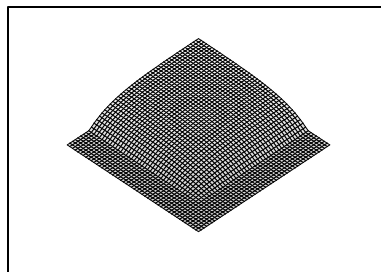
Exemplo 4.5 Considere o par de variáveis aleatórias (X, Y) com uma função de distribuição conjunta:

$$F_{XY}(x, y) = \begin{cases} 0, & x \leq 0 \vee y \leq 0 \\ (1 - e^{-x})(1 - e^{-y}), & x > 0 \wedge y > 0 \end{cases}$$

Determine as funções de distribuição marginais

$$z = \begin{cases} 0 & \text{if } x \leq 0 \vee y \leq 0 \\ (1 - e^{-x})(1 - e^{-y}) & \text{if } x > 0 \wedge y > 0 \end{cases}$$

A função de distribuição conjunta pode ser visualizada através do programa de visualização tridimensional que faz parte do Matlab, ou de outro semelhante:



$$\begin{aligned} F_X(x) &= \lim_{y \rightarrow \infty} F_{XY}(x, y) = \begin{cases} 0, & x \leq 0 \\ (1 - e^{-x}), & x > 0 \end{cases} \\ F_Y(y) &= \lim_{x \rightarrow \infty} F_{XY}(x, y) = \begin{cases} 0, & y \leq 0 \\ (1 - e^{-y}), & y > 0 \end{cases} \end{aligned}$$

Este exemplo permite compreender melhor a designação de "marginal" para estas funções. Como se pode ver são as funções determinadas nas "margens" da distribuição bidimensional.

4.4 Função de densidade de probabilidade conjunta

Definição 4.6 A função $f_{XY}(x, y)$, não negativa, é a função de densidade de probabilidade do par de variáveis aleatórias (X, Y) conjuntamente contínuas, se para qualquer acontecimento A a sua probabilidade for expressa por

$$P[A] = \int \int_A f_{XY}(x, y) dx dy$$

4.4.1 Propriedades

$$1. \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = 1$$

Demonstração. Os limites de integração correspondem a todo o plano, portanto ao acontecimento certo. ■

$$2. F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(x, y) dx dy$$

Demonstração. Os limites de integração correspondem ao acontecimento $\{X \leq x\} \cap \{Y \leq y\}$ que, por definição, é a função de distribuição em (x, y) ■

Se a função $F_{XY}(x, y)$ é contínua em (x, y) então $f(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$

4.4.2 Funções de densidade marginal

Como a função de densidade marginal se refere a uma variável aleatória, utilizando a definição geral 3.1

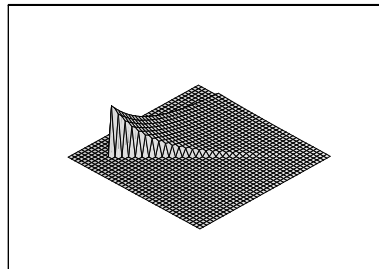
$$\begin{aligned} f_X(x) &= \frac{dF_X(x)}{dx} = \frac{d}{dx} \lim_{y \rightarrow \infty} F_{XY}(x, y) = \\ &= \frac{d}{dx} \int_{-\infty}^x \left(\int_{-\infty}^{+\infty} f_{XY}(x, y) dy \right) dx = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy \end{aligned}$$

e analogamente

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx$$

Exemplo 4.6 Um par de variáveis aleatórias (X, Y) tem uma função de densidade conjunta:

$$f_{XY}(x, y) = \begin{cases} 0, & x \leq 0 \vee y \leq 0 \vee x \leq y \\ ce^{-x-y}, & y > x > 0 \end{cases}$$

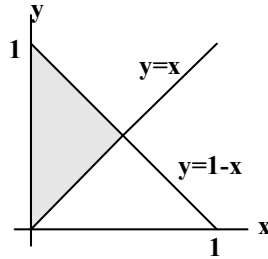


1. Determine o valor da constante c
2. Determine $P[X + Y < 1]$

Respostas:

$$1. 1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = \int_0^{+\infty} \int_0^y ce^{-x-y} dx dy = \frac{1}{2}c \Rightarrow c = 2$$

2.



$$\int_0^{0.5} \int_0^y 2e^{-x-y} dx dy + \int_{0.5}^1 \int_0^{1-y} 2e^{-x-y} dx dy = 0.26424 \text{ ou}$$

$$\int_0^{0.5} \int_x^{1-x} 2e^{-x-y} dy dx = 0.26424$$

Exemplo 4.7 Considere um par de variáveis aleatórias (X, Y) que é uniformemente distribuído no quadrado unitário, tendo portanto uma densidade conjunta definida por

$$f_{XY}(x, y) = \begin{cases} k, & 0 < x \leq 1 \wedge 0 < y \leq 1 \\ 0, & x \leq 0 \vee x > 1 \vee y \leq 0 \vee y > 1 \end{cases}$$

1. Determine a constante k
2. Determine as funções de densidade marginal
3. Determine a função de distribuição

Respostas:

1. $1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = \int_0^1 \int_0^1 k dx dy = \int_0^1 k dy = k$
2. $f_X(x) = \int_{-\infty}^{+\infty} k dy = \int_0^1 1 dy = 1$; $f_Y(y) = \int_{-\infty}^{+\infty} k dx = \int_0^1 1 dx = 1$;
3. No 2º, 3º e 4º quadrantes onde $f_{XY}(x, y) = 0$ a função de distribuição é nula;

$$\text{Se } x > 1 \text{ e } 0 < y < 1 \quad F_{XY}(x, y) = \int_0^y \int_0^1 1 dx dy = y$$

$$\text{Se } y > 1 \text{ e } 0 < x < 1 \quad F_{XY}(x, y) = \int_0^x \int_0^1 1 dy dx = x$$

$$\text{Se } 0 < x < 1 \text{ e } 0 < y < 1 \quad F_{XY}(x, y) = \int_0^y \int_0^x 1 dx dy = xy$$

Se $x \geq 1$ e $y \geq 1$ a zona de integração engloba o acontecimento certo e $F_{XY}(x, y) = 1$

4.5 Funções de probabilidade condicional

A probabilidade da variável Y num acontecimento A conhecendo um valor exato x que a variável X assume, pode ser determinada usando a definição de probabilidade condicional 2.10:

$$P[Y \text{ em } A | X = x] = \frac{P[Y \text{ em } A, X = x]}{P[X = x]}$$

Se X for uma variável discreta pode-se determinar a função condicional de distribuição de Y se $X = x_k$:

$$F_Y(y|x_k) = \frac{P[Y \leq y, X = x_k]}{P[X = x_k]}, P[X = x_k] \neq 0$$

Se $F_Y(y|x_k)$ for derivável, a função condicional de densidade de probabilidade será:

$$f(y|x_k) = \frac{d}{dy} F_Y(y|x_k)$$

Se ambas as variáveis são discretas pode-se definir a função condicional de probabilidade de Y se $X = x_k$:

$$p_Y(y_j|x_k) = P[Y = y_j | X = x_k] = \frac{P[X = x_k, Y = y_j]}{P[X = x_k]} = \frac{p_{XY}(x_k, y_j)}{p_X(x_k)}$$

se $P[X = x_k] \neq 0$

Se X for uma variável aleatória contínua surge um problema pois $P[X = x] = 0$. A definição de função condicional de distribuição será:

$$F_Y(y|x) = \lim_{h \rightarrow 0} F_Y(y|h < X \leq x+h)$$

$$\begin{aligned} F_Y(y|h < x \leq x+h) &= \frac{P[Y \leq y, h < X \leq x+h]}{P[h < X \leq x+h]} = \\ &= \frac{\int_x^{x+h} \int_{-\infty}^y f_{XY}(x, y) dy dx}{\int_x^{x+h} f_X(x) dx} \end{aligned}$$

e fazendo $h \rightarrow 0$

$$\lim_{h \rightarrow 0} \frac{\frac{1}{h} \int_x^{x+h} \int_{-\infty}^y f_{XY}(x, y) dy dx}{\frac{1}{h} \int_x^{x+h} f_X(x) dx} = \frac{\int_{-\infty}^y f_{XY}(x, y) dy}{f_X(x)}$$

se $f_X(x) > 0$, e portanto

$$F_Y(y|x) = \frac{\int_{-\infty}^y f_{XY}(x, y) dy}{f_X(x)}$$

Se $F_Y(y|x)$ for derivável, a função condicional de densidade de probabilidade de Y se $X = x$ será

$$f_Y(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

Analogamente se poderiam demonstrar

$$F_X(x|y) = \frac{\int_{-\infty}^y f_{XY}(x, y) dy}{f_Y(y)}$$

e

$$f_X(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

Como $f_{XY}(x, y) = f_Y(y|x)f_X(x) \Rightarrow f_X(x) = \int_{-\infty}^{+\infty} f_Y(y|x)f_X(x)dy$ substituindo na expressão anterior obtém-se

$$f_X(x|y) = \frac{f_Y(y|x)f_X(x)}{\int_{-\infty}^{+\infty} f_Y(y|x)f_X(x)dy}$$

resultado que é uma extensão do teorema de Bayes a variáveis aleatórias contínuas.

Exemplo 4.8 Considere que X e Y são a entrada e a saída, respetivamente, de um canal de comunicação. A entrada pode assumir os valores $+1V$ ou $-1V$ e a saída é igual à entrada somada a um ruído uniformemente distribuído entre $-2V$ e $+2V$.

Determine:

1. $P[X = 1, Y \leq 0]$

Se $X = 1$ a entrada é uma variável $U(-1, 3)$ e

$$F_Y(y|X = 1) = \frac{y+1}{4}, -1 \leq y \leq 3$$

$$\begin{aligned} P[X = 1, Y \leq 0] &= P[Y \leq 0|X = 1] P[X = 1] = \\ &= F_Y(0|X = 1) P[X = 1] = \frac{1}{4} \frac{1}{2} = \frac{1}{8} \end{aligned}$$

2. $P[Y \leq 0|X = -1]$

Se $X = -1$ a entrada é uma variável $U(-3, 1)$ e

$$f_Y(y|X = -1) = \frac{1}{4}, -3 \leq y \leq 1$$

$$P[Y \leq 0|X = -1] = \int_{-3}^0 f_Y(y|X = -1) dy = \frac{3}{4}$$

Exemplo 4.9 Um par de variáveis aleatórias (X, Y) tem uma função de densidade conjunta:

$$f_{XY}(x, y) = \begin{cases} 0, & x \leq 0 \vee y \leq 0 \vee x > y \\ 2e^{-x-y}, & y > x > 0 \end{cases}$$

Determine as funções de densidade marginal e as funções de densidade condicional

$$\begin{aligned} f_X(x) &= \int_x^\infty 2e^{-x-y} dy = 2e^{-2x}, x \geq 0 \\ f_Y(y) &= \int_0^y 2e^{-x-y} dx = 2e^{-y}(1 - e^{-y}), y \geq 0 \\ f_X(x|y) &= \frac{2e^{-x-y}}{2e^{-y}(1 - e^{-y})} = \frac{e^{-x}}{1 - e^{-y}}, 0 < x < y \\ f_Y(y|x) &= \frac{2e^{-x-y}}{2e^{-2x}} = e^{-(x-y)}, y \geq x \end{aligned}$$

Como $f_{XY}(x, y) = f_Y(y|x)f_X(x)$, para se determinar a $P[Y \text{ em } A]$ pode proceder-se do seguinte modo:

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_{y \text{ em } A} f_{XY}(x, y) dx dy &= \int_{-\infty}^{+\infty} \int_{y \text{ em } A} f_Y(y|x) f_X(x) dy dx = \\ &= \int_{-\infty}^{+\infty} \left(\int_{y \text{ em } A} f_Y(y|x) dy \right) f_X(x) dx \end{aligned}$$

que conduz ao resultado

$$P[Y \text{ em } A] = \int_{-\infty}^{+\infty} P[Y \text{ em } A | X = x] f_X(x) dx$$

que pode ser interpretado como uma versão "contínua" do teorema das probabilidades totais 2.12.

Exemplo 4.10 Considere uma variável aleatória $X = U(0, 1)$ e uma variável $Y = U(0, X)$, isto é, a gama de valores da segunda variável aleatória é uniformemente com o valor assumido pela primeira. Determine a função de distribuição e de densidade de probabilidade de Y

Usando o resultado anterior

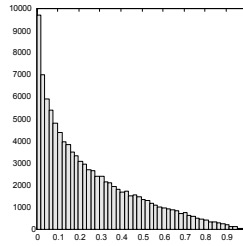
$$F_Y(y) = P[Y \leq y] = \int_0^1 P[Y \leq y | X = x] f_X(x) dx$$

$$P[Y \leq y | X = x] = \begin{cases} \frac{y}{x}, & 0 \leq y \leq x \\ 1, & y > x \end{cases}$$

$$\begin{aligned}
F_Y(y) &= \int_0^y 1 dx + \int_y^1 \frac{y}{x} dx = y - y \ln y \\
f_Y(y) &= -\ln y
\end{aligned}$$

Este resultado pode ser muito simplesmente verificado com a seguinte simulação com o Matlab

```
y=rand(1,100000).*rand(1,100000);
hist(y,50)
```



4.6 Independência

Definição 4.7 Duas variáveis aleatórias $(X; Y)$ são independentes se quaisquer acontecimentos do tipo produto "X em A_1 " e "Y em A_2 " forem independentes:

$$P[X \text{ em } A_1, Y \text{ em } A_2] = P[X \text{ em } A_1] P[Y \text{ em } A_2]$$

Se as duas variáveis aleatórias independentes forem discretas e para acontecimentos do tipo $A_1 = \{X = x_j\}$ e $A_2 = \{Y = y_k\}$ a condição de independência conduz a

$$\begin{aligned}
p_{XY}(x_j, y_k) &= P[X = x_j, Y = y_k] = P[X = x_j] P[Y = y_k] = \\
&= p_X(x_j) p_Y(y_k), j = 1, 2, \dots, k = 1, 2, \dots
\end{aligned}$$

isto é, a função de probabilidade conjunta é igual ao produto das funções de probabilidade.

