Introdução à Teoria das Probabilidades

Eduardo Esteves

23 de Outubro de 2009

Departamento de Engenharia Alimentar, Instituto Superior de Engenharia da Universidade do Algarve, *Campus* da Penha, 8005-139 Faro, Portugal. E-mail: eesteves@ualg.pt, URL: `http://w3.ualg.pt/~eesteves`

Resumo

Este "artigo" constitui o segundo (grande) "capítulo" das matérias leccionadas no âmbito da disciplina de Estatística Aplicada do 2º Ano do Curso de Engenharia Alimentar (*cf.* Esteves (2009)[1]). Pretende-se apresentar, explicar (ainda que de forma sucinta e relativamente informal) e exercitar os conceitos básicos e as distribuições (teóricas) de probabilidades mais relevantes.

Conteúdo

1 Introdução

A ESTATÍSTICA diz respeito à análise e interpretação de dados (geralemente amostrais) com vista à avaliação objectiva da validade das conclusões que se obtiveram (relativamente à população). Os diversos MÉTODOS ESTATÍSTICOS que se empregam em ESTATÍSTICA APLICADA servem para colher, organizar, resumir, apresentar e/ou analisar dados de modo a obter conclusões válidas. A selecção criteriosa de amostras representativas duma população ou a inferência estatística a partir dessas amostras baseiam-se em conceitos relacionados com probabilidades. Esta é uma das "aplicações" da teoria das probabilidades.

[1] Esteves E. (2009) *Introdução aos Métodos Estaísticos e Estatística Descritiva.* Instituto Superior de Engenharia, Universidade do Algarve, Faro, 14 p. [disponível em `http://w3.ualg.pt/~eesteves`]

Tabela 1: Acontecimentos possíveis no lançamento simultâneo de dois dados honestos de seis faces.

	1	2	3	4	5	6
1	11	12	13	14	15	16
2	21	22	23	24	25	26
3	31	32	33	34	35	36
4	41	42	43	44	45	46
5	51	2	53	54	55	56
6	61	62	63	64	65	66

2 Perspectiva histórica (muito) resumida

Noções correntes de "possibilidade", "previsibilidade" ou "certeza" que vulgarmente não apresentam dificuldade de compreensão ou de interpretação, são formalizadas e estudados pela TEORIA DAS PROBABILIDADES. Este ramo da matemática desenvolveu-se sobretudo nos séculos XVII e XVIII, fruto do empenho dos franceses Blaise Pascal [1623-1662], Pierre de Fermat [1601-1665], Abraham de Moivre [1667-1754] e Pierre Simon Laplace [1749-1827] e do suíço Jakob Bernoulli [1654-1705], com o intuito de "predizer" os resultados de jogos de azar, populares entre a nobreza francesa daquele tempo. Mais recentemente, na transição entre os séculos XIX e XX, o russo Andrei Nikolaevich Kolmogorov [1903-1987], entre outros, contribuiu muito para este ramo da matemática. Ainda hoje se utilizam exemplos de jogos (dados e cartas) na apresentação deste tópico por causa das primeiras investigações e aplicações.

3 Conceitos básicos

A ideia-base da Estatística é de que as observações individuais são naturalmente variáveis aleatórias, isto é, os seus valores oscilam devido aos efeitos do Acaso. Quando assim não se verifica, podemos pensar ou considerar que outras causas não-aleatórias estão a actuar (designados *tratamentos*, *factores*, etc.). Antes de prosseguir para a análise e inferência estatística, é importante sistematizar alguns dos conceitos mais simples relacionados com a teoria das probabilidades.

3.1 Provas aleatórias, acontecimentos possíveis, evento e espaço amostral

Uma PROVA ALEATÓRIA é uma actividade a que correspondem dois ou mais ACONTECIMENTOS POSSÍVEIS. Antes de realizar a prova aleatória é incerto o seu resultado, isto é, qual dos acontecimentos possíveis irá ocorrer.

> Exemplo 1. Quando se lança um dado, diz-se que se realiza uma prova aleatória. Nesta prova existem seis acontecimentos possíveis: "sair" a face com 1, ou 2, ou 3, ou 4, ou 5 ou 6 pontos.
>
> Exemplo 2. Quando um técnico de controlo da qualidade selecciona uma lata de sardinha (com o objectivo de verificar se a lata é defeituosa), diz-se que está a realizar uma prova aleatória, em que existem dois acontecimentos possíveis: a lata é (classificada como) defeituosa, ou não.
>
> Exemplo 3. Lançamento simultâneo de dois dados. Nesta provas aleatória existem 36 acontecimentos possíveis.

Designa-se por ESPAÇO AMOSTRAL, referido por S, o conjunto de todos os acontecimentos possíveis de uma prova aleatória (*e.g.* a_i, b_i e c_i designam os acontecimentos possíveis de outros tantos espaços amostrais).

> Exemplo 1. Lançamento de um dado, $S = \{1, 2, 3, 4, 5, 6\}$ ou genericamente $S = \{a_1, a_2, a_3, a_4, a_5, a_6\}$.
>
> Exemplo 2. Selecção de uma lata de sardinha, $S = \{defeituosa, não - defeituosa\}$ ou em termos genéricos $S = \{b_1, b_2\}$.
>
> Exemplo 3. Lançamento simultâneo de dois dados $S = \{11, 12, ..., 21, 22, ..., 66\}$ ou $S = \{c_1, c_2, ..., c_{36}\}$ (Tab. 1).

Frequentemente, interessam sub-conjuntos do espaço amostral. Cada sub-conjunto de acontecimentos possíveis de um espaço amostral é desginado por EVENTO e usualmente representado por uma letra maiúscula (*A, B, C,* etc.) diferente de S.

Tabela 2: Informação da tabela anterior, na qual os acontecimentos do evento favorável "soma dos pontos igual a 7" estão assinalados.

	1	2	3	4	5	6
1	11	12	13	14	15	**16**
2	21	22	23	24	**25**	26
3	31	32	33	**34**	35	36
4	41	42	**43**	44	45	46
5	51	**52**	53	54	55	56
6	**61**	62	63	64	65	66

No lançamento único de um dado (ver exemplos anteriores), considere-se o domínio do evento "números pares" $A = \{2, 4, 6\}$. Usualmente, designa-se A como evento favorável. Pelo contrário, os restantes resultados possíveis, complementares de A, referem-se por A (alguns autores utilizam a notação A^c) e constituem o evento complementar. Se recorremos ao exemplo que deu origem à Tab. 1, o evento $B =$ "soma dos pontas igual a sete" é o sub-conjunto de S constituido pelos acontecimentos favoráveis indicados a sombreado na Tab. 2.

3.2 Conceito de probabilidade

A Teoria das probabilidades pretende formular modelos de fenómenos (naturais) em que se supõe intervir o Acaso, isto é, a partir do passado não se pode prever deterministicamente o futuro mas podem encontrar-se taxas de realização constantes de certos fenómenos.

Definição clássica de probabilidade A PROBABILIDADE[2] de ocorrer um evento A, designada por $P\{A\}$, ou $P(A)$, é definida classicamente:

$$P(A) = \frac{n_A}{N}$$

em que n_A é o número de acontecimentos favoráveis que podem ocorrer e N o número de todos os acontecimentos possíveis, desde que todos os acontecimentos sejam igualmente possíveis ou equiprováveis. Esta definição de probabilidade (estabelecida por Pierre Laplace em 1812) e ambas as expressões salientam o seu carácter ***a priori***. Todavia, nem sempre os acontecimentos são igualmente prováveis (nem o espaço amostral é finito), pelo que a sua aplicação não é geral.

Qual a probabilidade de ocorrer um número impar num único lançamento de um dado não viciado? O número de acontecimentos possíveis é 6, enquanto o número de acontecimentos favoráveis é 3 (isto é, faces 1, 3 e 5). Assim, P(A) = 3/6 = 1/2.

Se o dado estivesse viciado em favor do 6, por exemplo, já se não poderia aplicar a "definição clássica" de probabilidade. Porquê?

Definição de probabilidade como frequência relativa De modo similar às tabelas de frequências poderemos definir a frequência relativa de um evento como a proporção do número total de acontecimentos possíveis que esse evento representa. Ou seja, por definição a PROBABILIDADE é o limite da frequência relativa de determinado evento, quando o número de observações, isto é, o número de provas aleatórias cresce infinitamente:

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

[2] Para além das definições clássica e "frequencista" (apresentadas aqui), existe uma definição (perspectiva) subjectiva da probabilidade, que se desenvolveu a partir do Teorema de Bayes (devido a Thomas Bayes [1702-1761]). Por razões várias não será abordada neste documento embora se possa adiantar que a "corrente bayesiana" defende que há medida que se adquire (alguma) informação se deve proceder a uma reavaliação das probabilidades consideradas a priori. Neste contexto, a atribuição de probabilidade a um dado evento, principalmente numa primeira fase, como sendo subjectiva, e por isso encarada como uma medida de credibilidade - a probabilidade é uma expressão do estado de conhecimento sobre um certo sistema (Maria de Fátima Brilhante, Universidade dos Açores, `http://www.uac.pt/~fbrilhante/MedicinaPagWeb/probabilidade_v2.pdf` em 24/09/2009). Dito de outro modo, o teorema de Bayes mostra como alterar as probabilidades ***a priori*** tendo em conta novas evidências de forma a obter probabilidades ***a posteriori***. Esta e as restantes definições aplicam-se, de forma "natural," em certas e determinadas situações contudo todas têm desvantagens.

em que n_A é o número de provas aleatórias em que o evento A ocorre e n o número total de provas aleatórias. Esta é uma definição *a posteriori* de probabilidade. Genericamente, podemos descrever esta definição, recorrendo aos conceitos de frequência relativa f e de frequência absoluta F de que se falou anteriormente:

$$f_A = \frac{F_A}{n}$$

Por exemplo, em mil lances de uma moeda, obtêm-se 529 "caras". A frequência relativa deste evento é $f = 529/1000 = 0,529$. Façam-se outros 1000 lances da moeda e obtêm-se 493 "caras". A frequência relativa do acontecimento "caras" será: $f = (529+493)/2000 = 0,511$. Ou seja, quanto maior o número de lances, mais próximo se estará da probabilidade de ocorrer "caras" num único lançamento de uma moeda. Nota: Actualmente, considera-se esta probabilidade como sendo 0,5 (com um único algarismo significativo).

