Textos Auxiliares
Probabilidades e Estatística
Eng$^{\underline{a}}$ Informática

Cecília Maria Fernandes Fonseca
UTC de Ciências Exactas e Experimentais - E.S.T.G

9 de dezembro de 2020

Nota Introdutória

Este documento trata os conteúdos programáticos da unidade curricular de Probabilidades e Estatística do curso de Engenharia Informática (ano lectivo 2020/2021), sendo de referir que a bibliografia consultado incluiu [1],[2], [3],[4],[5],[6],[7],[8],[9],[10].

Capítulo 1

Introdução à Teoria das Probabilidades

1.1 A Origem

Poderíamos fazer um percurso pelos últimos 4 séculos o que nos levaria a percorrer o nome de todos os que de alguma forma contribuíram para que se atingi-se o estado actual. Contudo apenas iremos referir, de forma breve, a origem da teoria das probabilidades.

Apesar de Gerolmo Cardano (1501-1576) ter apresentado um texto ("De ludo alea") onde tanto faz considerações morais sobre jogos de dados, discutindo os modos de ganhar ou a disposição dos jogadores, como calcula (por vezes de forma incorrecta) a probabilidade de ganhar um certo jogo, a origem real do estudo das probabilidades ocorrerá apenas no século XVII. Assim, considera-se que a origem real dos métodos probabilistas está na troca de correspondência entre Pascal e Fermat no ano de 1654 na tentativa de responder ao seguinte problema posto pelo cavaleiro De Méré sobre a divisão de apostas: "Dois jogadores (digamos A e B) apostam cada um 32 moedas num jogo de cara ou coroa; ficará com as 64 moedas aquele que, dos dois, conseguir obter em primeiro lugar 3 sucessos, consecutivos ou não. Fazem uma primeira jogada que A ganha; se nesse momento tiverem que se separar, não podendo terminar o jogo, como deverão dividir de modo equitativo o total apostado? Em termos gerais, como dar resposta à mesma questão, se o jogo for interrompido num qualquer momento antes do seu terminus?"A 29 de Julho de 1654, Pascal, doente, obrigado a permanecer na cama, estava tão entusiasmado com o método simples que tinha acabado de descobrir para dar resposta ao referido problema, que não esperou por se restabelecer e escreve de imediato a Fermat relatando a sua descoberta. Refira-se que esta carta vem na sequência de uma que Pascal recebeu de Fermat a 28 de Julho, na qual é proposta uma solução para o problema baseada no cálculo combinatório. A solução apresentada por Pascal consiste num método de divisão das apostas considerando a esperança de ganhar que cada um dos jogadores tem no momento da interrupção do jogo. Seguem-se algumas das passagens mais relevantes da carta:

"[...]O vosso método está certíssimo e foi aquele que primeiro me ocorreu ; mas uma vez que a dureza do cálculo das combinações é excessiva encontrei um outro mais curto e mais claro que gostaria de vos descrever aqui em poucas palavras [...]Veja-se, de modo aproximado, como faço para saber o valor em jogo de cada uma das partes, quando dois jogadores jogam, por exemplo, para três sucessos e cada um pôs em jogo 32 moedas: Suponhamos que o primeiro tem duas partidas ganhas e o outro uma; jogam de seguida uma partida, cujo resultado é tal que, se o primeiro ganha, ganha todo o dinheiro que está em jogo, a saber 64 moedas; se o outro ganha, ficam empatados, e por consequência, se pretenderem separar-se, devem retirar cada um o seu valor de entrada, a saber, cada um 32 moedas. Então se eles se quiserem separar sem jogar esta partida, o primeiro deve dizer : "Estou certo de ter 32 moedas, pois mesmo perdendo as consigo; mas para as 32 restantes, talvez eu as consiga ou talvez vós as consigais, a sorte é igual; dividamos então as 32 moedas em partes iguais e, além disso, dê-me as 32 que tenho certas."Ele terá então 48 moedas e o outro 16. Suponhamos agora que o primeiro tem duas partidas ganhas e o outro nenhuma, e que começam a jogar uma nova partida. O resultado é tal que, se o primeiro ganha, ganha as 64 moedas; se o outro ganha voltam ao caso precedente no qual o primeiro terá duas partidas e o outro uma. Ora, nós já mostramos que neste caso aquele que ganha duas partidas deve receber 48 moedas. Assim, se não quiserem jogar nova partida, dirá então (o que tem duas partidas ganhas): «Se eu ganhar, ganharei tudo, que é 64 moedas; se perder, pertencem-me legitimamente 48; então dê-me as 48 moedas que tenho certas mesmo no caso de perder, e dividamos, em partes iguais as restantes 16, pois temos iguais hipóteses de ganhar». Assim ele terá 48 mais 8, isto é 56 moedas. Suponhamos finalmente que o primeiro tem só uma partida e o outro nenhuma [...]

Este episódio encontra-se de forma detalhada em Jacquard (1974).
Em resumo, a teoria da probabilidade, no sentido moderno do termo, pode considerar-se fundada por Pascal (1623-1662) e Fermat (1608-1665), seguidos de perto por Huygens (1629-1695), na sequência de problemas sobre jogos propostos pelo cavaleiro De Méré, jogador inveterado da corte de Luís XIV para além de filósofo e homem de letras.

O problema acima apresentado teve uma solução, devida à troca de correspondência entre Pascal e Fermat, que constitui um método geral de análise de probabilidades (Jacquard (1986)). Pascal afirmou que «a aliança do rigor geométrico com a incerteza do acaso daria origem a uma nova ciência». Assim, esta nova ciência surge, sobretudo, com o objectivo de dar uma explicação racional a certos fenómenos observados em jogos, de dados e cartas, genericamente designados por jogos de azar e sorte. Estes fenómenos são o que hoje chamamos de "fenómenos de massa"ou "leis estatísticas empíricas", ou seja, propriedades que surgem quando jogamos um número suficientemente grande de partidas.

É ainda no século XVII que surge formalmente a Teoria das Probabilidades, estando ligado a nomes como Huygens (escreveu o primeiro livro sobre cálculo de probabilidades), Leibniz (1646-1716)(faz aplicações das probabilidades a questões financeiras) e Jacques Bernoulli (estudou tão intensamente o assunto que contribuiu para que o cálculo das probabilidades tivesse, finalmente, o estatuto de ciência).

1.2 Conceitos Fundamentais

Existem experiências cujos possíveis resultados são equiprováveis, de que é exemplo o lançamento de uma moeda equilibrada ao ar, onde temos dois possíveis resultados (*cara*,*coroa*) e a probabilidade de sair face *cara* é igual à probabilidade de sair face *coroa*. Contudo, existem experiências em que os resultados possíveis não são equiprováveis, sendo exemplo disto a probabilidade de uma máquina de café avariar ou não no decurso de uma utilização, neste caso temos dois resultados (a máquina *avaria*, a máquina *não avaria*), mas cada um deles terá diferente probabilidade de ocorrência.

A teoria das probabilidades, genericamente, pretende formular modelos de fenómenos [1] naturais em que se supõe intervir o acaso, ou seja, cujo futuro não pode ser previsto, com exactidão, a partir do seu passado. Tais fenómenos dizem-se **fenómenos aleatórios**. Para estes, mesmo que se observe os resultados obtidos num elevado número de experiências do mesmo fenómeno não se consegue prever deterministicamente o resultado de uma experiência futura. Podemos dar como exemplo de fenómeno aleatório o lançamento de uma moeda ao ar, a duração da vida de uma lâmpada, o resultado obtido por um aluno num determinado exame e o resultado do lançamento de um dado.

Contudo, na natureza existe outro tipo de **fenómenos** denominados de **determinísticos**, para os quais os resultados obtidos em qualquer experiência são sempre os mesmos, sendo por isso previsíveis. Se considerarmos

[1] Tudo o que impressiona os nossos sentidos ou a nossa consciência; tudo que é objecto de uma possível experiência(Kant)

a experiência de aquecer um determinado recipiente com água no estado liquido, sabemos que a determinada temperatura ocorrerá a passagem ao estado gasoso (100ºC); sendo este um exemplo de fenómeno determinístico. Conhece-se o valor da velocidade de propagação do som (340 m/s) sem realizar a experiência.

Sendo os fenómenos aleatórios os que nos interessam, vamos debruçar-nos um pouco mais sobre eles.

Os modelos que correspondem aos fenómenos aleatórios dizem-se **modelos aleatórios** e as experiências, cujos resultados são imprevisíveis, dizem-se experiências aleatórias. Estas, podem ser realizadas uma infinidade de vezes e para cada uma é impossível prever com precisão o resultado. Assim, pretende-se observar o comportamento do fenómeno aleatório em sucessivas experiências com o objectivo de analisar os possíveis resultados e suas correspondentes frequências.

As características da experiência aleatória resumem-se a (Murteira, 1979):

- Possibilidade de repetição da experiência em condições similares;
- Imprevisibilidade, ou seja, não se poder prever um resultado exacto antes da sua ocorrência;
- Apresentar uma certa regularidade estatística após um elevado número de repetições da experiência.

É a última característica que permite chegar e a um conjunto de modelos probabilísticos que pretendem explicar os fenómenos aleatórios e dar indicações sobre as probabilidades de ocorrência.

Exemplo: Realizando um elevado número de lançamentos com uma moeda equilibrada, verifica-se que a frequência relativa de ocorrência de uma qualquer das faces tende a estabilizar em torno do valor $1/2$.

1.2.1 Espaço de Resultados

Apesar de não ser possível saber o resultado de uma experiência aleatória interessa-nos conhecer os possíveis resultados, surgindo assim a definição de espaço de resultados ou espaço fundamental.

Definição 1.2.1 (Espaço de Resultados ou Espaço Fundamental) *É o conjunto de todos os resultados possíveis de obter numa determinada experiência aleatória. Este conjunto será denotado por Ω.*

Exemplos:

1. Quando se atira uma moeda ao ar o espaço de resultados será
$$\Omega = \{Cara, Coroa\}.$$

2. No lançamento de um dado ao ar tem-se
$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

3. Ao atirar um dado ao ar podemos ter interesse em analisar a paridade do resultado, sendo assim tem-se
$$\Omega = \{\text{par, ímpar}\}.$$

4. Registo do número de chamadas que chegam a uma central telefónica durante um certo período:
$$\Omega = \mathbb{N}_0.$$

5. Registo do estado civil da população adulta do distrito da Guarda:
$$\Omega = \{\text{S,C,D,V,U}\}$$
onde S= "Solteiro", C="Casado", D="Divorciado", V= "Viúvo"e U="União de facto".

Pelos exemplos apresentados vê-se que o espaço de resultados depende da experiência aleatória em causa, podendo até depender de alguma propriedade que nos interesse estudar. Além disso, qualquer resultado de uma experiência tem que corresponder exactamente a um elemento de Ω.

Dependendo do número de elementos do espaço de resultado, pode ser:

1. Discreto,
   - Finito,
   - Infinito Numerável,
2. Contínuo.

Considerando a natureza dos elementos que compõem o espaço de resultados, pode ser:

1. Quantitativo,

2. Qualitativo.

Por outro lado, a indicação dos elementos do espaço de resultados, pode ser por extensão ou por compreensão.

Exemplos:

1. O atendimento num determinado serviço de um hospital funciona entre as 09:00 e as 16:00 horas. Cada utente chegará ao serviço na hora X e deixará o serviço, após ter sido atendido, na hora Y. Assim, supondo que o objectivo é observar as entradas e as saídas dos utentes temos as variáveis X e Y que são contínuas e tomam valores no intervalo $[9, 16]$. Sabe-se ainda que a hora de chegada será inferior à hora de saída, sendo $X<Y$, pelo que o espaço de resultados pode ser descrito por $\Omega =\{(X,Y):9<X<Y<16\}$ (definição por compreensão)

2. Na experiência aleatória "lançamento de uma moeda ao ar"o conjunto de resultados possíveis é finito e pode representar-se por

$$\Omega =\{\text{cara,coroa}\}(\text{definição por extensão}).$$

1.2.2 Acontecimentos

Tendo sido a primeira etapa da modelação de uma experiência aleatória a noção de espaço de resultados (Ω), segue-se a noção de Acontecimento. Um acontecimento está associado a uma **propriedade sobre o espaço de resultados**. Toda a propriedade que se defina sobre o espaço de resultados determina um subconjunto A de Ω e todo o subconjunto A de Ω define uma propriedade por intermédio de uma relação de pertença. Assim, diz-se que um acontecimento A é realizado se e só se o resultado da experiência aleatória é um elemento pertencente a A.

Exemplos:

1. Se na experiência aleatória "lançamento de um dado e observação da face exposta", $\Omega = \{1, 2, 3, 4, 5, 6\}$, considerarmos a propriedade "a face exposta tem um número ímpar"podemos identificar esta propriedade com o acontecimento A_1, sendo $A_1 = \{1, 3, 5\}$.

2. Na experiência aleatória "registo do número de chamadas que chegam a uma central telefónica durante um certo período", $\Omega = \{\mathbb{N}_0\}$, se apenas tivermos interesse no "número de chamadas entre 50 e 100"teremos o acontecimento $A = \{50, 51, \ldots, 99, 100\}$.

3. Na experiência aleatória "registo da duração das chamadas telefónicas ($\Omega = \{\mathbb{R}_0^+\}$) podemos analisar o acontecimento A="duração de chamadas superior a 30 minutos"sendo $A =]30, +\infty]$.

4. Considerando a experiência aleatória "registo altura das pessoas com idade compreendida entre 15 e 21 anos", $\Omega = \{\mathbb{R}^+\}$, pode ter interesse analisar o acontecimento "altura entre $1,60m$ e $1,90m$", que corresponde a $A =]1,60; 1,90[$.

5. Sendo a experiência aleatória "registo do valor máximo da tensão arterial", $\Omega = [0, +\infty[$, e se pretendermos analisar os casos em que "valor máximo é superior ou igual a 14"teremos um acontecimento A a que corresponde $A = [14, +\infty[$.

Existem acontecimentos com denominações especificas:

1. *Acontecimento certo*: Ω, pois é sempre realizado.
   Exemplo: Na experiência aleatória "lançamento de um dado e observação da face exposta"($\Omega = \{1, 2, 3, 4, 5, 6\}$) o acontecimento "a face exposta tem um dos números $\{1, 2, 3, 4, 5, 6\}$"é sempre realizado.

2. *Acontecimento impossível*: $\emptyset$, dado que nunca se realiza.
   Exemplo: Considerando a experiência aleatória anterior e o acontecimento "a face exposta tem número 7", dado que o espaço de resultados da experiência não contempla o número 7, este nunca se realizará.

3. *Acontecimento elementar*: cada subconjunto singular de Ω.
   Exemplo: Continuando com a mesma experiência aleatória, mas considerando agora o acontecimento "a face exposta tem o número 1".Este acontecimento corresponde a um único elemento do espaço de resultados, sendo por isso um acontecimento elementar.

Vamos representar por $P(\Omega)$ o conjunto de todos os subconjuntos de Ω, ou seja, de todos os acontecimentos. Note-se que se Ω for finito tem-se $\#P(\Omega) = 2^{\#\Omega}$.

1.2.3 Álgebra dos Acontecimentos

Dada que à noção de acontecimento, apresentada anteriormente, está associada à noção de conjunto, podem-se utilizar as noções de teoria de conjuntos para tratar as operações que se podem definir no âmbito dos acontecimentos.

Tendo em conta que uma propriedade determina um conjunto podem relacionar-se propriedades e conjuntos da seguinte forma:

- Conjunção ($\wedge$) de propriedades $\longleftrightarrow$ Intersecção ($\cap$)de Conjuntos;
- Disjunção ($\vee$)de propriedades $\longleftrightarrow$ União ($\cup$) de conjuntos;
- Negação ($\sim$) de propriedade $\longleftrightarrow$ Conjunto Complementar.

Definição 1.2.2 (Intersecção ou produto lógico) *Dados dois acontecimentos A e B, define-se intersecção ou produto lógico e representa-se* $A \cap B$*, como o acontecimento que ocorre se e só se ocorrerem A e B, ou seja,* $A \cap B = \{x : x \in A \wedge x \in B\}$.

Definição 1.2.3 (União ou soma lógica) *Dados dois acontecimentos A e B, o acontecimento união de A com B, representa-se por* $A \cup B$*, consiste a realização de pelo menos um dos dois acontecimentos, ou seja,* $A \cup B = \{x : x \in A \vee x \in B\}$.

As duas definições anteriores podem generalizar-se para mais de dois conjuntos.

Definição 1.2.4 (Acontecimento complementar ou contrário) *Dado um acontecimento* $A \subset \Omega$ *designa-se por* $\overline{A}$ *ou* A^c *o acontecimento contrário ou complementar de A, isto é, o acontecimento que se realiza se e só se A não se realiza, que corresponde ao conjunto de elementos de* Ω *que não pertencem a A.*

Definição 1.2.5 (Diferença de Acontecimentos) *Dados dois acontecimentos A e B, o acontecimento diferença* $A-B$ *ou* $A \backslash B$ *é o acontecimento que se realiza quando ocorre A e não ocorre B, ou seja,* $A\text{-}B = A \backslash B = \{x : x \in A \wedge x \notin B\}$.

Note-se que a noção de acontecimento complementar é um caso particular de diferença de acontecimentos, pois $A^c = \Omega - A$.

Exemplo: Considere-se a experiência aleatória que consiste em medir o tempo médio mensal de utilização da internet por pessoa em Portugal (em horas) e A, B e C os seguintes acontecimentos:

A="utilização média mensal superior ou igual a 50 horas mas inferior a 100 horas", A=$[50, 100[$
B="utilização média mensal superior ou igual a 100 horas mas inferior a 200 horas", B=$[100, 200[$
C="utilização média mensal superior ou igual a 60 horas mas inferior a 100 horas", C=$[60, 100[$

- O acontecimento que consiste na ocorrência simultânea dos acontecimentos A e B é A$\cap$B=$\emptyset$, ou seja, o acontecimento impossível. A realização simultânea dos acontecimentos A e C não é mais do que a realização do acontecimento C, isto é, A$\cap$C=C.

- O acontecimento A$\cup$B="utilização média mensal superior ou igual a 50 horas mas inferior a 200 horas", A$\cup$B=$[50, 200[$.

- Sendo $\Omega = [0, 720]$, considerando o mês com 30 dias, o acontecimento contrário ou complementar de A é $A^c = \Omega$-A= $[0, 50[\cup[100, 720]$ ou seja A^c é o acontecimento "utilização média mensal superior ou igual a 0 horas e inferior a 50 horas ou superior ou igual a 100 horas e inferior ou igual a 720".

- O acontecimento A-C ou A\C é a "utilização média mensal superior ou igual a 50 horas mas inferior a 60 horas", ou seja, A-C= $[50, 60[$.

Definição 1.2.6 (Acontecimentos mutuamente exclusivos) *Dois acontecimentos A e B tais que A$\cap$B= $\emptyset$ dizem-se mutuamente exclusivos ou incompatíveis.*

Exemplo: Considere-se a experiência aleatória do exemplo anterior. Os acontecimentos A e B são mutuamente exclusivos ou incompatíveis, uma vez que não podem ocorrer em simultâneo: se ocorre A, isto é "utilização média mensal superior ou igual a 50 horas mas inferior a 100 horas", não pode ocorrer B, isto é "utilização média mensal superior ou igual a 100 horas mas inferior a 200 horas"e vice-versa.

Propriedades das Operações

Seguem-se algumas das propriedades das operações definidas anteriormente no âmbito dos acontecimentos.

1. Propriedade comutativa: $A \cup B = B \cup A$ e $A \cap B = B \cap A$

2. Propriedade associativa: $(A \cup B) \cup C = A \cup (B \cup C)$ e $(A \cap B) \cap C = A \cap (B \cap C)$

3. Propriedade distributiva: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ e $A \cup (B \cap C = (A \cup B) \cap (A \cup C)$

4. Dupla negação: $(A^c)^c = A$

5. Elemento absorvente: $A \cap \emptyset = \emptyset$ e $A \cup \Omega = \Omega$
6. Elemento neutro: $A \cap \Omega = A$ e $A \cup \emptyset = A$
7. Lei do complemento: $A \cap A^c = \emptyset$ e $A \cup A^c = \Omega$
8. Idempotência: $A \cap A = A$ e $A \cup A = A$
9. Lei de Morgan: $(A \cap B)^c = A^c \cup B^c$
10. Lei de Morgan: $(A \cup B)^c = A^c \cap B^c$
11. Lei de Morgan: $(\bigcup_{n=1}^{\infty} A_n)^c = \bigcap_{n=1}^{\infty} A_n^c$
12. Lei de Morgan: $(\bigcap_{n=1}^{\infty} A_n)^c = \bigcup_{n=1}^{\infty} A_n^c$

1.3 Teoria Elementar das probabilidades

Em qualquer experiência aleatória, há sempre uma incerteza quanto à ocorrência, ou não, de determinado evento. Medir a ocorrência de um evento ou a probabilidade com que ele se verifica implica a definição de um valor numérico. Quando, por exemplo, a partir do passado não se consegue prever deterministicamente o futuro, mas é necessário tomar decisões com a perspectiva do futuro, as probabilidades podem fornecer aos agentes de decisão dados que apoiem a tomada de decisões. Nas experiências ou fenómenos aleatórios não se pode prever o resultado de cada execução da experiência, mas é previsível a frequência relativa de ocorrência de qualquer um dos resultados, num grande número de provas, ou seja, observando-se o fenómeno um grande número de vezes , sob idênticas condições, a frequência relativa de cada resultado possível tende a estabilizar, aproximando-se de um valor constante no intervalo [0,1].

Exemplo: Se efectuarmos um grande número de lançamentos de uma moeda perfeita, verifica-se que a frequência relativa de ocorrência, de uma qualquer das faces, tende a estabilizar na vizinhança de $1/2$.

No que se segue vamos referir dois conceitos de probabilidade: conceitos clássico e frequencista, apesar de existirem outros.

1.3.1 Probabilidade: Teoria Clássica de Laplace

Segue-se a definição de probabilidade segundo Laplace.