Se as funções de probabilidade satisfazem a condição anterior, então para dois acontecimentos A_1 e A_2 da forma de produto

$$\begin{aligned}
P[A_1, A_2] &= \sum_{x_j \in A_1} \sum_{y_k \in A_2} p_{XY}(x_j, y_k) = \sum_{x_j \in A_1} p_X(x_j) \sum_{y_k \in A_2} p_Y(y_k) = \\
&= P[A_1] P[A_2]
\end{aligned}$$

o que significa que as duas variáveis são independentes

Em geral pode-se afirmar que duas variáveis são independentes se e só a sua função de distribuição conjunta iguala o produto das funções de distribuição marginal

$$F_{XY}(x, y) = F_X(x) F_Y(y)$$

e se as duas variáveis forem conjuntamente contínuas, se e só a sua função de densidade conjunta iguala o produto das funções de densidade marginal

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

4.7 Esperança matemática

Considere-se o par de variáveis aleatórias (X, Y) e uma função $Z = g(X, Y)$

Definição 4.8 *Esperança matemática, valor médio esperado ou média de Z é*

$$E[Z] = \begin{cases} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{XY}(x, y) dx dy & \text{se } X \text{ e } Y \text{ são} \\ & \text{conjuntamente contínuas} \\ \sum_m \sum_n g(x_m, y_n) p_{XY}(x_m, y_n) & \text{se } X \text{ e } Y \text{ são discretas} \end{cases}$$

Definição 4.9 *Os momentos conjuntos de ordem jk das variáveis X, Y definem-se como*

$$E[X^j Y^k] = \begin{cases} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^j y^k f_{XY}(x, y) dx dy & \text{se } X \text{ e } Y \text{ são} \\ & \text{conjuntamente contínuas} \\ \sum_m \sum_n x_m^j y_n^k f_{XY}(x_m, y_n) & \text{se } X \text{ e } Y \text{ são discretas} \end{cases}$$

Se $j = 1$ e $k = 0$ ou $j = 0$ e $k = 1$, obtem-se os valores médios de X e Y , e se $j = 2$ e $k = 0$ ou $j = 0$ e $k = 2$, obtem-se os valores quadráticos médios respetivos.

Definição 4.10 *Os momentos centrais conjuntos de ordem jk das variáveis X, Y definem-se como $E[(X - E[X])^j (Y - E[Y])^k]$*

Se $j = 2$ e $k = 0$ ou $j = 0$ e $k = 2$, obtem-se as variâncias de X e Y

4.7.1 Correlação

Definição 4.11 *A correlação das variáveis X e Y é o seu momento de ordem $j = k = 1$, $E[XY]$*

Quando $E[XY] = 0$ as variáveis X e Y são ortogonais

4.7.2 Covariância

Definição 4.12 *A covariância das variáveis X e Y é o seu momento central de ordem $j = k = 1$: $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$*

$$\begin{aligned} Cov(X, Y) &= E[(X - E[X])(Y - E[Y])] = \\ &= E[XY - XE[Y] - YE[X] - E[X]E[Y]] = \\ &= E[XY] - 2E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y] \end{aligned}$$

Se $E[X] = 0$ ou $E[Y] = 0 \Rightarrow Cov(X, Y) = E[XY]$

Teorema 4.1 Se X e Y são variáveis independentes $Cov(X, Y) = E[XY]$

Demonstração.

$$\begin{aligned} Cov(X, Y) &= E[(X - E[X])(Y - E[Y])] = \\ &= E[(X - E[X])E[(Y - E[Y])] = 0 \end{aligned}$$

■

Definição 4.13 O coeficiente de correlação de X e Y é $\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$

Como o valor médio quadrático de qualquer variável ou função de variável aleatória é não negativo

$$E \left[\left(\left(\frac{X - E[X]}{\sqrt{\sigma_X}} \right) \pm \left(\frac{Y - E[Y]}{\sqrt{\sigma_Y}} \right) \right)^2 \right] \geq 0$$

e portanto

$$\begin{aligned} 1 \pm 2\rho_{XY} + 1 &\geq 0 \\ -1 &\leq \rho_{XY} \leq 1 \end{aligned}$$

Demonstra-se que os valores extremos de $\rho_{XY} = \pm 1$ se obtêm para uma relação linear $Y = aX + b$ com $a > 0$ ou $a < 0$, respetivamente.

Se $\rho_{XY} = 0$ as variáveis dizem-se descorrelacionadas. Como se viu, se X e Y são independentes, a sua covariância é nula e, portanto são descorrelacionadas. Mas o contrário não é verdadeiro como se pode ver no seguinte exemplo

Exemplo 4.11 Seja $\Phi = U(0, 2\pi)$ e considerem-se as funções de variável aleatória $X = \cos \Phi$ e $Y = \sin \Phi$.

A função de densidade conjunta só é não nula sobre o círculo unitário, pois para um dado ϕ , o par $(X, Y) \in \{(x, y) : x^2 + y^2 = 1\}$ e como $f_X(x)$ e $f_Y(y)$ são não nulas no intervalo $[-1, 1]$, o seu produto é não nulo no interior do quadrado $[-1, 1] \times [-1, 1]$ e portanto $f_{XY}(x, y) \neq f_X(x)f_Y(y)$

$$\text{Como } f_\Phi(\phi) = \frac{1}{2\pi}, \phi \in [0, 2\pi]$$

$$E[X] = E[Y] = 0$$

$$E[XY] = E[\cos \Phi \sin \Phi] = \frac{1}{2\pi} \int_0^{2\pi} \sin \phi \cos \phi d\phi = 0$$

Então $Cov(X, Y) = 0$ e as variáveis são descorrelacionadas mas não independentes

Exemplo 4.12 Um par de variáveis aleatórias (X, Y) tem uma função de densidade conjunta:

$$f_{XY}(x, y) = \begin{cases} 0, & x \leq 0 \vee y \leq 0 \vee x \leq y \\ e^{-x-y}, & x > 0 \wedge y > 0 \end{cases}$$

Determine as densidades marginais, médias, variâncias, correlação, covariância e coeficiente de correlação

1. (a) $f_X(x) = \int_0^\infty e^{-x-y} dy = e^{-x}, x \geq 0; f_Y(y) = \int_0^\infty e^{-x-y} dx = e^{-y}, y \geq 0;$
- (b) $E[X] = E[Y] = \int_0^\infty xe^{-x} dx = 1;$
- (c) $E[X^2] = E[Y^2] = \int_0^\infty x^2 e^{-x} dx = 2;$
- (d) $Var(X) = Var(Y) = 2 - 1^2 = 1;$
- (e) $E[XY] = \int_0^\infty \int_0^\infty xye^{-x-y} dx dy = 1$
- (f) $Cov(X, Y) = E[XY] - E[Y]E[X] = 0;$
- (g) $\rho_{XY} = 0$

Teorema 4.2 Desigualdade de Schwarz

Se as variáveis aleatórias X e Y têm momentos de 2^o ordem finitos, verifica-se

$$E[XY]^2 \leq E[X^2] E[Y^2]$$

e a condição de igualdade verifica-se se existir uma constante a_0 tal que $P[a_0X = Y] = 1$

Demonstração. Definina-se a função

$$g(x) = E[(aX - Y)^2] = a^2 E[X^2] - 2a E[XY] + E[Y^2] \geq 0.$$

Para que esta condição se verifique é necessário que a equação de 2^o grau em a tenha raízes nulas ou complexas pelo que $E[XY]^2 - E[X^2] E[Y^2] \leq 0$.

No caso de igualdade existe um a_0 tal que $E[(a_0X - Y)^2] = 0$ o que implica que a variável aleatória $a_0X - Y$ seja nula com probabilidade unitária e portanto $P[a_0X - Y = 0] = 1$ ■

4.7.3 Média condicional

Definição 4.14 A média condicional de Y se $X = x$ é

$$E[Y|x] = \int_{-\infty}^{+\infty} f_Y(Y|x) dy$$

A média condicional $E[Y|x]$ é uma função real da variável real x , fazendo portanto sentido considerar a função de variável aleatória $E[Y|X]$, função de X . O seu valor médio será:

$$\begin{aligned} E[E[Y|X]] &= \int_{-\infty}^{+\infty} E[Y|x] f_X(x) dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_Y(Y|x) y f_X(x) dx dy = \\ &= \int_{-\infty}^{+\infty} y \int_{-\infty}^{+\infty} f_Y(Y|x) f_X(x) dx dy = \\ &= \int_{-\infty}^{+\infty} y \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = \int_{-\infty}^{+\infty} y f_Y(y) dy = E[Y] \end{aligned}$$

4.8 Extensão ao caso multivariável

Sejam X_1, \dots, X_n variáveis aleatórias componentes de um vetor aleatório \mathbf{X} de dimensão n . As seguintes definições são extensões do que se viu para duas variáveis.

Definição 4.15 A função de distribuição conjunta de X_1, \dots, X_n é

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = P[X_1 \leq x_1, \dots, X_n \leq x_n]$$

Definição 4.16 A função de probabilidade conjunta das variáveis X_1, \dots, X_n discretas é

$$p_{X_1 \dots X_n}(x_1, \dots, x_n) = P[X_1 = x_1, \dots, X_n = x_n]$$

A probabilidade de um acontecimento A n -dimensional obtém-se somando a função de probabilidade para todos os pontos $\mathbf{x} = (x_1, \dots, x_n)$ pertencentes a A :

$$P[A] = P[\mathbf{x} \in A] = \sum_{\mathbf{x} \in A} \dots \sum_{x_n} p_{X_1 \dots X_n}(x_1, \dots, x_n)$$

As funções marginais podem ser determinadas somando a função conjunta em ordem a uma ou mais variáveis. Desta forma, a função de probabilidade marginal da variável X_j será:

$$p_{X_j}(x_j) = \sum_{x_1} \dots \sum_{x_{j-1}} \sum_{x_{j+1}} \dots \sum_{x_n} p_{X_1 \dots X_n}(x_1, \dots, x_n)$$

e a função de probabilidade marginal das variáveis X_1, \dots, X_{n-1} será:

$$p_{X_1 \dots X_{n-1}}(x_1, \dots, x_{n-1}) = \sum_{x_n} p_{X_1 \dots X_n}(x_1, \dots, x_n)$$

Definição 4.17 Se as variáveis aleatórias $X_1 \dots X_{n-1}$ forem conjuntamente contínuas, a função de densidade probabilidade conjunta é a função $f_{X_1 \dots X_n}(x_1, \dots, x_n)$ que para todo o acontecimento A :

$$P[A] = P[\mathbf{x} \in A] = \int_{\mathbf{x} \in A} \dots \int f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n$$

Se as funções forem contínuas e deriváveis verificar-se-ão as igualdades

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n$$

$$f_{X_1 \dots X_n}(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_{X_1 \dots X_n}(x_1, \dots, x_n)$$

As funções marginais podem ser determinadas integrando a função de densidade conjunta em ordem a uma ou mais variáveis. Desta forma, a função de probabilidade marginal da variável X_j será:

$$f_{X_j}(x_j) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_{j-1} dx_{j+1} \dots dx_n$$

e a função de probabilidade marginal das variáveis X_1, \dots, X_{n-1} será:

$$f_{X_1 \dots X_{n-1}}(x_1, \dots, x_{n-1}) = \int_{-\infty}^{+\infty} f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_n$$

Podem ainda definir-se funções de densidade condicional a partir da densidade conjunta. A função de densidade condicional de X_j dadas as variáveis $X_1 \dots X_{j-1}, X_{j+1} \dots X_n$, será:

$$f_{X_j}(x_j | x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) = \frac{f_{X_1 \dots X_n}(x_1, \dots, x_n)}{f_{X_1 \dots X_{j-1}, X_{j+1} \dots X_n}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n)}$$

Definição 4.18 As variáveis aleatórias X_1, \dots, X_n conjuntamente contínuas são mutuamente independentes se e só se

$$f_{X_1 \dots X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$

4.8.1 Variáveis conjuntamente gausseanas

Definição 4.19 As variáveis X_1, \dots, X_n são conjuntamente gausseanas se a sua densidade conjunta for

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1 \dots X_n}(x_1, \dots, x_n) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \Sigma^{-1}(\mathbf{x} - \mathbf{m})\right)}{(2\pi)^{n/2} |\Sigma|^{1/2}}$$

em que

$$\mathbf{x} = (x_1, \dots, x_n)^T, \mathbf{m} = (m_1, \dots, m_n)^T$$

e

$$\Sigma = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_n) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(X_n, X_1) & \dots & \dots & \text{Var}(X_n) \end{bmatrix}$$

A matriz Σ é designada por matriz de correlação e é real e simétrica pois $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$ para $i \neq j$ e é definida não negativa, isto é, os seus valores próprios são não negativos.