FIGURA?

A interpretação das probabilidades como frequência-limite, corresponde ao Teorema de Bernoulli que, em resumo, diz o seguinte: num fenómeno aleatório não se pode prever o resultado da próxima prova aleatória, mas pode prever-se globalmente a frequência da sua observação numa longa série de provas.

De facto, a frequência (de um evento) deve entender-se como uma medição física de uma grandeza teórica, a PROBABILIDADE associada a esse evento.

"Definição" de probabilidade relacionada com a Teoria dos conjuntos Modernamente, desde a axiomatização em 1933 por Andrei N. Kolmogorov, prefere-se fundamentar os teoremas das probabilidades na teoria dos conjuntos, pois recorre a menos e mais simples axiomas. Aliás, os conceitos básicos iniciais que se introduziram anteriormente derivam dessa abordagem ao problema (Fig. 1). Alguns dos axiomas e teoremas mais importantes são abordados na secção seguinte.

Neste texto (e na aplicação prática da Estatística), serão utilizadas as várias definições de probabilidades consoante o contexto do problema e a informação disponível.

3.3 Postulados das probabilidades

Os conceitos anteriores e a teoria das probabilidades baseiam-se em POSTULADOS[3], ou axiomas, que se exigem pragmáticos e consistentes (ou coerentes e compatíveis) dos quais três são:

1. Para qualquer acontecimento a_i, de um espaço amostral S, a probabilidade de ocorrer esse resultado favorável varia entre zero e um, inclusivamente:
$$0 \leq P(a_i) \leq 1$$
2. Para qualquer evento A do espaço amostral S, a probabilidade desse evento é o somatório das probabilidades dos acontecimentos a_i favoráveis nele incluídos:
$$P(A) = \sum P(a_i)$$
3. A probabilidade do espaço amostral S é igual a um e a probabilidade de acontecimentos impossíveis (isto é, daqueles que ocorrem fora de S, e que se incluem no conjunto ϕ, em que ϕ lê-se "Fi") é zero:
$$P(S) = 1$$
$$P(\phi) = 0$$

3.4 Teoremas das probabilidades

A utilização prática dos axiomas, permitiu o desenvolvimento de conclusões fundamentais ou TEOREMAS[4], auxiliares preciosos em estudos de probabilidades, designadamente[5]:

[3]Entede-se por postulado, ou axioma (um sinónimo relativamente caído em desuso), qualquer proposição aceite sem demonstração (Weisstein, Eric W. *"Axiom." From MathWorld–A Wolfram Web Resource* `http://mathworld.wolfram.com/Axiom.html`, consultado em 21/10/2009).

[4]Teoremas são proposições (ou afirmações) que se podem provar como verdadeiras, usando operações e argumentos matemáticos, numa determinada estrutura lógica (ou sistema axiomático) (Weisstein, Eric W. *"Theorem."* From MathWorld–A Wolfram Web Resource `http://mathworld.wolfram.com/Theorem.html`, consultado em 21/10/2009).

[5]Outros teorema(s), relacionados com probabilidades condicionais por exemplo, não estão, ainda, contemplados neste texto.

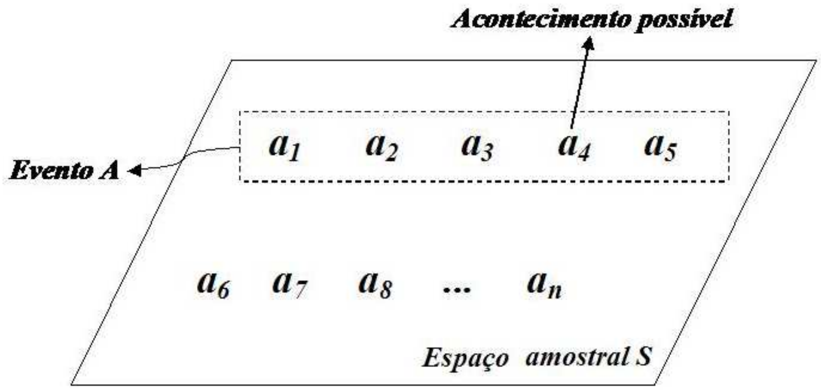

Figura 1: Representação esquemática dum espaço amostral S associado a uma prova aleatória. Dos vários acontecimentos possíveis a_i, apenas alguns pertencem ao evento A.

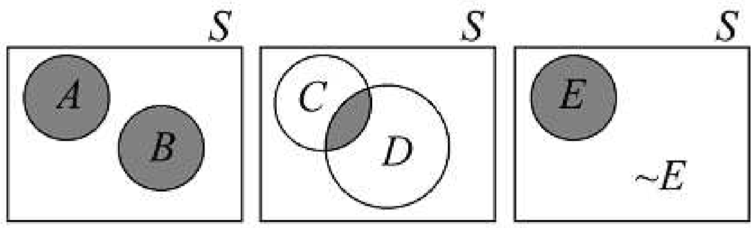

Figura 2: Diagramas de Venn, representando eventos mutuamente exclusivos (A e B), eventos independentes (C e D) e complementares (E e ˜E) nos respectivos espaços amostrais S. Área sombreada refere-se às probabilidades que se pretendem determinar de acordo com os Teoremas de probabilidade.

Teorema da adição Para dois eventos (ou acontecimentos) MUTUAMENTE EXCLUSIVOS A e B de um espaço amostral S, a probabilidade de ocorrer UM OU O OUTRO evento é igual à soma das respectivas probabilidades individuais, ou seja:

$$P(A \cup B) = P(A) + P(B)$$

em que $\cup$ se lê "ou". Podemos extender este teorema a mais do que dois eventos mutuamente exclusivos. Entende-se que eventos mutuamente exclusivos são aqueles que não ocorrem simultaneamente: se ocorre A, não ocorre B. Dito de outro modo, a intersecção dos conjuntos (eventos) A e B no espaço amostral S é um conjunto nulo (Fig. 2, painel da esquerda).

Teorema da multiplicação Para dois eventos INDEPENDENTES C e D de um espaço amostral S a probabilidade de OCORREREM SIMULTANEAMENTE é igual ao produto das probabilidades, isto é:

$$P(C \cap D) = P(C) \cdot P(D)$$

em que $\cap$ se lê "e". Por eventos independentes, entende-se que um dos eventos não determina ou influencia o resultado do(s) outro(s) (Fig. 2, painel central).

Teorema da complementaridade Para qualquer evento E de um espaço amostral S, a probabilidade de não ocorrer E, designado por $P(\sim E)$ ou $P(E^c)$, é igual a:

$$P(\sim E) = 1 - P(E)$$

sendo que $P(\sim E)$, representa o evento complementar de E (Fig. 2, painel da direita).

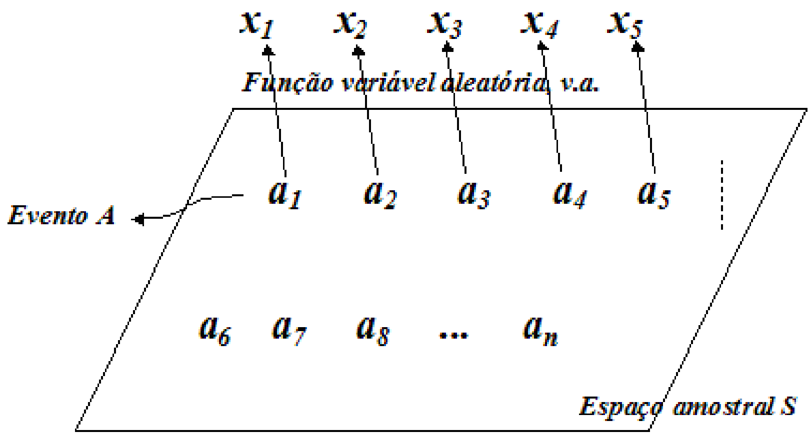

Figura 3: Representação esquemática do conceito de função variável aleatória, v.a. (linhas contínuas no topo da figura). S - espaço amostral duma prova aleatória, A - evento, a_i - acontecimento possível, e x_i - resultado possível.

4 Distribuição de probabilidades

4.1 Variável aleatória

A "ligação" entre a teoria das probabilidades e a estatística aplicada é "materializada" pela noção de variável aleatória. Até agora, têm-se considerado genericamente espaços de acontecimentos. Porém, certas experiências (ou provas) aleatórias podem dar origem a resultados numéricos (por exemplo, número de reprovações na disciplina por ano lectivo, duração do efeito de um desinfectante, volume duma embalagem, etc.). Noutros casos, substituem-se os resultados não-quantitativos por números para simplificar ou facilitar a análise desses resultados. Recordem-se os conceitos de variável contínua e discreta, e de amostragem aleatória, como método necessário para a análise estatísticamente válida dos assuntos. Assim sendo, é importante definir o que é uma variável aleatória e as suas principais características (ou propriedades): X é uma VARIÁVEL ALEATÓRIA quando o seu valor (numérico) é determinado pelo acontecimento possível de uma prova aleatória. A definição que se encontra nos manuais de estatística é, mais ou menos, a seguinte: sejam ϵ (lê-se "épsilon") uma prova aleatória (ou experiência) e S um espaço amostral associado a essa prova aleatória; uma função X que associe a cada acontecimento possível (elemento) a_i desse espaço amostral um número real $X(a_i)$, ou mais simplesmente x_i, é denominada VARIÁVEL ALEATÓRIA (Fig. 3).