Definição 1.3.1 (Probabilidade dedutiva ou a priori) *A probabilidade de um acontecimento é a razão do número de resultados favoráveis a esse acontecimento sobre o número total de resultados possíveis, desde que todos os acontecimentos elementares sejam igualmente prováveis.*

Assim $P(A)$, a probabilidade do acontecimento A, é uma função de conjunto, isto é, a cada conjunto $A \subset \Omega$ faz corresponder um número real que é a sua probabilidade. Genericamente, dado um espaço de resultados finito

$$\Omega = \{x_1, x_2, \ldots, x_n\}, \text{ com } n \in \mathbb{N},\ \#\Omega = n \text{ e com}$$

$$P(\{x_i\}) = \frac{1}{n}, i \in \{1, \ldots, n\}$$

e dado um acontecimento A, que está associado o conjunto $\{x_1, x_2, \ldots, x_m\}$ com $m \in \mathbb{N}$, tem-se

$$P(A) = \frac{\#A}{\#\Omega} = \frac{m}{n}.$$

Exemplos:

1. Considere-se o lançamento de um dado perfeito (todos os resultados são equiprováveis), com $\Omega = \{1, 2, 3, 4, 5, 6\}$, e seja A o acontecimento "sair face múltipla de 2". A probabilidade de A ou $P(A)$ é a razão entre o número de casos favoráveis a A, o que corresponde ao conjunto $\{2, 4, 6\}$, e o número de casos igualmente possíveis, ou seja, $P(A) = 3/6 = 1/2$.

2. Considere o lançamento de duas moedas iguais ao ar, sendo co="coroa"e ca="cara"tem-se

   $$\Omega = \{(ca, ca), (ca, co), (co, co), (co, ca)\},$$

   e o acontecimento A="saírem duas caras"ao qual está associado o conjunto$\{(ca, ca)\}$. A probabilidade do acontecimento A é então

   $$P(A) = 1/4.$$

A definição de Laplace é demasiado restritiva, pois não abrange os fenómenos aleatórios em que os possíveis resultados ou casos não são igualmente prováveis.

Se assumirmos que os possíveis resultados ou casos de uma experiência aleatória não são igualmente prováveis, genericamente $\#\Omega = n$ e

$$\Omega = \{x_1, x_2, \ldots, x_n\}$$

com $P(\{x_i\}) = p_i$ com $i \in \{1, \ldots, n\}$ e $p_i \geq 0$ com $\sum_{i=1}^{n} p_i = 1$, dada a normalização em probabilidades.

Exemplos:

1. Suponhamos que temos uma moeda desiquilibrada em que
$$p_1 = P(\{Ca\}) = \frac{1}{3}, p_2 = P(\{Co\}) = \frac{2}{3}$$
mas com
$$\sum_{i=1}^{2} p_i = 1.$$
Considerando o lançamento de duas destas moedas ao ar tem-se
$$\Omega = \{(ca, ca), (ca, co), (co, co), (co, ca)\},$$
e o acontecimento A="saírem duas caras"ao qual está associado o conjunto$\{(ca, ca)\}$. A probabilidade do acontecimento A é então
$$P(A) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}.$$

2. Um exemplo de aplicação de pesos diferentes é o cálculo da média de curso. Neste caso, consideram-se pesos diferentes para unidades curriculares com ECTS (European Credit Transfer System) diferentes.

1.3.2 Probabilidade:Teoria Frequencista

A regularidade estatística asociada aos fenómenos aleatórios permite formular a teoria frequencista da probabilidade. Segue-se a definição da denominada probabilidade estatística ou empírica ou indutiva ou à posteriori.

Definição 1.3.2 (Probabilidade estatística) *Seja f o número de vezes que um acontecimento A se realiza em n provas de uma experiência aleatória. O inteiro f é a frequência absoluta de A, nas n provas, e $F_n(A) = f/n$ corresponde à frequência relativa, sendo $F_n(A) \leq 1$.*

Pela regularidade estatística é de esperar que em sucessões de provas, formadas por um grande número de repetições da mesma experiência aleatória onde se pretende observar a ocorrência de um dado acontecimento A, se constate que as frequências relativas de A sejam aproximadamente iguais. Ou seja, à medida que o número de provas aumenta a frequência relativa do acontecimento A tende a estabilizar na vizinhança de determinado valor.

Assim, numa tentativa de conceptualização, associa-se a cada acontecimento A um número real p, $(0 \leq p \leq 1)$ que se diz probabilidade teórica ou matemática de A, que corresponde ao valor limite em torno do qual as frequências relativas tendem a estabilizar quando $n \to \infty$.

Assim, o conceito teórico de probabilidade aparece ligado à noção experimental de frequência relativa e tomar $P(A)$ como probabilidade de A equivale a aceitar que na sucessão numerosa de provas é praticamente certo que $F_n(A)$ é aproximadamente igual a $P(A)$. Quando a definição de Laplace é aplicável, as flutuações das frequências dão-se em torno do número p dado pela definição, o que traduz a compatibilidade entre as duas teorias.

Exemplo: Ao realizar várias vezes a experiência "atirar uma moeda ao ar"calcularam-se as frequências do acontecimento A = "sair cara"e registaram-se os resultados que se apresentam na tabela 1.1.

Nº de lançamentos	1	5	10	20	50	100	200	500	1000
Nº de realizações de A	1	1	6	10	23	63	105	237	479
frequência relativa de realizações de A	1	0,2	0,6	0,5	0,46	0,63	0,525	0,474	0,479

Tabela 1.1: Realizações do acontecimento A no lançamento de uma moeda equilibrada.

Na Figura 1.1 representa-se a frequência relativa ou proporção do número de caras, realização do acontecimento A, em função do número de lançamentos da moeda equilibrada, sendo ilustrado o conceito frequencista de probabilidade. Assim, o número para o qual tende a frequência relativa

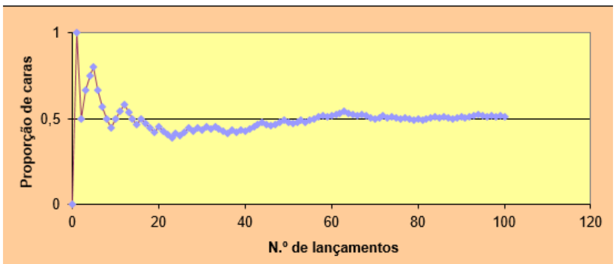

Figura 1.1: Representação da proporção do número de caras em função do número de lançamentos.

do acontecimento A quando o número de lançamentos aumenta consideravelmente designa-se probabilidade do acontecimento A e é 0,5.

Note-se que os conceitos apresentados, conceitos clássico e frequencista, não são mais do que instrumentos de medida para avaliar a probabilidade.

1.4 Axiomas da Probabilidade

A construção axiomática da teoria da probabilidade respeita as propriedades básicas das definições clássica e frequencista, sendo os axiomas decorrentes das referidas definições. Considerando uma qualquer experiência aleatória e $F_n(A)$ a frequência absoluta de ocorrência de um acontecimento A em n provas da experiência aleatória. Temos $0 \leq F_n(A) \leq n$ e portanto $0 \leq F_n(A)/n \leq 1$ e para n grande $F_n(A)/n$ é um valor vizinho da probabilidade do acontecimento A $P(A)$. Assim, tal como já referimos anteriormente, considere-se $P(A)$ uma função de conjunto, isto é, a cada conjunto $A \subset \Omega$ faz corresponder um número real que é a sua probabilidade, que satisfaz os seguintes axiomas:

- **Axioma1:** $P(A) \geq 0, \forall A \subseteq \Omega$

- **Axioma2:** A probabilidade do acontecimento certo é dada por

$$P(\Omega) = 1$$

- **Axioma3:** Sendo A e B acontecimentos mutuamente exclusivos definidos em Ω, ou seja $A \cap B = \emptyset$, tem-se que

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

  De forma mais geral, sendo $A_1, A_2, A_3, \ldots, A_n$ acontecimentos mutuamente exclusivos definidos em Ω tem-se que:

$$P[\bigcup_{i=1}^{n} A_i] = \sum_{i=1}^{n} P[A_i].$$

Note-se que estamos a considerar que Ω é finito, caso seja infinito o axioma 3 terá de sofrer alterações (ver F. Galvão de Mello, 2000 página 103). Pelos axiomas tem-se que $\forall A \subseteq \Omega$ verifica-se $0 < P(A) \leq 1$.
Seguem-se alguns teoremas que são consequências imediatas dos axiomas.

Teorema 1.4.1 *Sendo $\emptyset$ o acontecimento impossível tem-se $P(\emptyset) = 0$.*

Demostração: Sendo $\Omega \cup \emptyset = \Omega$, aplicando probabilidades vem

$$P(\Omega \cup \emptyset) = P(\Omega)$$

e sendo acontecimentos mutuamente exclusivos

$$P(\Omega) + P(\emptyset) = P(\Omega)$$

de onde vem $P(\emptyset) = 0$.$\blacksquare$

Teorema 1.4.2 *Dado um acontecimento $A \subseteq \Omega$ com probabilidade $P(A)$, a probabilidade do seu complementar é $P(A^c) = 1 - P(A)$.*

Demostração: Sendo A e A^c acontecimentos mutuamente exclusivos em que $A \cup A^c = \Omega$, então aplicando probabilidades

$$P(A \cup A^c) = P(\Omega)$$

pelo segundo e terceiro axiomas temos

$$P(A) + P(A^c) = 1$$

surgindo

$$P(A^c) = 1 - P(A).\blacksquare$$

Exemplo: Considere a experiência aleatória que consiste na extracção de uma carta de um baralho de 52 cartas e o acontecimento A="sair naipe naipe de espadas". Dado que existem 13 cartas do naipe de espadas, sendo os casos favoráveis ao acontecimento

$$\{2, 3, 4, 5, 6, 7, 8, 9, 10, \grave{a}s, dama, valete, rei\}$$

sendo

$$P(A) = \frac{13}{52}.$$

Sendo a probabilidade do acontecimento complementar

A^c="não sair naipe de espadas", $P(A^c) = 1 - P(A) = 1 - \frac{13}{52} = \frac{39}{52}$.

Este resultado confirma-se facilmente, pois no baralho de 52 cartas são 39 as que não têm naipe de espadas.

Teorema 1.4.3 *Dados dois quaisquer acontecimentos A e B, a probabilidade do acontecimento diferença, $B - A$, é*

$$P(B - A) = P(B) - P(A \cap B).$$

Demonstração: Sendo A e B acontecimentos quaisquer tem-se

$$B = B \cap \Omega = B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c),$$

sendo os acontecimentos $B \cap A$ e $B \cap A^c$ mutuamente exclusivos pelo terceiro axioma temos

$$P(B) = P(B \cap A) + P(B \cap A^c) = (B \cap A) + P(B - A)$$

surgindo

$$P(B - A) = P(B) - P(A \cap B).\blacksquare$$

Exemplo: Uma empresa que produz software analisou os erros do seu software e concluiu que em 10% dos produtos produzidos os erros se deviam a uma insuficiente validação do software, em 5% dos produtos a erros algorítmicos devidos a uma deficiente especificação e em 1% verificavam-se as duas situações. Considerando os acontecimentos:

$A =$"software com erros de validação",
$B =$"software com erros de algorítmicos"

o acontecimento diferença

$A - B =$"software que apenas apresentou erros de validação".

Assim, a probabilidade de um pacote de software, retirado ao acaso da produção da empresa, apenas ter erros de validação será dada por

$$P(A - B) = P(A) - P(A \cap B) = 0,1 - 0,01 = 0,09.$$

Teorema 1.4.4 *Dados dois quaisquer acontecimentos A e B, a probabilidade do acontecimento união, $A \cup B$, é $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.*

Demonstração: Sendo A e B acontecimentos quaisquer tem-se

$$A \cup B = (A \cup B) \cap \Omega = (A \cup B) \cap (A \cup A^c) = A \cup (B \cap A^c) = A \cup (B - A),$$

aplicando probabilidades

$$P(A \cup B) = P(A \cup (B - A)),$$

sendo A e $(B-A)$ mutuamente exclusivos

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

e utilizando o teorema anterior

$$P(A \cup B) = P(A) + P(B) - P(B \cap A).\blacksquare$$

Exemplo: Em determinada loja que vende produtos informáticos observou-se que dos computadores vendidos 30% são portáteis e 50% são computadores de secretária, havendo uma franja de 10% de compradores que adquiriu os dois tipos. Se do universo de clientes da referida loja escolhermos um ao acaso, qual a probabilidade de ter comprado pelo menos um dos dois tipos de computadores?
Considerando os acontecimentos:

$A =$"compra de portátil"
$B =$"compra de um computador de secretária"

o que se pretende é determinar $P(A \cup B)$. Assim,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.5 - 0.1 = 0.7$$

ou seja 70% dos clientes compra pelo menos um dos dois tipos de computadores.

Teorema 1.4.5 (generalização do teorema 1.4.4) *Dados quaisquer acontecimentos $A_1, A_2, A_3, \ldots, A_n$ definidos em Ω tem-se*

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) =$$

$$= P(\bigcup_{i=1}^{n} A_i) =$$

$$= \sum_{i=1}^{n} P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_i \cap A_j) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} P(A_i \cap A_j \cap A_k) +$$

$$+ \cdots + (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n).$$

Considerando no teorema n=3 tem-se:

$$P(A_1 \cup A_2 \cup A_3) =$$

$$= \sum_{i=1}^{3} P(A_i) - \sum_{i=1}^{2} \sum_{j=i+1}^{3} P(A_i \cap A_j) + (-1)^4 P(A_1 \cap A_2 \cap A_3) =$$

$$= P(A_1)+P(A_2)+P(A_3)-P(A_1\cap A_2)-P(A_1\cap A_3)-P(A_2\cap A_3)+P(A_1\cap A_2\cap A_3)$$

Considerando no teorema n=4 tem-se:

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) =$$
$$= \sum_{i=1}^{4} P(A_i) - \sum_{i=1}^{3} \sum_{j=i+1}^{4} P(A_i \cap A_j) +$$
$$+ \sum_{i=1}^{2} \sum_{j=i+1}^{3} \sum_{k=j+1}^{4} P(A_i \cap A_j \cap A_k) - P(A_1 \cap A_2 \cap A_3 \cap A_4).$$

1.5 Probabilidade Condicionada

Em muitas situações, informação adicional sobre um dado fenómeno pode implicar alterações nas probabilidades de certos acontecimentos. Considere-se, por exemplo, a experiência aleatória que consiste em lançar dois dados ao ar e na qual se observam os seguintes acontecimentos:

A="soma das faces igual a 8"e B="duas faces iguais".

dado1/dado2	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Tabela 1.2: Resultados possíveis no lançamento simultâneo de dois dados equilibrados.

Sendo $\#\Omega = 36$ e os casos favoráveis do acontecimento

$$A = \{(2,6),(6,2),(4,4),(3,5),(5,3)\}$$

e os favoráveis do acontecimento

$$B = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$$

conclui-se que

$$P(A) = \frac{5}{36}$$

e

$$P(B) = \frac{6}{36}.$$

Tendo a informação que A se realizou, pretende-se saber, com esta condição, qual a probabilidade do acontecimento B se realizar? Nesta situação o conjunto de resultados possíveis deixou de ser Ω passando agora a ser o conjunto A, sendo neste conjunto que vamos procurar a realização do acontecimento B. Obteremos desta forma a probabilidade de B ocorrer sabendo que A já se realizou, representa-se por $P(B|A)$ e será

$$P(B|A) = \frac{1}{5} = \frac{\#A \cap B}{\#A} = \frac{\frac{\#A\cap B}{\#\Omega}}{\frac{\#A}{\#\Omega}} = \frac{P(A \cap B)}{P(A)}.$$

Definição 1.5.1 (Probabilidade Condicional) *Sejam A e B dois acontecimentos de um espaço Ω e $P(B) > 0$. A probabilidade condicional de A dado B ($P(A|B)$ ou $P_B(A)$) é definida por*

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

Da definição apresentada podemos escrever:

$$P(B|A) \times P(A) = P(A \cap B)$$

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

Note-se que o conceito de probabilidade condicional ou condicionada satisfaz os axiomas da teoria das probabilidades, apresentados anteriormente. Sendo B um acontecimento tal que $P(B) > 0$ provam-se os seguintes resultados:

1. Para qualquer acontecimento A tem-se $P(A|B) \geq 0$;

2. $P(\Omega|B) = 1$;

3. Se A_1 e A_2 são acontecimentos mutuamente exclusivos, $A_1 \cap A_2 = \emptyset$ então $P((A_1 \cup A_2)|B) = P(A_1|B) + P(A_2|B)$.

Exemplo: Um jogador comprou 5 bilhetes de lotaria dos quais dois eram números pares e três ímpares, compreendidos entre 0000 e 5000, sendo o conjunto de resultados possíveis para o sorteio $\Omega = \{0, \ldots, 5000\}$. Considerando o acontecimento $A =$"O jogador tem o bilhete do primeiro prémio", o conjunto de resultados favoráveis é constituído pelos bilhetes adquiridos, assim

$$P(A) = \frac{5}{5001}$$

. Supondo que ocorreu alguma irregularidade e que antes do sorteio se soube que o número premiado seria ímpar, qual a probabilidade de A se realizar? Considerando o acontecimento B="o bilhete sorteado tem número ímpar", a probabilidade pedida é

$$P(A|B) = \frac{3}{2500}$$

, sendo agora o conjunto de resultados possíveis não Ω mas o conjunto de números ímpares contido em Ω, com cardinal 2500.

1.5.1 Teoremas relacionados com Probabilidade Condicionada

Teorema 1.5.1 (Teorema da Probabilidade Composta) *Sejam* $A_0, A_1, \cdots, A_n$, $n+1$ *acontecimentos tais que* $P(A_0 \cap A_1 \cap A_2 \cap \ldots \cap A_{n-1}) > 0$ *então*

$$P(A_0 \cap A_1 \cap A_2 \cap \ldots \cap A_{n-1} \cap A_n) =$$

$$= P(A_0) \times P(A_1|A_0) \times P(A_2|A_0 \cap A_1) \times P(A_3|A_0 \cap A_1 \cap A_2) \times \ldots$$
$$\ldots \times P(A_n|A_0 \cap A_1 \cap \ldots \cap A_{n-1}).$$

Demonstração: Pela definição de probabilidade condicionada tem-se $P(A_0 \cap A_1 \cap A_2 \cap \ldots \cap A_{n-1} \cap A_n) = P(A_n|A_0 \cap A_1 \cap \ldots \cap A_{n-1}) \times P(A_0 \cap A_1 \cap \ldots \cap A_{n-1})$ aplicando a definição de probabilidade condicionada ao segundo factor $P(A_0 \cap A_1 \cap A_2 \cap \ldots \cap A_{n-1} \cap A_n) = P(A_n|A_0 \cap A_1 \cap \ldots \cap A_{n-1}) \times P(A_0 \cap A_1 \cap \ldots \cap A_{n-1}) = P(A_n|A_0 \cap A_1 \cap \ldots \cap A_{n-1}) \times P(A_{n-1}|A_0 \cap A_1 \cap \ldots \cap A_{n-2}) \times P(A_0 \cap A_1 \cap \ldots \cap A_{n-2})$, aplicando sucessivamente a definição de probabilidade condicionada ao último factor, em cada momento, obtém-se o resultado pretendido.■

Exemplo: Determine a probabilidade de em 4 extracções sucessivas e sem reposição de uma carta de um baralho de 52 cartas, se obterem 4 cartas do naipe espadas. Definindo o acontecimento $A_i =$"sai naipe de espadas na i^{sima} extracção", com $i \in \{1, 2, 3, 4\}$, a probabilidade pretendida é

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) =$$

$$= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap_2)P(A_4|A_1 \cap A_2 \cap A_3) =$$

$$= \frac{13}{52} \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49} \cong 0.0026.$$

Teorema 1.5.2 (Teorema da Probabilidade Absoluta) *Sejam $A_1, A_2, \cdots, A_n$, n subconjuntos disjuntos de Ω,*

$$\bigcup_{i=1}^{n} A_i = \Omega$$

e

$$P(A_i) \geq 0, i \in \{1, \ldots, n\}.$$

Então para qualquer acontecimento B tem-se

$$P(B) = \sum_{i=1}^{n} P(A_i) \times P(B|A_i).$$

Demonstração: Atendendo a que

$$\bigcup_{i=1}^{n} A_i = \Omega$$

e que

$$B = B \cap \Omega = B \cap (\bigcup_{i=1}^{n} A_i) =$$

$$= \bigcup_{i=1}^{n} (B \cap A_i)$$

e como

$$(B \cap A_i) \cap (B \cap A_j) = \emptyset, i \neq j$$

então

$$P(B) = P(\bigcup_{i=1}^{n} (B \cap A_i)) =$$

$$= \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i). \blacksquare$$

Exemplo: Numa empresa têxtil estão a produzir-se três tipos de peças de vestuário. Da produção total da empresa 20% é da peça A, 40% da B e 40% da C. Sabe-se que algumas peças têm pequenos defeitos sendo o seu número de 1% para a peça A, 2% para a peça B e 1% para a peça C. Seleccionou-se um peça de vestuário ao acaso, de entre as produzidas pela empresa. Pretende-se saber a probabilidade da peça ser defeituosa. Definindo os acontecimentos

D="peça de vestuário com defeito",
A="peça de vestuário do tipo A",
B="peça de vestuário do tipo B"
C="peça de vestuário do tipo C"

temos

$$P(A) = 0.2, P(B) = 0.4, P(C) = 0.4,$$

$$P(D|A) = 0.01, P(D|B) = 0.02 e P(D|C) = 0.01.$$

Aplicando o teorema da probabilidade absoluta tem-se

$$P(D) = \sum_{i=1}^{3} P(A_i) \times P(D|A_i) =$$

$$= 0.2 \times 0.01 + 0.4 \times 0.02 + 0.4 \times 0.01$$

Teorema 1.5.3 (Teorema de Bayes) *Sejam $A_1, A_2, \cdots, A_n$, n subconjuntos disjuntos de Ω,*

$$\bigcup_{i=1}^{n} A_i = \Omega$$

e tais que

$$P(A_i) \geq 0, i \in \{1, \ldots, n\}.$$

Então para qualquer acontecimento B, tal que $P(B) > 0$, tem-se

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^{n} P(B|A_k)P(A_k)}.$$

Demonstração: Pela definição de probabilidade condicionada temos

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

mas a probabilidade do numerador pode escrever-se

$$P(B \cap A_i) = P(B|A_i)P(A_i)$$

e a do denominador, pelo teorema anterior é

$$P(B) = \sum_{k=1}^{n} P(B|A_k) \times P(A_k)$$

surgindo naturalmente

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^{n} P(B|A_k)P(A_k)}.\blacksquare$$

Exemplo:Uma fábrica de bolachas utiliza 3 máquinas de empacotamento cuja produção diária é 10 000, 15 000 e 20 000 unidades, respectivamente. Pelo histórico das máquinas, sabe-se que cada máquina origina pequenos defeitos de empacotamento em 0.5%, 0.4% e 0.1% dos pacotes, respectivamente. Encontrou-se um pacote com defeito, pretende saber-se qual a máquina que tem maior probabilidade de o ter originado. Definindo o acontecimento

$$A_i = \text{"Pacote originário da máquina } i\text{", com } i \in \{1,2,3\}$$
$$D = \text{"Pacote com defeito"}$$

o que se pretende é determinar

$$P(A_i|B), i \in \{1,2,3\}$$

e proceder à sua ordenação. Sendo

$$\bigcup_{i=1}^{3} A_i = \Omega,$$

$$A_i \cap A_j = \emptyset, i \neq j, i, j \in \{1,2,3\}$$

e dado que todas as máquinas estão activas sendo $P(A_i) > 0$ pode-se utilizar o teorema anterior. Sabendo que:

$$P(A_1) = \frac{10000}{45000} \cong 0.222$$

$$P(A_2) = \frac{15000}{45000} \cong 0.333$$

$$P(A_3) = \frac{20000}{45000} \cong 0.445$$

$$P(D|A_1) = 0.005$$

$$P(D|A_2) = 0.004$$

$$P(D|A_3) = 0.001$$

temos

$$P(D) = \sum_{k=1}^{3} P(B|A_k)P(A_k) =$$

$$= 0.222 \times 0.005 + 0.333 \times 0.004 + 0.444 \times 0.001 \cong 0.0029$$

pelo que:

$$P(A_1|D) = \frac{P(D|A_1)P(A_1)}{P(D)} \cong 0.383,$$

$$P(A_2|D) = \frac{P(D|A_2)P(A_2)}{P(D)} \cong 0.459,$$

$$P(A_3|D) = \frac{P(D|A_3)P(A_3)}{P(D)} \cong 0.153.$$

Assim a máquina 2 é a que apresenta maior probabilidade de ter originado um pacote com defeito.