Teorema 4.3 Se $\forall i \neq j \text{Cov}(X_i, X_j) = 0$ as variáveis X_1, \dots, X_n são independentes

Demonstração. $\Sigma = \text{diag}(\text{Var}(X_i)) = \text{diag}(\sigma_i^2) \Rightarrow \Sigma^{-1} = \text{diag}\left(\frac{1}{\sigma_i^2}\right) \Rightarrow$
 $|\Sigma| = \prod_{i=1}^n \sigma_i^2$
 $(\mathbf{x} - \mathbf{m})^T \Sigma^{-1} (\mathbf{x} - \mathbf{m}) = \sum_{i=1}^n \left(\frac{x_i - m_i}{\sigma_i} \right)^2$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - m_i}{\sigma_i} \right)^2\right)}{(2\pi)^{n/2} \prod_{i=1}^n \sigma_i} = \prod_{i=1}^n \frac{\exp\left(-\frac{1}{2} \left(\frac{x_i - m_i}{\sigma_i} \right)^2\right)}{\sqrt{2\pi} \sigma_i} = \prod_{i=1}^n f_{X_i}(x_i) \blacksquare$$

Teorema 4.4 Se X_1, \dots, X_n forem variáveis conjuntamente gausseanas com valor médio \mathbf{m}_X e matriz de correlação Σ_X , as variáveis Y_1, \dots, Y_n obtidas pela transformação linear $Y = AX$ com A não singular, são também variáveis conjuntamente gausseanas com valor médio $\mathbf{m}_Y = A\mathbf{m}_X$ e matriz de correlação $\Sigma_Y = A\Sigma_X A^T$

Demonstração. Como¹

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{f_{\mathbf{X}}(A^{-1}\mathbf{y})}{|A|},$$

$$(A^{-1}\mathbf{y} - \mathbf{m}_X) = A^{-1}(\mathbf{y} - A\mathbf{m}_X)$$

e

$$(A^{-1}\mathbf{y} - \mathbf{m}_X)^T = (\mathbf{y} - A\mathbf{m}_X)^T A^{-1T} :$$

será

$$\begin{aligned} (\mathbf{y} - \mathbf{m}_X)^T \Sigma_X^{-1} (\mathbf{y} - A\mathbf{m}_X) &= (\mathbf{y} - A\mathbf{m}_X)^T A^{-1T} \Sigma_X^{-1} A^{-1} (\mathbf{y} - A\mathbf{m}_X) = \\ &= (\mathbf{y} - A\mathbf{m}_X)^T (A\Sigma_X A^T)^{-1} (\mathbf{y} - A\mathbf{m}_X) \end{aligned}$$

e ainda como $|A\Sigma_X A^T| = |A|^2 |\Sigma_X|$ vem

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \frac{\exp\left(-\frac{1}{2}(\mathbf{y} - A\mathbf{m}_X)^T (A\Sigma_X A^T)^{-1} (\mathbf{y} - A\mathbf{m}_X)\right)}{(2\pi)^{n/2} |A| |\Sigma_X|^{1/2}} = \\ &= \frac{\exp\left(-\frac{1}{2}(\mathbf{y} - A\mathbf{m}_X)^T (A\Sigma_X A^T)^{-1} (\mathbf{y} - A\mathbf{m}_X)\right)}{(2\pi)^{n/2} |A\Sigma_X A^T|^{1/2}} \end{aligned}$$

■

Este teorema vai permitir gerar amostras de variáveis aleatórias conjuntamente gausseanas com uma matriz de correlação arbitrária como se mostra no anexo B.

¹Para demonstração deste resultado consultar A.Leon-Garcia: *Probability and Random Processes for Electrical Engineering*, p225-226, Addison-Wesley, 1994

Uma vez que Σ é uma matriz simétrica, é sempre possível encontrar uma matriz A tal que $A\Sigma A^T = \Lambda$ em que Λ é uma matriz diagonal. Portanto será sempre possível transformar as variáveis conjuntamente gausseanas e dependentes X_1, \dots, X_n nas variáveis Y_1, \dots, Y_n também conjuntamente gausseanas mas independentes.

4.8.2 O caso bidimensional

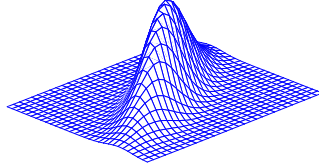
Para o par de variáveis X, Y conjuntamente gausseanas com uma matriz de correlação

$$\begin{bmatrix} \sigma_X^2 & \frac{\sigma_X \sigma_Y}{\rho_{XY}} \\ \frac{\sigma_X \sigma_Y}{\rho_{XY}} & \sigma_Y^2 \end{bmatrix}$$

a densidade de probabilidade será:

$$f_{XY}(x, y) = \frac{\exp\left(\frac{-1}{2(1-\rho_{XY}^2)} \left[\left(\frac{x-m_X}{\sigma_X}\right)^2 - 2\rho_{XY}^2 \left(\frac{x-m_X}{\sigma_X}\right) \left(\frac{y-m_Y}{\sigma_Y}\right) + \left(\frac{y-m_Y}{\sigma_Y}\right)^2 \right]\right)}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}$$

A representação gráfica 3-D desta superfície para $m_X = m_Y = 0, \sigma_X = 1, \sigma_Y = 0.6$ e $\rho_{XY} = 0.8$ é a seguinte.

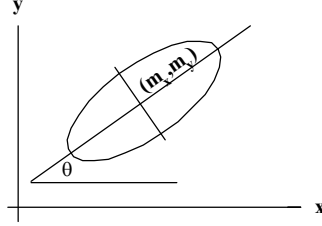


A função de densidade de probabilidade conjunta está centrada em (m_X, m_Y) . Se se pretender determinar as curvas de igual densidade basta determinar o lugar geométrico dos pontos (x, y) que tornam constante o argumento da exponencial

$$\left[\left(\frac{x-m_X}{\sigma_X}\right)^2 - 2\rho_{XY}^2 \left(\frac{x-m_X}{\sigma_X}\right) \left(\frac{y-m_Y}{\sigma_Y}\right) + \left(\frac{y-m_Y}{\sigma_Y}\right)^2 \right] = \text{const.}$$

A equação anterior representa uma família de elipses centradas em (m_X, m_Y) , com semieixos proporcionais a σ_X e a σ_Y e com uma orientação de eixos como

se indica na figura seguinte:



com $\theta = \frac{1}{2} \arctan \left(\frac{2\rho_{XY}\sigma_X\sigma_Y}{\sigma_X^2 - \sigma_Y^2} \right)$ que mostra que para $\rho_{XY} = 0$ (variáveis independentes ou descorrelacionadas) a elipse está alinhada com os eixos coordenados, pois $\theta = 0$ e para $\sigma_X = \sigma_Y$, $\theta = 45^\circ$.

As densidades marginais são facilmente determinadas como sendo as das variáveis $X = N(m_X, \sigma_X)$ e $Y = N(m_Y, \sigma_Y)$.

As probabilidades condicionais podem ser obtidas usando as definições gerais já referidas:

$$f_X(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{\exp \left(\frac{-1}{2\sigma_X^2(1 - \rho_{XY}^2)} \left(x - \rho_{XY} \frac{\sigma_X}{\sigma_Y} (y - m_Y) - m_X \right)^2 \right)}{\sqrt{2\pi\sigma_X^2(1 - \rho_{XY}^2)}}$$

Como se vê, a probabilidade condicional ainda é gausseana com um valor médio $E[X|y] = m_X + \rho_{XY} \frac{\sigma_X}{\sigma_Y} (y - m_Y)$ e uma variância $\sigma_X^2(1 - \rho_{XY}^2)$. Note que se as variáveis são descorrelacionadas ou independentes, como $\rho_{XY} = 0$, a média e variância condicionais são iguais à média e variância marginais.

O valor mais provável de X dado $Y = y$, isto é, conhecido um valor de Y é o valor que maximiza a função de densidade condicional $f_X(x|y)$ e, neste caso, sendo a função gausseana o seu valor máximo coincide com o seu valor médio $E[X|y]$.

Exemplo 4.13 *Suponha que X e Y são variáveis conjuntamente gausseanas que representam as temperaturas medidas em duas cidades diferentes e que se conhece $m_X = 15^\circ\text{C}$, $m_Y = 14^\circ\text{C}$, $\sigma_X = \sigma_Y = 4^\circ\text{C}$ e $\rho_{XY} = 0.85$. Se num dado dia se observar a temperatura de 20 na cidade Y , qual o valor mais provável para a temperatura na cidade X .*

$$\text{Valor mais provável de } X = E[X|y = 20] = 15 + 0.85 \frac{4}{4} (20 - 14) = 20.1^\circ\text{C}$$

A esta estimativa do valor de uma variável aleatória e que corresponde à maior probabilidade, chama-se estimativa de máxima verosimilhança.

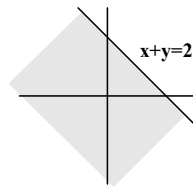
4.9 Exercícios

Exercício 4.1 Para um par de variáveis aleatórias (X, Y) indique a região do plano equivalente aos seguintes acontecimentos:

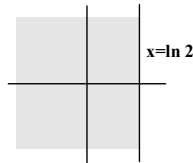
$$\begin{aligned} A_1 &= \{X + Y \leq 2\}, A_2 = \{e^X > 2\}, A_3 = \{\max(X, Y) < 6\} \\ A_4 &= \{|X + Y| \leq 2\}, A_5 = \{|X| > |Y|\}, A_6 = \{Y/X < 1\} \\ A_7 &= \{X^2 < Y\}, A_8 = \{XY < 2\}, A_9 = \{\max(|X|, |Y|) < 3\} \end{aligned}$$

Respostas:

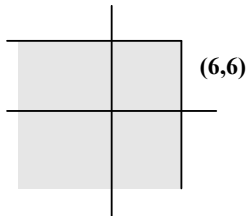
Exercício 4.2 $A_1 = \{(x, y) \in \mathbb{R}^2 : x + y < 2\}$



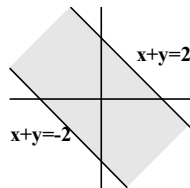
$$A_2 = \{(x, y) \in \mathbb{R}^2 : x < \ln 2\}$$



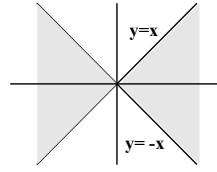
$$A_3 = \{(x, y) \in \mathbb{R}^2 : x < 6 \vee y < 6\}$$



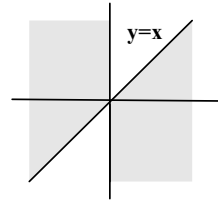
$$A_4 = \{(x, y) \in \mathbb{R}^2 : x + y < 2 \wedge x + y > -2\}$$



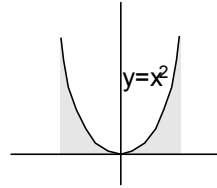
$$A_5 = \{(x, y) \in 1^{\circ}quad : x > y\} \cup \{(x, y) \in 2^{\circ}quad : -x > y\} \cup \\ \{(x, y) \in 3^{\circ}quad : -x > -y\} \cup \{(x, y) \in 4^{\circ}quad : x > -y\}$$



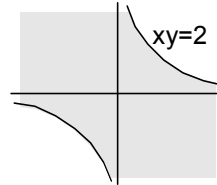
$$A_6 = \{(x, y) \in \mathbb{R}^2 : y < x \text{ se } x > 0\} \cup \{(x, y) \in \mathbb{R}^2 : y > x \text{ se } x < 0\}$$



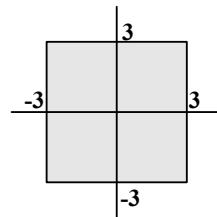
$$A_7 = \{(x, y) \in \mathbb{R}^2 : x^2 < y\}$$



$$A_8 = \{(x, y) \in \mathbb{R}^2 : xy < 2\}$$



$$A_9 = \{(x, y) \in 1^{\circ}quad : x < 3 \vee y < 3\} \cup \{(x, y) \in 2^{\circ}quad : -x < 3 \vee y < 3\} \cup \\ \{(x, y) \in 3^{\circ}quad : -x < 3 \vee -y < 3\} \cup \{(x, y) \in 4^{\circ}quad : x < 3 \vee -y < 3\}$$



Exercício 4.3 Dadas duas variáveis aleatórias X e Y independentes e exponenciais com $\lambda = 1$, determine $E[|X + Y|]$

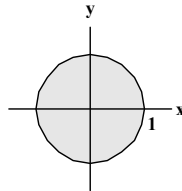
$$f_{XY}(x, y) = e^{-(x+y)}, x, y \geq 0$$

$$\begin{aligned} E[|X - Y|] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = \int_0^{+\infty} \int_0^{+\infty} |x - y| e^{-(x+y)} dx dy = \\ &= \int_0^{+\infty} \int_0^x (x - y) e^{-(x+y)} dy dx + \int_0^{+\infty} \int_x^{+\infty} (-x + y) e^{-(x+y)} dy dx = \\ &= \frac{1}{2} + \frac{1}{2} = 1 \end{aligned}$$

Exercício 4.4 Considere um par de variáveis aleatórias (X, Y) que é uniformemente distribuída no círculo unitário, tendo portanto uma densidade conjunta definida por

$$f_{XY}(x, y) = \begin{cases} k, & x^2 + y^2 \leq 1 \\ 0, & x^2 + y^2 > 1 \end{cases}$$

1. Determine a constante k
2. Determine as funções de densidade marginal
3. Determine as médias e variância de X e Y
4. Determine a correlação, covariância e coeficiente de correlação
5. Indique face aos resultados se as variáveis são independentes, ortogonais e correlacionadas



Exercício 4.5 1.