A terminologia usada é um tanto infeliz mas é "universalmente" aceite. Torna-se claro que X é uma função, contudo denominamo-la variável (aleatória)! É evidente que nem todas as funções imaginárias se podem considerar variáveis aleatórias. Um requisito importante é que as probabilidades dos acontecimentos e respectivos resultados (da função variável) sejam bem definidos e consistentes com os axiomas básicos (ver tópico anterior). Na maior parte das utilizações não se indica a natureza funcional da variável aleatória X (neste texto, usaremos v.a. para referir uma variável aleatória). Geralmente, interessam mais os valores possíveis da v.a. X do que "a sua origem".

> Exemplo 1: Quando se lançam simultaneamente dois dados, existem 36 acontecimentos possíveis diferentes (ver Tab. 1). Se interessar "a soma dos pontos" podem obter-se 36 resultados x_i (v.a. X - "soma dos pontos"). Se, por outro lado, se pretender estudar o produto dos pontos, então para os mesmos acontecimentos possíveis obtêm-se outros resultados y_i (agora da v.a. Y - "produto dos pontos").
>
> Exemplo 2: Se se pretender estudar determinado sector de actividade, por exemplo a indústria conserveira, o espaço amostral é composto algumas das empresas do sector, seleccionadas com determinado critério (p. ex. amostragem aletatória), ou seja S={algumas das empresas conserveiras}. Para cada empresa (acontecimento possível) é possível estudar diferentes aspectos (qualidades ou características), por exemplo o nº de empregados (v.a. W), o volume de negócios (v.a. V), etc.

As probabilidades dos resultados possíveis duma v.a. X, i.e. $P(x_i)$, também se podem estudar. Para esse fim podem utilizar-se as distribuições (ou leis) de probabilidades.

4.2 Distribuições de probabilidades de variáveis discretas

No caso das VARIÁVEIS DISCRETAS, a variável aleatória X pode tomar valores x_i, com $i = 1, 2, ..., n$. À função que estabelece uma correspondência entre o resultado da variável aleatória $X(a_i)$, ou x_i, e a respectiva probabilidade denomina-se FUNÇÃO DENSIDADE DE PROBABILIDADE, e representa-se por $P(X = x_i)$ ou $P(x_i)$.

> Consideremos uma variável X que pode assumir os valores 0, 1 ou 2 (equiprováveis). Considere-se, aínda, o espaço amostral S={0,0,0,0,0,0,0,0,1,1,1,2}. A função densidade da probabilidade $P(X = x_i)$ faz corresponder a cada valor de x_i (resultado do acontecimento possível) uma probabilidade $P(x_i)$. Neste caso, $P(X = 0) = 8/12 = 0,6667$, $P(X = 1) = 3/12 = 0,2500$ e $P(X = 2) = 1/12 = 0,0833$.

A função densidade de probabilidade (de variáveis discretas) é semelhante às frequências relativas. Poderemos representar graficamente esta função através dum histograma de frequências relativas (Fig. 4a).

Contudo, poderemos estudar outras questões relativas aos mesmos resultados, para além de saber qual é a probabilidade $P(X = x_i)$ para determinado x_i. A probabilidade da variável aleatória X tomar um valor inferior ou igual a x_i é uma nova função, que se representa por $P(X \leq x_i)$ e se designa por FUNÇÃO DISTRIBUIÇÃO DE PROBABILIDADE. Na prática, esta função é similar às frequências relativas acumuladas.

> Consideremos o exemplo anterior. A função distribuição da probabilidade $P(X \leq x_i)$ faz corresponder a cada valor de x_i (resultado possível) o somatório das probabilidades para os casos em que $X \leq x_i$, ou seja, $P(X \leq 0) = 8/12 = 0,6667$, $P(X \leq 1) = 11/12 = 0,9167$ e $P(X \leq 2) = 12/12 = 1,000$.

Como anteriormente, podemos representar graficamente esta função, agora por um polígono (de frequências) (Fig. 4b).

Propriedades da função densidade

a) Para cada resultado de um acontecimento possível x_i, com $i = 1, 2, ..., n$, a probabilidade pode variar entre zero e um:

$$0 \leq P(X = x_i) \leq 1$$

b) O somatório de todas as probabilidades dos resultados dos acontecimentos é igual a um:

$$\sum P(X = x_i) = 1$$

Propriedades da função distribuição

a) $P(X \leq x_i)$, com $i = 1, 2, ..., n$, é sempre um valor entre zero e um:

$$0 \leq P(X \leq x_i) \leq 1$$

b) A função distribuição (das probabilidades) nunca decresce à medida que x_i aumenta, *i.e.* função monótona crescente.

Caraterísticas da distribuição de probabilidades É possível definir nas distribuições de probabilidade de determinada variável aleatória (v.a.), alguns "pontos" ou características com interesse estatístico. Estas características são similares a conceitos de estatística descritiva que se abordaram anteriormente para amostras (medidas de localização e de dispersão), nomeadamente o valor médio, a variância e o desvio-padrão duma distribuição de probabilidades.

O VALOR MÉDIO da distribuição de probabilidades é também designado por VALOR ESPERADO ou ESPERANÇA MATEMÁTICA, e representado por $E\{X\}$. No caso de uma v.a. discreta X, tal que os resultados dos acontecimentos possíveis x_i, com $i = 1, 2, ..., n$, o valor médio será:

$$E\{X\} = \sum [x_i \cdot P(X = x_i)]$$

O valor médio obtém-se a partir da importância relativa do resultado de cada acontecimento possível. Enquanto a média é empírica, i.e. experimental, por se obter dos valores observados, o valor médio é uma noção teórica visto ser calculado a partir da distribuição de probabilidades dos valores observáveis (e não dos resultados observados!!). Pode interpretar-se como o valor teórico (em geral desconhecido) de que as médias são medições bastante próximas,

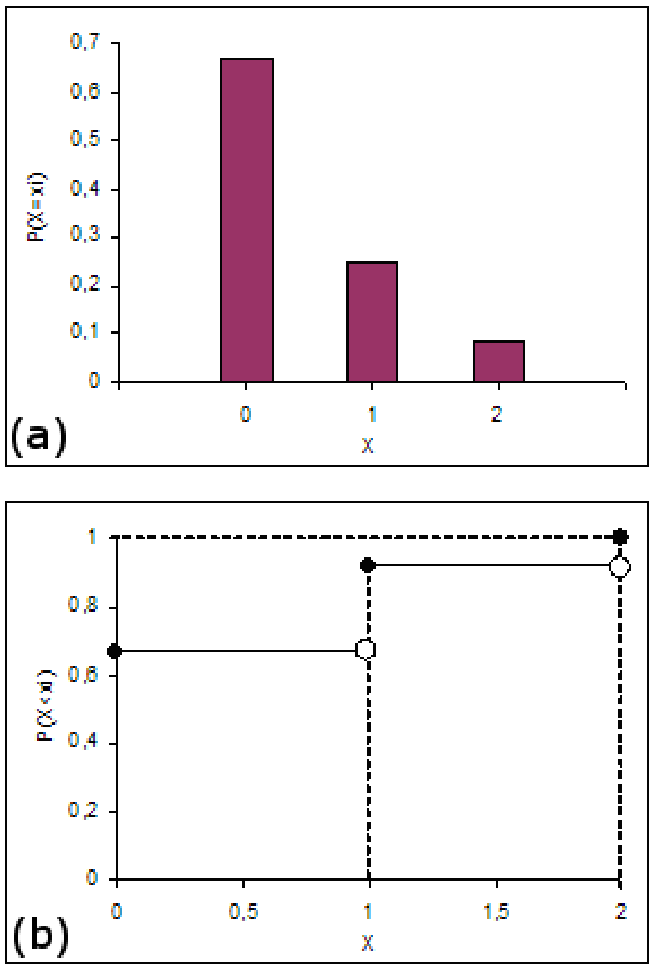

Figura 4: (a) Função densidade de probabilidade $P(X = x_i)$ e (b) função distribuição de probabilidade $P(X \leq x_i)$ duma v.a. discreta.

se o número de observações (ou provas aleatórias) for bastante grande. O exemplo seguinte mostra que o valor médio não é necessariamente um valor assumido pela v.a.

A variabilidade (ou a forma) da distribuição de probabilidades de uma v.a. discreta X pode ser quantificada (ou descrita) pela VARIÂNCIA, $Var\{X\}$ ou $V\{X\}$, que é dada para os resultados dos acontecimentos possíveis x_i por:

$$V\{X\} = \sum[(x_i - E\{X\})^2 \cdot P(X = x_i)]$$

A raíz quadrada positiva da variância $\sqrt{V\{X\}}$ é designada por DESVIO-PADRÃO da distribuição de probabilidades.

> Após o lançamento dum dado não-viciado, a variável aleatória X - número de pontos na face visível, pode tomar os valores $x_i = 1, 2, 3, 4, 5, 6$. As respectivas probabilidades são $P(X = 1) = P(X = 2) = ... = P(X = 6) = 1/6$. Então, teremos que o valor médio[6] será: $E\{X\} = 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) = 21/6 = 3,5$. Por outro lado, a variância é $V\{X\} = (1 - 3,5)^2(1/6) + (2 - 3,5)^2(1/6) + (3 - 3,5)^2(1/6) + (4 - 3,5)^2(1/6) + (5 - 3,5)^2(1/6) + (6 - 3,5)^2(1/6) = 2,92$. Então $\sqrt{V\{X\}} = \sqrt{2,92} = 1,71$ é o desvio-padrão.

Existem casos de distribuições (teóricas) de probabilidades de variáveis discretas que merecem atenção especial e estudo particular, nomeadamente as distribuições Binomial (Secção 3.3) e de Poisson (Secção 3.4) - estas distribuições constituem o "cenário probabilístico" de várias técnicas de controlo estatístico da qualidade por exemplo.