1.6 Independência de Acontecimentos

As noções de dependência e independência desempenham um papel fundamental na teoria da probabilidade. Intuitivamente dois acontecimentos dizem-se independentes se a realização de qualquer um deles não influência nem é influenciada pela realização do outro. Surgindo a seguinte definição formal de acontecimentos independentes.

Definição 1.6.1 (Acontecimentos Independentes) *Dois acontecimentos A e B de um espaço Ω dizem-se independentes se*

$$P(A \cap B) = P(A) \times P(B).$$

Sendo A e B acontecimentos independentes tem-se:

1. $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \times P(B)}{P(B)} = P(A);$$

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

Exemplo: De um baralho de 52 cartas extrai-se uma carta, verifica-se de que carta se trata, e repõe-se no baralho. Procede-se a nova tiragem de uma carta do baralho. Pretende-se saber a probabilidade de termos como resultado uma Dama e um Valete. Sendo D ="sair Dama"e V ="sair Valete"o que se pretende é

$$P(D \cap V) = P(D) \times P(V) = \frac{4}{52} \times \frac{4}{52},$$

dada a independência dos acontecimentos.

Definição 1.6.2 (Mais de dois Acontecimentos Independentes) *Os acontecimentos $A_1, A_2, \ldots, A_n$ são independentes se dados quaisquer k índices, $i_1, i_2, \ldots, i_k$, todos distintos se tiver*

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \times P(A_{i_2}) \times \ldots \times P(A_{i_k}).$$

Exemplo: Um saco contém 15 bolas pretas e 6 brancas. Fazem-se 8 extracções com reposição de uma bola. Determine a probabilidade de se obterem 8 bolas pretas. Definindo o acontecimento A_i ="A bola da i^{sima} extracção é preta"o que se pretende é determinar

$$P(A_1 \cap A_2 \cap \ldots \cap A_8) = \prod_{i=1}^{8} P(A_i) = (\frac{15}{21})^8 \cong 0.06.$$

1.6.1 Algumas propriedades associadas aos acontecimentos independentes

Se A e B são acontecimentos independentes então também o são os seguintes pares de acontecimentos:

1. A e B^c;
2. A^c e B;
3. A^c e B^c.

Exemplo: Considerando o exemplo anterior, se pretendermos que as cartas seleccionadas não sejam nem Dama nem Valete tem-se

$$P(D^c \cap V^c) = P(D^c) \times P(V^c) = \frac{48}{52} \times \frac{48}{52}.$$

@BookletID, title = title, OPTkey = key, OPTauthor = author, OPThowpublished = howpublished, OPTaddress = address, OPTmonth = month, OPTyear = year, OPTnote = note, OPTannote = annote,

Capítulo 2

Variáveis Aleatórias e suas Distribuições

Existem experiências aleatórias em que o espaço de resultados Ω é, à partida, um conjunto numérico que pode ser constituído por números reais como por exemplo: a medição da temperatura em doentes, o tempo de execução de um algoritmo, o tempo de vida de uma máquina, etc. Contudo, existem experiências aleatórias em que Ω não é um conjunto numérico. Nestes casos, pode proceder-se à atribuição de um número real ou conjunto ordenado de números reais a cada elemento $x \in \Omega$, podendo esta atribuição ser unicamente convencional. Isto, porque tendo-se como objectivo determinar a probabilidade de ocorrência de um ou mais acontecimentos associados a uma qualquer experiência aleatória e a eventual aplicação de métodos estatísticos, os cálculos inerentes a estes procedimentos tornam-se facilitados se os acontecimentos forem identificados com valores numéricos. Para dar resposta a esta necessidade de estabelecer uma correspondência surge o conceito de variável aleatória, que introduziremos de seguida.
Como exemplo de espaço de resultados não numérico associado a uma experiência aleatória considere-se a seguinte situação: supondo que se pretende controlar a qualidade de componentes electrónicas e que para tal se selecionam, ao acaso, 4 componentes num lote. Designando

$D =$"Componente defeituosa"
$D^c =$"Componente não defeituosa"

o espaço de resultados é

$$\Omega = \{(D, D, D, D), (D^c, D, D, D), (D, D^c, D, D), (D, D, D^c, D),$$

$$(D, D, D, D^c), (D^c, D^c, D, D), (D^c, D, D^c, D), (D^c, D, D, D^c),$$

$$(D^c, D^c, D^c, D), (D^c, D^c, D, D^c), (D^c, D^c, D^c, D^c), (D, D^c, D^c, D^c),$$
$$(D, D^c, D^c, D), (D, D^c, D, D^c), (D^c, D, D^c, D^c), (D, D, D^c, D^c)\}$$
$$\#\Omega = 2^4 = 16.$$

2.1 Variável Aleatória Unidimensional

O valor de uma variável aleatória é determinado pelo acaso ou fenómeno aleatório.

Definição 2.1.1 (Variável Aleatória Unidimensional) *Considere-se uma função numérica $X(w)$, real e finita, cujo domínio é Ω e cujo contradomínio é $\mathbb{R}$. Assim, a aplicação*

$$X : \Omega \longrightarrow \mathbb{R}$$

formaliza matematicamente a atribuição de um número real a cada elemento de Ω.

Dada a definição, uma variável aleatória é uma aplicação do espaço de resultados num conjunto de números reais que a cada acontecimento $A \subseteq \Omega$ faz corresponder $X(A)$ (conjunto de números reais)

$$X(A) = \{x(w) : w \in A\}.$$

Dado que a cada acontecimento, A, é possível fazer corresponder um número real, através do conceito de variável aleatória, pode-se também proceder ao cálculo de probabilidades com base nas imagens dos acontecimentos, $X(A)$, passando a ter a $P(A) = P(X = x)$.

Exemplos:

1. Considere o espaço de resultados do lançamento de um dado perfeito, $\Omega = \{1, 2, 3, 4, 5, 6\}$, e definindo os acontecimentos

   $A =$"saída de face par"e $B =$"saída de face ímpar".

   Aos acontecimentos estão associados os seguintes conjuntos

   $A = \{2, 4, 6\}$ e $B = \{1, 3, 5\}$,

definindo a aplicação

$$X : \Omega \longrightarrow \mathbb{R}$$

em que:

$$X(1) = X(3) = X(5) = 0; X(2) = X(4) = X(6) = 1$$

$$X(A) = \{1\} e X(B) = \{0\}$$

ou seja, tem-se uma aplicação tal que

$$X^{-1}(1) = A$$

e

$$X^{-1}(0) = B.$$

Esta aplicação é uma variável aleatória que toma valores $\{0, 1\}$ com probabilidade

$$P(X = 0) = P(X^{-1}(0)) = P(B) = \frac{1}{2}$$

e

$$P(X = 1) = P(X^{-1}(1)) = P(A) = \frac{1}{2}.$$

2. Considere o universo das famílias com três filhos e considere o espaço formado pelas possíveis sequências ordenadas de nascimentos, em função do sexo das crianças, considerando M="sexo masculino"e F="sexo feminino",

$$\Omega = \{MMM, MMF, MFM, FMM, FFM, FMF, MFF, FFF\}$$

supõe-se que estas 8 sequências têm a mesma probabilidade de ocorrer. Definindo a variável X="número de rapazes"os possíveis resultados desta são 0,1,2,3 em que

$$X^{-1}(0) = \{FFF\} = A_1; X(A_1) = \{0\}$$

$$X^{-1}(1) = \{MFF, MFF, FMF\} = A_2; X(A_2) = \{1\}$$

$$X^{-1}(2) = \{MMF, MFM, FMM\} = A_3; X(A_3) = \{2\}$$

$$X^{-1}(3) = \{MMM\} = A_4; X(A_4) = 3$$

sendo

$$P(X=0)=P(X^{-1}(0)=P(A_1)=\frac{1}{8}$$
$$P(X=1)=P(X^{-1}(1)=P(A_2)=\frac{3}{8}$$
$$P(X=2)=P(X^{-1}(2)=P(A_3)=\frac{3}{8}$$
$$P(X=3)=P(X^{-1}(3)=P(A_4)=\frac{1}{8}.$$

3. Considere-se, por exemplo, que o mercado de telecomunicações em Portugal é controlado pelas seguintes empresas:

$$\{XPTO, YPTO, ZPTO, WPTO\}$$

suponhamos que se pretende avaliar o volume de vendas de cada uma das empresas em 2004. Definindo uma variável aleatória $X=$"volume de vendas, em 2004, de cada uma das empresas"(em 10^{10} euros) observaram-se os seguintes valores para cada uma das empresas:

$$X(XPTO)=200, X(YPTO)=100, X(ZPTO)=50, X(WPTO)=150$$

sendo

$$X:\Omega\longrightarrow\{50,100,150,200\}$$

Considerando a experiência aleatória que consiste em escolher, ao acaso, uma das quatro empresas de telecomunicação, pode perguntar-se: qual será a probabilidade de escolher uma empresa com um volume de vendas superior ou igual a 100×10^{10} de euros?
A resposta é

$$P(X\geq 100)=P(X=100\cup X=150\cup X=200)=$$
$$=P(X=100)+P(X=150)+P(X=200)=$$
$$=\frac{1}{4}+\frac{1}{4}+\frac{1}{4}=\frac{3}{4}.$$

2.1.1 Tipos de Variáveis Aleatórias

As variáveis aleatórias podem ser classificadas em discretas ou contínuas.

Definição 2.1.2 (Variável Aleatória Discreta) *Uma variável aleatória diz-se discreta se o conjunto de valores possíveis de X for finito ou infinito numerável.*

Exemplos:

1. Contagem das partículas emitidas por uma fonte radioactiva durante um intervalo de tempo. Define-se a variável $X =$"número de partículas radioactivas contadas"que é uma variável discreta com valores
$$\{0, 1, 2, 3, \cdots\} = \mathbf{N}_0.$$

2. Registo do nascimentos no país num determinado horizonte temporal. A variável aleatória que se pode considerar nesta situação é $X =$"número de nascimentos observados"que é uma variável discreta com valores
$$\{0, 1, 2, 3, \cdots\} = \mathbf{N}_0(!).$$

3. Considerando livros com 100 páginas procura-se o número de páginas com erros. Neste caso pode definir-se a variável aleatória discreta $X =$"número de páginas com erros"que é uma variável discreta com valores
$$\{0, 1, 2, 3, \cdots 100\}.$$

Definição 2.1.3 (Variável Aleatória Contínua) *Uma variável aleatória X diz-se contínua se tomar valores num intervalo ou numa colecção de intervalos.*

Exemplos:

1. Se observarmos o tempo entre avarias de uma máquina de empacotamento, em funcionamento numa determinada fábrica, podemos definir uma variável aleatória $X =$"tempo entre avarias"que é uma variável aleatória contínua tomando valores em
$$[0, +\infty[.$$

2. Se pretendermos estimar o consumo de gasolina de um determinado modelo automóvel, pode-se encher o depósito com apenas 10 litros de gasolina e rodar com o carro até esgotar a gasolina, registando a distância percorrida (admitindo ser realista um máximo de $200km$). Se definir a variável aleatória $X =$"distância percorrida com 10 litros de gasolina"será contínua e toma valores em
$$[0, 200].$$

3. Considerando a altura da população adulta para diferentes gerações, pode-se definir uma variável aleatória $X =$"altura da população adulta em diferentes gerações"é contínua, admitindo um mínimo de $0,57m$ máximo de $2,72m$ (Guinness Records!), toma valores em
$$[0,57; 2,72].$$

Variáveis Aleatórias Discretas:Função de Probabilidade

A distribuição de probabilidade de uma variável aleatória X não é mais do que uma descrição formal da verosimilhança associada à ocorrência dos possíveis valores de X.

Definição 2.1.4 (Função de Probabilidade) *sendo X uma variável aleatória discreta cujos possíveis valores são $\{x_1, x_2, \ldots, x_n\}$ com $n \in IN$, a função de probabilidade de X é uma função f que associa a cada valor $x_i \in X$ a sua probabilidade $f(x_i) = P(X = x_i)$, com $i \in \{1, 2, \ldots, n\}$ e $f(x_i) = 0$ se $i \notin \{1, 2, \ldots, n\}$, com $n \in IN$. A função f satisfaz as seguintes propriedades:*

1. $f(x) \geq 0; \forall x \in \mathbb{R}$
2. $\sum_{x \in \mathbb{R}} f(x) = 1.$

O gráfico da função de probabilidade, f, é um gráfico de barras ou diagrama de traços.

Exemplos:

1. Considere-se a experiência aleatória que consiste no lançamento de três moedas iguais e equilibradas e a variável aleatória $X =$"número de ocorrências da face cara". Esta variável é discreta e vamos ver quais os seus valores. Sendo
$$\Omega = \{(ca, ca, ca)(ca, ca, co), (ca, co, ca), (ca, co, co),$$
$$(co, ca, ca), (co, ca, co), (co, co, ca), (co, co, co)\}$$
os possíveis valores de X são:
$$\begin{array}{l} 0 \rightarrowtail A_0 = \{(co, co, co)\} \\ 1 \rightarrowtail A_1 = \{(ca, co, co), (co, ca, co), (co, co, ca)\} \\ 2 \rightarrowtail A_2 = \{(ca, ca, co), (ca, co, ca), (co, ca, ca)\} \\ 3 \rightarrowtail A_3 = \{(ca, ca, ca)\} \end{array}$$

assim X toma valores em $\{0,1,2,3\}$. Sendo

$$f(0) = P(X=0) = P(A_0) = \frac{1}{8}$$
$$f(1) = P(X=1) = P(A_1) = \frac{3}{8}$$
$$f(2) = P(X=2) = P(A_2) = \frac{3}{8}$$
$$f(3) = P(X=3) = P(A_3) = \frac{1}{8}$$

para todos os valores $x \notin \{0,1,2,3\}$ a função $f(x)$ vale 0. Pode-se representar a função graficamente (gráfico 2.1) ou por:

$$f(x) = \begin{cases} \frac{1}{8} & se \quad x=0 \\ \frac{3}{8} & se \quad x=1 \\ \frac{3}{8} & se \quad x=2 \\ \frac{1}{8} & se \quad x=3 \\ 0 & \quad \text{para outros valores de} x \end{cases}$$

ou através da tabela: é fácil provar que f verifica as propriedades de

x	0	1	2	3	$\notin \{0,1,2,3\}$
$f(x)$	1/8	3/8	3/8	1/8	0

uma função de probabilidade, pois

(a) $f(x) \geq 0; \forall x \in \mathbb{R}$

(b) $\sum_x f(x) = f(0) + f(1) + f(2) + f(3) = 1$

2. Consideraram-se casais com três filhos e registou-se o sexo de cada um dos três filhos por ordem crescente de idade. Definindo uma variável aleatória X ="número de rapazes"sendo o espaço amostral

$$\Omega = \{MMM, MMF, MHM, MHH, HMM, HMH, HHM, HHH\}$$

em função da característica que se pretende observar temos os seguintes valores de X:

$$0 \rightarrowtail A_0 = \{HHH\}$$
$$1 \rightarrowtail A_1 = \{MMH, MHM, HMM\}$$
$$2 \rightarrowtail A_2 = \{MHH, HMH, HHM\}$$
$$3 \rightarrowtail A_3 = \{HHH\}$$

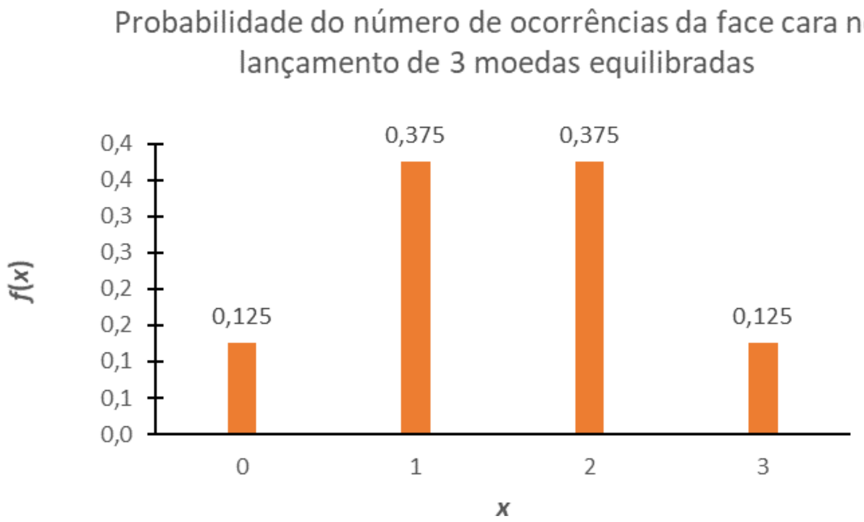

Figura 2.1: Função de probabilidade do número de caras no lançamento de 3 moedas equilibradas.

assim X toma valores em $\{0, 1, 2, 3\}$. Sendo

$$f(0) = P(X = 0) = P(A_0) = \frac{1}{8}$$

$$f(1) = P(X = 1) = P(A_1) = \frac{3}{8}$$

$$f(2) = P(X = 2) = P(A_2) = \frac{3}{8}$$

$$f(3) = P(X = 3) = P(A_3) = \frac{1}{8}$$

Esta função de probabilidade é igual à do exemplo anterior.

3. Para controlar a qualidade dos iogurtes até 5 dias após a data de validade, analisam-se 4 retirados aleatoriamente. Designando por

$$B = \text{"iogurte bom"e } E = \text{"iogurte estragado"},$$

os resultados possíveis são:
Definindo a variável aleatória

$$X = \text{"numero de iogurtes estragados "},$$

tendo em conta que se consideraram amostras com 4 elementos. Esta variável pode assumir valores em $\{0, 1, 2, 3, 4\}$, vamos admitir que, nas

1ª tiragem	2ª tiragem	3ª tiragem	4ª tiragem	sequência
B	B	B	B	BBBB
B	B	B	E	BBBE
B	B	E	B	BBEB
B	B	E	E	BBEE
B	E	B	B	BEBB
B	E	B	E	BEBE
B	E	E	B	BEEB
B	E	E	E	BEEE
E	B	B	B	EBBB
E	B	B	E	EBBE
E	B	E	B	EBEB
E	B	E	E	EBEE
E	E	B	B	EEBB
E	E	B	E	EEBE
E	E	E	B	EEEB
E	E	E	E	EEEE

condições apresentadas, a probabilidade de um iogurte estar Bom, B, ou de estar estragado,E, é igual. A correspondência entre valores de X e conjuntos de resultados (acontecimentos) é:

$0 \rightarrowtail A_0 = \{BBBB\}$
$1 \rightarrowtail A_1 = \{BBBE, BEBB, BBEB, EBBB\}$
$2 \rightarrowtail A_2 = \{BBEE, BEBE, EEBB, EBEB, EBBE, BEEB\}$
$3 \rightarrowtail A_3 = \{EEEB, BEEE, EBEE, EEBE\}$
$4 \rightarrowtail A_4 = \{EEEE\}$

Sendo

$$f(0) = P(X = 0) = P(A_0) = \frac{1}{16}$$

$$f(1) = P(X = 1) = P(A_1) = \frac{1}{4}$$

$$f(2) = P(X = 2) = P(A_2) = \frac{3}{8}$$

$$f(3) = P(X = 3) = P(A_3) = \frac{1}{4}$$

$$f(4) = P(X = 4) = P(A_4) = \frac{1}{16}$$

a função de probabilidade.