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} k dx dy = \int_{-1}^{+1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} k dy dx = \pi k \Rightarrow k = \frac{1}{\pi}$$

2.

$$f_X(x) = \int_{-\infty}^{+\infty} k dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = 2 \frac{\sqrt{1-x^2}}{\pi} \text{ se } |x| \leq 1;$$

$$f_X(x) = 0 \text{ se } |x| > 1$$

$$f_Y(y) = \int_{-\infty}^{+\infty} k dx = \frac{1}{\pi} dx = 2 \frac{\sqrt{1-y^2}}{\pi} \text{ se } |y| \leq 1;$$

$$f_Y(y) = 0 \text{ se } |y| > 1$$

3.

$$\begin{aligned}
E[X] &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_{-1}^{+1} 2x \frac{\sqrt{1-x^2}}{\pi} dx \\
&= \left[-\frac{2}{3} \frac{(\sqrt{1-x^2})^3}{\pi} \right]_{-1}^{+1} = 0 \\
E[X^2] &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_{-1}^{+1} 2x^2 \frac{\sqrt{1-x^2}}{\pi} dx = \\
&= \frac{2}{\pi} \left[-\frac{1}{4} x (\sqrt{1-x^2})^3 + \frac{1}{8} x \sqrt{1-x^2} + \frac{1}{8} \arcsin x \right]_{-1}^{+1} = \frac{1}{4} \\
\text{Var}(X) &= E[X^2] = \frac{1}{4}
\end{aligned}$$

Valores iguais para Y .

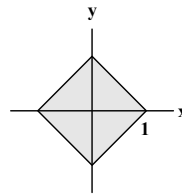
4.

$$\begin{aligned}
E[XY] &= \int_{-1}^{+1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} xy \frac{1}{\pi} dy dx = 0 \\
\text{Cov}(X, Y) &= E[XY] - E[X] E[Y] = 0 \\
\rho_{XY} &= \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = 0
\end{aligned}$$

5. Como $f_{XY}(x, y) \neq f_X(x)f_Y(y)$ as variáveis não são independentesComo $E[XY] = 0$ as variáveis são ortogonaisComo $\text{Cov}(X, Y) = 0$ as variáveis não são correlacionadas

Exercício 4.6 Repita as questões anteriores para um par de variáveis aleatórias (X, Y) que é uniformemente distribuída num quadrado com vértices em $(0, 1)$, $(0, 1)$, $(-1, 0)$ e $(0, -1)$, tendo portanto uma densidade conjunta definida por

$$f_{XY}(x, y) = \begin{cases} k, & |x+y| \leq 1 \\ 0, & |x+y| > 1 \end{cases}$$



1.

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} k dx dy = \int_{-1}^0 \int_{-1-x}^{1+x} k dy dx + \int_0^1 \int_{-1+x}^{1-x} k dy dx = 2k$$

$$k = \frac{1}{2}$$

2.

$$f_X(x) = \int_{-\infty}^{+\infty} k dy = \begin{cases} \int_{-1-x}^{1+x} \frac{1}{2} dy = x+1 & -1 < x \leq 0 \\ \int_{-1+x}^{1-x} \frac{1}{2} dy = 1-x & 0 < x \leq 1 \\ 0 & |x| > 1 \end{cases}$$

Definição análoga para $f_Y(y)$

3.

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_{-1}^0 x(x+1) dx + \int_0^1 x(1-x) dx = 0$$

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_{-1}^0 x^2(x+1) dx + \int_0^1 x^2(1-x) dx = \frac{1}{6}$$

$$Var(X) = E[X^2] = \frac{1}{6}$$

Valores iguais para Y .

4.

$$E[XY] = \int_{-1}^0 \int_{-1-x}^{1+x} xy \frac{1}{2} dy dx + \int_0^1 \int_{-1+x}^{1-x} xy \frac{1}{2} dy dx = 0$$

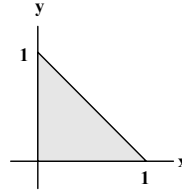
$$Cov(X, Y) = E[XY] - E[X]E[Y] = 0$$

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = 0$$

5. Como $f_{XY}(x, y) \neq f_X(x)f_Y(y)$ as variáveis não são independentesComo $E[XY] = 0$ as variáveis são ortogonaisComo $Cov(X, Y) = 0$ as variáveis não são correlacionadas

Exercício 4.7 Repita as questões anteriores para um par de variáveis aleatórias (X, Y) que é uniformemente distribuída num triângulo com vértices em $(0, 0)$, $(0, 1)$ e $(1, 0)$, tendo portanto uma densidade conjunta definida por

$$f_{XY}(x, y) = \begin{cases} k, & x \geq 0 \wedge y \geq 0 \wedge (x+y) \leq 1 \\ 0, & x < 0 \wedge y < 0 \wedge (x+y) > 1 \end{cases}$$



1.

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} k dx dy = \int_0^1 \int_0^{1-x} k dy dx = \frac{1}{2} k \Rightarrow k = 2$$

2.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} k dy = \begin{cases} \int_0^{1+x} 2 dy = 2(1+x) & 0 < x \leq 1 \\ 0 & x \leq 0 \vee x > 1 \end{cases} \\ f_Y(y) &= \int_{-\infty}^{+\infty} k dx = \begin{cases} \int_0^{1+y} 2 dx = 2(1+y) & 0 < y \leq 1 \\ 0 & y \leq 0 \vee y > 1 \end{cases} \end{aligned}$$

3.

$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^1 x 2(x+1) dx = \left[\frac{2}{3} x^3 + x^2 \right]_0^1 = \frac{5}{3} \\ E[X^2] &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_0^1 x^2 2(x+1) dx = \left[\frac{7}{12} x^4 + \frac{7}{9} x^3 \right]_0^1 = \frac{7}{6} \\ Var(X) &= E[X^2] - E[X]^2 = \frac{5}{3} - \left(\frac{5}{3}\right)^2 = \frac{11}{36} \end{aligned}$$

Valores iguais para Y .

4.

$$\begin{aligned} E[XY] &= \int_0^1 \int_0^{1+x} xy 2 dy dx = \frac{17}{12} \\ Cov(X, Y) &= E[XY] - E[X] E[Y] = \frac{17}{12} - \left(\frac{5}{3}\right)^2 = \frac{1}{18} \\ \rho_{XY} &= \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{1/18}{11/36} = \frac{2}{11} \end{aligned}$$

5. Como $f_{XY}(x, y) \neq f_X(x)f_Y(y)$ as variáveis não são independentesComo $E[XY] \neq 0$ as variáveis não são ortogonaisComo $Cov(X, Y) \neq 0$ as variáveis são correlacionadas

Exercício 4.8 Um par de variáveis aleatórias (X, Y) tem uma densidade de probabilidade definida por

$$f_{XY}(x, y) = \begin{cases} k e^{-2x} e^{-2y}, & x \geq 0 \wedge y \geq 0 \\ 0, & x < 0 \vee y < 0 \end{cases}$$

1. Determine o valor de k
2. Calcule a probabilidade de $\{X + Y \leq 4\}$
3. Calcule a probabilidade de $\{X + Y > 2\}$
4. Calcule a probabilidade de $\{X \leq 2\}$
5. Calcule a probabilidade de $\{X < Y\}$

Exercício 4.9 *Respostas:*

1.

$$\begin{aligned} 1 &= \int_0^{+\infty} \int_0^{+\infty} k e^{-2x} e^{-2y} dx dy = \int_0^{+\infty} \left[-\frac{1}{2} k e^{-2x} e^{-2y} \right]_0^{+\infty} dy = \\ &= \int_0^{+\infty} \frac{1}{2} k e^{-2y} dy = \frac{1}{4} k \Rightarrow k = 4 \end{aligned}$$

2.

$$P[X + Y \leq 4] = \int_0^4 \int_0^{4-y} 4e^{-2x} e^{-2y} dx dy = -9e^{-8} + 1 = .99698$$

3.

$$\begin{aligned} P[X + Y > 2] &= 1 - P[X + Y \leq 2] = 1 - \int_0^2 \int_0^{2-y} 4e^{-2x} e^{-2y} dx dy \\ &= 5e^{-4} = 9.1578 \times 10^{-2} \end{aligned}$$

4.

$$P[X \leq 2] = \int_0^\infty \int_0^2 4e^{-2x} e^{-2y} dx dy = 1 - e^{-4} = .98168$$

5.

$$P[X < Y] = \int_0^\infty \int_y^\infty 4e^{-2x} e^{-2y} dx dy = \frac{1}{2}$$

Exercício 4.10 *O par de variáveis aleatórias (X, Y) tem uma densidade de probabilidade definida por*

$$f_{XY}(x, y) = \begin{cases} k e^{-x(1+y)}, & x \geq 0 \wedge y \geq 0 \\ 0, & x < 0 \vee y < 0 \end{cases}$$

Determine as densidades marginais

$$f_Y(y) = \int_0^{+\infty} e^{-x(1+y)} dx = \left[-\frac{1}{1+y} e^{-x(1+y)} \right]_0^{+\infty} = \frac{1}{1+y}$$

$$f_X(x) = \int_0^{+\infty} e^{-x(1+y)} dy = \left[-\frac{1}{x} e^{-x(1+y)} \right]_0^{+\infty} = \frac{1}{x} e^{-x}$$

Exercício 4.11 Se o par de variáveis aleatórias (X, Y) forem independentes e sabendo que X é uma variável gaussiana de média e variância unitárias, e Y é uniformemente distribuída em $[0, 1]$, determine o valor de $E[X^2Y]$

Se (X, Y) são independentes $f_{XY}(x, y) = f_X(x)f_Y(y)$

$$\begin{aligned}
 E[Y] &= 0.5 \\
 E[X^2] &= \text{Var}(X) + E[X]^2 = 1 + 1 \\
 E[X^2Y] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 y f_{XY}(x, y) dx dy = \\
 &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx \int_{-\infty}^{+\infty} y f_Y(y) dy = \\
 &= E[X^2] E[Y] = 2 \cdot 0.5 = 1
 \end{aligned}$$

Chapter 5

Somas de variáveis aleatórias

5.1 Média e variância da soma

Sejam X_1, X_2, \dots, X_n , n variáveis aleatórias e seja $S_n = X_1 + X_2 + \dots + X_n$, a sua soma.

Teorema 5.1 *A média da soma de n variáveis aleatórias é igual à soma das médias*

Demonstração.

$$\begin{aligned} E[S_n] &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \left(\sum_{j=1}^n x_j \right) f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n = \\ &= \sum_{j=1}^n \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} x_j f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n = \\ &= \sum_{j=1}^n \int_{-\infty}^{+\infty} x_j f_{X_j}(x_j) dx_j = \sum_{j=1}^n E[X_j] \end{aligned}$$

■

Teorema 5.2 *A variância da soma de n variáveis aleatórias é dada por:*

$$Var(S_n) = \sum_{i=1}^n Var(X_i) + \sum_{j=1}^n \sum_{\substack{k=1 \\ j \neq k}}^n Cov(X_j, X_k)$$

Demonstração.

$$\begin{aligned} \text{Var}(S_n) &= E \left[\sum_{j=1}^n (X_j - E[X_j]) \sum_{k=1}^n (X_k - E[X_k]) \right] = \\ &= \sum_{j=1}^n \sum_{k=1}^n E[(X_j - E[X_j]) E[(X_k - E[X_k])]] \end{aligned}$$

■

Se as variáveis são independentes, para todo o $j \neq k$, $\text{Cov}(X_j, X_k) = 0$ e

$$\text{Var}(S_n) = \sum_{i=1}^n \text{Var}(X_i)$$

Quando as variáveis são independentes e identicamente distribuídas (**IID**), se $E[X_i] = \mu$ e $\text{Var}(X_i) = \sigma^2, i = 1, 2, \dots, n$, a média e a variância da soma valem

$$E[S_n] = n\mu \text{ e } \text{Var}(S_n) = n\sigma^2$$

5.2 Densidade de probabilidade da soma

Se X_1, X_2, \dots, X_n são variáveis independentes com funções de densidade de probabilidade $f_{X_1}(x_1), \dots, f_{X_n}(x_n)$, para se determinar a função de densidade de probabilidade da soma pode-se recorrer à sua de função característica:

$$\begin{aligned} \Phi_{S_n} &= E[\exp(jwS_n)] = E \left[\exp(jw(\sum_{i=1}^n X_i)) \right] = \\ &= E \left[\prod_{i=1}^n \exp(jwX_i) \right] = \prod_{i=1}^n E[\exp(jwX_i)] \end{aligned}$$

e portanto

$$\Phi_{S_n}(w) = \prod_{i=1}^n \Phi_{X_i}(w) \Leftrightarrow f_{S_n}(x_1, \dots, x_n) = f_{X_1}(x_1) * \dots * f_{X_n}(x_n) \quad (5.1)$$

em que $*$ simboliza a operação de convolução e, portanto, a densidade de probabilidade da soma de variáveis aleatórias independentes é a convolução das densidades das parcelas.