4.3 Distribuição Binomial

Considere-se uma prova (experiência) aleatória que tem apenas dois acontecimentos possíveis: um que se designa por "sucesso" (a_1) e o seu complementar designado por "insucesso" (a_0). Em cada p.a. a probabilidade de ocorrer a_1 é p e $q = (1-p)$ é a probabilidadade do "insucesso", ambas constantes - prova aleatória de Bernoulli. A distribuição binomial é o modelo probabilístico adequado para os casos em que se consideram repetidas p.a. independentes como a descrita. Nestes casos, o conjunto de resultados nas sucessivas provas constitui uma variável aleatória discreta que segue a distribuição binomial. De facto, as probabilidades de observar a v.a. X igual a $0, 1, 2, ..., n$ são dados por$(1 - p)^n$, $n(p)(1 - p)^{n-1}$, $n!/2!(n - 2)!(p^2)(1 - p)^{n-2}$, ..., p^n, em que p é a probabilidade de realização do acontecimento em cada prova. Aquelas quantidades correspondem ao desenvolvimento do binómio $[(1-p)+p]^n = 1$, daí a designação distribuição binomial.

Na DISTRIBUIÇÃO BINOMIAL, devida a Jakob Bernoulli [1654-1705], numa sequência de n provas aleatórias com reposição, a FUNÇÃO DENSIDADE DE PROBABILIDADE da v.a. discreta X - número de "sucessos" - que pode tomar os valores $x_i = 0, 1, 2, ..., n$ é dada por:

$$P(X = x_i) = C^n_{x_i} \cdot p^{x_i} \cdot (1 - p)^{n - x_i}$$

em que p é a probabilidade de ocorrer o resultado favorável (ou sucesso) em cada prova e se mantém constante de experiência para experiência, n é o número máximo de tentativas (ou provas aleatórias) independentes e C^n_x (designado por vezes como coeficiente binomial) é dado pela expressão de cálculo combinatório:

$$C^n_x = \frac{n!}{x!(n - x)!}$$

i.e. em n p.a. independentes existem C^n_x maneiras diferentes de se obterem x "sucessos"[7]. A função densidade de probabilidades da distribuição binomial depende dos parâmetros p e n, e é usada quando cada prova aleatória tem apenas dois acontecimentos possíveis e de natureza qualitativa: a_1 - sucesso ou 1; e a_0 - insucesso ou 0. Uma v.a. X que segue esta distribuição representa-se por $X \cap Binomial(n, p)$ ou $X \sim Binomial(n, p)$.

> Em 3 lançamentos de um dado não-viciado ($n = 3$ provas aleatórias repetidas), a variável aleatória X - "número faces com um ponto", pode tomar os valores $x_i = 1, 2, 3$ (Fig. 5). Mas, para cada resultado x_i existem vários casos (acontecimentos) possíveis! De facto, temos de considerar a ordem de saída das faces com um ponto. Vamos abordar a questão considerando que o resultado "saída da face com

[6] Curiosamente, podemos verificar que a designação valor esperado não faz muito sentido uma vez que 3,5 não pode realmente ocorrer neste exemplo! Talvez seja mais adequado utilizar a designação valor médio! Se se imaginar a situação de lançamentos sucessivos dum dado, em cada lançamento o resultado pode ser 1, 2, etc. Se se anotar o resultado após cda lançamento, a média do conjunto dos resultados possíveis após muitas (infinitas) p.a. será o valor médio.

[7] Também se pode "ler" como o nº de combinações de ordem x de um conjunto de cardinal igual a n (com $x \leq n$).

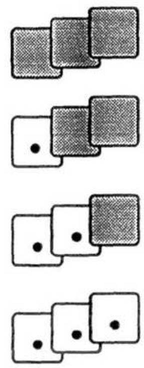

Figura 5: Representação esquemática dos possíveis resultados favoráveis ("face com um ponto") no lançamento simultâneo de três dados não-viciados. No entanto, nos casos em que $x_i = 1$ e $x_i = 2$ não estão representados todas as situações possíveis, pois não?

Tabela 3: Casos possíveis no exemplo da página anterior (*cf.* Fig. 5). Probabilidades calculadas por aproximação e recorrendo à função densidade de probabilidades da distribuição binomial $P(X = x_i)$.

x_i	Casos (acontecimentos) possíveis	Por aproximação	$P(X = x_i)$
0	000	1(5/6)(5/6)(5/6)	0,5787
1	100 ou 010 ou 001	3(1/6)(5/6)(5/6)	0,3472
2	110 ou 101 ou 011	3(1/6)(1/6)(5/6)	0,0694
3	111	1(1/6)(1/6)(1/6)	0,0046

um ponto" consitui um sucesso e que todos os outros resultados possíveis constituem um insucesso. Adicionalmente, vamos quantificar os sucessos com 1 e os insucessos com 0. Assim, para cada valor de x_i (0, 1, 2 ou 3 faces com um ponto) o número total de casos possíveis será 1, 3, 3 e 1, respectivamente (Fig. 5 e Tab. 3).

Em cada uma das provas aleatórias sucessivas do exemplo anterior, a probabilidade de sair uma face com um ponto (um "sucesso") é igual a $P(X = 1) = 1/6$, sendo que a $P(X \neq 1) = 1 - 1/6 = 5/6$ (T. complementaridade). Então a probabilidade de se realizar o acontecimento "101", por exemplo, é $p(q)p = 1/6(5/6)1/6 = 5/216$ (T. multiplicação). Contudo, nem todas situações permitem esquematizar e contabilizar todos os casos possíveis! Podemos utilizar técnicas matemáticas para calcular aquelas quantidades - combinações. Ou seja, para um dado valor X o número de combinações possíveis em n provas aleatórias sucessivas é dado por C_x^n (também designado coeficiente binomial). Assim, podemos calcular $P(X = x_i)$ para cada um dos resultados possíveis através da função densidade de probabilidades binomial (*cf.* Tab. 3).

A FUNÇÃO DISTRIBUIÇÃO DE PROBABILIDADES $P(X \leq x_i)$ duma v.a. com distribuição binomial obtém-se somando as probabilidades de cada um dos resultados em que $X \leq x_i$. Existem tabelas com as probabilidades para determinados valores de n e x que facilitam a resolução de problemas envolvendo esta distribuição teórica. As máquinas de calcular mais recentes e as folhas-de-cálculo (e.g. Microsoft©Excel ou OpenOffice Calc) permitem calcular as probabilidades com facilidade.

Na distribuição binomial, também podemos calcular o valor médio e a variância da distribuição de probabilidades para descrever teoricamente a sua "localização" e "forma". O VALOR MÉDIO da distribuição binomial é dado por:

$$E\{X\} = n \cdot p$$

em que p é a probabilidade de ocorrer o resultado favorável ("sucesso") em cada prova aleatória e n é o número provas aleatórias (independentes) realizadas (ou a realizar). É possível (com)provar a validade desta formulação recorrendo ao exemplo anterior. A VARIÂNCIA da distribuição binomial pode calcular-se através de:

$$V\{X\} = n \cdot p \cdot (1 - p)$$

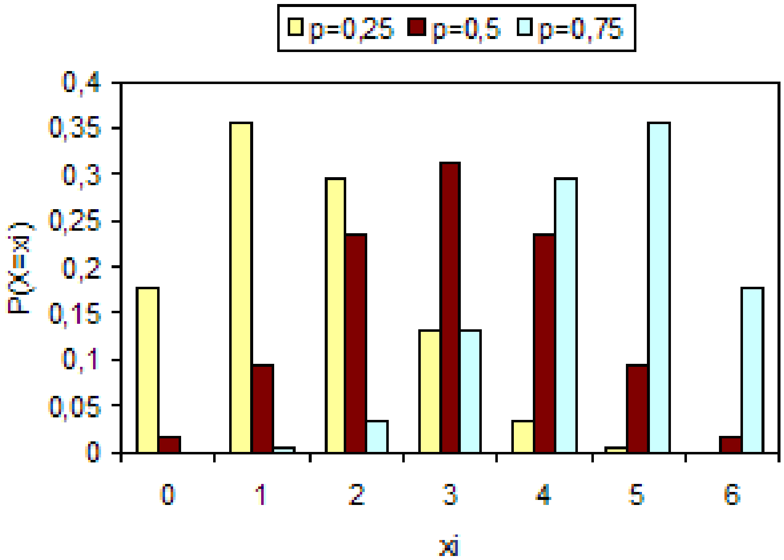

Figura 6: Modificação da forma da função de densidade de probabilidades da distribuição Binomial com a alteração de p. Variável aleatória X com resultados possíveis $x_i = 1, 2, ..., 6$.

Como se provou para $E\{X\}$, também é possível (com)provar a aplicação desta formulação para calcular a variância da distribuição binomial. Tente demonstrar esta afirmação! A distribuição binomial é desviada para a esquerda quando $p < 0,5$, é simétrica quando $p = 0,5$ e é desviada para a direita quando $p > 0,5$ (Fig. 6).

> Considere-se que o valor médio da distribuição de probabilidades duma v.a. discreta é dado por $E\{X\} = \sum[x_i \cdot P(X = x_i)]$, logo para o caso descrito no exemplo anterior, $E\{X\} = 0(0,5787) + 1(0,3472) + 2(0,0694) + 3(0,0046) = 0,5$, o mesmo que $E\{X\} = n \cdot p = 3(1/6) = 0,5$. Verifique que, para este exemplo, $V\{X\} = 0,4167$.

4.4 Distribuição de Poisson

A distribuição de Poisson (devida a Simon Poisson [1781-1840]) pode entender-se como um caso particular da distribuição binomial e aplica-se nas situações em que a probabilidade p de ocorrer determinado evento é muito pequena ou quando n é bastante grande (em estatística $n > 30$ é considerado como "grande"!) ou seja, quando estamos a estudar acontecimentos "raros" (vários autores sugerem que, se $n \cdot p < 5$, será adequado usar esta distribuição em vez da binomial). A designação "distribuição dos acontecimentos raros", utilizada por alguns autores, advém das primeiras aplicações, do princípio do séc. XX e devidas a von Bortkiewicz. Aquele matemático utilizou a distribuição de Poisson para descrever o número anual de mortos por coice de cavalo nos regimentos Prussianos de cavalaria.