Pode-se representar a função através de uma representação gráfica (gráfico 2.2) ou por:

$$f(x) = \begin{cases} \frac{1}{16} & se \quad x = 0 \\ \frac{1}{4} & se \quad x = 1 \\ \frac{3}{8} & se \quad x = 2 \\ \frac{1}{4} & se \quad x = 3 \\ \frac{1}{16} & se \quad x = 4 \\ 0 & \quad \text{para outros valores de} x \end{cases}$$

ou através da tabela:

x	0	1	2	3	4	$\notin \{0,1,2,3\}$
$f(x)$	1/16	1/4	3/8	1/4	1/16	0

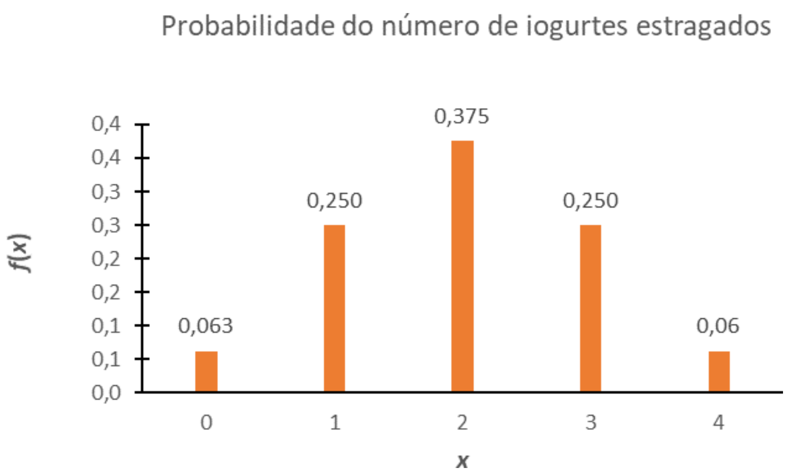

Figura 2.2: Função de probabilidade do número de iogurtes estragados numa qualquer amostra aleatória de 4 iogurtes.

É fácil de verificar, tal como nos exemplos anteriores, que f verifica as propriedades de uma função de probabilidade.

Variáveis Aleatórias Discretas:Função de Distribuição

Definição 2.1.5 (Função de Distribuição) *seja X uma variável aleatória discreta com função de probabilidade f. A sua função de distribuição acumulada F está definida para qualquer valor real de x e é dada por*

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u)$$

onde u toma todos os valores possíveis da variável aleatória X não superiores a x.

Note-se que usualmente utiliza-se a letra minúscula f para representar função de probabilidade e a maiúscula F para representar a função de distribuição acumulada ou função de distribuição.

Considerando, genericamente, uma v.a. discreta X com valores

$$\{x_1, x_2, \ldots, x_n\}$$

sendo $n \in IN$ e com função de probabilidade $f(x)$ a função de distribuição será:

$$F(x) = \begin{cases} 0 & se \quad x < x_1 \\ f(x_1) & se \quad x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & se \quad x_2 \leq x < x_3 \\ f(x_1) + f(x_2) + f(x_3) & se \quad x_3 \leq x < x_4 \\ \vdots & \\ f(x_1) + f(x_2) + \cdots + f(x_n) & se \quad x \geq x_n \end{cases}$$

Teorema 2.1.1 *As Propriedades da função de distribuição acumulada F da v.a. discreta X são:*

1. *Função com descontinuidades e em escada.*

2. *Se $x \leq y$ então $F(x) \leq F(y)$, ou seja, a função é crescente em sentido lato.*

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

4. $$\lim_{x \rightarrow +\infty} F(x) = 1.$$

A função de distribuição de uma v.a. discreta X obtém-se a partir da sua função de probabilidade por :

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u).$$

Por outro lado, a função de probabilidade de X pode obter-se a partir da função de distribuição:

$$f(x) = P(X = x) = F(x) - F(x^-),$$

onde $F(x)$ representa o valor da função de distribuição no ponto x e $F(x^-)$ o valor da função de distribuição à esquerda do ponto x, podendo dizer-se que corresponde ao valor da função no ponto imediatamente anterior a x.
Exemplo: Uma máquina de jogo tem 200 bolas (110 azuis, 60 brancas e 30 cor de rosa), que são baralhadas permanentemente, à espera que alguém introduza 1 euro numa ranhura e carregue num botão. Nesse momento, é seleccionada aleatoriamente uma das bolas e lançada para o exterior. Entretanto, essa bola é substituída automaticamente, na máquina, por outra da mesma cor. O jogador tem direito a um prémio de 1 euro, se a bola for branca, de 2 euros, se for cor de rosa. Determine:

a) A função de probabilidade do valor do prémio obtido por um jogador.
O prémio conseguido pelo jogador pode ser $\{0, 1, 2\}$ euros, pelo que a v.a. $X =$ "valor do prémio conseguido por um jogador"tem função de probabilidade

$$f(0) = \frac{110}{200} = 0,55$$

$$f(1) = \frac{60}{200} = 0,3$$

$$f(2) = \frac{30}{200} = 0,15.$$

b) A correspondente função de distribuição acumulada e a sua representação gráfica (gráfico 2.3).

$$F(x) = \begin{cases} 0 & se \quad x < 0 \\ f(0) = 0,55 & se \quad 0 \leq x < 1 \\ f(0) + f(1) = 0,85 & se \quad 1 \leq x < 2 \\ f(0) + f(1) + f(2) = 1 & se \quad x \geq 2 \end{cases}.$$

c) Mostre como obteria $f(x)$ a partir de $F(x)$

$$f(0) = F(0) - F(0^-) = 0,55 - 0 = 0,55$$
$$f(1) = F(1) - F(1^-) = 0,85 - 0,55 = 0,3$$
$$f(2) = F(2) - F(2^-) = 1 - 0,85 = 0,15.$$

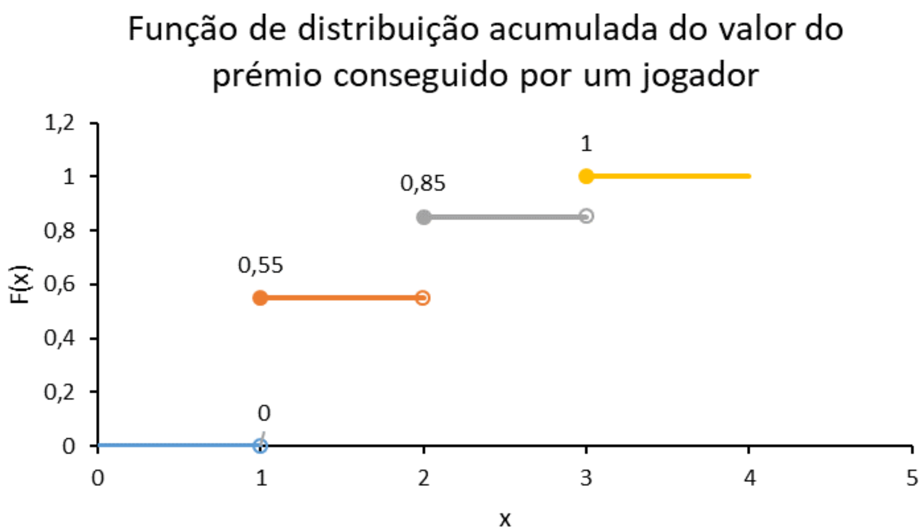

Figura 2.3: Função de distribuição acumulada do valor do prémio conseguido por um jogador.

Variáveis Contínuas: Função de Densidade de Probabilidade

Considerando uma variável aleatória X que representa, por exemplo, a altura dos alunos da ESTG, a probabilidade de X tomar exactamente o valor 1,60m em vez de um valor próximo deste, ou seja um valor numa vizinhança arbitrariamente proxima de 1,60m, é zero. Pois, para uma variável aleatória contínua a probabilidade de tomar exactamente um qualquer valor (um ponto), no seu domínio, é nula. Uma variável aleatória contínua é definida num intervalo com amplitude não nula. Pelo que, considerando a distribuição da probabilidade total, ou do ponto de vista físico massa probabilística total (1), pelo número infinito não numerável de pontos do intervalo a probabilidade associada a cada ponto pode considerar-se nula. Contudo, isto não significa que cada um dos resultados não pode ocorrer, apenas indica que antes de se efectuar a experiência aleatória a probabilidade de ocorrer exactamente um valor (e não a restante infinidade de valores no intervalo) é nula. Se pretendermos a probabilidade da variável tomar valores num sub-intervalo (em vez de um ponto desse sub-intervalo) do seu domínio, onde tem uma infinidade de resultados possíveis, a probabilidade já não será nula. A função que dá indicação da intensidade relativa, ou em termos físicos densidade de massa probabilística, denomina-se função de densidade de probabilidade (equivale à função de probabilidade das variáveis discretas) que se representa por f e é uma curva de frequência da variável a que as observações dizem respeito (que corresponde ao modelo matemático que melhor se ajusta às observações). Para um dado valor x a função $f(x)$ indica a probabilidade de ocorrerem os valores numa vizinhança de x.

Genericamente, sendo X uma variável aleatória que pode tomar todos os valores num intervalo(a,b) podendo a e b serem ou não finitos. Registando, para n observações, a frequência relativa em que a realização da variável X cai no intervalo $(x, x+h)$, com h pequeno, é possível construir um histograma. Se considerarmos intervalos com amplitude, h, então a área de cada rectângulo de base $(x, x+h)$ é aproximadamente igual à probabilidade $P(x < X < x+h)$. Fazendo $n \rightarrowtail \infty$ e $h \rightarrowtail 0$ o perfil em escada do histograma tende para uma curva contínua que será a curva de frequência de X que se denomina de função de densidade de probabilidade, já referida anteriormente.

Definição 2.1.6 (Função de Densidade de Probabilidade) *sendo X uma variável aleatória contínua a função de densidade de probabilidade (f.d.p.) de X é uma função f tal que*

$$P(a \leq X \leq b) = \int_a^b f(x)dx, \forall a, b \in \mathbb{R} : a \leq b,$$

e que tem as seguintes propriedades:

1. $f(x) \geq 0; \forall x \in \mathbb{R}$
2. $\int_{-\infty}^{+\infty} f(x)dx = 1.$

Exemplos:

1. Considerando uma v.a. X que representa, por exemplo, a altura (em centímetros) dos alunos do IPG com a função de densidade representada graficamente na Figura 2.4.

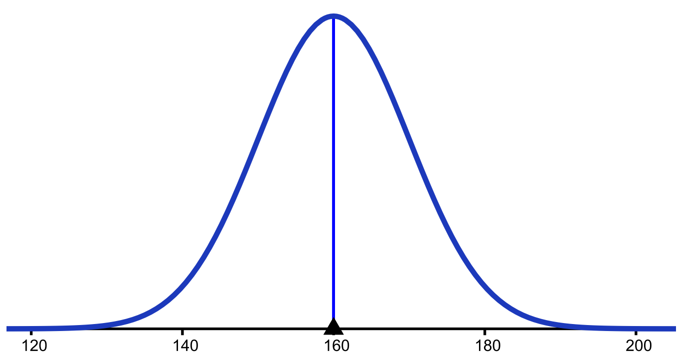

Figura 2.4: Função de densidade de probabilidade da v.a. X.

A probabilidade da altura de um aluno do IPG, escolhido ao acaso, ser inferior ou igual a 170 cm corresponde à área sombreada ($\approx 0,8413$?) que é limitada superiormente pela função densidade, inferiormente pelo eixo das abcissas e por a à direita pela reta $x = 170$, como se mostra na Figura 2.5.

$$P(X \leq 170) = \int_{-\infty}^{170} f(x)dx = \text{área sombreada na Figura 2.5.}$$

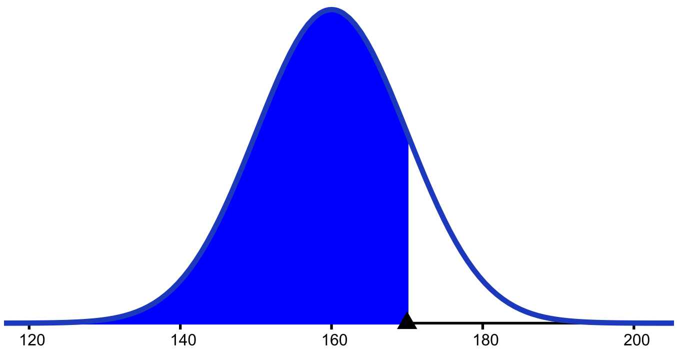

Figura 2.5: Probabilidade de a v.a. tomar valores no intervalo $]-\infty, 175]$

A probabilidade da altura de um aluno do IPG, escolhido ao acaso, estar compreendido entre 170 cm e 175 cm corresponde à área sombreada ($\approx 0,0918$?) que é limitada superiormente pela função de densidade, inferiormente pelo eixo das abcissas à esquerda pela reta $x = 170$ e à direita pela reta $x = 175$ (Figura 2.6).

$$P(170 \leq X \leq 175) = \int_{170}^{175} f(x)dx = \text{área sombreada na Figura 2.6.}$$

A área limitada superiormente pela função densidade e inferiormente pelo eixo das abcissas é igual a 1 (Figura 2.7), pois corresponde à probabilidade de se observar um qualquer valor de altura para um aluno escolhido ao acaso, no universo dos alunos do IPG.

$$P(-\infty < X < +\infty) = 1.$$

2. Dada a seguinte função

$$f(x) = \begin{cases} kx & se \quad 0 < x \leq 1 \\ 0 & se \quad x \leq 0 ou x > 1 \end{cases}$$

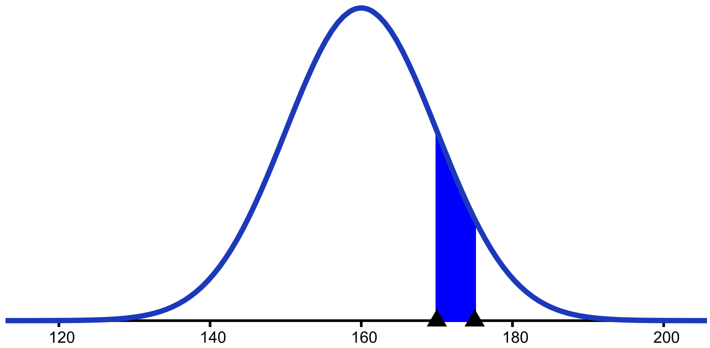

Figura 2.6: Probabilidade de a v.a. tomar valores no intervalo [170, 175]

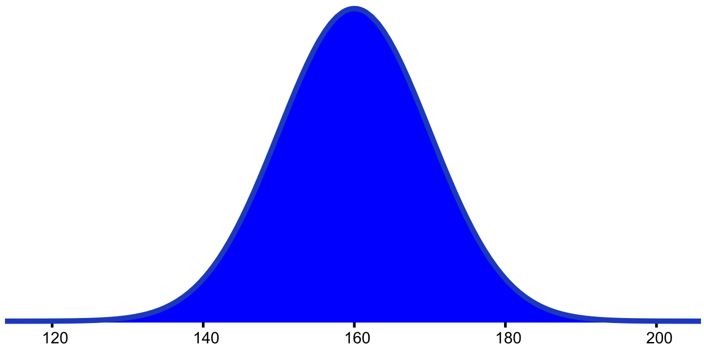

Figura 2.7: Probabilidade total.

a) Determine k de modo que $f(x)$ seja uma função de densidade de probabilidade.
Dada a definição de função de densidade de probabilidade $f(x)$ tem de verificar:

$$f(x) \geq 0 \forall x \in \mathbb{R} \text{ e } \int_{-\infty}^{+\infty} f(x)dx = 1$$

pela última condição tem-se:

$$\begin{array}{l}
\int_{-\infty}^{+\infty} f(x)\,dx = 1 \Leftrightarrow \\
\Leftrightarrow \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} (kx)\,dx + \int_{1}^{+\infty} 0\,dx = 1 \\
\Leftrightarrow k\int_{0}^{1} x\,dx = 1 \\
\Leftrightarrow k\left[\frac{x^2}{2}\right]_0^1 = 1 \\
\Leftrightarrow\Leftrightarrow k\left(\frac{1}{2} - 0\right) = 1 \\
\Leftrightarrow k = 2
\end{array}$$

considerando $k = 2$ na função $f(x)$ é fácil verificar que a primeira condição, ou seja, $f(x) \geq 0, \forall x \in \mathbb{R}$ é satisfeita.

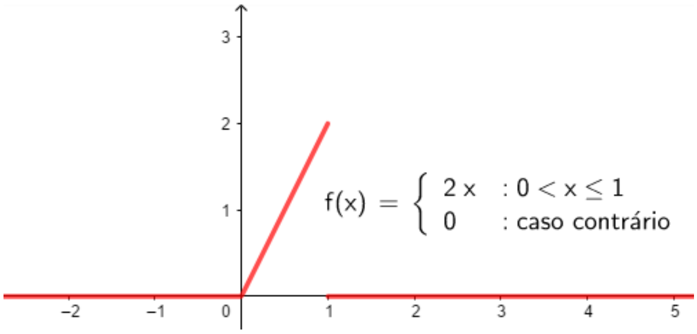

Figura 2.8: Representação gráfica da função de densidade de probabilidade.

b) Determine $P(0 \leq X \leq 1/2)$.

$$P(0 \leq X \leq 1/2) = \int_0^{\frac{1}{2}} (2x)dx = \left[x^2\right]_0^{\frac{1}{2}} = \frac{1}{4}$$

3. Considere a variável aleatória X, cuja função de densidade de probabilidade (f.d.p.) é

$$f(x) = \begin{cases} x & se \quad x \in [0, 1[\\ -x + 2 & se \quad x \in [1, 2] \\ 0 & se \quad x \notin [0, 2] \end{cases}$$

e verifique as propriedades da definição de função de densidade de probabilidade (Figura 2.9).

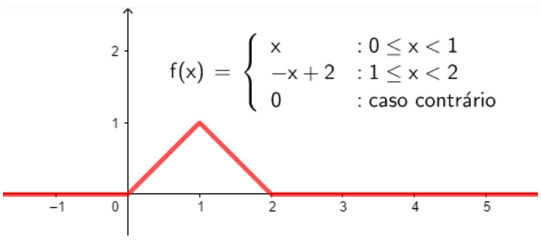

Figura 2.9: Representação gráfica da função de densidade de probabilidade.

Como

$$f(x) = x \geq 0, \forall x \in [0, 1]$$

e

$$f(x) = -x + 2 \geq 0, \forall x \in [1, 2]$$

então
$$f(x) \geq 0 \forall x \in \mathbb{R}$$
ficando verificada a propriedade (i). Quanto à última propriedade tem-se:

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \Leftrightarrow$$
$$\Leftrightarrow \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} x\,dx + \int_{1}^{2} (-x+2)\,dx + \int_{2}^{+\infty} 0\,dx = 1$$
$$\Leftrightarrow \left[\frac{x^2}{2}\right]_0^1 + \left[-\frac{x^2}{2} + 2x\right]_1^2 = 1$$
$$\Leftrightarrow 1 = 1 \text{(Proposição Verdadeira)}.$$

Variáveis Contínuas: Função de Distribuição

Definição 2.1.7 (Função de Distribuição) *seja X uma variável aleatória contínua com função de densidade de probabilidade f. A sua função de distribuição acumulada F está definida para qualquer valor real de x e é dada por*
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du.$$

À semelhança do que se verifica nas v.a. discretas, é usual utilizar-se a letra minúscula f para representar função de probabilidade e a maiúscula F para representar a função de distribuição acumulada ou função de distribuição.

Teorema 2.1.2 *As Propriedades da função de distribuição acumulada F da v.a. contínua X são:*

1. *É uma função contínua.*
2. *Se $x \leq y$ então $F(x) \leq F(y)$, ou seja, a função é crescente em sentido lato.*
3. $$\lim_{x \rightarrowtail -\infty} F(x) = 0.$$
4. $$\lim_{x \rightarrowtail +\infty} F(x) = 1.$$

5. $f(x) = \frac{dF(x)}{dx}$, *se a função de distribuição for derivável.*
6. $P(a \leq X \leq b) = F(b) - F(a)$.

Para qualquer ponto x de uma variável aleatória contínua tem-se

$$P(X = x) = 0,$$

já referido anteriormente, pois

$$P(X = x) = P(]-\infty, x] -]-\infty, x[) = P(X \leq x) - P(X < x) = F(x) - F(x^-)$$

dada a continuidade de $F(x)$ temos

$$\lim_{x \rightarrow x_0^-} F(x) = \lim_{x \rightarrow x_0^+} F(x) = F(x_0)$$

de onde se concluiu que:

$$F(x) - F(x^-) = F(x) - F(x) = 0.$$

Note-se que:

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

que dada a continuidade de $F(x)$ vem

$$F(b) - F(a^-) = F(b) - F(a)$$

também

$$\begin{aligned} P(a < X \leq b) &= P(] - \infty, b] -] - \infty, a]) \\ &= P(] - \infty, b]) - P(] - \infty, a]) \\ &= P(X \leq b) - P(X \leq a) = F(b) - F(a), \end{aligned}$$

de forma análoga temos

$$P(a \leq X < b) = P(a \leq X < b) = F(b) - F(a).$$

Exemplos:

1. Considere a v.a. contínua X com f.d.p.

$$f(x) = \begin{cases} 2x & se \quad 0 < x \leq 1 \\ 0 & se \quad x \leq 0 \text{ ou } x > 1 \end{cases} .$$

a) Defina a função F(x).