A um resultado semelhante se pode chegar usando a função geradora de probabilidade. É trivial verificar que:

$$G_{S_n} = \prod_{i=1}^n G_{X_i} \quad (5.2)$$

Pode-se agora analisar alguns exemplos de somas de variáveis aleatórias:

Exemplo 5.1 *Sejam $X_i, i = 1, \dots, n$ variáveis binomiais independentes com parâmetros $(p, n_i), i = 1, \dots, n$, respectivamente.*

Como para uma variável binomial X_i a função geradora vale

$$G_{X_i}(z) = \sum_{k=0}^{n_i} C_k^{n_i} p^k (1-p)^{n_i-k} z^k = (1-p+pz)^{n_i}$$

atendendo a 5.2:

$$G_{S_n} = \prod_{i=1}^n (1-p+pz)^{n_i} = (1-p+pz)^{\sum_{i=1}^n n_i}.$$

donde se conclui que a soma é uma variável binomial com parâmetros $(p, \sum_{i=1}^n n_i)$

Exemplo 5.2 *Sejam $X_i, i = 1, \dots, n$ variáveis geométricas independentes com parâmetro p .*

Como para uma variável geométrica X_i a função geradora vale

$$G_{X_i}(z) = \sum_{k=0}^{\infty} p(1-p)^{k-1} z^k = \frac{pz}{1-(1-p)z}$$

atendendo a 5.2:

$$G_{S_n} = \prod_{i=1}^n \frac{pz}{1-(1-p)z} = \left(\frac{pz}{1-(1-p)z} \right)^n$$

donde se conclui que a soma é uma variável binomial negativa com parâmetros (p, n)

Exemplo 5.3 *Sejam $X_i, i = 1, \dots, n$ variáveis dePoisson independentes com parâmetros $\alpha_i, i = 1, \dots, n$, respectivamente.*

Como para uma variável de Poisson X_i a função geradora vale

$$\begin{aligned} G(z) &= E[z^{X_i}] = \sum_{k=0}^{\infty} \frac{\alpha^k e^{-\alpha}}{k!} z^k = \\ &= e^{-\alpha} \sum_{k=0}^{\infty} \frac{(\alpha z)^k}{k!} = e^{-\alpha} e^{\alpha z} = e^{-\alpha(z-1)} \end{aligned}$$

atendendo a 5.2:

$$G_{S_n} = \prod_{i=1}^n e^{-\alpha_i(z-1)} = e^{-\sum_{i=1}^n \alpha_i(z-1)}.$$

donde se conclui que a soma é uma variável de Poisson com um parâmetro igual à soma dos parâmetros.

Exemplo 5.4 Sejam $X_i, i = 1, \dots, n$ variáveis gaussianas $N(m_i, \sigma_i^2)$ independentes

Atendendo a 5.1:

$$\Phi_{S_n}(w) = \prod_{i=1}^n \exp(jwm_i - w^2 \sigma_i^2 / 2) = \exp(jw \sum_{i=1}^n m_i - w^2 \sum_{i=1}^n \sigma_i^2 / 2)$$

donde se conclui que a soma é uma variável gaussiana com média igual à soma das médias e variância igual à soma das variâncias.

Exemplo 5.5 Sejam $X_i, i = 1, \dots, n$ variáveis gama independentes com parâmetros (α_i, λ) , respetivamente.

Como para uma variável gama X_i a função característica é

$$\begin{aligned} \Phi_X(w) &= \int_0^\infty e^{jwx} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-jw)x} dx = \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\lambda-jw)^\alpha} = \frac{1}{(1-jw/\lambda)^\alpha} \end{aligned}$$

Então sendo $\Phi_{X_i}(w) = \frac{1}{(1-jw/\lambda)^{\alpha_i}}$ atendendo a 5.1:

$$\Phi_{S_n}(w) = \prod_{i=1}^n \frac{1}{(1-jw/\lambda)^{\alpha_i}} = \frac{1}{(1-jw/\lambda)^{\sum_{i=1}^n \alpha_i}}$$

donde se conclui que a soma é uma variável gama com parâmetros $(\sum_{i=1}^n \alpha_i, \lambda)$.

Se as variáveis forem IID com parâmetros (α, λ) a variável soma terá parâmetros $(n\alpha, \lambda)$.

Exemplo 5.6 Sejam $X_i, i = 1, \dots, n$ variáveis gaussianas IID $N(0, 1)$. Caracterize a sua soma quadrática

$$S_n = X_1^2 + \dots + X_n^2.$$

Como se viu no exemplo 3.?? a transformação $Y_i = X_i^2$ transforma uma variável gaussiana $N(0, 1)$ numa variável χ^2 com $k = 1$, o que é equivalente a uma variável gama com parâmetros $(\alpha = 1/2, \lambda = 1/2)$. A variável em estudo é pois, a soma de n variáveis gama IID e, atendendo ao exemplo anterior, será uma variável gama com parâmetros $(n/2, 1/2)$, ou seja, uma variável χ^2 com n graus de liberdade.

Exemplo 5.7 Sejam $X_i, i = 1, \dots, n$ variáveis gaussianas IID $N(0, 1)$. Caracterize a variável:

$$Y = \sqrt{S_n} = \sqrt{X_1^2 + \dots + X_n^2}.$$

Do exemplo anterior sabe-se que S_n é uma variável χ^2 com n graus de liberdade. Atendendo a ??? a transformação $Y = \sqrt{X}$ conduz a $f_Y(y) = 2yf_X(y^2), y > 0$ e portanto como para a variável χ^2 com n graus de liberdade

$$f_X(x) = \frac{x^{(n-2)/2} e^{-x/2}}{2^{n/2} \Gamma(n/2)}$$

será

$$f_Y(y) = 2y \frac{y^{n-2} e^{-y^2/2}}{2^{n/2} \Gamma(n/2)} = \frac{y^{n-1} e^{-y^2/2}}{2^{n/2-1} \Gamma(n/2)}$$

Dois casos particulares:

1. $n = 2$

$$f_Y(y) = \frac{y e^{-y^2/2}}{\Gamma(1)} = y e^{-y^2/2}, y > 0$$

que é uma variável de Rayleigh

2. $n = 3$

$$f_Y(y) = \frac{y^2 e^{-y^2/2}}{\sqrt{2} \Gamma(3/2)} = \sqrt{\frac{2}{\pi}} y^2 e^{-y^2/2}, y > 0$$

que é uma variável de Maxwell¹

5.3 Leis dos grandes números

Seja X uma variável aleatória com valor médio $E[X] = \mu$ desconhecido, e sejam X_1, X_2, \dots, X_n , n variáveis aleatórias que representam n observações/medições da variável X . Admitindo que as observações são feitas em experiências independentes e idênticas, as variáveis X_1, X_2, \dots, X_n são IID e com valor médio igual ao X , isto é, $E[X_i] = \mu$ e variância σ^2 . Para se estimar² o valor médio μ usa-se o valor médio das amostras definido por

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i$$

A média das amostras é uma variável aleatória e o seu valor médio será

$$E[M_n] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \mu$$

Como $M_n = S_n/n$, $Var(M_n) = Var(S_n)/n^2$ e $Var(S_n) = n\sigma^2$, será:

$$Var(M_n) = \frac{\sigma^2}{n}$$

donde se pode concluir que, à medida que se aumenta o número de experiências/medições se diminui a variância da estimativa da média das observações, significando isso que a dispersão das médias obtidas em séries de experiências se vai reduzindo.

¹A variável ou distribuição de Maxwell é muito utilizada para modelar as características estatísticas da distribuição das moléculas de um gás.

²Este assunto será retomado com mais detalhe no capítulo 6.

Como se viu, a variância da média das estimativas tende para 0 à medida que o número de experiências cresce, o que se pode interpretar como a probabilidade da média das amostras se aproximar do valor médio, ser cada vez maior, aproximando-se de 1. Este resultado é facilmente demonstrável recorrendo à desigualdade de Chebyshev 3.5:

$$\begin{aligned} P[|M_n - E[M_n]| \geq \epsilon] &\leq \frac{\text{Var}(M_n)}{\epsilon^2} \\ P[|M_n - \mu| \geq \epsilon] &\leq \frac{\sigma^2}{\epsilon^2} \\ P[|M_n - \mu| < \epsilon] &\geq 1 - \frac{\sigma^2}{n\epsilon^2} \end{aligned}$$

o que no limite será

$$\lim_{n \rightarrow \infty} P[|M_n - \mu| < \epsilon] = 1 \quad (5.3)$$

resultado que é conhecido como **lei fraca dos grandes números**. Existe um segundo enunciado que não está dentro dos objetivos do presente curso e que afirma

$$P\left[\lim_{n \rightarrow \infty} M_n = \mu\right] = 1 \quad (5.4)$$

resultado que é conhecido como **lei forte dos grandes números**.

A lei fraca dos grandes números afirma que para um valor de n suficientemente elevado a média das amostras estará muito próxima do valor médio esperado com grande probabilidade, enquanto que a lei forte garante que é certo que o limite para que tende a média das amostras é o valor médio esperado.

Considere-se o caso particular em que se está a considerar experiências de Bernouilli. Neste caso a variável soma refere-se a uma soma de variáveis de Bernouilli e fornece o número de vezes que um certo acontecimento em teste se realiza em n experiências, sendo portanto a frequência absoluta de ocorrência. A média das amostras é então a frequência relativa de ocorrência $f_n(A)$ e o valor limite para que tende é a probabilidade do acontecimento $P[A]$, Substituindo nas expressões 5.3 e 5.4:

$$\lim_{n \rightarrow \infty} P[|f_n(A) - P[A]| < \epsilon] = 1 \quad , \quad P\left[\lim_{n \rightarrow \infty} f_n(A) = P[A]\right] = 1$$

Note que, para chegar a este resultado, se partiu de um conjunto de axiomas e sobre eles se desenvolveu uma teoria. Mas este resultado mostra que a frequência relativa se aproxima quasi de certeza da probabilidade, noção que foi o ponto de partida para o estabelecimento das bases axiomáticas da Teoria de Probabilidades.

Exemplo 5.8 *Uma tensão constante V e desconhecida tem de ser medida, mas o processo de medição introduz ruído N de valor médio nulo mas de variância igual a 16 (mV)^2 . Quantas medições deve fazer para que a média das medições difira do verdadeiro valor menos do que 1m com uma probabilidade de 99%*

$P[|M_n - \mu| < \epsilon] \geq 1 - \frac{\sigma^2}{n\epsilon^2}$ com $\epsilon = 1$ e $\sigma^2 = 16$ virá $0.99 \geq 1 - \frac{16}{n}$ e portanto $n \leq 1600$

Exemplo 5.9 Para se estimar a $P[A] = p$ uma sequência de experiências de Bernoulli é executada e observa-se frequência relativa de A . Determine o número de experiências a realizar para termos 0.95 de probabilidade de a frequência relativa estar a menos de 1% da probabilidade de A ?

Seja $X = I_A$: $E[I_A] = p$ e $\text{Var}(I_A) = \sigma^2 = p(1-p)$ são desconhecidas por se desconhecer p . No entanto é trivial mostrar que $p(1-p) \leq 1/4$. $P[|f_n(A) - p| < \epsilon] \geq 1 - \frac{\sigma^2}{n\epsilon^2} \geq 1 - \frac{1}{4n\epsilon^2}$ e portanto $0.95 \geq 1 - \frac{1}{4n \cdot 0.01^2}$ e portanto $n \leq 50000$

Os limites encontrados nos exemplos anteriores foram obtidos a partir da desigualdade de Chebyshev e, como já foi referido, este teorema fornece limites bastante pouco apertados. Mais à frente se verá que os resultados pretendidos se obtêm para um valor do número de experiências muito menor.

5.4 Teorema do limite central

Seja X_1, X_2, \dots uma sequência de variáveis aleatórias I.I.D. com média finita μ e variância finita σ^2 , seja S_n a variável soma das n primeiras variáveis e seja Z_n é a variável aleatória de média nula e variância unitária que se obtém a partir de S_n usando a transformação $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. O teorema do limite central afirma³

Teorema 5.3

$$\lim_{n \rightarrow \infty} P[Z_n \leq z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx$$

isto é, a função de distribuição de Z_n tende para a função de distribuição de uma variável gaussiana normalizada $N(0, 1)$.

Demonstração. $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu)$ e calculando a função característica

$$\begin{aligned} \Phi_{Z_n}(w) &= E \left[\exp \left(\frac{iw}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \right) \right] = \\ &= E \left[\prod_{i=1}^n \exp \left(\frac{iw}{\sigma\sqrt{n}} (X_i - \mu) \right) \right] = \\ &= \prod_{i=1}^n E \left[\exp \left(\frac{iw}{\sigma\sqrt{n}} (X_i - \mu) \right) \right] = \\ &= \prod_{i=1}^n \Phi_{X_i}(w / (\sigma\sqrt{n})) = (\Phi_X(w / (\sigma\sqrt{n})))^n \end{aligned}$$

³O resultado que se apresenta conhecido por teorema de Lindberg-Levy é um caso particular do teorema do limite central.