A FUNÇÃO DENSIDADE DE PROBABILIDADE da distribuição de Poisson é dada por:

$$P(X = x_i) = \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda}$$

em que (λ lê-se "lambda"), $\lambda > 0$ e $x_i = 0, 1, ..., n$ (Fig. 7). Alguns autores referem-se a λ como valor médio da v.a. X, que é representada por $X \cap Poisson(\lambda)$. Como no caso da distribuição binomial, a função distribuição de probabilidades $P(X \leq x_i)$ obtém-se por adição das probabilidades dos resultados de $X \leq x_i$.

O VALOR MÉDIO e a VARIÂNCIA da distribuição de Poisson têm valor igual, ou seja,

$$E\{X\} = \lambda = n \cdot p = V\{X\}$$

Note-se que o valor médio e a variância são iguais. Esta é uma propriedade muito interessante da distribuição de Poisson, se se atender às condições iniciais, $n \rightarrow \infty$ e $p \rightarrow 0$. Para um dado valor de λ, é possível consultar tabelas

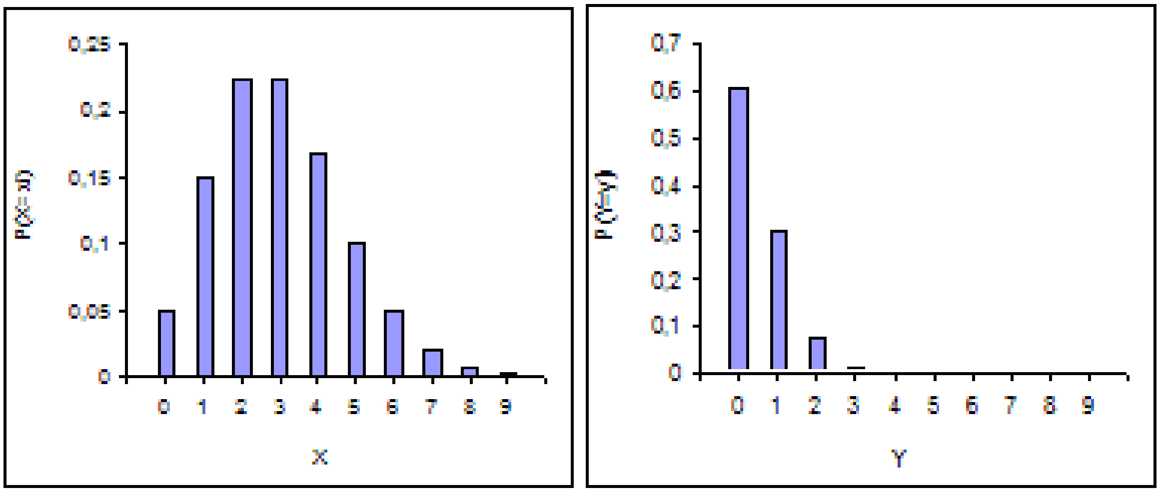

Figura 7: Funções densidade de probabilidade de v.a. X e Y com distribuição de Poisson com (a) $\lambda = 3$ (ver exemplo anterior) ou (b) $\lambda = 0,5$, em que $x_i = 0, 1, 2, ..., 9$ e $y_i = 0, 1, 2, ..., 9$. Observe-se a assimetria em ambas as distribuições de probabilidades.

de probabilidades para a função densidade de probabilidades da distribuição de Poisson (tabelas com x linhas e λ colunas) e obter a probabilidade pretendida. Como no caso da distribuição binomial (e outras...) as calculadoras científicas ou as folhas-de-cálculo de uso comum permitem determinar as probabilidades com facilidade. O exemplo seguinte facilita a compreensão e sugere a utilidade desta distribuição teórica.

> Uma fábrica de conservas produz, continua e cadenciadamente, cerca de 2330 latas de sardinha em molho de tomate por periodo de 8 horas de laboração e em média cerca de 7 latas são defeituosas. Qual a probabilidade de encontrarmos 3 latas defeituosas num lote de $n = 1000$ latas adquiridas àquela fábrica? Pode-se obter $p = 7/2330 = 0,0030$ e $\lambda = n \cdot p = 1000(0,003) = 3$. Portanto a probabilidade de três latas defeituosas num lote de 1000 latas é $P(X = 3) = \lambda^x/x! \cdot e^{-\lambda} = 3^3/3! \cdot e^{-3} = 0,2240$. Para calcular a probabilidade de ocorrerem até duas latas defeituosas nesse lote de 1000 latas, determina-se $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$.

4.5 Distribuições de probabilidades de variáveis contínuas

Em muitos problemas, torna-se necessário ou matematicamente mais simples considerar um espaço amostral para uma variável aleatória X, no qual todos os números reais possíveis (num intervalo especificado ou conjunto de intervalos) possam ser considerados como resultados possíveis. Daí ser necessário utilizar variáveis aleatórias contínuas.

Em contraste com as v.a. discretas, diz-se que uma v.a. X é contínua quando: (i) o seu valor numérico x_i é determinado pelo resultado de uma prova aleatória; e (ii) $x_i \in \mathbb{R}$ (conjunto dos números reais), ou seja, x_i pode tomar qualquer um dos infinitos (ou não-enumeráveis) valores num certo intervalo em $\mathbb{R}$.

> Exemplo 1. O consumo anual de energia eléctrica para fins industriais, numa determinada região (em 10^9 kW), v.a. W, é uma v.a. contínua.
>
> Exemplo 2. O tempo de prateleira (em dias) de determinado produto alimentar, v.a. T, também é uma v.a. contínua.

No caso das v.a. contínuas, a densidade de probabilidades está continuamente dispersa pelo espaço amostral S ao invés de se concentrar num conjunto discreto de resultados como acontece com as v.a. discretas (Fig. 8). Enquanto no caso de v.a. discretas a probabilidade do espaço amostral é "dividida" pelos resultados possíveis, nas v.a. contínuas aquela probabilidade "está amassada" e distribuida pelos (não-enumeráveis) resultados possíveis x_i. Por esse motivo, a probabilidade da v.a. contínua X tomar um valor particular x_i é nula. Contudo, um determinado acontecimento possível (ou resultado) x_i é quase-impossível mas não é impossível pois em cada realização da prova aleatória (ou experiência) obtém-se sempre um resultado e, por conseguinte, é possível obter x_i. Por outras palavras, dizer que "a probabilidade pontual é sempre nula" quer somente traduzir que é nula a probabilidade de "acertar" exactamente no resultado x_i. Logo, no caso das v.a. contínuas, as probabilidades estudam-se para intervalos de valores de X e não para valores "exactos" de X.

Se se dividir a probabilidade de X tomar um valor do intervalo $[x; x + \Delta x]$ pela amplitude desse intervalo Δx, obtém-se de forma aproximada a densidade de probabilidade da v.a. X tomar um valor qualquer do intervalo $[x; x + \Delta x]$, ou seja,

$$f(x) = \frac{P(x < X < x + \Delta x)}{\Delta x}$$

Rigorosamente, para um intervalo infinitesimal dx de valores de X, a função densidade de probabilidade (f.d.p.) dada por:

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x < X < x + \Delta x)}{\Delta x} = \frac{dP(x < X < x + dx)}{dx}$$

ou seja, a FUNÇÃO DENSIDADE DE PROBABILIDADE de uma v.a. contínua, designada por $f(x)$ em vez de $\Pr(X = x_i)$, é genericamente uma função matemática para a qual a àrea limitada pela curva $y = f(x)$, pelo eixo das abcissas e pelas rectas $Y = x$ e $Y = x + \Delta x$ é igual à probabilidade da v.a. contínua X assumir um valor do intervalo $[x; x + \Delta x]$ (Fig. 9). Podemos re-escrever a equação anterior na forma de:

$$dP(x < X < x + dx) = f(x)dx$$

A probabilidade pretendida (graficamente corresponde à área mencionada acima e representada na Fig. 9a, uma vez que se obtém de um integral definido de uma função não-negativa) pode ser obtida por integração[8] de $f(x)$:

$$P(x < X < x + \Delta x) = \int_x^{x+\Delta x} f(x)dx$$

Um modo de calcular a probabilidade de X tomar um valor do intervalo $[x; x + \Delta x]$ é resolver o integral por primitivação, isto é,

$$P(x < X < x + \Delta x) = \int_x^{x+\Delta x} f(x)dx = F(x + \Delta x) - F(x)$$

em que $F(x+\Delta x)$ e $F(x)$ constituem as soluções da primitiva de $f(x)$ para os limites do intervalo considerado (Fig. 9b).

Introduz-se, assim, a FUNÇÃO DISTRIBUIÇÃO DE PROBABILIDADES, que se designa $F(x)$ em vez de $\Pr(X \leq x_i)$, e que corresponde à probabilidade da v.a. contínua X tomar um valor igual ou inferior a x_i e, portanto, corresponde à área sob a curva *y=f(x)* à esquerda de x_i (Fig. 9a).

Genericamente,

$$F(x) = P(X \leq x_i) = \int_{-\infty}^{x} f(u)du$$

(note-se que o limite superior de integração é *x*, pelo que há que considerar uma variável de integração distinta de *x*, no caso a variável u).

Enquanto no caso discreto, a função distribuição de probabilidades era obtida somando $P(X = x_i)$, em domínios contínuos aquela função é calculada por integração da função densidade de probabilidades (Fig. 9).

> A f.d.p. duma v.a. contínua T, que representa o tempo de funcionamento sem avarias (expresso em dias) dum determinado equipamento, é dada por $f(t) = 0,5e^{-0,5t}$. Qual é a probabilidade desse equipamento funcionar sem avarias por um periodo de 1 a 3 dias? $P(1 < X < 3) = \int_1^3 f(t)dt = \int_1^3 0,5e^{-0,5t}dt = [-e^{-0,5}]_1^3 = e^{-0,5} - e^{-1.5} = 0,3834$. A função distribuição pode ser obtida directamente por integração de $f(x)$, ou seja: $F(t) = \int_0^t 0,5e^{-0,5t}du = [0,5e^{-0,5t}]_0^t = 1 - e^{-0,5t}$.