Para $x \leq 0$ vem

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(u)\, du \\ &= \int_{-\infty}^{x} 0\, du = 0 \end{aligned}$$

Para $0 < x < 1$ vem

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(u)\, du \\ &= \int_{0}^{x} (2u)\, du \\ &= 2\left[\tfrac{u^2}{2}\right]_0^x = x^2 \end{aligned}$$

Para $x \geq 1$ vem

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(u)\, du \\ &= \int_{-\infty}^{0} 0\, du + \int_{0}^{1} (2u)\, du + \int_{1}^{x} 0\, du \\ &= \textstyle\int_0^1 2u\, du \\ &= 2\left[\tfrac{u^2}{2}\right]_0^1 = 1 \end{aligned}$$

a função pedida é

$$F(x) = \begin{cases} 0 & se \quad x \leq 0 \\ x^2 & se \quad 0 < x < 1 \\ 1 & se \quad x \geq 1 \end{cases} .$$

b) Determine P(1/4<X<3/4).

$$P(1/4 < X < 3/4) = F(3/4) - F(1/4) = \frac{9}{16} - \frac{1}{16} = \frac{8}{16} = \frac{1}{2}.$$

c) Mostre como obteria $f(x)$ a partir de $F(x)$.

$$f(x) = \frac{dF(x)}{dx} = 2x, x \in]0, 1]$$

e

$$f(x) = 0, x \notin]0, 1].$$

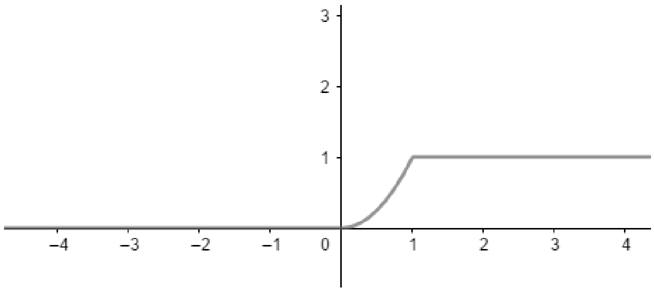

Figura 2.10: Representação gráfica da função dedistribuição acumulada.

2. Numa experiência laboratorial, mede-se a temperatura (^{0}**C**) de uma reacção química. O erro dessa medição é uma variável aleatória contínua X com a seguinte função densidade de probabilidade (Figura 2.11):

$$f(x) = \begin{cases} \frac{3}{4}(1-x^2) & se \quad x \in [-1,1] \\ 0 & se \quad x \notin []-1,1] \end{cases}.$$

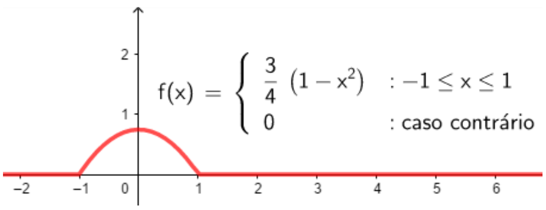

Figura 2.11: Representação gráfica da função de densidade de probabilidade.

a) Verifique que

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

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x)\, dx &= \int_{-\infty}^{-1} 0\, dx + \int_{-1}^{1} \frac{3}{4}\left(1-x^2\right)\, dx + \int_{1}^{+\infty} 0\, dx \\ &= \frac{3}{4}\left[x - \frac{x^3}{3}\right]_{-1}^{1} = 1. \end{aligned}$$

b) Determinar a função de distribuição acumulada de X.

Para $x \leq -1$ vem

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(u)\, du \\ &= \int_{-\infty}^{x} 0\, du = 0 \end{aligned}$$

Para $-1 \leq x \leq 1$ vem

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(u)\, du \\ &= \int_{-1}^{x} \left(\frac{3}{4}\, (1-u^2) \right) du \\ &= \tfrac{3}{4} \left[u - \frac{u^3}{3} \right]_{-1}^{x} \\ &= \frac{2+3x-x^3}{4} \end{aligned}$$

Para $x > 1$ vem

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(u)\, du \\ &= \int_{-\infty}^{-1} 0\, du + \int_{-1}^{1} \left(\frac{3}{4}(1-u^2) \right) du + \int_{1}^{x} 0\, du \\ &= \int_{-1}^{1} \left(\frac{3}{4}\, (1-u^2) \right) du \\ &= \tfrac{3}{4}\, \left[u - \frac{u^3}{3} \right]_{-1}^{1} = 1 \end{aligned}$$

a função pedida é

$$F(x) = \begin{cases} 0 & se \quad x < -1 \\ \frac{2+3x-x^3}{4} & se \quad -1 \leq x \leq 1 \\ 1 & se \quad x > 1 \end{cases}$$

com representação gráfica na Figura 2.12.

c) Determinar $P(0,5 < X \leqslant 1)$.
Utilizando o resultado da alínea anterior temos:

$$P(0,5 < X \leqslant 1) = F(1) - F(0,5) = \frac{5}{32}$$

com representação na Figura 2.13.

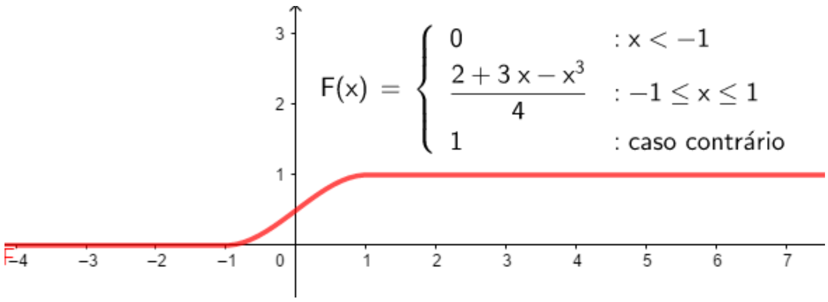

Figura 2.12: Representação gráfica da função de distribuição acumulada.

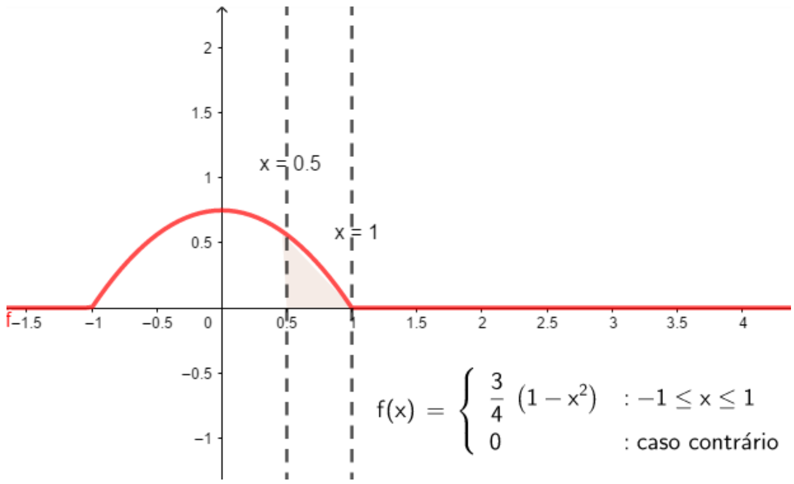

Figura 2.13: Representação $P(0,5 < X \leqslant 1)$.

2.1.2 Média ou Valor Esperado

Associado a cada distribuição de probabilidade podemos associar algumas medidas que dão informação relevante sobre o comportamento das variáveis aleatórias. Uma dessas medidas é a média ou valor esperado ou esperança matemática de uma variável aleatória unidimensional (apenas estamos a tratar as unidimensionais) que é um parâmetro de localização indicando onde a distribuição de probabilidade está centrada.

Definição 2.1.8 (Média ou valor esperado) *seja X uma variável aleatória. A média, valor esperado ou esperança matemátixa de X $E(X)$, representa-se por μ ou μ_X, é tal que:*

i) *se X é uma v.a. discreta então*

$$\mu = \mu_X = E[X] = \sum_x xf(x)$$

ii) *se X é uma v.a. contínua então*

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

Note-se que o valor obtido para o parâmetro média de uma v.a. pode não pertencer ao conjunto de possíveis valores da variável, veja-se o o exemplo que se segue.

Exemplo: Uma *Rent Car* tem uma procura semanal de carros de alta cilindrada que é aleatória e que pode ser descrita pela variável aleatória $X =$"procura semanal de carros com alta cilindrada"com função de probabilidade:

$$f(x) = \begin{cases} 0,25 & se \quad x = 0 \\ 0,35 & se \quad x = 1 \\ 0,3 & se \quad x = 2 \\ 0,1 & se \quad x = 3 \\ 0 & se \quad x \notin \{0,1,2,3\} \end{cases}$$

Pretende-se estimar quantos carros de alta cilindrada, em média, são procurados semanalmente.

Para responder a esta questão vamos calcular a média da variável

$$E[X] = \sum_x x f(x) = 0 \times 0,25 + 1 \times 0,35 + 2 \times 0,3 + 3 \times 0,1 = 1,25.$$

Podemos concluir que a média do número de carros de alta cilindrada procurados por semana é de aproximadamente 1, já que 1,25 não corresponde a um valor assumido pela variável X.

Segue-se um exemplo de calculo da média no caso da variável ser contínua.

Exemplo: De acordo com a especificação de um determinado tipo de lâmpada, a sua duração (em horas) é uma v.a. X com f.d.p. dada por:

$$f(x) = \begin{cases} \frac{1}{60} e^{-\frac{x}{60}} & se \quad x \geq 0 \\ 0 & se \quad x < 0 \end{cases}$$

Que duração média se pode esperar para este tipo de lâmpada?

Vamos determinar a média da variável

$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} x \times f(x)\, dx \\ &= \int_{0}^{+\infty} x \times \left(\frac{1}{60} \times e^{-\frac{x}{60}} \right) dx \\ &= \lim_{t \to +\infty} \left[(-x) \times e^{-\frac{x}{60}} \right]_0^t - \left[60 \times e^{-\frac{x}{60}} \right]_0^t = 60, \end{aligned}$$

que indica uma duração média de 60 horas.

Teorema 2.1.3 *Sejam X v.a. e $g(X)$ uma função com contradomínio em $\mathbb{R}$ com função de probabilidade (X v.a.discreta) ou f.d.p. (X v.a. contínua) f. A média ou valor esperado da v.a. $g(X)$ é:*

i) *se X é uma v.a. discreta então*

$$E[g(X)] = \sum_x g(x)f(x)$$

ii) *se X é uma v.a. contínua então*

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)\, dx.$$

Exemplo: Considerando o exemplo anterior da empresa *Rent Car*, se definirmos a variável $C(x)$ ="custo semanal de manutenção de x carros de alta cilindrada"(em 10^3 u.m.) por:

$$C(x) = \begin{cases} 50 & se \quad x = 0 \\ 100 & se \quad x = 1 \\ 175 & se \quad x = 2 \\ 230 & se \quad x = 3 \end{cases}$$

Então, como sabemos que em média são alugados $E[X] = 1.25$ carros por semana, torna-se imediato calcular o custo médio semanal que a empresa tem com a manutenção deste tipo de carros. Note-se que sendo:

$P(x = 0) = 0.25$ então $P(C(x) = 50) = 0.25$
$P(x = 1) = 0.35$ então $P(C(x) = 100) = 0.35$
$\vdots$

o custo médio semanal de manutenção dos carros de alta cilindrada é

$$E[C(x)] = \sum_x C(x)f(x) = 50 \times 0.25 + 100 \times 0.35 + 175 \times 0.3 + 230 \times 0.1 =$$

$$= 123(10^3 u.m.).$$

Teorema 2.1.4 *Sendo X e Y duas v.a. e K uma constante real, a média ou valor esperado verifica as seguintes propriedades:*

1. $E[K] = K$.
2. $E[KX] = KE[X]$.
3. $E[X \pm Y] = E[X] \pm E[Y]$.
4. $E[X \times Y] = E[X] \times E[Y]$ *se X e Y forem v.a. independentes.*

2.1.3 Variância

A variância de uma v.a. X é uma medida de dispersão (ou variabilidade) em relação à sua média, em que quanto menor for o número de observações da variável que estão afastadas da média menor será a referida dispersão.

Definição 2.1.9 (Variância) *seja X uma variável aleatória. A variância X $Var(X) = V(X)$, representa-se por σ^2 ou σ^2_X, é tal que:*

i) *se X é uma v.a. discreta então*

$$\sigma^2 = \sigma^2_X = Var(X) = V(X) = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x)$$

ii) *se X é uma v.a. contínua então*

$$\sigma^2 = \sigma^2_X = Var(X) = V(X) = E[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx$$

A variância também se pode determinar através do seguinte resultado:

Teorema 2.1.5 *Sendo X uma v.a. a sua variância é:*

$$\sigma^2 = \sigma^2_X = Var(X) = V(X) = E[X^2] - E[X]^2 = E[X^2] - \mu^2.$$

Outra medida de dispersão ou variabilidade da distribuição de uma v.a. é o desvio padrão, que é a raiz quadrada da variância.

Definição 2.1.10 (Desvio Padrão) *seja X uma variável aleatória. O desvio padrão de X σ ou σ_X, é tal que:*

$$\sigma = \sigma_X = \sqrt{E[(X-\mu)^2]} = \sqrt{Var(X)}.$$

As medidas de variabilidade dependem da unidade de medida utilizada, o que dificulta a comparação da variabilidade de diferentes distribuições.
Exemplos:

1. O serviço de cardiologia de um determinado hospital, regulariza a pulsação dos doentes que sofrem de ataque cardíaco com um determinado medicamento. Seja X a v.a. que representa o número de pulsações por minuto de um doente, após o tratamento, e cuja distribuição de probabilidade é dada por:

$$f(x) = \begin{cases} \frac{x-45}{750} & se \quad x \in \{45, 46, \ldots, 75\} \\ \frac{95-x}{500} & se \quad x \in \{76, 77, \ldots, 95\} \end{cases}$$

Determine a variância e o desvio padrão de X.
O valor esperado de pulsação dos doentes sujeitos ao tratamento é:

$$E[X] = \sum_x x f(x) = \frac{215}{3}$$

e

$$\begin{aligned} E[X^2] &= \sum_x x^2 f(x) \\ &= \sum_{45}^{75} x^2 \frac{x-45}{750} + \sum_{76}^{95} x^2 \frac{95-x}{500} \\ &= \tfrac{10483}{2} \end{aligned}$$

sendo

$$\sigma^2 = \sigma_X^2 = Var(X) = E[X^2] - E[X]^2 = \frac{1897}{18} \approx 105$$

e o desvio é:

$$\sigma = \sqrt{Var(X)} = 10.3.$$

2. Sendo X uma v.a contínua com f.d.p dada por

$$f(x) = \begin{cases} x - \frac{1}{2} & se \quad 1 \leq x \leq 2 \\ 0 & se \quad x < 1 \vee x > 2 \end{cases}$$

Determine a variância e o desvio padrão de X.
O valor esperado é:

$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} x \times f(x)\, dx \\ &= \int_1^2 x \times \left(x - \frac{1}{2}\right) dx \\ &= \left[\tfrac{x^3}{3}\right]_1^2 - \tfrac{1}{2}\left[\tfrac{x^2}{2}\right]_1^2 \\ &= \tfrac{19}{12} \end{aligned}$$

e

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{+\infty} \left[x^2 \times f(x)\right] dx \\ &= \int_1^2 \left[x^2 \times (x - \frac{1}{2})\right] dx \\ &= \int_1^2 \left(x^3 - \frac{x^2}{2}\right) dx \\ &= \left[\tfrac{x^4}{4}\right]_1^2 - \tfrac{1}{2}\left[\tfrac{x^3}{3}\right]_1^2 \\ &= \tfrac{62}{24} \end{aligned}$$

sendo a variância

$$\sigma^2 = \sigma_X^2 = Var(X) = E[X^2] - E[X]^2 = \frac{62}{24} - \frac{361}{144} = \frac{11}{144}$$

e o desvio padrão

$$\sigma = \sqrt{Var(X)} = \sqrt{\frac{11}{144}}.$$

Teorema 2.1.6 *Sendo X uma v.a. e a, b duas constantes reais, a variância verifica as seguintes propriedades:*

1. $Var(aX) = a^2 Var(x)$.
2. $Var(a) = 0$.
3. $Var(aX + b) = a^2 Var(X)$.

2.2 Variáveis Aleatórias Bidimensionais

Existem experiências aleatórias em que se pretende estudar mais do que um resultado, por exemplo pode ter interesse estudar o salário e o tempo de serviço dos operários numa fábrica. Para tal, teremos uma variável X que representa o salário e uma v.a Y que representa o tempo de serviço, ou seja, uma variável bidimensional. Representando as duas variáveis aleatórias por (X, Y), a distribuição de probabilidade deste par aleatório corresponde à verosimilhança associada aos valores possíveis de (X, Y), cuja descrição formal para as v.a. discretas é diferente das v.a. contínuas.

2.2.1 Variáveis Aleatórias Discretas

Definição 2.2.1 (Função de Probabilidade Conjunta) *sejam X e Y variáveis aleatórias discretas, a função de probabilidade conjunta de (X, Y) é a função f_{XY} que a cada par ordenado (x, y) de (X, Y) associa a sua probabilidade*

$$f_{XY}(x, y) = P(X = x, Y = y).$$

Esta função verifica

i)

$$f_{XY}(x, y) \geq 0$$

ii)

$$\sum_x \sum_y f_{XY}(x,y) = 1.$$

Exemplo: Duas linhas de produção, linha I e II, fabricam uma determinada componente electrónica. A capacidade máxima de produção por minuto de ambas as linhas é 3 componentes. Definindo a v.a. X para representar o número de peças produzidas na linha I e a Y o número de peças produzidas na linha II, a função de probabilidade conjunta é:

Y|X	0	1	2	3
0	0	0.04	0.09	0.12
1	0.01	0.06	0.08	0.11
2	0.01	0.08	0.08	0.08
3	0.01	0.06	0.09	0.08

Interessa trabalhar a variável bidimensional (X, Y) que descreve a ocorrência simultânea da produção da linhas I e II. A variável bidimensional (X,Y) pode assumir todos os pares ordenados possíveis de construir com os valores de X e de Y. Em que, por exemplo, o par (2,1) corresponde à produção de 2 peças na linha I e 1 peça na linha II, num minuto. Dada a função de probabilidade conjunta apresentada anteriormente pode-se saber, por exemplo, a probabilidade de ocorrência simultânea de $X = 2$ e $Y = 1$ ou seja $P(X = 2, Y = 1) = 0.08$.

1. A probabilidade de a linha I produzir mais peças do que a linha II será

$$\begin{aligned} P(X > Y) &= \sum_{y=0}^{2} \sum_{x=y+1}^{3} f_{XY}(x,y) \\ &= (0.04 + 0.09 + 0.12) + (0.08 + 0.11) + 0.08 = 0.52 \end{aligned}$$

2. A probabilidade de a linha I produzir tantas peças como a II é dada por

$$\begin{aligned} P(X = Y) &= \sum_{x=y=0}^{3} f_{XY}(x,y) \\ &= 0 + 0,06 + 0,08 + 0,08 = 0,22 \end{aligned}$$

3. A probabilidade da linha I produzir menos peças que a linha II é

$$P(X < Y) = 1 - P(X \geq Y) = 1 - (0,22 + 0,52)$$

4. A probabilidade de a linha I produzir menos de 2 peças, e a linha II produzir 2 ou 3 peças é dada por

$$\begin{aligned} P(0 \leq X < 2, 2 \leq Y \leq 3) &= \sum_{x=0}^{1}\sum_{y=2}^{3} f_{XY}(x,y) \\ &= 0,01 + 0,01 + 0,08 + 0,06 = 0,16. \end{aligned}$$

Genericamente, admitindo que um v.a. X discreta toma valores $\{x_1, x_2, \ldots, x_n\}$ e uma v.a. Y discreta toma valores $\{y_1, y_2, \ldots, y_m\}$ então a função de probabilidade conjunta $f_{XY}(x,y) = f(x,y)$ pode representar-se através de um quadro de dupla entrada:

$X\|Y$	y_1	y_2	$\cdots$	y_m	Totais
x_1	$f(x_1,y_1)$	$f(x_2,y_2)$	$\cdots$	$f(x_1,y_m)$	$f_X(x_1)$
x_2	$f(x_2,y_1)$	$f(x_2,y_2)$	$\cdots$	$f(x_2,y_m)$	$f_X(x_2)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_n	$f(x_n,y_1)$	$f(x_n,y_2)$	$\cdots$	$f(x_n,y_m)$	$f_X(x_n)$
Totais	$f_Y(y_1)$	$f_Y(y_2)$	$\cdots$	$f_Y(y_m)$	1

onde $f_X(x)$ representa a função de probabilidade marginal de X e $f_Y(y)$ a função de probabilidade marginal de Y.

Definição 2.2.2 (Funções de Probabilidade Marginais) *sejam X e Y variáveis aleatórias discretas com função de probabilidade conjunta f_{XY}, a função de probabilidade marginal de X são:*

$$f_X(x) = \sum_y f_{XY}(x,y)$$

e a função de probabilidade marginal de Y são:

$$f_Y(y) = \sum_x f_{XY}(x,y).$$

Pela definição tem-se:

$$\begin{aligned} f_X(x_i) &= P\left(X = x_i, -\infty < Y < +\infty\right) \\ &= \sum_y f(x_i, y) \\ &= f(x_i,y_1) + f(x_i,y_2) + \ldots + f(x_i,y_m) \\ f_Y(y_j) &= P\left(-\infty < X < +\infty, Y = y_j\right) \\ &= \sum_x f(x, y_j) \\ &= f(x_1,y_j) + f(x_2,y_j) + \ldots + f(x_n,y_j) \end{aligned}$$

para algum $i \in \{1, \ldots, n\}$ e $j \in \{1, \ldots, m\}$.
Exemplo: Considerando o exemplo anterior, a função de probabilidade marginal de X , $f_X(x)$, é

x	0	1	2	3	Total
$f_X(x)$	0,03	0,24	0,34	0,39	1

pois,

$$f_X(0) = \sum_{y=0}^{3} f(0, y) = f(0, 0) + f(0, 1) + f(0, 2) + f(0, 3) = 0,03$$

$$f_X(1) = \sum_{y=0}^{3} f(1, y) = f(1, 0) + f(1, 1) + f(1, 2) + f(1, 3) = 0,24$$

$$\vdots$$

e a função de probabilidade marginal de Y , $f_Y(y)$, é pois,

y	0	1	2	3	Total
$f_Y(y)$	0,25	0,26	0,25	0,24	1

$$f_Y(0) = \sum_{x=0}^{3} f(x, 0) = f(0, 0) + f(1, 0) + f(2, 0) + f(3, 0) = 0,25$$

$$f_Y(1) = \sum_{x=0}^{3} f(x, 1) = f(0, 1) + f(1, 1) + f(2, 1) + f(3, 1) = 0,26$$

$$\vdots$$

Função de Distribuição Acumulada Conjunta

Segue-se a definição de função de distribuição acumulada conjunta de uma variável aleatória bidimensional (X,Y) discreta.