Como $E[X_i - \mu] = 0$ e $\text{Var}(X_i - \mu) = \sigma^2$, a aplicação do teorema dos momentos 3.2 a $\Phi_X(w/(\sigma\sqrt{n}))$ conduzirá a:

$$\Phi_X(w/(\sigma\sqrt{n})) = 1 - \frac{w^2}{2n} + R\left(\frac{w^2}{\sigma^2 n}\right)$$

em que $R(t^2)/t^2 \rightarrow 0$ com t .

Então

$$\Phi_{Z_n}(w) = \left(1 - \frac{w^2}{2n} + R\left(\frac{w^2}{\sigma^2 n}\right)\right)^n$$

e

$$\ln \Phi_{Z_n}(w) = n \ln \left(1 - \frac{w^2}{2n} + R\left(\frac{w^2}{\sigma^2 n}\right)\right)$$

Atendendo a que $\ln(1+z) = z \cdot \epsilon$ com $\lim_{z \rightarrow 0} \epsilon = 1$:

$$\begin{aligned} \ln \Phi_{Z_n}(w) &= n \left(-\frac{w^2}{2n} + R\left(\frac{w^2}{\sigma^2 n}\right) \right) = -\frac{w^2}{2} + nR\left(\frac{w^2}{\sigma^2 n}\right) \\ &= -\frac{w^2}{2} + R\left(\frac{w^2}{\sigma^2 n}\right) / (1/n) \end{aligned}$$

e no limite

$$\lim_{n \rightarrow \infty} \ln \Phi_{Z_n}(w) = -\frac{w^2}{2}$$

ou seja, Z_n é uma variável gaussiana de valor médio nulo e variância unitária.

■

Este é um resultado extremamente importante pois afirma que a soma de **quaisquer** variáveis I.I.D. de média e variância finitas aproxima-se de uma distribuição normal à medida que o seu número aumenta, abrindo caminho a numerosas aplicações.

Exemplo 5.10 Suponha que as despesas feitas por cada cliente de um restaurante são variáveis aleatórias I.I.D. com $\mu = 6.5$ Euros e $\sigma = 2.5$ Euros

1. Estime a probabilidade de que os primeiros 100 clientes gastem um total superior a 600 Euros e de que gastem entre 640 e 660 Euros

Considere-se a variável $S_{100} = X_1 + \dots + X_{100}$. Como a sua média e variância são $100\mu = 650$ e $n\sigma^2 = 625$, a correspondente variável normalizada será $Z_{100} = \frac{S_{100} - 650}{25}$.

Admitindo que Z_{100} segue uma lei normal $N(1, 0)$:

$$\begin{aligned} P[S_{100} > 600] &= P\left[Z_{100} > \frac{600 - 650}{25}\right] = P[Z_{100} > -2] = \\ &= Q(-2) = 0.9773 \end{aligned}$$

$$\begin{aligned} P[640 < S_{100} < 660] &= P\left[\frac{640 - 650}{25} < Z_{100} < \frac{660 - 650}{25}\right] = \\ &= P[-0.4 < Z_{100} < 0.4] = 1 - 2Q(0.4) = 0.3108 \end{aligned}$$

2. Quantas encomendas têm de ser feitas para que se tenha 90% de probabilidade dos clientes gastarem mais do que 375?

Como S_n tem média igual a $6.5n$ e variância de $2.5\sqrt{n}$

$$\begin{aligned} P[S_n > 375] &= P\left[Z_n > \frac{375 - 6.5n}{2.5\sqrt{n}}\right] > 0.9 \\ \frac{375 - 6.5n}{2.5\sqrt{n}} &< -1.2815 \Rightarrow n \geq 62 \end{aligned}$$

O teorema do limite central permite também aproximar a distribuição binomial pela distribuição gaussiana. Com efeito uma variável binomial X_n é equivalente a uma soma de variáveis (IID) de Bernoulli $X_n = \sum_{i=1}^n I_i$. Se se definir a variável normalizada $Y_n = \frac{X_n - E[X_n]}{\sqrt{\text{Var}(X_n)}} = \frac{X_n - np}{\sqrt{np(1-p)}}$ pode adivinhar-se que, pelo teorema do limite central, a variável Y_n se aproxima de uma $N(1, 0)$. O seguinte teorema conhecido por Teorema de Moivre-Laplace traduz este facto ⁴:

Teorema 5.4 *Seja $I_i, i = 1, \dots, n$ uma sequência de variáveis de Bernoulli com probabilidade de sucesso p e seja $X_n = \sum_{i=1}^n I_i$. Então para $a < b$ e n suficientemente grande*

$$P\left[a \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq b\right] \simeq \Phi(b) - \Phi(a) \quad (5.5)$$

Este teorema é também apresentado das seguintes maneiras equivalentes

Teorema 5.5

$$P[X = k] = C_k^n p^k (1-p)^{n-k} \simeq \frac{1}{\sqrt{np(1-p)}} \phi\left(\frac{k - np}{\sqrt{np(1-p)}}\right) \quad (5.6)$$

ou

$$P[X \leq l] = \sum_{k=0}^l C_k^n p^k (1-p)^{n-k} \simeq \Phi\left(\frac{l - np}{\sqrt{np(1-p)}}\right) \quad (5.7)$$

Exemplo 5.11 *Para se estimar a $P[A] = p$ uma sequência de experiências de Bernoulli é executada e observa-se frequência relativa de A . Determine o número de experiências a realizar para termos 0.95 de probabilidade de a frequência relativa estar a menos de 1% da probabilidade de A ?*

No exemplo nº ??? usou-se a desigualdade de Chebyshev para estimar uma solução. Aplique-se o resultado agora referido, e tendo em consideração que

⁴A demonstração deste teorema fica um pouco para além do âmbito deste curso; o leitor mais interessado poderá encontrar uma demonstração em L.Helms: Probability Theory with Contemporary Applications, Freeman, 1997.

$f_A(n)$ tem média p e variância $p(1-p)/n$, a variável normalizada será $Y_n = \frac{f_A(n) - p}{\sqrt{p(1-p)/n}}$

$$\begin{aligned} P[-\epsilon \leq f_A(n) - p \leq \epsilon] &= P\left[-\epsilon \frac{\sqrt{n}}{p(1-p)} \leq Y_n \leq \epsilon \frac{\sqrt{n}}{\sqrt{p(1-p)}}\right] \simeq \\ &\simeq 1 - 2Q\left(\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) \leq 1 - 2Q(2\epsilon\sqrt{n}) \end{aligned}$$

pois $p(1-p) \leq 1/4$

$$\begin{aligned} 1 - 2Q(2\epsilon\sqrt{n}) &\geq 0.95 \\ Q(2\epsilon\sqrt{n}) &\leq (1 - 0.95)/2 \\ 2\epsilon\sqrt{n} &\geq 1.96 \\ n &\geq ((1.96/2)/0.01)^2 = 9604 \end{aligned}$$

número significativamente menor que a anterior estimativa

5.5 Exercícios

Exercício 5.1 É necessário fazer um inquérito de opinião para determinar qual a percentagem de uma população (p) é favorável ao candidato A numa dada eleição. Determine o número de entrevistas a realizar pro forma determinar p a menos de 0.05 com uma probabilidade de 0.95.

A resposta procurada é a solução da inequação: $P\left[\left|\frac{S_n}{n} - p\right| \leq 0.05\right] \geq 0.95$
Resolvendo recorrendo à desigualdade de Chebyshev:

$$P\left[\left|\frac{S_n}{n} - p\right| \geq 0.05\right] \leq \frac{1}{4n(0.05)^2} \leq 0.05 \Rightarrow n \geq 2000$$

Utilizando os resultados do teorema do limite central e tendo em atenção que $p(1-p) \leq 1/4$

$$\begin{aligned} P\left[\left|\frac{S_n}{n} - p\right| \leq 0.05\right] &= P\left[\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| \leq 0.05\sqrt{n/(p(1-p))}\right] = \\ &= 1 - 2Q\left(\frac{0.05\sqrt{n}}{\sqrt{p(1-p)}}\right) \leq 1 - 2Q(0.1\sqrt{n}) \geq 0.95 \end{aligned}$$

$$Q(0.1\sqrt{n}) \geq 0.025 \Rightarrow 0.1\sqrt{n} \geq 1.96 \Rightarrow n \geq 19.6^2 = 384.16$$

ou seja $n = 385$ é o número mínimo procurado.

Exercício 5.2 Determine aproximadamente o valor de

$$\sum_{i=26}^{36} C_i^{64} 2^{-64}$$

O somatório corresponde a determinar a probabilidade de uma variável binomial de parâmetros $n = 64$ e $p = 1/2$ estar no intervalo $26 \leq i \leq 36$. Então $np = 32$ e $\sqrt{np(1-p)} = 4$

$$\begin{aligned} \sum_{i=26}^{36} C_i^{64} 2^{-64} &= P[26 \leq X_{64} \leq 36] = P\left[\frac{26-32}{4} \leq \frac{X_{64}-32}{4} \leq \frac{36-32}{4}\right] = \\ &= P\left[-1.5 \leq \frac{X_{64}-32}{4} \leq 1\right] = Q(-1.5) - Q(1) = 0.7745 \end{aligned}$$

Exercício 5.3 Um avião com capacidade para 360 passageiros tem reservado 27500Kg para os passageiros. Supondo que a população tem um peso médio de 75kg com um desvio padrão de 22kg determine a probabilidade de, numa viagem com 360 passageiros, o limite de peso seja ultrapassado.

$$\begin{aligned} P[S_{360} \geq 27500] &= P\left[\frac{S_{360} - 360 \cdot 75}{22\sqrt{360}} \geq \frac{27500 - 360 \cdot 75}{22\sqrt{360}}\right] = \\ &= Q\left(\frac{27500 - 360 \cdot 75}{22\sqrt{360}}\right) = Q(1.1978) = 0.1155 \end{aligned}$$

Exercício 5.4 Um programa de cálculo foi feito por forma a manter apenas m algarismos da parte decimal, efetuando-se uma truncatura sempre que uma operação aritmética é executada. Admitindo que o erro de truncatura é uma variável aleatória uniformemente distribuída no intervalo $[-0.5 \cdot 10^{-m}, 0.5 \cdot 10^{-m}]$ e que o erro total é a soma de todos os erros, determine a probabilidade do erro de truncatura não exceder $0.5 \cdot 10^{-m+3}$ se se efetuarem 10^6 operações.

Como a média é nula e desvio padrão vale $10^{-m}/\sqrt{12}$:

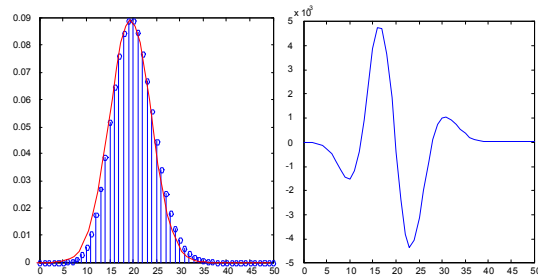
$$\begin{aligned} P[|E_{final}| \leq 0.5 \cdot 10^{-m+3}] &= P\left[\left|\frac{E_{final}}{\frac{10^{-m}}{\sqrt{12}}\sqrt{10^6}}\right| \leq \frac{0.5 \cdot 10^{-m+3}}{\frac{10^{-m}}{\sqrt{12}}\sqrt{10^6}}\right] = \\ &= P\left[\left|\frac{E_{final}}{\frac{10^{-m}}{\sqrt{12}}\sqrt{10^6}}\right| \leq \sqrt{3}\right] = \\ &= Q(-\sqrt{3}) - Q(\sqrt{3}) = 0.9167 \end{aligned}$$

Exercício 5.5 Seja S_{100} a soma de 100 variáveis IID de Poisson com média 0.2. Faça um programa que compare o valor exato com o valor aproximado dado pela aplicação do teorema do limite central.

A soma de 100 variáveis de Poisson com $\alpha = 0.2$ é uma variável de Poisson com $\alpha_{100} = 0.2 \cdot 100 = 20$. Como para a variável de Poisson a variância iguala o valor médio tem-se $\sigma_{100}^2 = 20$. Pelo teorema do limite central a variável gaussiana normalizada será uma boa aproximação a $\frac{S_{100} - a_{100}}{\sigma_{100}} = \frac{S_{100} - 20}{\sqrt{20}}$, ou seja, $S_{100} \simeq N(20, \sqrt{20})$.