Neste contexto, é possível relacionar as funções densidade e distribuição de probabilidades de v.a. contínuas, $\frac{dF(x)}{dx} = f(x)$, ou então, $F'(x) = f(x)$ em que $F'(x)$ é a derivada de $F(x)$. Desta relação resultam algumas conclusões importantes:

1. $f(x) \geq 0$ (equação não-negativa) e $0 \leq F(x) \leq 1$ (função monótona crescente e contínua)

2. $\int_{-\infty}^{+\infty} f(x)dx = 1$ (a área ou probabilidade total é igual a 1)

[8] O integral $\int$ é um objecto matemático que se pode interpretar como uma área, ou uma generalização de uma área. Juntamente com as derivadas são os objectos fundamentais do Cálculo.

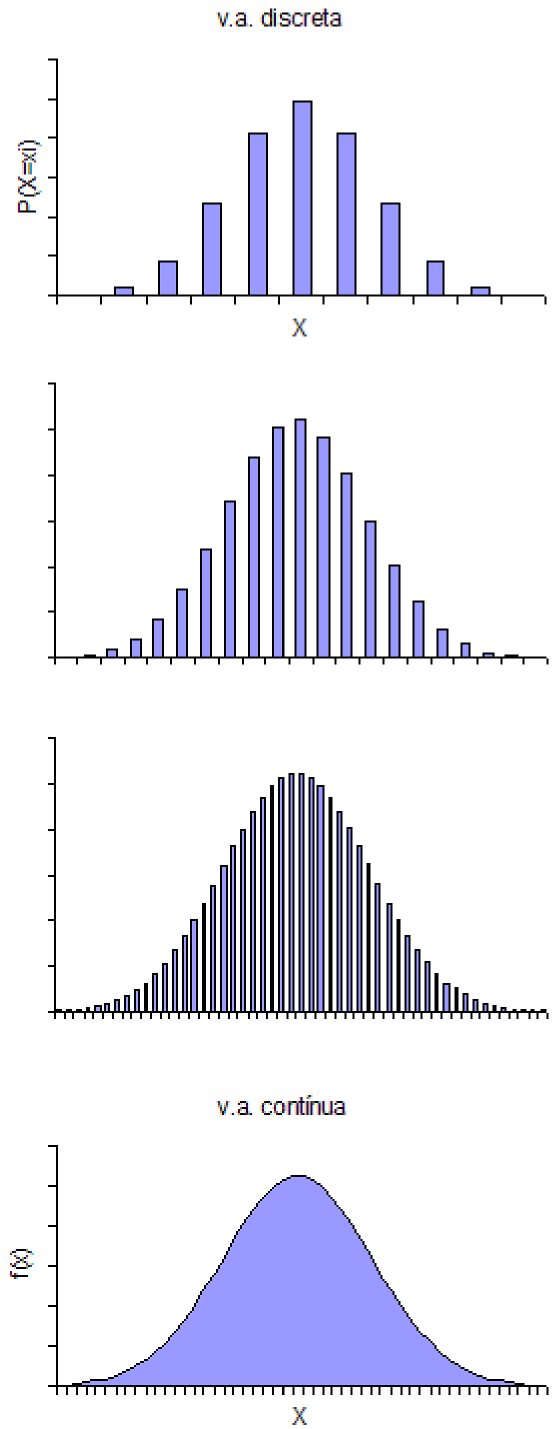

Figura 8: Ilustração da "relação" entre distribuições de probabilidades de v.a. discretas e contínuas (e.g. função densidade de probabilidades $f(x)$).

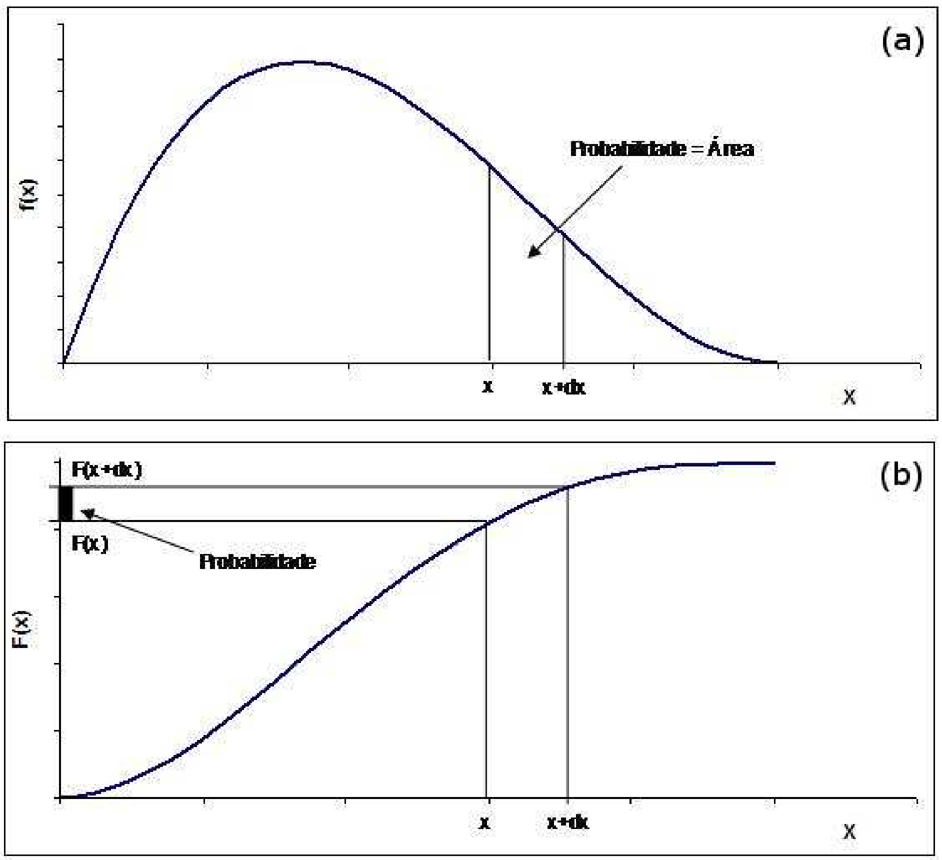

Figura 9: Representação gráfica: (a) duma função densidade de probabilidades $f(x)$, em que a área assinalada sob curva $y = 12x(1-x)^2$, acima dos eixo das abcissas e entre as semi-rectas verticais $Y = x$ e $Y = x + \Delta x$ corresponde à probabilidade da v.a. contínua X assumir um valor do intervalo; e (b) da respectiva função distribuição de probabilidades $F(x) = 6x^2 - 8x^3 + 3x^4$.

3. $\int_{x}^{x+\Delta x} f(x)dx = F(x+\Delta x) - F(x) = \Pr(x < X < x + \Delta x)$

4. $P(X = x) = \Pr(X = x + \Delta x) = 0$.

Como no caso das variáveis aleatórias discretas, é possível caracterizar resumidamente a distribuição de probabilidades recorrendo aos conceitos de valor médio e variância das probabilidades. Assim, para uma v.a. contínua X o VALOR MÉDIO é dado por

$$E\{X\} = \int_{-\infty}^{+\infty} x \cdot f(x)dx$$

e a VARIÂNCIA por

$$V\{X\} = \int_{-\infty}^{+\infty} (x - E\{X\})^2 \cdot f(x)dx$$

Do ponto de vista formal, a passagem do caso discreto para o caso contínuo (ou vice-versa) faz-se por "dualidade", substituindo-se os somatórios por integrais e as probabilidades $P(x)$ por densidades $f(x)$. Atente-se que no caso contínuo $P(X = x_i) = 0$ para todo o real x_i e, portanto, o que se calcula recorrendo a $f(x)$ é a área correspondente ao intervalo $[x, x + \Delta x]$ - a "probabilidade média" - nessa porção do *continuum*.

É importante, como se fez no caso das probabilidades de v.a. discretas, estudar algumas distribuições teóricas de probabilidades (para v.a. contínuas) de utilização muito generalizada.

4.6 Distribuição Normal

Entre as distribuições teóricas de probabilidades de v.a. contínuas destaca-se a DISTRIBUIÇÃO NORMAL, ou curva normal, ou curva de Gauss (em homenagem a Carl Friedrich Gauss [1777-1855] que foi pioneiro na sua utilização, apesar da distribuição se dever a Abraham de Moivre [1667-1754] que a desenvolveu em 1733 como aproximação à distribuição Binomial e a Pierre-Simon Laplace [1749-1827] que a formalizou).

Curiosamente, verifica-se que em muitas situações amostrais, a distribuição das variáveis aleatórias contínuas parece "concentrar-se" perto da média e "dispersar-se, diminuindo", em direcção aos extremos, de acordo com esta distribuição teórica. Por outro lado, a distribuição normal é de manipulação matemática fácil, o que tem contribuido para o número apreciável de testes estatísticos dela derivados. A curva normal - FUNÇÃO DENSIDADE DE PROBABILIDADES DA DISTRIBUIÇÃO NORMAL (Fig. 11) - é definida pela seguinte expressão:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)/2\sigma]^2}$$

com os parâmetros μ e σ e em que $-\infty < x < +\infty$,$-\infty < \mu < +\infty$, $\sigma > 0$, $\pi = 3,141659...$ e e a função exponencial. Para indicar que a v.a. contínua X tem distribuição normal usa-se $X \cap N(\mu, \sigma)$.

Dado que a função densidade de probabilidades $f(x)$ da distribuição normal tem dois parâmetros, média μ e desvio-padrão σ, cada par de valores de μ e σ origina uma curva com "forma diferente". Contudo, $f(x)$ da distribuição normal é sempre simétrica, em forma de sino, centrada em μ (que determina a posição da distribuição no eixo das abcissas que corresponde ao "universo" de valores de X), e dispersa relativamente a μ de acordo com o desvio-padrão σ (Fig. 11).