Definição 2.2.3 (Função de Distribuição Acumulada Conjunta) *sejam X e Y variáveis aleatórias discretas com função de probabilidade conjunta*

f_{XY}, a função de distribuição acumulada conjunta de (X, Y) está definida para todo o par ordenado $(x, y) \in \mathbb{R}^2$ e é dada por:

$$F_{XY}(x,y) = P(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} f_{XY}(u,v)$$

onde u e v tomam todos os valores possíveis das v. a. X e Y não superiores a x e a y, respectivamente.

Exemplo: Continuando com o mesmo exemplo, vamos deduzir a função de distribuição conjunta para o par (X, Y):

$Y \mid X$	$X < 0$	$0 \leq X < 1$	$1 \leq X < 2$	$2 \leq X < 3$	$X \geq 3$
$Y < 0$	0	0	0	0	0
$0 \leq Y < 1$	0	0	0.04	0.13	0.25
$1 \leq Y < 2$	0	0,01	0,11	0,28	0,51
$2 \leq Y < 3$	0	0,02	0,2	0,45	0,76
$Y \geq 3$	0	0,03	0,27	0,61	1

$$
\begin{aligned}
F_{XY}(0,0) &= P(X \leq 0, Y \leq 0) = 0 \\
F_{XY}(0,1) &= P(X \leq 0, Y \leq 1) \\
&= f(0,0) + f(0,1) \\
&= 0 + 0.01 \\
&= 0.01 \\
F_{XY}(0,2) &= P(X \leq 0, Y \leq 2) \\
&= f(0,0) + f(0,1) + f(0,2) \\
&= 0 + 0,01 + 0,01 \\
&= 0,02 \\
F_{XY}(0,3) &= P(X \leq 0, Y \leq 3) \\
&= f(0,0) + f(0,1) + f(0,2) + f(0,3) \\
&= 0 + 0.01 + 0.01 + 0.01 \\
&= 0.03 \\
F_{XY}(1,0) &= P(X \leq 1, Y \leq 0) \\
&= f(0,0) + f(1,0) \\
&= 0.04 \\
F_{XY}(1,1) &= P(X \leq 1, Y \leq 1) \\
&= f(0,0) + f(0,1) + f(1,0) + f(1,1) \\
&= 0.11 \\
F_{XY}(1,2) &= P(X \leq 1, Y \leq 2) \\
&= f(0,0) + f(0,1) + f(0,2) + f(1,0) + f(1,1) + f(1,2) \\
&= 0.2 \\
F_{XY}(1,3) &= P(X \leq 1, Y \leq 3) \\
&= f(0,0) + f(0,1) + f(0,2) + f(0,3) + f(1,0) + f(1,1) \\
&\quad + f(1,2) + f(1,3) \\
&= 0,27 \\
F_{XY}(2,0) &= P(X \leq 2, Y \leq 0) \\
&= f(0,0) + f(1,0) + f(2,0) \\
&= 0,13 \\
F_{XY}(2,1) &= P(X \leq 2, Y \leq 1) \\
&= f(0,0) + f(0,1) + f(1,0) + f(1,1) + f(2,0) + f(2,1) \\
&= 0,28 \\
F_{XY}(2,2) &= P(X \leq 2, Y \leq 2) \\
&= f(0,0) + f(0,1) + f(0,2) + f(1,0) + f(1,1) + f(1,2) \\
&\quad + f(2,0) + f(2,1) + f(2,2) \\
&= 0,45
\end{aligned}
$$

$$\begin{aligned}
F_{XY}(2,3) &= P(X \leq 2, Y \leq 3) \\
&= f(0,0)+f(0,1)+f(0,2)+f(0,3)+f(1,0)+f(1,1) \\
&\quad +f(1,2)+f(1,3)+f(2,0)+f(2,1)+f(2,2)+f(2,3) \\
&= 0,61 \\
F_{XY}(3,0) &= P(X \leq 3, Y \leq 0) \\
&= f(0,0)+f(1,0)+f(2,0)+f(3,0) \\
&= 0,25 \\
F_{XY}(3,1) &= P(X \leq 3, Y \leq 1) \\
&= f(0,0)+f(0,1)+f(1,0)+f(1,1)+f(2,0)+f(2,1) \\
&\quad +f(3,0)+f(3,1) \\
&= 0,51 \\
F_{XY}(3,2) &= P(X \leq 3, Y \leq 2) \\
&= f(0,0)+f(0,1)+f(0,2)+f(1,0)+f(1,1)+f(1,2)+ \\
&\quad +f(2,0)+f(2,1)+f(2,2)+f(3,0)+f(3,1)+f(3,2) \\
&= 0,76 \\
F_{XY}(3,3) &= P(X \leq 3, Y \leq 3) = 1.
\end{aligned}$$

Definição 2.2.4 (Função de Distribuição Acumulada Marginal) *sejam X e Y variáveis aleatórias discretas com função de probabilidade conjunta f_{XY}, a função de distribuição acumulada marginal de X é:*

$$F_X(x) = \sum_{u \leq x} \sum_{y} f_{XY}(u,y)$$

e a função de distribuição acumulada marginal de Y é:

$$F_Y(y) = \sum_{x} \sum_{v \leq y} f_{XY}(x,v).$$

2.2.2 Variáveis Aleatórias Contínuas

Definição 2.2.5 (Função de Densidade Conjunta) *sejam X e Y variáveis aleatórias contínuas, a função de probabilidade conjunta de (X,Y) é a função f_{XY} dada por*

$$f_{XY}(x,y) = P((X,Y) \in A) = \int\int_A f_{XY}(x,y)\, dxdy,$$

onde A é uma região do espaço bidimensional. Esta função verifica

i)

$$f_{XY}(x,y) \geq 0\ ,\quad \forall (x,y) \in \mathbb{R}^2$$

ii)

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)\, dxdy = 1.$$

Geometricamente $z = f_{XY}(x,y)$ representa uma superficie, chamada superficie de probabilidade em que o volume total delimitado por esta superficie e pelo plano XOY é 1. Pela definição anterior,

$$P\left(a \leq X \leq b, c \leq Y \leq d\right) = \int_c^d \int_a^b f_{XY}(x,y)\, dxdy.$$

Exemplo: Uma fábrica de produtos alimentares produz cereais de milho e aveia cobertos ou não com mel, e que são misturados e embalados em caixas. As v.a. X e Y representam, respectivamente, as proporções de cereais de milho e aveia, cobertos com mel, existentes numa caixa seleccionada ao acaso. A função de densidade de probabilidade conjunta é:

$$f_{XY}(x,y) = \begin{cases} x+y & se \quad x \in [0,1] \wedge y \in [0,1] \\ 0 & \text{outros valores de x} \end{cases}$$

1. Determine $P(0 < X < 1/2, 1/4 < Y < 1/2)$.

$$\begin{aligned} P(0 < X < 1/2, 1/4 < Y < 1/2) &= \int_{1/4}^{1/2}\int_0^{1/2} (x+y)\, dxdy \\ &= \int_{1/4}^{1/2}\left[\frac{x^2}{2} + xy\right]_0^{1/2} dy \\ &= \int_{1/4}^{1/2}\left(\frac{1}{8} + \frac{y}{2}\right) dy = \frac{5}{64}. \end{aligned}$$

2. Verifique a condição $ii)$ da definição de função de densidade conjunta.

$$\begin{aligned} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)\, dxdy &= \int_0^1\int_0^1 (x+y)\, dxdy \\ &= \int_0^1 \left[\frac{x^2}{2} + xy\right]_0^1 dy \\ &= \int_0^1 \left(\frac{1}{2} + y\right) dy = 1. \end{aligned}$$

As funções de densidade de probabilidade marginais de X e Y obtêm-se da função de densidade de probabilidade conjunta e correspondem à função de densidade se as variáveis forem tratadas como unidimensionais.

Definição 2.2.6 (Funções de Densidade de Probabilidade Marginais) *sejam X e Y variáveis aleatórias contínuas com função de densidade de probabilidade conjunta f_{XY}, a função de densidade de probabilidade marginal de X são:*

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy$$

e a função de densidade de probabilidade marginal de Y são:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx.$$

Exemplo: Considerando o exemplo da fábrica de produtos alimentares, vamos determinar as funções de densidade de probabilidade marginais de X e Y. Considerando a definição anterior tem-se:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy = \int_0^1 (x+y)dy = x + \frac{1}{2}$$

para $0 \leq x \leq 1$ e $f_X(x) = 0$ para os restantes valores de x;

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx = \int_0^1 (x+y)dx = y + \frac{1}{2}$$

para $0 \leq y \leq 1$ e $f_X(x) = 0$ para os restantes valores de y.

Função de Distribuição Acumulada Conjunta

Segue-se a definição de função de distribuição acumulada conjunta de uma variável aleatória bidimensional (X,Y) contínua.

Definição 2.2.7 (Função de Distribuição Acumulada Conjunta) *sejam X e Y variáveis aleatórias contínuas com função de densidade de probabilidade conjunta f_{XY}, a função de distribuição acumulada conjunta de (X, Y) é:*

$$F_{XY}(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u,v)\, dvdu$$

que está definida para todo o par ordenado $(x, y) \in \mathbb{R}^2$.

Prova-se que se pode calcular $P(x_1 \leq X \leq x_2; y_1 \leq Y \leq y_2)$ através de:

$$P(x_1 \leq X \leq x_2; y_1 \leq Y \leq y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(u,v)\, dvdu$$
$$= F_{XY}(x_2,y_2) - F_{XY}(x_2,y_1) - F_{XY}(x_1,y_2) + F_{XY}(x_1,y_1).$$

Teorema 2.2.1 *sejam X e Y variáveis aleatórias contínuas com função de densidade de probabilidade conjunta f_{XY}. Se*

$$\frac{\partial^2 F_{XY}(x,y)}{\partial_x \partial_y}$$

existir então

$$f_{XY} = \frac{\partial^2 F_{XY}(x,y)}{\partial_x \partial_y}.$$

Exemplo: Considerando uma v.a. bidimensional contínua (X,Y) com função de densidade de probabilidade conjunta:

$$f_{XY}(x,y) = \begin{cases} xy & se \quad x \in [0,2] \wedge y \in [0,1] \\ 0 & \text{outros valores de x} \end{cases}$$

Vamos determinar a função de distribuição acumulada conjunta de (X,Y).

Para $x < 0 \vee y < 0$:

$$\begin{aligned} F_{XY}(x,y) &= P(X \leq x, Y \leq y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u,v)\, dvdu = 0 \end{aligned}$$

Para $0 \leq x \leq 2 \wedge 0 \leq y \leq 1$:

$$\begin{aligned} F_{XY}(x,y) &= F_{XY}(0,0) + \int_{0}^{x} \int_{0}^{y} (uv)\, dvdu \\ &= 0 + \int_{0}^{x} u \left[\frac{v^2}{2}\right]_{0}^{y} du \\ &= \int_{0}^{x} u\frac{y^2}{2}\, du = \frac{y^2}{2}\left[\frac{u^2}{2}\right]_{0}^{x} \\ &= \frac{1}{4}x^2y^2 \end{aligned}$$

Para $0 \leq x \leq 2 \wedge y > 1$:

$$\begin{aligned} F_{XY}(x,y) &= \int_{0}^{x} \int_{0}^{1} (uv)\, dvdu \\ &= F_{XY}(x,1) = \frac{1}{4}x^2 1^2 \\ &= \frac{1}{4}x^2 \Leftrightarrow F_X(x) \end{aligned}$$

Para $x > 2 \wedge 0 \leq y \leq 1$:

$$\begin{aligned} F_{XY}(x,y) &= \int_0^2 \int_0^y (uv)\, dvdu \\ &= F_{XY}(2,y) \\ &= \frac{1}{4}2^2y^2 = y^2 \Leftrightarrow F_Y(y) \end{aligned}$$

Para $x > 2 \wedge y > 1$:

$$\begin{aligned} F_{XY}(x,y) &= \int_0^2 \int_0^1 (uv)\, dvdu \\ &= F_{XY}(2,1) = \frac{1}{4}2^2 1^2 = 1 \end{aligned}$$

Assim,

$$F_{XY}(x,y) = \begin{cases} 0 & se \quad x < 0 \vee y < 0 \\ \frac{1}{4}x^2y^2 & se \quad 0 \leq x \leq 2 \wedge 0 \leq y \leq 1 \\ \frac{1}{4}x^2 & se \quad 0 \leq x \leq 2 \wedge y > 1 \\ \frac{1}{4}2^2y^2 = y^2 & se \quad x > 2 \wedge 0 \leq y \leq 1 \\ 1 & se \quad x > 2 \wedge y > 1 \end{cases}$$

Definição 2.2.8 (Função de Distribuição Acumulada Marginal) *sejam X e Y variáveis aleatórias contínuas com função de densidade de probabilidade conjunta f_{XY}, a função de distribuição acumulada marginal de X é:*

$$F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f_{XY}(u,y)dy\, du = \int_{-\infty}^{x} f_X(u)\, du$$

ou ainda

$$F_X(x) = \lim_{y \to +\infty} F_{XY}(x,y)$$

e a função de distribuição acumulada marginal de Y é:

$$F_Y(y) = \int_{-\infty}^{y} \int_{-\infty}^{+\infty} f_{XY}(x,v)\, dxdv = \int_{-\infty}^{y} f_Y(v)\, dv$$

ou ainda

$$F_Y(y) = \lim_{x \to +\infty} F_{XY}(x,y)$$

Exemplo: Considerando o exemplo anterior, vamos determinar as funções de distribuição acumulada marginais.

Para $x < 0 \wedge y > 1$:

$$F_X(x) = \int_{-\infty}^{x} \int_{0}^{1} 0 \; dvdu = 0$$

Para $0 \leq x \leq 2 \wedge y > 1$:

$$F_X(x) = F_{XY}(x, 1) = \int_{-\infty}^{x} \int_{0}^{1} (uv) \; dvdu = \frac{1}{4}x^2$$

Para $x > 2 \wedge y > 1$:

$$F_X(x) = F_{XY}(2, 1) = \int_{0}^{2} \int_{0}^{1} (uv) \; dvdu = 1$$

Assim,

$$F_X(x) = \begin{cases} 0 & se \quad x < 0 \\ \frac{1}{4}x^2 & se \quad 0 \leq x \leq 2 \\ 1 & se \quad x > 2 \end{cases}$$

De forma análoga temos:

Para $y < 0 \wedge x > 2$:

$$F_Y(y) = 0$$

Para $0 \leq y \leq 1 \wedge x > 2$:

$$F_Y(y) = F_{XY}(2, y) = \int_{0}^{2} \int_{-\infty}^{y} (uv) \; dvdu = y^2$$

Para $y > 1 \wedge x > 2$:

$$F_Y(y) = F_{XY}(2, 1) = \int_{0}^{2} \int_{0}^{1} (uv) \; dvdu) = 1$$

Assim,

$$F_Y(y) = \begin{cases} 0 & se \quad y < 0 \\ y^2 & se \quad 0 \leq y \leq 1 \\ 1 & se \quad x > 1 \end{cases} .$$

Como exercício, determine estas funções utilizando directamente as funções de densidade de probabilidade marginais.

Teorema 2.2.2 *Sejam X e Y variáveis aleatórias contínuas, as funções de densidade marginais podem obter-se através das funções de distribuição marginais, por derivação. Se*

$$\frac{\partial F_X(x)}{\partial_x}$$

existir então

$$f_X(x) = \frac{\partial F_X(x)}{\partial_x}.$$

Se

$$\frac{\partial F_Y(y)}{\partial_y}$$

existir então

$$f_Y(y) = \frac{\partial F_Y(y)}{\partial_y}.$$

Exemplo: Continuando com o exemplo anterior, vamos determinar as funções de densidade marginais utilizando o teorema anterior.

$$f_X(x) = \frac{\partial F_X(x)}{\partial_x} = \frac{1}{2}\, x$$

se $0 \leq x \leq 2$ e temos

$$f_X(x) = 0$$

para os restantes valores de x e

$$f_Y(y) = \frac{\partial F_Y(y)}{\partial_y} = 2y$$

se $0 \leq y \leq 1$ e temos

$$f_Y(y) = 0$$

para os restantes valores de y. Verifique que as funções encontradas também se podem obter por

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)\, dy$$

e

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y)\, dx.$$

2.2.3 Independência de Variáveis Aleatórias

Tendo em conta o conceito já introduzido de independência de dois acontecimentos ou eventos, torna-se fácil perceber o conceito de variáveis aleatórias independentes dado que são análogos. Duas variáveis aleatórias são independentes se o resultado de uma variável não influencia o resultado da outra . Do comportamento conjunto de duas variáveis aleatórias, X e Y, pode inferir-se acerca da independência entre elas.

Definição 2.2.9 (Variáveis Aleatórias Independentes) *As X e Y variáveis aleatórias discretas (ou contínuas) com função de probabilidade conjunta (ou função de densidade de probabilidade conjunta) f_{XY}, são independentes se e só se*

$$f_{XY}(x,y) = f_X(x) f_Y(y) \ , \ \ \forall (x,y) \in \mathbb{R}^2.$$

Note-se que, no caso das variáveis bidimensionais discretas, basta que a igualdade da definição anterior não se verifique para um par de valores (x,y) para que X e Y não sejam independentes.

Exemplos:

1. Considere os doentes que realizam tratamentos de emagrecimento. Sejam X e Y duas variáveis aleatórias descritas pela função de densidade de probabilidade conjunta:

$$f_{XY}(x,y) = \begin{cases} \frac{8}{3}(1-xy) & se \quad 0 \leq x \leq y \leq 1 \\ 0 & \text{outros valores de x e y} \end{cases}$$

e que representam, respectivamente, a fracção de doentes sem curso superior e a fracção de doentes com curso superior que têm sucesso nesses tratamentos.

a) Determine $f_X(x)$ e $f_Y(y)$.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{XY}(x,y)\, dy \\ &= \int_x^1 \left(\frac{8}{3}(1-xy)\right)\, dy \\ &= \frac{8}{3}\left(1 - \frac{3}{2}x + \frac{1}{2}x^3\right) \ , \ \ 0 \leq x \leq 1 \end{aligned}$$

e

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dx \\ &= \int_0^y \left(\frac{8}{3}(1-xy)\right)\,dy \\ &= \frac{8}{3}\left(y-\frac{1}{2}y^3\right)\ ,\ \ 0 \le y \le 1 \end{aligned}$$

.

b) Verifique que X e Y não são independentes.
Na alínea anterior determinamos as funções de densidade marginais e sem proceder ao produto

$$f_X(x)f_Y(y)$$

é fácil concluir

$$f_X(x)f_Y(y) \neq f_{XY}(x,y)$$

pelo que X e Y são v.a. dependentes. Poderíamos também justificar a dependência das variáveis utilizando o domínio da f.d.p conjunta, pois dado um qualquer valor $Y = y$ a v.a. X só pode tomar valores inferiores a y, logo os valores que X pode assumir dependem dos valores da variável Y.

2. Considerando o exemplo das duas linhas de produção de uma determinada componente electrónica, averigúe a independência das duas variáveis.

   Considerando o par (2,3) constatamos que $f_{XY}(2,3) = 0,09$ e $f_X(2) = 0,34$ e $f_Y(3) = 0,24$ pelo que

   $$f_X(x)f_Y(y) = 0,34 \times 0,24 = 0,0816 \neq f_{XY}(2,3) = 0,09.$$

   Assim, conclui-se que as v.a. não são independentes, ou seja, são dependentes.

2.2.4 Covariância e Coeficiente de correlação

A covariância entre duas variáveis aleatórias é uma medida da relação linear entre elas. Pode ainda dizer-se que a covariância, das v.a X e Y, é uma medida de distribuição conjunta dos valores de X e Y em termos de desvios em relação às respectivas médias. O sinal da covariância indica se a relação linear das variáveis é positiva (as variáveis variam na mesma direcção) ou negativa (as variáveis variam em direcções opostas).

Definição 2.2.10 (Covariância) *A covariância entre as variáveis X e Y é a constante:*

$$\begin{aligned} cov(X,Y) &= \sigma_{XY} \\ &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f_{XY}(x,y) \end{aligned}$$

se as v.a. são discretas, e

$$\begin{aligned} cov(X,Y) &= \sigma_{XY} \\ &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x-\mu_X)(y-\mu_Y) f_{XY}(x,y)\, dx\, dy \end{aligned}$$

se as v.a. são contínuas.

A covariância também pode ser determinada por:

Teorema 2.2.3 *A covariância entre as variáveis X e Y é a constante:*

$$cov(X,Y) = \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X\mu_Y.$$

Se X e Y são independentes $cov(X,Y) = 0$ (em geral, o inverso não é verdadeiro).

A covariância depende das unidades em que as v.a. forem medidas, se estas unidades mudarem também muda a covariância. Assim, a covariância dá-nos o sinal da relação das v.a., mas não nos dá indicação da intensidade. Para obtermos esta informação utiliza-se o coeficiente de correlação, que é uma medida independente das unidades utilizadas nas v.a..

Definição 2.2.11 (Coeficiente de Correlação) *O coeficiente de correlação entre as variáveis X e Y é a constante:*

$$corr(X,Y) = \rho_{XY} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sqrt{var(X)var(Y)}} = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}.$$

O valor deste coeficiente varia entre -1 e 1.

Exemplo: Calcule a covariância entre as v.a.

X="procura de aviões para executivos"
Y="procura de aviões para transporte de correio rápido",

$Y\|X$	0	1	2	3	$f_Y(y)$
0	0.05	0.07	0.06	0.02	
1	0.125	0.175	0.15	0.05	
2	0.075	0.105	0.09	0.03	
$f_X(x)$					1

sendo a função de probabilidade conjunta dada por:

$$cov(X,Y) = \sigma_{XY} = E[XY] - \mu_X\mu_Y.$$

sendo

$$E[XY] = \sum_x \sum_y xyf_{XY}(x,y) = 1.375$$

e

$$E[X] = \sum_x xf_X(x) = 1.25$$

$$E[Y] = \sum_y yf_Y(y) = 1.1$$

logo

$$cov(X,Y) = \sigma_{XY} = 1.375 - 1.25 \times 1.1 = 0$$

o que seria de esperar dado que as v.a. são independentes.

2.3 Momentos

Seguem-se as definições de momentos populacionais.