O seguinte programa em Matlab permite calcular a função de probabilidade da variável de Poisson com $\alpha_{100} = 20$, a densidade de probabilidade de $N(20, \sqrt{20})$, desenhar os respectivos gráficos e o gráfico do erro $S_{100} - N(20, \sqrt{20})$

```
t=(0:50);
s100=poisspdf(t,20);
y=normpdf(t,20,sqrt(20));
stem(t,s100),hold on, plot(t,y,'r'),hold off
figure(2)
plot(t,s100-y)
```



Appendix A

Análise Combinatória

Considerem-se n objetos distintos a_1, \dots, a_n e designe-se por $A = \{a_1, \dots, a_n\}$ o seu agrupamento. Seja $B = \{b_1, \dots, b_m\}$ um outro agrupamento. Podemos formar um novo agrupamento $A \times B$ constituído pelos pares ordenados (a_i, b_j) , $i = 1, \dots, n$ e $j = 1, \dots, m$. Como estes pares ordenados se podem organizar numa matriz retangular de n linhas por m colunas o número total de elementos de $A \times B$ é $n \times m$.

Generalizando, se A_1, \dots, A_r forem agrupamentos com n_1, \dots, n_r membros respetivamente, pode-se formar um agrupamento de $A_1 \times \dots \times A_r$ de r -tuplos ordenados (grupos ordenados com r elementos) (a_1, \dots, a_r) , $a_i \in A_i$, $i = 1, \dots, r$ cujo número será $n_1 \times \dots \times n_r$. (a demonstração poderá ser feita por indução sobre r).

$$\text{número de objetos de } (A_1 \times \dots \times A_r) = n_1 \times \dots \times n_r \quad (\text{A.1})$$

Um caso de particular interesse é o de todos os A_i serem o mesmo agrupamento A , contendo n objetos. Neste caso o número de r -tuplos ordenados é de n^r . A estes r -tuplos ordenados, simbolizados por $A_r'^n$, chamam-se os arranjos com repetição de ordem r de uma população A com n objetos, ou arranjos de n a r .

$$A_r'^n = n^r \quad (\text{A.2})$$

Exemplo A.1 *Considere-se o lançamento de um dado três vezes consecutivas. O resultado destes lançamentos podem ser vistos como os arranjos de ordem 3 com repetição de uma população $A = \{1, 2, 3, 4, 5, 6\}$. Portanto, neste caso, $n = 6$ e $r = 3$ resultando num número total de resultados $6^3 = 216$.*

Pode-se formar grupos de uma população sem que haja repetição: cada objeto de A entrará uma e apenas uma vez em cada agrupamento.

Pode-se considerar que se está perante um problema com diferentes coleções de objetos: a primeira $A_1 = A$ com n objetos; a segunda A_2 terá $n - 1$ objetos;

e a de ordem r , A_r terá $n - r + 1$ objetos. Então de acordo com a equação A.1 os arranjos sem repetição de n objetos em grupos de r serão:

$$A_r^n = n(n-1) \times \dots \times (n-r+1) \quad (\text{A.3})$$

Um caso particular será quando $n = r$, isto é, quando se pretende formar grupos em que entrem todos os objetos da coleção, habitualmente designados por permutações dos n objetos:

$$P_n = A_n^n = n(n-1) \times \dots \times 1 = n! \quad (\text{A.4})$$

Nalguns casos a ordem em que aparecem os objetos é irrelevante (por exemplo num jogo de cartas, uma "mão" - conjunto de cartas atribuído a um jogador - não interessa qual a ordem em que recebeu as cartas, apenas vale apenas considerar quais as cartas que a constituem). A este tipo de agrupamentos não ordenados de dimensão k construídos a partir de uma coleção de n objetos chamam-se as combinações de n objetos k a k e representam-se simbolicamente por C_k^n ou $\binom{n}{k}$.

Para determinar o seu número pode seguir-se o seguinte raciocínio: para cada grupo não ordenado de k objetos pode-se construir $k!$ grupos ordenados permutando os k objetos de todas as maneiras possíveis. Então os arranjos sem repetição de n , k a k podem ser determinados pelo produto das combinações de n , k a k por o número de permutações de k , ou seja, $A_k^n = C_k^n \cdot P_k$. Então:

$$C_k^n = \frac{A_k^n}{P_k} = \frac{n(n-1) \times \dots \times (n-k+1)}{k!} = \frac{n!}{k!(n-k)!} \quad (\text{A.5})$$

Note que por definição $C_0^n = 1$.

Uma vez que para cada agrupamento de k objetos corresponde sempre um agrupamento de $n - k$ objetos tem-se

$$C_k^n = C_{n-k}^n$$

Aos C_k^n chamam-se também coeficientes binomiais por estarem associados à fórmula do binómio.

$$(a+b)^n = \sum_{k=0}^n C_k^n a^k b^{n-k}$$

Se $a = b = 1$ $(1+1)^n = \sum_{k=0}^n C_k^n 1^k 1^{n-k}$ e portanto

$$2^n = \sum_{k=0}^n C_k^n$$

Se $a = -1$ e $b = 1$ $(1-1)^n = \sum_{k=0}^n C_k^n (-1)^k 1^{n-k}$ e portanto

$$0 = \sum_{k=0}^n (-1)^k C_k^n = 1 - C_1^n + C_2^n - \dots + (-1)^k C_k^n$$

Considere-se que se divide um conjunto de n objetos em subconjuntos B_1, B_2, \dots, B_m em que a B_j se atribuem k_j objetos com $j = 1, 2, \dots, m$, $k_j > 0$ e $k_1 + k_2 + \dots + k_m = n$.

Quantos grupos se podem construir destes n objetos assim repartidos?

Seja $N_j, j = 1, \dots, m$ o número de maneiras como se pode formar o subconjunto B_j . N_1 será então de $C_{k_1}^n$. Dos restantes $n - k_1$ pode-se agora escolher k_2 objetos para construir B_2 . E assim sucessivamente até que por fim teremos k_m objetos que constituem B_m e portanto, $N_m = 1$. Tem-se então $N_1 = C_{k_1}^n, N_2 = C_{k_2}^{n-k_1}, N_3 = C_{k_3}^{n-k_1-k_2}, \dots, N_{m-1} = C_{k_{m-1}}^{n-k_1-k_2-\dots-k_{m-2}}, N_m = 1$.

O número total de maneiras de criar este tipo de agrupamentos, e que se representará por $P_{k_1, k_2, \dots, k_m}^n$ será então de

$$N_1 \times N_2 \times \dots \times N_{m-1} \times N_m = \frac{n!}{k_1!(n-k_1)!} \times \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \times \dots \times \frac{(n-k_1-\dots-k_{m-2})!}{k_{m-1}!\underbrace{(n-k_1-\dots-k_{m-2}-k_{m-1})!}_{=k_m}}$$

$$P_{k_1, k_2, \dots, k_m}^n = \frac{n!}{k_1! k_2! \dots k_m!} \quad (\text{A.6})$$

Também se chamam a estes agrupamentos os coeficientes polinomiais e se $m = 2$ reduzem-se aos valores anteriores das combinações dados por A.5.

As combinações de objetos que têm vindo a considerar-se pressupõem que os n objetos são distinguíveis entre si. No caso de não se poder fazer essa distinção (situação que surge por exemplo na física moderna em que em certas condições n partículas não se podem diferenciar do ponto de vista das suas características quânticas) tem que se considerar outro processo de contagem das combinações que se podem construir. Suponhamos que se tem n objetos e que se retiram k objetos com substituição e que se faz o seguinte registo das observações. Por exemplo, se $n = 6$ e $k = 5$, e admitindo que os objeto 2 e 6 saíram duas vezes e o objeto 4 saiu 1 vez:

$$\begin{array}{cccccc} \text{Obj. 1} & & \text{Obj. 2} & & \text{Obj. 3} & & \text{Obj. 4} & & \text{Obj. 5} & & \text{Obj. 6} \\ & / & \times \times & / & & / & \times & / & & / & \times \times \end{array}$$

A observação pode então ser representada por $/ \times \times // \times // \times \times$. E qualquer outra observação pode ser representada por $k = 5$ símbolos \times e por $n - 1 = 5$ símbolos $/$. Mas então o número de possíveis extrações que se podem obter iguala o número de agrupamentos que se podem fazer com estes símbolos em número total de $n - 1 + k$ em grupos de k (ou de $n - 1$). Este tipo de agrupamento é por vezes designado por combinações com repetição de n em grupos de k e o seu número pode ser calculado por

$$C_k'^n = C_k^{n-1+k} = C_{n-1}^{n-1+k}$$

Exemplo A.2 *5 amigos vão a um restaurante que apresenta na lista 3 tipos de refeição. De quantas maneiras poderão ser efetuadas as encomendas de refeição?*

Neste caso 3 objetos podem ser escolhidos com repetição pelos 5 indivíduos. O número de possibilidades será $C_5^3 = C_5^{3-1+5} = C_5^7 = 21$.

Appendix B

Geração de variáveis aleatórias

B.1 Métodos diretos

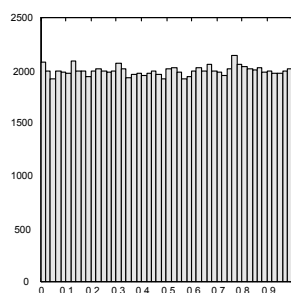
B.1.1 Variável uniforme

Conforme foi afirmado no capítulo inicial, a base da geração em computador de séries de números pseudo-aleatórios, é a geração de M números com igual frequência e isso consegue-se usando o método do resíduo.

No Matlab está disponível a subrotina `rand(.)` que permite gerar não números inteiros, mas sim números uniformemente distribuídos no intervalo $[0, 1)$.

Simples operações de multiplicação, adição e arredondamento permitem gerar distribuições uniformes na gama pretendida

.Exemplo para $U(0,1)$ e 100000 amostras: `hist(rand(1,100000),50)`



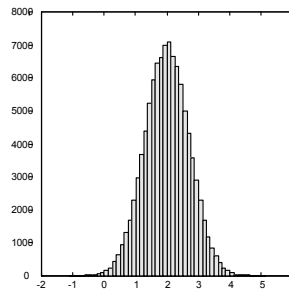
B.1.2 Variável gausseana

O método habitualmente aconselhado para gerar variáveis deste tipo consiste em gerar duas variáveis uniformemente distribuídas $U_1(0,1)$ e $U_2(0,1)$ e aplicar as seguintes fórmulas obtendo-se duas variáveis gausseanas X e Y independentes de média nula e variância unitária $X = \sqrt{-2\ln(U_1)} \cos(2\pi U_2)$ e $Y = \sqrt{-2\ln(U_1)} \sin(2\pi U_2)$ (note que a multiplicação pelos valores sinusoidais serve para efetuar uma ortogonalização das variáveis garantindo a independência).

No entanto com o Matlab não é necessário recorrer a este tipo de transformação pois está disponível a subrotina `randn(.)` que gera números aleatórios com uma distribuição gausseana de valor médio nulo e variância unitária. Utilizando transformações já referidas $E[X+c] = E[X]+c$ e $Var(cX) = c^2 Var(X)$ pode-se gerar valores de distribuições gausseanas com média e variância arbitrárias.

Exemplo para $m = 2$ e $\sigma^2 = 1/2$ e 100000 amostras:

```
hist(sqrt(.5)*randn(1,100000)+2,50)
```



B.1.3 Variável de Bernoulli

Para gerar uma variável de Bernoulli parte-se de uma variável uniforme $U(0,1)$ e constroi-se a variável I_A fazendo

$$I_A = \begin{cases} 0, & U \leq p \\ 1, & U > p \end{cases} \quad \text{que pode ser implementado com a subrotina } \text{bern}(n,p):$$

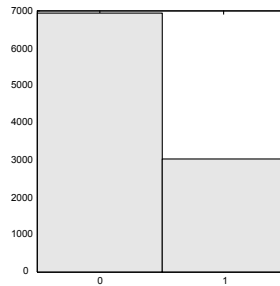
%função que gera n valores de uma variável de Bernoulli com

%probabilidade p para o valor 1 e 1-p para 0

```
function y=bern(n,p)
```

```
y=round(rand(1,n)-0.5+p);
```

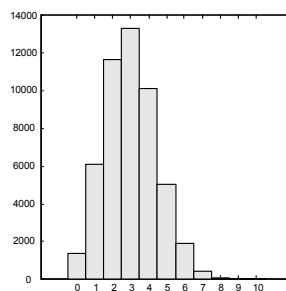
Exemplo para $p=0.3$ e 10000 amostras: `hist(bern(10000,.3),(0:1)')`



B.1.4 Variável binomial

Para gerar uma variável binomial de parâmetros n e p pode-se gerar e somar n variáveis de Bernoulli com probabilidade p que pode ser implementado com a subrotina `binomial(namostras,n,p)`:

```
%função que gera namostras de uma variável binomial com
%com parâmetros n e p
function y=binomial(namostras,n,p)
y=sum(round(rand(n,namostras)-0.5+p));
Exemplo para  $n = 10, p = 0.3$  e 50000 amostras: hist(binomial(50000,10,.3),(0:10)')
```



B.2 Método da transformação

Teorema B.1 Se X é uma variável aleatória de tipo contínuo com função de distribuição $F_X(x)$ a variável $Y = F_X(X)$ é $U(0, 1)$.

Teorema B.2 $0 \leq Y \leq 1$ porque $F_X(x)$ é uma probabilidade.