No caso da distribuição normal, também se pode resumir a informação acerca das probabilidades recorrendo a "medidas de localização e dispersão", já referidas anteriormente - VALOR MÉDIO EX e VARIÂNCIA VX da distribuição de probabilidades:

$$E\{X\} = \int_{-\infty}^{+\infty} x \cdot f(x)dx = \mu$$

$$V\{X\} = \int_{-\infty}^{+\infty} (x - E\{X\})^2 \cdot f(x)dx = \sigma$$

Existem algumas características com importância nas funções densidade e distribuição de probabilidades de variáveis normais, designadamente aquelas representados na Fig. 12. A média μ corresponde à mediana (quantil de 50%). Verifica-se, aínda, que os intervalos $\mu \pm \sigma$, $\mu \pm 2\sigma$ e $\mu \pm 3\sigma$ incluem, respectivamente, 68,27%, 95,45% e 99,73% das possíveis observações duma variável contínua normal. Por outras palavras, a probabilidade da v.a. contínua X tomar um valor desses intervalos é 0,6827, 0,9545 e 0,9973 respectivamente.

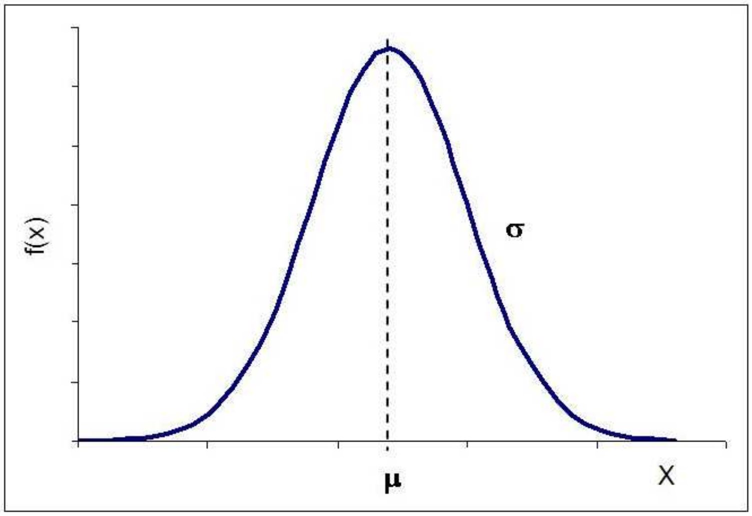

Figura 10: Representação esquemática da função densidade de probabilidade da distribuição normal $f(x)$, com parâmetros μ e σ, de uma varíavel aleatória contínua. A "curva" está centrada em μ e a sua forma (está relacionada com) depende de σ.

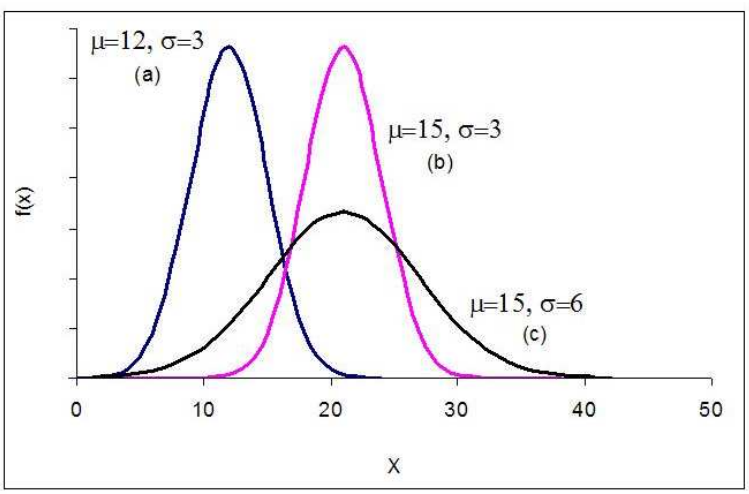

Figura 11: Comparação entre a forma de três curvas normais (funções densidade de probabilidades) com diferentes parâmetros μ e σ. As "curvas" (a) e (b) com média diferente mas com igual desvio-padrão. Pelo contrário, as "curvas (b) e (c) possuem igual média mas diferente desvio-padrão.

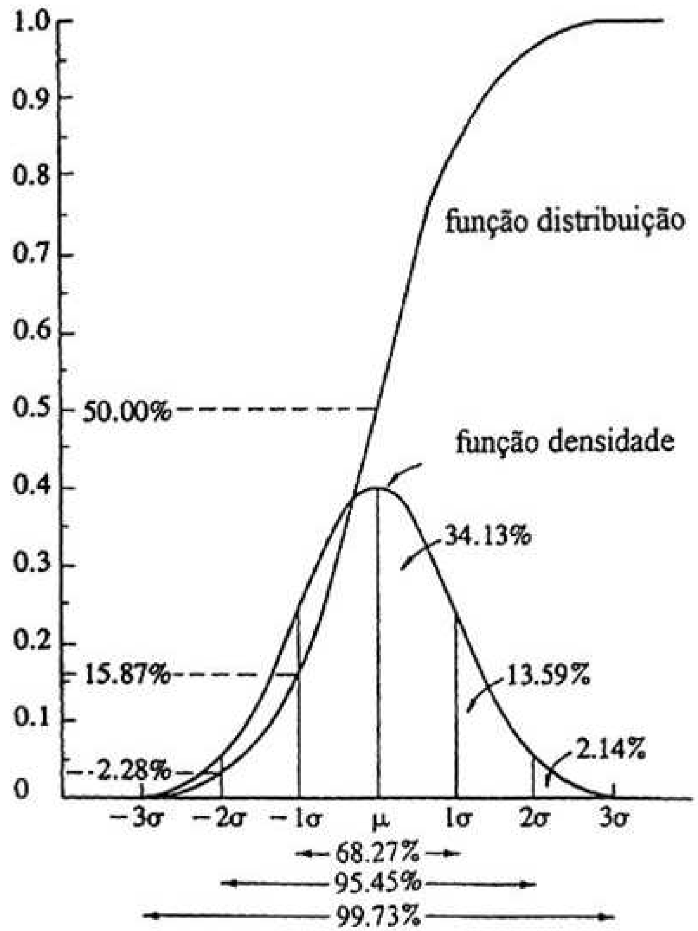

Figura 12: Pontos importantes das funções densidade e distribuição de probabilidades da distribuição normal (adaptado de Sokal & Rohlf). Por exemplo, cerca de 68% dos resultados possíveis dessa v.a. com distribuição normal incluem-se no intervalo $[\mu - 1\sigma; \mu + 1\sigma]$.

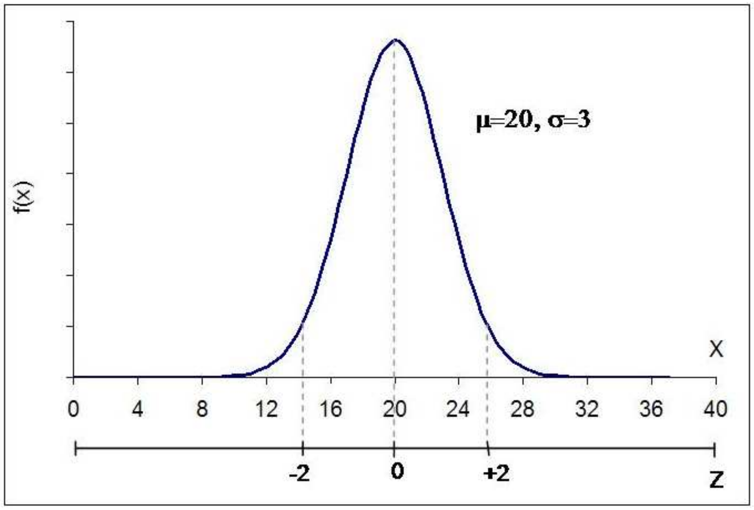

Figura 13: Ilustração da relação entre a distribuição dos resultados possíveis xi da v.a. contínua X e os valores de Z correspondentes. Tente determinar os valores x_i correspondentes a $z_i = -2$ e $z_i = -2$.

4.7 Distribuição Normal Reduzida[9] Z

O cálculo da área sob a curva normal para determinado intervalo da v.a. contínua, que se efectua por integração da função densidade para os valores de μ e de σ pretendidos, seria um trabalho "fastidioso" e, infelizmente, não é possível tabelar todas as combinações possíveis de μ e σ. Como resolver este "problema"?

Um modo é padronizar os resultados recorrendo à transformação da variável aleatória X numa nova variável Z com média igual a zero e desvio-padrão igual a um, ou seja, $Z \cap N(0, 1)$. Deste modo, se define a DISTRIBUIÇÃO NORMAL REDUZIDA Z transformando a v.a. $X \cap N(\mu, \sigma)$, da seguinte forma:

$$Z = \frac{X - \mu}{\sigma}$$

Esta transformação permite "reduzir" (ou sintetizar ou padronizar) qualquer distribuição normal desde que se conheçam μ e σ daquelas distribuições (Fig. 13). A f.d.p. é dada por:

$$f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

A tabulação da distribuição de probabilidades de Z, comum em qualquer manual de estatística, permite obter com facilidade $P(X \leq x_i)$ para qualquer v.a. $X \cap N(\mu, \sigma)$. Basta calcular z_i e, posteriormente, consultar a tabela de Z (que geralmente apresenta a função distribuição de probabilidades ou "probabilidades acumuladas") para o valor obtido (ver Tabela A). Por outro lado, a maioria das "calculadoras científicas" e do *software* (folhas de cálculo) permite obter com facilidade a probabilidade de $Z \leq z_i$. É possível demonstrar matematicamente que qualquer função linear de uma variável aleatória com distribuição normal é, também, uma v.a. com distribuição normal, isto é, $Z \cap N(0, 1)$[10].

> Por exemplo, a variação diária da temperatura de determinada câmara de refrigeração pode ser razoavelmente aproximada por uma distribuição normal com média de 0,2% e desvio-padrão de 1,6%. a) Qual a probabilidade da variação da temperatura ultrapassar 1%? b) E qual a probabilidade dessa variação se situar entre 1% e 1,4%?
>
> a) $P(X > 1\%) = P(Z > 0,5) = 1 - P(Z < 0,5) = 1 - (0,5 + 0,1915) = 0,3085$ porque, de acordo com a transformação de X em Z, obtém-se que $z_i = \frac{X-\mu}{\sigma} = \frac{1\% - 0,2\%}{1,6\%} = 0,5$ e da Tabela A pode-se obter $P(Z < z_i)$.