Definição 2.3.1 (Momento Simples de Ordem r) *O momento simples de ordem r (r=0,1,2,...) é a constante*

$$\mu_r^{'} = E\left[X^r\right] = \sum_x x^r f(x),$$

sendo X uma v.a. discreta e

$$\mu_r^{'} = E\left[X^r\right] = \int_{-\infty}^{+\infty} x^r f(x)dx,$$

se X uma v.a. contínua.

Assim, o momento simples de ordem r é a média da variável X^r. Quando $r=1$ tem-se a média de X, $E[X]=\mu$ e quando $r=0$ o momento simples é 1.

Definição 2.3.2 (Momento Centrado de Ordem r) *O momento centrado de ordem r (r=0,1,2,...) é a constante*

$$\mu_r = E\left[(X-\mu)^r\right] = \sum_x (x-\mu)^r f(x),$$

sendo X uma v.a. discreta e

$$\mu_r = E\left[(X-\mu)^r\right] = \int_{-\infty}^{+\infty} (x-\mu)^r f(x)dx,$$

se X uma v.a. contínua.

A variância é um momento centrado de ordem 2

$$\sigma^2 = E\left[(X-\mu)^2\right] = \mu_2,$$

no teorema seguinte apresentam-se duas propriedades verificadas por este momento centrado.

Teorema 2.3.1 *Sendo $X e Y$ v.a. discretas (ou contínuas) verifica-se:*

i) $Var(X \pm Y) = Var(X) + Var(Y) \pm 2\, cov(X,Y)$;

ii) *Se X e Y são v.a. independentes então*

$$Var(X \pm Y) = Var(X) + Var(Y).$$

2.4 Variáveis Aleatórias Estandardizadas

Quando se pretende comparar distribuições de duas ou mais variáveis aleatórias com unidades diferentes, por exemplo quando uma das variáveis representa uma distância e a outra um tempo, não estando relacionadas. Para conseguir efectuar a referida comparação torna-se necessário estandardizar ou padronizar as variáveis aleatórias em causa.

Definição 2.4.1 (V.A. Estandardizada) *Seja X uma variável aleatória com média μ_X e desvio padrão σ_X, a variável aleatória estandardizada é*

$$U = \frac{X-\mu_X}{\sigma_X}.$$

A variável U é a forma estandardizada de X.

Este procedimento analítico pode facilitar o cálculo de probabilidades, como veremos posteriormente. Com base na definição apresentada, onde a variável aleatória estandardizada é uma função linear da variável aleatória original, prova-se que uma variável aleatória estandardizada tem média igual a 0 e desvio padrão igual a 1 e é adimensional, dado que o numerador e denominador estão na mesma unidade de medida.

Teorema 2.4.1 *Uma variável aleatória estandardizada*

$$U = \frac{X - \mu_X}{\sigma_X}$$

tem média

$$\mu_U = E(U) = E\left[\frac{X - \mu_X}{\sigma_X}\right] = 0$$

e variância

$$\sigma_U = Var\left(U\right) = Var\left[\frac{X - \mu_X}{\sigma_X}\right] = 1.$$

2.5 Desigualdade de Chebyshev

Com esta desigualdade é possível obter uma estimativa da probabilidade de uma variável aleatória tomar um valor distanciado da sua média em pelo menos k desvios padrões, mesmo quando não se conhece a distribuição da variável aleatória.

Considere-se o seguinte resultado:

Teorema 2.5.1 *Seja X uma variável aleatória e $h(x)$ uma função dessa variável aleatória tal que $h(x)$ é não negativa. Se existe $E[h(x)]$ então:*

$$P[h(x) \geq c] \leq \frac{E[h(x)]}{c}, \ \ \forall c > 0.$$

Deste teorema deduz-se a desigualdade de Chebyshev, a qual se apresenta no próximo corolário.

Corolário 2.5.1 (Desigualdade de Chebyshev) *sendo $\mu = E[X]$, $\sigma^2 = Var(X)$, $h(x) = (X - \mu)^2$ e $c = k^2\sigma^2$ tem-se:*

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

ou em termos complementares

$$P\left[|X - \mu| < k\sigma\right] > 1 - \frac{1}{k^2}.$$

Com base nesta desigualdade, dada uma v.a. com média e variância finitas, pode afirmar-se que:

i) A probabilidade de uma variável aleatória tomar u valor que diste da média menos de 2 desvios padrões é superior a 75%

ii) probabilidade de uma variável aleatória tomar u valor que diste da média menos de 3 desvios padrões é superior a 89%

Exemplos:

1. Uma variável aleatória X tem média 15, variância 16 e uma distribuição de probabilidade desconhecida. Estime o valor de:

   a) $P(|X-15| \leq 12)$.

   $$\begin{aligned} P(|X-15| \leq 12) &= P(-12 \leq X-15 \leq 12) \\ &= P(15-3\times 4 \leq X \leq 15+3\times 4) \\ &\geq 1-\tfrac{1}{3^2}=\tfrac{8}{9}. \end{aligned}$$

   b) $P(7 \leq X \leq 23)$.

   $$\begin{aligned} P(7 \leq X \leq 23) &= P(15-2\times 4 < X < 15+2\times 4) \\ &\geq 1-\tfrac{1}{2^2}=\tfrac{3}{4}. \end{aligned}$$

2. A distribuição dos salários mensais de uma empresa de consultores na Holanda tem média 110 (10^2 euros) e desvio padrão 8(10^2 euros). A empresa tem 500 funcionários. Quantos dos funcionários têm salário igual ou superior a 120 (10^2 euros) mensais? (Admita que a distribuição dos salários é simétrica)
   Sendo

   X="Salário mensal de um funcionário (10^2 euros)"

   tem-se $\mu_X = 110$ e $\sigma_X = 8$ e

   $$P(X \geq 120) = P(X-\mu \geq 120-110) = P(X-\mu \geq 10)$$

   e dado que a distribuição é simétrica

   $$P(X-\mu \geq 10) = \frac{1}{2}P(|X-\mu| \geq 10)$$

   então pela desigualdade de Chebyshev vem

   $$P(|X-\mu| \geq 10) = P\left(|X-\mu| \geq \frac{10}{8} \times 8\right) \leq \frac{1}{\left(\frac{10}{8}\right)^2}$$

de onde se conclui que $k = \frac{10}{8} = 0.64$ pelo que

$$P(X \geq 120) = \frac{1}{2} P(|X - \mu| \geq 10) \leq \frac{1}{2} \times 0.64 = 0.32$$

o que nos dá o limite superior para a probabilidade do salário mensal de um funcionário ser superior a 120 (10^2 euros). Assim, no máximo $500 \times 0.32 = 160$ funcionários têm um salário igual ou superior a 120(10^2 euros).

2.5.1 Lei dos Grandes Números

Teorema 2.5.2 *Sejam $X_1, X_2, \ldots$ variáveis aleatórias (discretas ou contínuas) mutuamente independentes cada uma com média μ e variância σ^2 finitas. Então se*

$$S_n = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i$$

tem-se

$$\lim_{n\to\infty} P\left[|\frac{S_n}{n} - \mu| \geq \delta\right] = 0.$$

Este resultado é uma consequência da desigualdade de Chebyshev. Sendo $\frac{S_n}{n}$ a soma aritmética de $X_1, X_2, \ldots, X_n$ o teorema diz que a probabilidade da média aritmética diferir do seu valor esperado μ em pelo menos δ, tende para zero quando n tende para infinito.

Capítulo 3

Distribuições Especiais

As variáveis associadas às experiências aleatórias têm distribuições empíricas às quais, muitas vezes, se pode associar um modelo matemático que permite dar resposta a problemas de aplicação da teoria das probabilidades. Estes modelos matemáticos, que usualmente se designam de distribuições estatísticas serão o nosso objecto de estudo no presente capítulo. Neste, iremos abordar as distribuições discretas e contínuas que são consideradas mais importantes, por abrangerem um número significativo de fenómenos aleatórios.

3.1 Distribuições Discretas

As distribuições dependem de um ou mais parâmetros, os quais determinam a localização (por exemplo o ponto médio), a variabilidade (determina a dispersão da distribuição) ou a forma da distribuição (estabelece a forma básica da distribuição e numa dada distribuição este parâmetro pode não existir ou podem existir vários).

3.1.1 Distribuição Uniforme

Nesta distribuição todos os valores possíveis da variável são equiprováveis.

Pode aplicar-se nas experiências aleatórias em que todos os resultados têm a mesma probabilidade de ocorrer ou quando se analisa uma quantidade inteira que varia entre dois inteiros, em que a informação que se conhece sobre esta quantidade não permite outra análise.

Definição 3.1.1 (Distribuição Uniforme Discreta) *Uma v.a. discreta X tem distribuição uniforme em n pontos, escreve-se $X \sim DU(n)$, se a sua* *função de probabilidade for dada por*

$$f(x) = \begin{cases} \frac{1}{n} & se \quad x = 1, 2, \ldots, n \\ 0 & \quad para\ outros\ valores\ de\ x \end{cases}$$

e cuja função de distribuição é

$$F(x) = \begin{cases} 0 & se \quad x < 1 \\ \frac{x_i}{n} & se \quad x_i \leq x < x_{i+1} \quad com\ i \in \{1, \ldots, n-1\} \\ 1 & se \quad x \geq n \end{cases} .$$

Assim, a v.a. X assume um conjunto finito de valores estando associado a cada um deles a probabilidade constante $\frac{1}{n}$. Refira-se que existem outras formas de definir as funções de probabilidade e de distribuição desta variável, mas que são equivalentes a esta.

Existem algumas características da distribuição que são fundamentais e por isso serão referidas de seguida, este procedimento será adoptado para todas as distribuições que iremos apresentar.
O parâmetro caracterizador desta distribuição é n, um valor inteiro, que é o maior valor assumido pela variável.

Teorema 3.1.1 *Se X é uma v.a. discreta com distribuição uniforme tem-se que:*

$$\mu_X = E[X] = \frac{n+1}{2}$$

$$\sigma_X^2 = Var(X) = \frac{n^2-1}{12}.$$

Exemplos:

1. Uma empresa produtora de energia eléctrica pretende construir no próximo ano uma nova central térmica. Ao planear a sua estratégia de produção, concluiu que é igualmente provável que a procura seja de 100000, 110000, 120000 ou 130000 Kilowatts. A distribuição de probabilidade da v.a. X = "procura de energia eléctrica", em kilowatts, pode ser descrita da seguinte forma:

$$f(x) = \begin{cases} \frac{1}{4} = 0,25 & se \quad x = 100000, 110000, 120000, 130000 \\ 0 & \quad \text{para outros valores de} x \end{cases}$$

   e representada pelo Gráfico 3.1, trata-se de uma distribuição uniforme discreta em que $n = 4$.

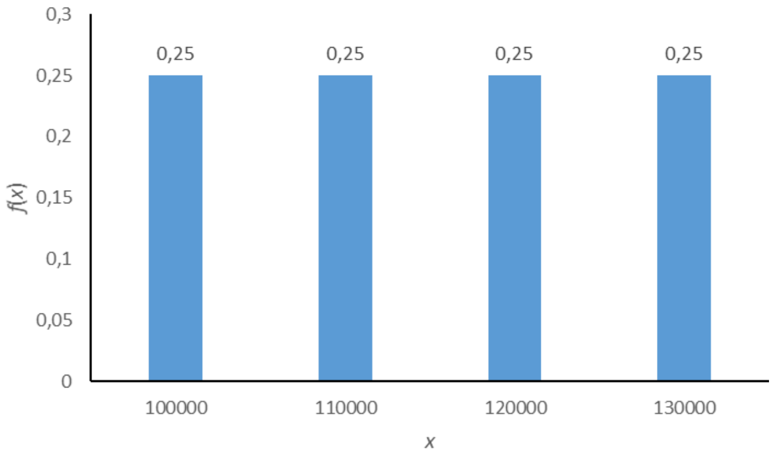

Figura 3.1: Função de probabilidade de X.

2. Uma empresa importadora de cafés estudou o lançamento de um novo lote de café de elevada qualidade e está disposta a comercializá-lo em 5 composições diferentes que vamos denominar de A, B, C, D e E, caso os consumidores revelem preferências diferenciadas. Recolheu-se uma amostra aleatória de 1000 potenciais consumidores aos quais se ofereceram 5 chávenas de café, com as referidas composições, sem identificação da composição. Obtiveram-se os seguintes resultados:

Composição Preferida	Identificação Numérica da Composição	nº de consumidores
A	1	200
B	2	200
C	3	200
D	4	200
E	5	200

Destes resultados,ou seja desta distribuição empírica, pode dizer-se que as preferências dos consumidores são diferenciadas. Considerando

$$X = \text{"composição preferida de certo consumidor"}$$

pode dizer-se que X tem distribuição

$$f(x) = \begin{cases} \frac{1}{5} & se \quad x = 1, 2, 3, 4, 5 \\ 0 & \quad \text{para outros valores de} x \end{cases}$$

sendo representada pelo Gráfico **??**.

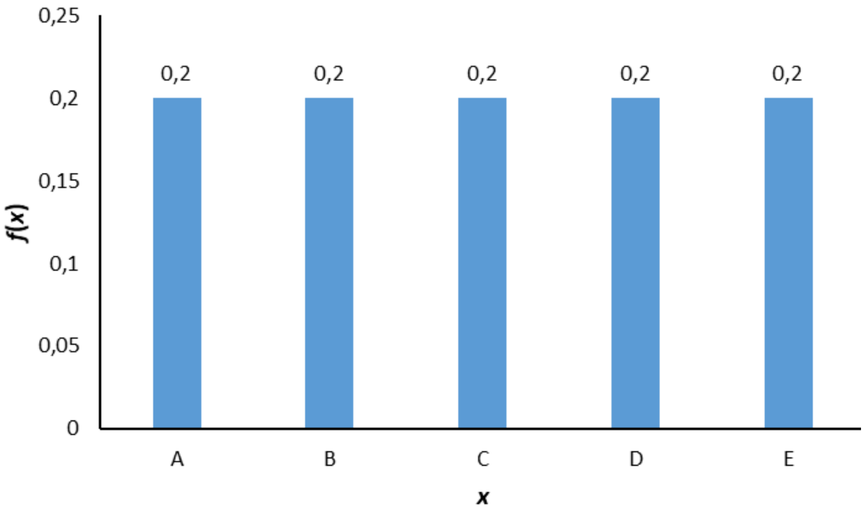

Figura 3.2: Função de probabilidade de X.

3.1.2 Distribuição Binomial

Esta distribuição corresponde ao modelo probabilístico adequado para descrever processos em que se realizam repetidas provas de Bernoulli, ou seja uma sucessão de provas de Bernoulli.
Uma prova de Bernoulli consiste numa experiência aleatória com a particularidade de apenas ter dois resultados possíveis, que se podem caracterizar por:

1. realização de A, que corresponde ao *sucesso*,
2. não se realização de A ou realização de A^c, que corresponde ao *insucesso*.

Sendo
$$\Omega = \{A, A^c\}$$
supondo
$$P(sucesso) = P(A) = p$$
tem-se
$$P(insucesso) = P(A^c) = 1 - p = q.$$

Uma sucessão de provas de Bernoulli não é mais do que uma sucessão de experiências aleatórias independentes, em cada uma das quais se observa a realização ou não realização de um determinado acontecimento A. A probabilidade de A se realizar é igual em todas as experiências.

Um exemplo concreto de aplicação desta distribuição é quando se pretende contar o número de artigos defeituosos num lote de dimensão n.

A distribuição binomial baseia-se no conceito de provas de Bernoulli e é utilizada como modelo teórico adequado a um grande número de situações práticas, podendo mesmo considerar-se a distribuição discreta mais importante. Sendo um esquema probabilístico adequado para as situações em que se pretende analisar um conjunto finito de indivíduos ou objectos, sabendo-se que cada elemento possui uma determinada característica A com $P(A) = p$, logo $P(A^c) = 1 - p = q$. É uma distribuição muito utilizada em amostragem e em situações em que se conhece o tamanho da amostra e o número de vezes em que um acontecimento ocorreu.

Segue-se um exemplo simples que tornará natural a definição de distribuição binomial. Considere-se a experiência aleatória "lançamento de uma moeda ao ar" e os acontecimentos

$$A=\text{"sair face cara" e } A^c=\text{"sair face coroa"},$$

procedendo a cinco experiências aleatórias independentes (cinco provas de Bernoulli) podendo, em cada uma delas, realizar-se ou não o acontecimento A. O acontecimento A é o *sucesso* e tem-se

$$P(A) = p = \frac{1}{2}$$

e

$$P(A^c) = 1 - p = \frac{1}{2}.$$

O espaço de resultados associado às cinco provas de Bernoulli é

$$\Omega = \{(A, A, A, A, A)(A^c, A, A, A, A)(A, A^c, A, A, A), (A, A, A^c, A, A), \ldots\}$$

sendo $\#\Omega = 2^5 = 32$.
Se pretendermos saber qual a probabilidade de em 5 provas de Bernoulli se obterem 3 *sucessos* têm-se as seguintes possibilidades:

$$\begin{array}{ll}
(A, A, A, A^c, A^c) & (A^c, A, A, A, A^c) \\
(A, A^c, A, A, A^c) & (A, A, A^c, A, A^c) \\
(A^c, A^c, A, A, A) & (A^c, A, A^c, A, A) \\
(A, A, A^c, A^c, A) & (A^c, A, A, A^c, A) \\
(A, A^c, A^c, A, A) & (A, A^c, A, A^c, A)
\end{array}$$

temos

$$C_3^5 = \frac{5!}{3!2!} = 10$$

maneiras de obter 3 *sucessos* em 5 provas de Bernoulli mas, dada a independência das provas, para todas elas a probabilidade é

$$\left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right)^2$$

sendo o primeiro factor a probabilidade associada aos 3 *sucessos* e o segundo factor a probabilidade associada a 2 *insucessos*. Assim, a probabilidade de obter 3 *sucessos* em 5 provas de Bernoulli é $\frac{1}{32}$, que se poderia ter obtido através da distribuição binomial.
A distribuição Binomial aparece associada à seguinte questão genérica:

Qual a probabilidade de, em n provas de Bernoulli, se obterem x *sucessos*?
O *sucesso* corresponde à realização de um acontecimento A e o insucesso corresponde à realização de A^c, pelo que uma das possíveis sequências que verifica a condição imposta é

$$\overbrace{\underbrace{A, A, A, \ldots, A,}_{x\ sucessos} \underbrace{A^c, A^c, A^c, \ldots, A^c}_{n-x\ insucessos}}^{n\ provas\ de\ Bernoulli}$$

mas o número de sequências distintas com x *sucessos* e $n - x$ *insucessos* é

$$C_x^n = \frac{n!}{x!(n-x)!},$$

que corresponde às possibilidades distintas de combinar x *sucessos* em n provas. Contudo, dada a independência das provas ou experiências aleatórias, todas as sequências têm a mesma probabilidade de ocorrência

$$p^x\ (\ 1 - p)^{n-x}\ .$$

Definição 3.1.2 (Distribuição Binomial) *Uma v.a. discreta*

$$X = \text{"}número\ de\ \text{sucessos}\ em\ n\ provas\ de\ Bernoulli\text{"}$$

com distribuição binomial, escreve-se $X \sim B(n,p)$, tem função de probabilidade

$$f(x) = P(X = x) = \begin{cases} C_x^n p^x\,(1-p)^{n-x} & se \quad x = 0, 1, 2, \ldots, n\ \wedge\ 0 < p < 1 \\ 0 & para\ outros\ valores\ de\ x \end{cases}$$

e função de distribuição é

$$F(x) = \begin{cases} 0 & se \quad x < 0 \\ \displaystyle\sum_{x_i=0}^{x} C_{x_i}^n p_i^x\,(1-p)^{n-x_i} & se \quad 0 \le x \le n \\ 1 & se \quad x \ge n \end{cases}.$$

Os parâmetros caracterizadores desta distribuição são

n="número de provas de Bernoulli a efectuar",
p="probabilidade associada ao *sucesso*".

Dado que a utilização da definição da distribuição binomial envolve cálculos trabalhosos, basta ver as funções apresentadas, estão disponíveis tabelas com valores para as referidas funções. Vamos utilizar uma tabela da função de probabilidade com valores $n \leq 20$ e $0.05 \leq p \leq 0.5$ em múltiplos de 0.05.

Teorema 3.1.2 *Se X é uma v.a. discreta com distribuição binomial tem-se que:*

$$\mu_X = E[X] = n\,p$$

$$\sigma_X^2 = Var(X) = n\,p\,q.$$

Exemplos:

1. Um técnico dos serviços de prevenção e segurança rodoviária afirma que 1 em 10 acidentes rodoviários é devido a cansaço. Determine a probabilidade de, em 5 acidentes, ocorrerem 0,1,2,3,4 e 5 devidos a cansaço.
   Sendo X="número de acidentes, em 5, devidos a cansaço", com

   $$X \sim B(n = 5, p = 0,1)$$

   $p = \frac{1}{10}$, dado que 1 acidente em cada 10 é devido a cansaço, vem

   $$\begin{array}{lcll} P(X=0) & = & C_0^5\, 0,1^0\,(0,9)^5 & = & 0,5905 \\ P(X=1) & = & C_1^5\, 0,1^1\,(0,9)^{5-1} & = & 0,328 \\ P(X=2) & = & C_2^5\, 0,1^2\,(0,9)^{5-2} & = & 0,0729 \\ P(X=3) & = & C_3^5\, 0,1^3\,(0,9)^{5-3} & = & 0,0081 \\ P(X=4) & = & C_4^5\, 0,1^4\,(0,9)^{5-4} & = & 0,0004 \\ P(X=5) & = & C_5^5\, 0,1^5\,(0,9)^{5-5} & = & 0,00001. \end{array}$$

   representada pelo Gráfico 3.3.