Demonstração.

$$\begin{aligned}
 F_Y(y) &= P[Y \leq y] = P[F_X(X) \leq y] = \\
 &= P[F_X^{-1}(F_X(X)) \leq F_X^{-1}(y)] = \\
 &= P[X \leq F_X^{-1}(y)]
 \end{aligned}$$

porque $F_X(x)$ é monótona crescente. Então:

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ P[X \leq F_X^{-1}(y)] = F_X(F_X^{-1}(y)) = y & 0 \leq y \leq 1 \\ 1 & y > 1 \end{cases}$$

■

Inversamente, se Y for $U(0, 1)$, a variável $X = F_X^{-1}(Y)$ terá a uma função de distribuição $F_X(x)$.

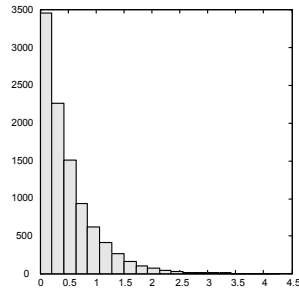
Portanto pode-se usar o seguinte método para a geração de variáveis aleatórias contínuas com uma função de distribuição arbitrária $F_X(x)$

1. Gerar uma variável $Y = U(0, 1)$
2. Determinar a função inversa $F_X^{-1}(x)$
3. Efetuar a transformação $F_X^{-1}(Y)$

B.2.1 Variável exponencial

Fazendo $u = F_X(x) = 1 - e^{-\lambda}$ e resolvend em ordem a x : $X = -\frac{1}{\lambda} \ln(1 - U)$ que pode ser implementado com a subrotina `exponencial(n,lam)`

```
%função que gera namostras de uma variável exponencial
%com com valor médio lam
function y=exponencial(namostras,lam)
y=-lam*log(1-rand(1,namostras));
Exemplo para  $\lambda = 0.5$  e 100000 amostras: hist(exponencial(10000, .5), 20)
```



B.3 Método da rejeição

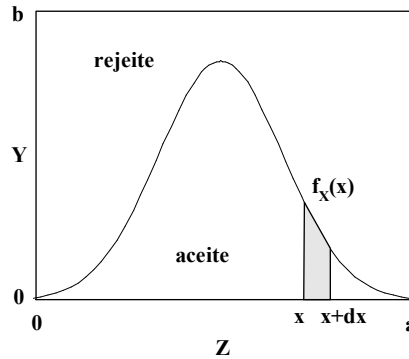
Para gerar amostras de uma variável aleatória X com uma densidade $f_X(x)$ que obedeça às condições

- $f_X(x) = 0$ para $x < 0$ e $x > a$

- $f_X(x)$ tem valores no intervalo $[0, b]$

pode-se usar o seguinte método:

1. Gere Z uniformemente distribuída no intervalo $[0, a]$
2. Gere Y uniformemente distribuída no intervalo $[0, b]$
3. Se $Y \leq f_X(Z)$ faça $X = Z$; se $Y > f_X(Z)$ rejeite Z e regresse ao passo 1



A probabilidade de aceitar Z é igual à área abaixo da curva $f_X(x)$ a dividir pela área do retângulo ab . Como a área abaixo da curva é unitária porque $f_X(x)$ é uma densidade de probabilidade, $P[Z \text{ aceite}] = \frac{1}{ab}$.

Então

$$\begin{aligned}
 P[x < Z \leq x + dx | Z \text{ aceite}] &= \frac{P[x < Z \leq x + dx \cap Z \text{ aceite}]}{P[Z \text{ aceite}]} = \\
 &= \frac{\text{área sombreada}}{1/ab} = \frac{f_X(x)dx/ab}{1/ab} = \\
 &= f_X(x)dx
 \end{aligned}$$

isto é, a variável X assim gerada tem a desejada densidade de probabilidade.

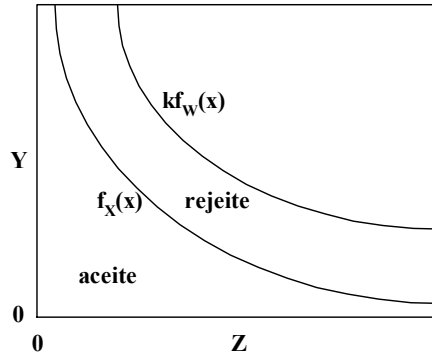
É trivial modificar o algoritmo para variáveis definidas não nulas num intervalo $[x_1, x_2]$.

Este método para além de estar limitado a variáveis definidas em intervalos finitos tem o inconveniente de obrigar muitas iterações que conduzem a resultados rejeitados, basta ver que se ab for elevado a probabilidade de aceitação é baixa. Para ultrapassar este problema pode-se generalizar o método envolvendo a curva de densidade de probabilidade por uma densidade de probabilidade fácil de gerar.

Seja W uma variável aleatória com uma densidade $f_W(w)$ tal que $kf_W(x) \geq f_X(x)$ para todo x ,

O método geral (que não será aqui demonstrado) é:

1. Gere Z com uma densidade $f_W(w)$ e defina $b(Z) = k f_W(Z)$
2. Gere Y uniformemente distribuída no intervalo $[0, b(Z)]$
3. Se $Y \leq f_X(Z)$ faça $X = Z$; se $Y > f_X(Z)$ rejeite Z e regresse ao passo 1



B.3.1 Variável com uma distribuição sinusoidal

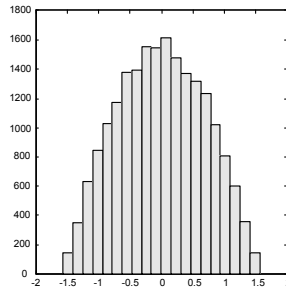
Pretende-se gerar uma variável aleatória com uma densidade de probabilidade

$$f_X(x) = \begin{cases} \frac{1}{2} \cos x, & |x| \leq \pi/2 \\ 0 & |x| > \pi/2 \end{cases}$$

o que pode ser feito usando o método da rejeição apresentado e implementado com a subrotina

```
sinusoidal(namostras):
%função que gera namostras de uma variável com uma densidade
%de probabilidade sinusoidal entre -pi/2 e pi/2
%usando o método da rejeição
function [z,ef]=sinusoidal(namostras)
if nargin<1
    namostras=1;
end
i=0;k=0;
while i<namostras
    x=pi*(rand-0.5);y=rand;
    if y<=0.5*cos(x)
        i=i+1;
        z(i)=x;
    else
        k=k+1;
    end
end
ef=i/(i+k);
```

Exemplo para 20000 amostras: `hist(sinusoidal(20000),20)`



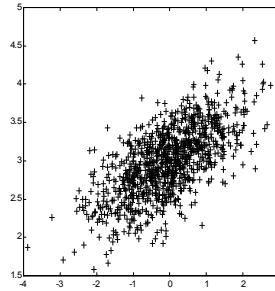
B.4 Geração de variáveis multidimensionais gausseanas

Utilizando os resultados apresentados para a transformação linear de variáveis conjuntamente gausseanas $\Sigma_Y = A\Sigma_X A^T$, se se pretende gerar amostras de um vetor \mathbf{Y} de variáveis aleatórias gausseanas com matriz de correlação Σ_Y e valor médio \mathbf{m}_Y pode proceder-se do seguinte modo: gerar um vetor X de variáveis aleatórias gausseanas com matriz de correlação $\Sigma_X = \mathbf{I}$ e valor médio nulo, determinar A tal que $\Sigma_Y = A\Sigma_X A^T = AA^T$ e efetuar a transformação linear $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{m}_Y$, tal como se implementa com o seguinte programa em Matlab.

```
%subrotina para gerar n amostras de um vetor
%de m variáveis conjuntamente gausseanas com
% matriz de correlação sigma e média med
function y=multigauss(n,m,sigma,med)
[i,j]=size(sigma);
k=max(size(med));
if m~=i|m~=j|m~=k
    error('dimensões erradas')
end
x=randn(m,n);
a=sqrtm(sigma);
y=a*x;
z=ones(m,n);
for i=1:k
    z(i,:)=z(i,:).*med(i);
end
y=y+z;
%visualização para m=2
if m==2
    close
    axis equal
    plot(y(1,:),y(2,:),'+k')
end
```

Exemplo mostrando o diagrama de espalhamento para 1000 amostras bidimensionais com $\Sigma = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 0.2 \end{bmatrix}$ e $\mathbf{m} = \begin{bmatrix} 0 & 3 \end{bmatrix}$

```
y=multigauss(1000,2,[1 0.3;0.3 0.2],[0 3]);
```



Appendix C

Formulário

Somas de sucessões

1. $\sum_{n=1}^N (a + nb) = Na + \frac{1}{2}N(N+1)b$
2. $\sum_{n=1}^N (a + nb)x^n = \frac{a - (a+bN)x^{N+1}}{1-x} + \frac{bx(1-x^N)}{(1-x)^2}$

Sucessões geométricas: $b = 0 \Rightarrow \sum_{n=1}^N ax^n = \frac{a - ax^{N+1}}{1-x}$

Outras sucessões

1. $\sum_{k=1}^n k = \frac{n(n+1)}{2}$
2. $\sum_{k=1}^n k^2 = \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n$
3. $\sum_{k=1}^n k^3 = \frac{1}{4}n^4 + \frac{1}{2}n^3 + \frac{1}{4}n^2$

Séries

Séries de potências

1. $\sum_{n=1}^{\infty} ax^n = \lim_{n \rightarrow \infty} \frac{a - ax^{n+1}}{1-x} = \frac{a}{1-x}, |x| < 1$
2. $\sum_{n=1}^{\infty} x^n = \frac{1}{1-x}, |x| < 1$
3. $\sum_{n=1}^{\infty} nx^{n-1} = \frac{d}{dx} \frac{1}{1-x} = \frac{1}{(1-x)^2}, |x| < 1$
4. $\sum_{n=1}^{\infty} nx^n = \frac{x}{(1-x)^2}, |x| < 1$

$$5. \sum_{n=1}^{\infty} n(n-1)x^{n-2} = \frac{d}{dx} \frac{1}{(1-x)^2} = \frac{2}{(1-x)^3}, |x| < 1$$

$$6. \sum_{n=1}^{\infty} \frac{x^n}{n!} = e^x$$

Função gama

$$\Gamma(x) = \int_0^{\infty} x^{t-1} e^{-t} dt, x > 0$$

Alguns valores e propriedades úteis:

1. $\Gamma(1/2) = \sqrt{\pi}$
2. $\Gamma(x+1) = x\Gamma(x)$
3. $\Gamma(k+1) = k!, k = 0, 1, 2, \dots$

Fórmulas trigonométricas

1. $\sin(a+b) = \sin a \cos b + \cos a \sin b$
2. $\sin(a-b) = \sin a \cos b - \cos a \sin b$
3. $\cos(a+b) = \cos a \cos b - \sin a \sin b$
4. $\cos(a-b) = \cos a \cos b + \sin a \sin b$
5. $\cos a \cos b = \frac{1}{2} \cos(a+b) + \frac{1}{2} \cos(a-b)$
6. $\sin a \sin b = \frac{1}{2} \cos(a-b) - \frac{1}{2} \cos(a+b)$

Transformada de Fourier

$$H(f) = \mathcal{F}(h(t)) = \int_{-\infty}^{\infty} h(t) e^{-j2\pi ft} dt$$

$$h(t) = \mathcal{F}^{-1}(H(f)) = \int_{-\infty}^{\infty} H(f) e^{j2\pi ft} df$$

Propriedades

- | | |
|-----------------------------|--|
| 1. Linearidade | $\mathcal{F}(k_1 a_1(t) + k_2 a_2(t)) = k_1 A_1(f) + k_2 A_2(f)$ |
| 2. Escalonamento no tempo | $\mathcal{F}(h(at)) = H(f/a)/ a $ |
| 3. Translação no tempo | $\mathcal{F}(h(t-\tau)) = H(f) e^{-j2\pi f\tau}$ |
| 4. Translação na frequência | $\mathcal{F}(h(t) e^{j2\pi f_0 t}) = H(f-f_0)$ |
| 5. Diferenciação | $\mathcal{F}(h'(t)) = j2\pi f H(f)$ |
| 6. Multiplicação no tempo | $\mathcal{F}(h(t).g(t)) = H(f) * G(f)$ |
| 7. Convolução no tempo | $\mathcal{F}(h(t) * g(t)) = H(f).G(f)$ |

Tabela de transformadas

$h(t)$	$H(f)$
1	$\delta(f)$
$\delta(t)$	1
$\delta(t - \tau)$	$e^{-j2\pi f\tau}$
$u(t)$	$\frac{1}{2}\delta(f) + \frac{1}{j2\pi f}$
$u(t + \tau) - u(t - \tau)$	$\frac{2\tau \sin 2\pi f\tau}{2\pi f\tau}$
$\frac{2f_0 \sin 2\pi f_0 t}{2\pi f_0 t}$	$u(f + f_0) - u(f - f_0)$
$e^{-at}u(t), a > 0$	$\frac{1}{a + j2\pi f}$
$e^{-a t }, a > 0$	$\frac{2a}{a^2 + (2\pi f)^2}$
e^{-at^2}	e^{-af^2}
$\cos 2\pi f_0 t$	$\frac{1}{2}\delta(f + f_0) + \frac{1}{2}\delta(f - f_0)$
$\sin 2\pi f_0 t$	$\frac{1}{2j}(\delta(f + f_0) - \delta(f - f_0))$
$1 - t /\tau, t \leq \tau$ 0 $ t > \tau$	$\tau \left(\frac{\sin \pi f\tau}{\pi f\tau} \right)^2$