[9] A designação *"Standard Normal Curve"* foi usada pela primeira vez por W.F. Sheppard em 1899 mas só na década de 1950 se banalizou (Miller J. *Earliest Known Uses of Some of the Words of Mathematics* disponível em `http://members.aol.com/jeff570/mathword.html`, consultado em 7/10/2009).

[10] Observe-se com maior atenção a equação associada a esta transformação. Se re-arranjarmos os termos daquela expressão, teremos que: $Z \cdot \sigma + \mu = X$, ou seja, uma "equação da recta" do tipo $bx + a = y$.

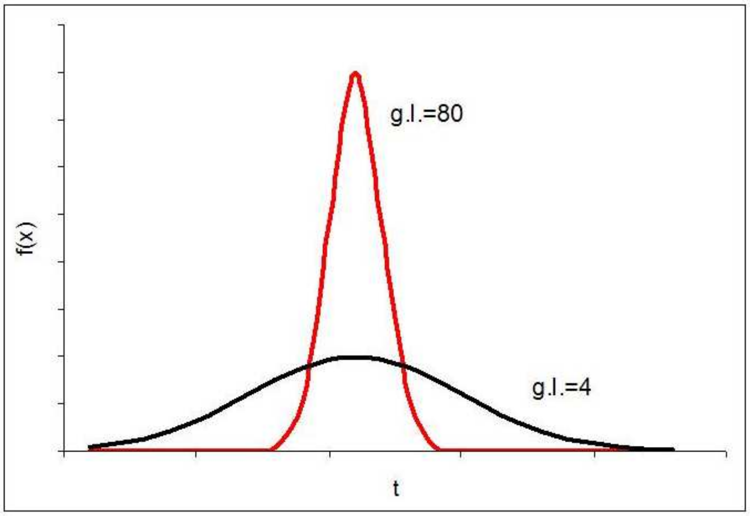

Figura 14: Representação de duas distribuições t de Student para diferentes graus de liberdade (g.l.).

b) $P(1\% < X < 1,4\%) = P(0,5 < Z < 0,75) = P(Z < 0,75) - P(Z < 0,5) = 0,2734 - 0,1915 = 0,0819$ uma vez que $z_i = \frac{1\% - 0,2\%}{1,6\%} = 0,5$ e $z_i = \frac{1,4\% - 0,2\%}{1,6\%} = 0,75$. Novamente, as respectivas probabilidades podem obter-se da Tabela A.

4.8 Distribuição t de Student

Até agora, consideraram-se como conhecidos os parâmetros da distribuição normal μ e σ, claramente relacionados com a população, ou universo estatístico. Na realidade, raramente se conhece μ ou σ (ou ambos) ou, então, não é possível recolher tantos dados que permitam assumir que as "estatísticas" da amostra (média $\bar{x}$ e desvio-padrão s) sejam estimadores correctos dos parâmetros da população, pois geralmente o tamanho da amostra é muito reduzido comparativamente à dimensão da população em estudo. Amostras de tamanho $n < 30$ podem considerar-se como "pequenas amostras".

O desenvolvimento da DISTRIBUIÇÃO t DE STUDENT, pelo inglês William Gosset [1876-1937] em 1908 (que por motivos profissionais/legais utilizou o pseudónimo *Student* para publicar os seus trabalhos), como alternativa a Z, constituiu um dos maiores avanços nas metodologias estatísticas. Aquele autor, propôs a transformação da v.a. contínua X na variável t da seguinte forma:

$$t = \frac{X - \bar{x}}{s}$$

em que $s = \sqrt{\sum(x_i - \bar{x})^2/(n-1)}$ é o desvio-padrão da amostra.

A distribuição de t depende dum único parâmetro, o número de GRAUS DE LIBERDADE ν (que se lê "niú"), com $\nu = n - 1$. Como se prova para Z, também a variável aleatória t se distribui "normalmente", ou seja, $t \cap N(\nu)$. A cada valor de ν, corresponde uma curva diferente dentro da família das distribuições t de Student (Fig. 14).

O procedimento para se obter t e consultar a respectiva tabela de probabilidades é idêntico ao descrito para Z, considerando-se, neste caso, $\nu = n - 1$ graus de liberdade da amostra (ver Tabela B).

Exemplo 1. A v.a. V segue distribuição de t com 7 g.l. a) Determine o valor v_0, tal que $P(V > v_0) = 1\%$; b) Qual a $P(-0,711 < V < 2,998)$?

a) v_0=2,998 (obtém-se directamente da tabela para p = 0,99); b) Uma vez que $P(V < 2,998) = 0,99$ e $P(V < -0,711) = P(V > 0,711) = 0,25$, então a probabilidade pretendida é $P = 0,99 - 0,25 = 0,74$.

Exemplo 2. O tempo (em minutos) que um grupo de operários leva a executar determinada tarefa, v.a. X, tem distribuição normal. Numa semana de trabalho seleccionada aleatoriamente, realizaram-se 12 medições daquela variável e o tempo médio foi de 107 min com um desvio-padrão de 23 min. Quanto tempo, no máximo, levam 90% dos operários a executar a dita tarefa ?

Uma vez que $P(t < t_0) = 90\%$ logo da tabela vem que $t_0 = 1,363$. Como $t = (X - \bar{x})/s \Leftrightarrow 1,363 = (x_i - 107)/23 \Leftrightarrow x_i = 138,4$ min.

5 Exercícios

1. Determine a probabilidade de cada um dos seguintes eventos: a) surgir um número ímpar num único lance de um dado honesto; b) ocorrer, pelo menos, uma "cara" em dois lançamentos de uma moeda honesta; c) aparecer o total 7 num único lançamento de dois dados; d) surgir um total diferente de 2 ou 6 ou 10 num único lançamento de dois dados; e) aparecer o total 11 num único lançamento de dois dados, em que um deles está viciado em favor do “seis” $[P(1) = P(2) = P(3) = P(4) = P(5) = 0,16$ e $P(6) = 0,20$ em vez de $P(i) = 0,166(6)]$.

2. Considere uma população em que a variável Y tem a distribuição de probabilidades descrita na tabela seguinte. a) Represente graficamente as funções densidade e distribuição de probabilidades. b) Calcule $E\{X\}$, $V\{Y\}$ e $\sqrt{V\{X\}}$. c) Indique todas as amostras possíveis, seleccionadas com reposição, de tamanho $n = 2$. d) Calcule a média de cada amostra e a probabilidade de ocorrer cada um desses valores.

Tabela 4: Distribuição de probabilidades da v.a. Y.

y_i	0	1	2	3	4
$P(Y = y_i)$	0,15	0,25	0,30	0,20	0,10

3. Duas máquinas A e B funcionam de forma independente uma da outra. No quadro seguinte, indicam-se as probabilidades de se verificar o número referido de avarias para cada máquina, no decurso de um dia de trabalho. Calcule: a) a probabilidade do nº de avarias em A ser superior a dois; b) a probabilidade do nº total de avarias num dia de trabalho ser inferior a três; e c) o nº médio de avarias em cada uma das máquinas.

Tabela 5: Probabilidades do nº de avarias registado em duas máquinas A e B.

Nº de avarias	0	1	2	3	4	5	6
Máquina A	0,10	0,20	0,30	0,30	0,09	0,07	0,04
Máquina B	0,30	0,32	0,10	0,08	0,10	0,05	0,05

4. Uma empresa comercializa garrafas de vinho de 1 litro. Supõe-se, no entanto, que 40% dessas garrafas contém realmente uma menor quantidade de líquido do que o volume indicado no rótulo. Tendo adquirido 6 dessas garrafas, qual a probabilidade de: a) Duas delas conterem menos de um litro? b) No máximo 2 conterem menos de um litro? c) Pelo menos 2 conterem menos de um litro? d) Todas conterem menos de um litro? e) Todas conterem o volume indicado no rótulo? f) Represente a função densidade de probabilidade da variável em questão.

5. Uma fábrica de embalagens, utilizadas para determinado produto alimentar, sabe que em cada 1000 produz 20 defeituosas. a) Qual é a probabilidade de um cliente ao comprar 100 embalagens receber todas sem defeito? b) Qual a probabilidade de receber, nessa mesma compra, pelo menos 3 embalagens defeituosas?

6. Calcule, com a ajuda da Tabela A (em anexo), a probabilidade de Z ser: a) Menor do que +1; b) Menor do que 0; c) Menor do que +1,96; d) $P\{1 < Z < 2\}$; e) $P\{Z > +0,84\}$; f) $P\{Z < -0,84\}$.

7. Considere que os diâmetros de ameixas seguem uma distribuição normal de média $\mu = 5$ cm e variância $\sigma^2 = 5,25$. Calcule: a) A probabilidade das ameixas terem diâmetros compreendidos entre 3,5 cm e 8 cm. b) A probabilidade das ameixas terem diâmetro maior do que 9,5 cm. c) A probabilidade das ameixas terem diâmetro menor do que 3,5 cm. d) O diâmetro abaixo do qual se encontram 95% das ameixas. e) O diâmetro acima do qual se situa metade da composição de tamanhos das ameixas. f) O diâmetro que corresponde ao quantil de ordem 28%.

8. Uma fábrica produz equipamento cujo tempo-de-vida é uma v.a. com distribuição normal de parâmetros $\mu = 10$ anos e $\sigma^2 = 2$ anos. A fábrica quer estabelcer o periodo de garantia de forma a que não mais de 3% dos equipamentos tenham de ser substituidos. Qual deverá ser o periodo de garantia máximo oferecido pela empresa?

9. Admite-se que o tempo de espera num determinado consultório (v.a. X) se distribui normalmente. Num dia, seleccionado aleatoriamente, registaram-se os tempos de espera de onze utentes, calculando-se um tempo

médio de $\bar{x} = 41$ min e $\sum(x_i - \bar{x})^2 = 1690$. a) No máximo, quanto tempo esperará 90% dos utentes daquele consultório? b) E qual é o tempo mínimo de espera para 95% dos utentes?