2. O Manuel e o João são amigos e gostam muito de jogar damas, o João joga muito bem e ganha 70 % dos jogos. Os dois amigos resolveram efectuar um campeonato de 15 jogos nos próximos dias. Qual a probabilidade de:

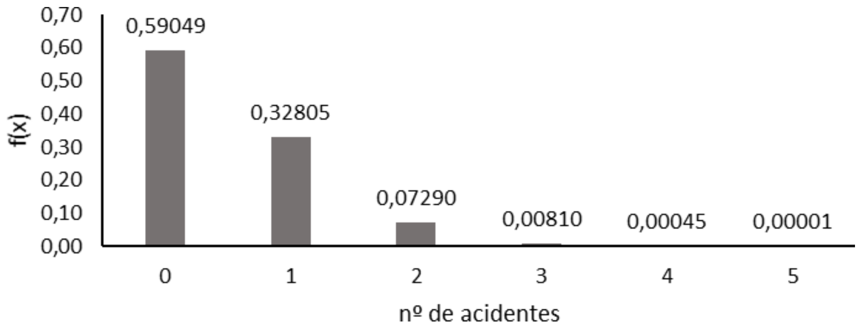

Figura 3.3: Função de probabilidade da ocorrência de acidentes rodoviários devidos a cansaço num conjunto de 5 acidentes.

a) O João ganhar exactamente 11 jogos?
Seja X="número de jogos, em 15, que o João ganha", então

$$X \sim B(n = 15, p = 0,7)$$

pelo que a resposta pedida é

$$\begin{aligned} P(X = 11) &= C^{15}_{11}\, 0,7^{11}\, (0,3)^{15-11} \\ &= \end{aligned}$$

b) O João ganhar entre 5 e 10 jogos?

$$\begin{aligned} P(5 \leq X \leq 10) &= \sum_{i=5}^{10} P(X = i) \\ &= \sum_{i=5}^{10} C^{15}_{i}\, 0,7^{i}\, (0,3)^{15-i} \\ &= 0,48384 \end{aligned}$$

c) O João ganhar pelo menos 12 jogos?

$$\begin{aligned} P(X \geq 12) &= \sum_{i=12}^{15} P(X = i) \\ &= \sum_{i=12}^{15} C^{15}_{i}\, 0,7^{i}\, (0,3)^{15-i} \\ &= 0,29687 \end{aligned}$$

3. Uma grande empresa tem 4 UPS's independentes para garantir que em caso de falha de energia o seu sistema informático se mantém activo

durante 48 horas, para tal é necessário que pelo menos duas não tenham avarias. A probabilidade de cada UPS avariar no período de um ano é 0.01. Qual a probabilidade, de durante um ano, se cumprirem as 48 horas de manutenção do sistema em caso de falha de energia?
Seja X="número de UPS, em 4, com avarias", então

$$X \sim B(n = 4, p = 0,01).$$

A resposta à questão é

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= C_0^4 \, 0,01^0 \, (0,99)^{4-0} + C_1^4 \, 0,01^1 \, (0,99)^{4-1} + C_2^4 \, 0,01^2 \, (0,99)^{4-2} \\ &= 1 \end{aligned}$$

4. Num determinado supermercado está para entrar em promoção arroz de determinada marca do qual têm armazenadas 500 embalagens, destas exactamente 50 expiraram o prazo de validade. É efectuada uma inspecção sobre uma amostra de 10 embalagens escolhidas ao acaso com reposição. Sabendo que a inspecção admite, no máximo, 3 embalagens fora da validade, qual a probabilidade de a inspecção rejeitar a distribuição do lote?
Neste caso

$$\begin{aligned} &sucesso\text{="embalagem com validade expirada"} \\ &X\text{="número de embalagens, em 10, com validade expirada",} \end{aligned}$$

então sendo

$$p = \frac{50}{500} = 0,1$$

tem-se

$$X \sim B(n = 10, p = 0,1).$$

A resposta à questão é

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) \\ &= 1 - f(0) - f(1) - f(2) - f(3) \\ &= 0,0128. \end{aligned}$$

5. Sabe-se que com determinado tratamento se alcançam 70 % de curas para certa doença, quando o mesmo é administrado a pacientes em condições bem definidas. Se o tratamento é aplicado a 20 pacientes em tais condições, qual a probabilidade de obter, no máximo, 15 curas?
Considerando

A=*sucesso*="o doente cura-se",
X="número de doentes, em 20, que se curam",

tem-se

$$P(A) = \frac{70}{100} = 0,7$$

$$P(A^c) = \frac{30}{100} = 0,3$$

com

$$X \sim B(n = 20, p = 0,7).$$

Para responder à questão é necessário determinar $P(X \leq 15)$, contudo não é possível utilizar directamente a tabela dada, já que a mesma apenas contempla valores de p até 0.5. Se considerarmos

A = *sucesso*="o doente não se cura",
Y="número de doentes, em 20, que não se curam"=$20 - X$,

tem-se

$$P(A) = \frac{30}{100} = 0,3$$

$$P(A^c) = \frac{70}{100} = 0,7$$

com

$$Y \sim B(n = 20, p = 0,3),$$

passando a ser possível utilizar directamente a tabela da binomial para determinar

$$\begin{aligned} P(Y \geq 5) &= 1 - P(Y < 5) \\ &= 1 - f_Y(0) - f_Y(1) - f_Y(2) - f_Y(3) - f_Y(4) \cong 0,7626. \end{aligned}$$

Apesar das considerações anteriores, pode-se sempre determinar a probabilidade pretendida recorrendo à função de probabilidade adequada.

3.1.3 Distribuição Geométrica

Se considerarmos sucessivas provas independentes de uma experiência aleatória, sendo que em cada prova se admite *sucesso*(realização do acontecimento A) com probabilidade p e *insucesso* (realização do acontecimento A^c) com probabilidade q verificando-se $p + q = 1$. Em resumo, estamos na presença de uma sucessão de provas de Bernoulli. Seja

X="número de provas a realizar até se obter um sucesso".

Esta distribuição pode ser o modelo probabilístico de v.a. qur representem, por exemplo, o número de peças inspeccionadas até encontrar uma defeituosa.

Supondo que realizamos x provas até ter como resultado da experiência aleatória um *sucesso*, ou seja, na x^{esima} prova ocorreu *sucesso* e nas restantes $x-1$ ocorreram *insucessos*. Neste caso, não existem diferentes possibilidades de ter $x-1$ *insucessos* em $x-1$ provas

$$\overbrace{\underbrace{A^c, A^c, A^c, \ldots, A^c}_{x-1\ provas}, A}^{x\ provas}.$$

Considere-se,genericamente, que $P(sucesso) = P(A) = p$ e a do *insucesso* é $P(A^c) = 1 - p = q$. A v.a. X pode assumir os seguintes valores:

i) $X = 1$, corresponde a sucesso na $1^{\underline{a}}$ prova com $P(X = 1) = p$;

ii) $X = 2$, corresponde a sucesso na $2^{\underline{a}}$ prova e insucesso na $1^{\underline{a}}$ prova com $P(X = 2) = P(A^c \cap A) = q \times p$;

ii) $X = 3$, corresponde a sucesso na $3^{\underline{a}}$ prova e insucessos na $2^{\underline{a}}$ prova e na $1^{\underline{a}}$ prova com $P(X = 3) = P(A^c \cap A^c \cap A) = q \times q \times p$

iii) $X = 4$, corresponde a sucesso na $4^{\underline{a}}$ prova e insucessos na $3^{\underline{a}}$ prova, na $2^{\underline{a}}$ prova e na $1^{\underline{a}}$ prova com $P(X = 4) = P(A^c \cap A^c \cap A^c \cap A) = q \times q \times q \times p$ e assim sucessivamente.

Genericamente, sendo x o número de provas a realizar até obter *sucesso* tem-se

$$P(X = x) = P(\underbrace{\overbrace{A^c \cap A^c \cap A^c \cap \ldots}^{x-1\ insucessos} \cap A}_{x\ provas}) = (q)^{x-1}\ p.$$

Definição 3.1.3 (Distribuição Geométrica) *Uma v.a. discreta*

X="número de provas a realizar até obter o primeiro sucesso*"*

tem distribuição geométrica, escreve-se $X \sim G(p)$, se a função de probabilidade é dada por

$$f(x) = P(X = x) = \begin{cases} (1-p)^{x-1}\ p & se \quad x = 1, 2, 3, \ldots\ 0 < p < 1 \\ 0 & para\ outros\ valores\ de\ x \end{cases}.$$

e cuja função de distribuição é

$$F(x) = \begin{cases} 0 & se \quad x < 1 \\ 1 - (1-p)^{\lfloor x \rfloor} & se \quad x \geq 1 \end{cases}.$$

O parâmetro caracterizador da distribuição é p, a probabilidade do acontecimento *sucesso*.

Pode ver-se esta distribuição como um caso particular da distribuição binomial negativa, na qual se considera $k = 1$. Pois, a função de probabilidade da distribuição geométrica pode ser obtida através da função de probabilidade da binomial negativa com $k = 1$.

Teorema 3.1.3 *Se X é uma v.a. discreta com distribuição geométrica tem-se que:*

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

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

Exemplos:

1. A probabilidade de se encontrar aberto um sinal de trânsito numa esquina é $0,2$. Qual a probabilidade de que seja necessário passar pelo local 5 vezes para encontrar o sinal aberto pela primeira vez?
   Sendo

   > *sucesso*$=A=$"sinal aberto",
   > X="número de passagens no local até obter o sinal aberto pela 1ª vez"

   com $P(A) = 0,2$ o que se pretende é $P(X = 5)$, dada a função de probabilidade da v.a. $X \sim G(0,2)$ (Figura 3.4) tem-se

   $$P(X = 5) = 0,2 \times (0,8)^4 \cong 0,0819.$$

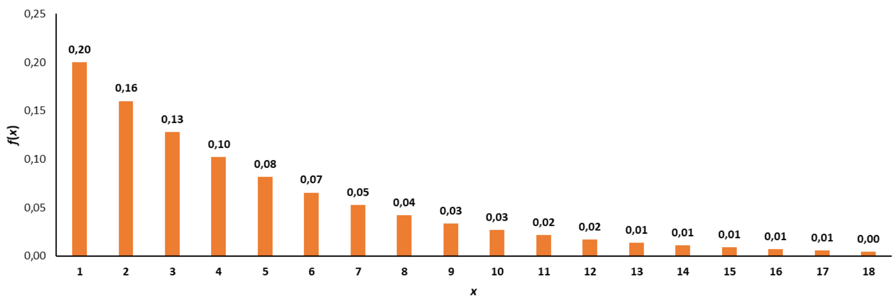

Figura 3.4: Função de probabilidade de $X =$ "número de passagens no local até obter o sinal aberto pela 1ª vez".

2. A probabilidade de um automóvel efectuar uma lavagem automática quando vai ser abastecido de combustível numa bomba de gasolina é de 0,1. Determine a probabilidade de nenhum dos próximos 6 automóveis a abastecer efectuarem lavagem.
   Sendo

   > $sucesso = A =$"Automóvel, que abasteceu, efectuou lavagem",
   > X="número de automóveis que entram na bomba, para abastecer, até que um também efectue lavagem"

   com $P(A) = 0,1$, o que se pretende é $P(X > 6)$. Dada a função de probabilidade da v.a. $X \sim G(0,1)$ (Figura 3.5) ou a função de distribuição tem-se

   $$P(X > 6) = 1 - P(X \leq 6) \cong 0,531.$$

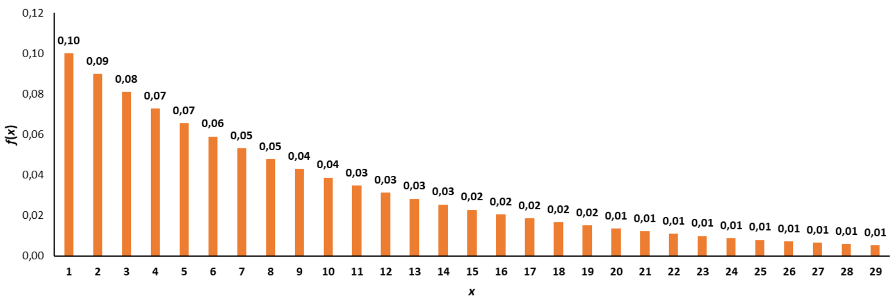

Figura 3.5: Função de probabilidade de $X =$ "número de automóveis que entram na bomba, para abastecer, até que um também efectue lavagem".

3.1.4 Distribuição de Poisson

Esta distribuição é adequada para descrever uma grande variedade de situações resultantes das mais diversas áreas. Seguem-se exemplos de situações concretas que podem ser descritas por v.a. que seguem uma distribuição de Poisson são:

1. número de chamadas telefónicas que chegam, num determinado período temporal, a uma central telefónica;

2. número de doentes que chegam, por unidade de tempo, a um determinado hospital central;

3. número de avarias que ocorrem numa máquina, num determinado intervalo de tempo,

4. número de micro-organismos em determinada quadrícula;

5. número de partículas defeituosas num certo volume de líquido;

6. número de deficiências, num dado comprimento, de fio produzido por uma máquina têxtil

7. número de peixes que existem num lago ... ...

Em todos os casos tem-se o número de ocorrências de um qualquer evento num intervalo temporal, numa região espacial, num volume ou num comprimento. Os exemplos apresentados partilham uma característica: podem ser descritos por uma v.a. discreta que toma valores inteiros não negativos. Além desta característica é necessário que verifiquem as exigidas para que se esteja perante um processo de Poisson, que são:

i) Não ter memória (o número de ocorrências num intervalo é independente do número de ocorrências em outro intervalo disjunto.

ii) A probabilidade de ocorrer um evento num intervalo 'muito pequeno' é proporcional ao comprimento do intervalo.

iii) A probabilidade de ocorrer mais do que um evento num intervalo 'muito pequeno' é nula

Podem existir fenómenos que originem uma maior densidade de ocorrências em determinados períodos, áreas, volumes, etc. Considerem-se as seguintes situações:

1. número de doentes que chegam a um hospital, por hora, se ocorreu um incêndio numa zona habitacional;

2. número de carros que vão abastecer a uma bomba, entre as 23 e 24 horas, após os jornais vespertinos informarem de um aumento de preços nos combustíveis.

Nos fenómenos apresentados a v.a. que os representa não deverá ter distribuição de Poisson, pois não é um modelo adequado às situações por violar a característica iii). Para estes casos existem outras distribuições, denominadas de agregativas, que reflectem as situações referidas.

Definição 3.1.4 (Distribuição Poisson) *Uma v.a. discreta*

X="número de sucessos num intervalo (área,...)"

tem distribuição de Poisson, escreve-se $X \sim P(\lambda)$, quando a sua função de probabilidade é dada por

$$f(x) = P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\,\lambda^x}{x!} & se \quad x = 0, 1, 2, \ldots \\ 0 & para\ outros\ valores\ de\ x \end{cases}.$$

e cuja função de distribuição é

$$F(x) = \begin{cases} 0 & se \quad x < 0 \\ e^{-\lambda} \displaystyle\sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} & se \quad x \geq 0 \end{cases}.$$

A distribuição tem como parâmetro λ que representa o número médio de eventos que ocorrem num intervalo de tempo (numa área ...). Tem-se $\lambda \in]0, +\infty[$.

Teorema 3.1.4 *Se X é uma v.a. discreta com distribuição geométrica tem-se que:*

$$\mu_X = E[X] = \lambda$$
$$\sigma_X^2 = Var(X) = \lambda.$$

Exemplos:

1. Admita que o número de camiões TIR que, em 30 minutos, atravessam a ponte 25 de Abril segue uma distribuição de Poisson com variância 8. Qual a probabilidade de que, em 30 minutos, exactamente 4 camiões TIR atravessem a ponte?
   Sendo

   > X = "número de camiões de TIR que, em 30 minutos, atravessam a ponto 25 de Abril"

   tem-se (Figura 3.6)

   $$X \sim P(\lambda = 8)$$
   $$P(X = 4) = \frac{e^{-8}\,8^4}{4!} \cong 0,0573.$$

2. Em média, numa hora, chegam 3 automóveis a uma bomba de gasolina.

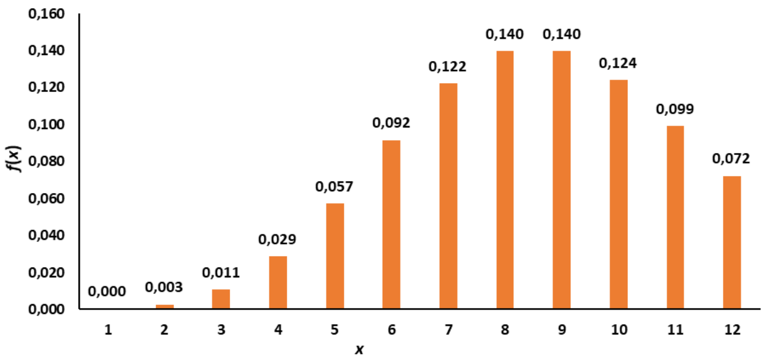

Figura 3.6: Função de probabilidade de X = "número de camiões de TIR que, em 30 minutos, atravessam a ponto 25 de Abril".

a) Defina a função de probabilidade que retrata a situação em causa. Sendo

X="número de automóveis que chegam à bomba numa hora"(Figura 3.7)

$$X \sim P(\lambda = 3)$$

$$f(x) = P(X = x) = \begin{cases} \dfrac{e^{-3}\, 3^x}{x!} & se \quad x = 0, 1, 2, \ldots \\ 0 & \quad \text{para outros valores de } x \end{cases} .$$

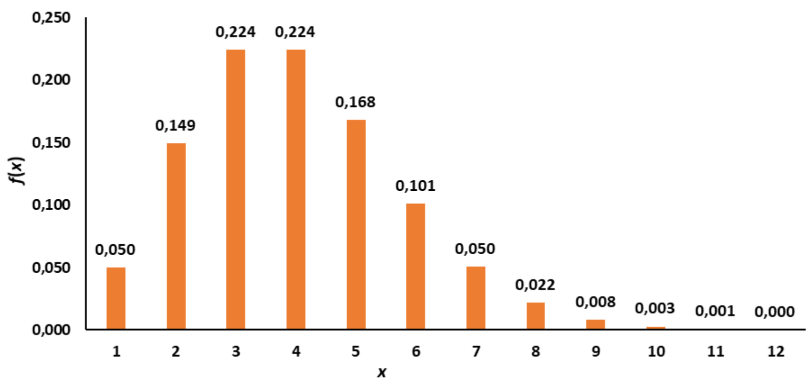

Figura 3.7: Função de probabilidade de X = "número de automóveis que chegam à bomba numa hora".

b) Determine a probabilidade de, numa hora, chegarem menos de 2 automóveis para se abastecerem.

$$\begin{aligned} P(X<2) &= f(0)+f(1) \\ &= \frac{3^0\, e^{-3}}{0!}+\frac{e^{-3}\, 3}{1!} \cong 0,199 \end{aligned}$$

Os valores $f(0)$ e $f(1)$ também podem ser obtidos pelas tabelas da distribuição de Poisson.

3. A uma central telefónica chegam 300 telefonemas por hora. Qual a probabilidade de:

a) Num minuto não haver nenhuma chamada.
Sendo

X="número de chamadas por minuto"

em média, por minuto, tem-se $\lambda = \frac{300}{60} = 5$ com

$$X \sim P(\lambda = 5).$$

A função de probabilidade (Figura 3.8) é:

$$f(x) = P(X=x) = \begin{cases} \dfrac{e^{-5}\, 5^x}{x!} & se \quad x = 0,1,2,\ldots \\ 0 & \text{para outros valores de } x \end{cases}.$$

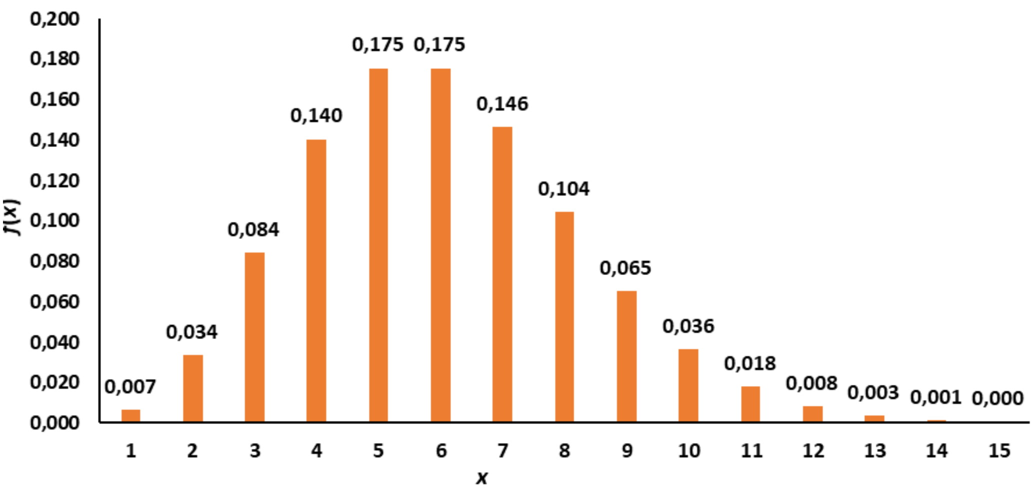

Figura 3.8: Função de probabilidade de X = número de chamadas por minuto".

$$\begin{aligned} P(X=0) &= f(0) \\ &= \frac{5^0\, e^{-5}}{0!} \cong 0,00674 \end{aligned}$$

b) Em 2 minutos ocorrerem 2 chamadas.
Definindo

Y="número de chamadas em 2 minutos"

em média, por minuto, tem-se $\lambda = \frac{300}{60} \times 2 = 10$ com

$$Y \sim P(\lambda = 10).$$

A função de probabilidade é:

$$f(y) = P(Y=y) = \begin{cases} \dfrac{e^{-10}\, 10^x}{x!} & se \quad x = 0,1,2,\ldots \\ 0 & \quad \text{para outros valores de } x \end{cases}.$$

$$\begin{aligned} P(Y=2) &= f(2) \\ &= \frac{10^2\, e^{-10}}{2!} \\ &= \frac{50}{e^{10}} \end{aligned}$$

c) Em t minutos não ocorrerem chamadas.
Definindo

Z="número de chamadas em t minutos"

em média, por minuto, tem-se $\lambda = \frac{300}{60} \times t = 5t$ com

$$Z \sim P(\lambda = 5t).$$

A função de probabilidade é:

$$f(z) = P(Z=z) = \begin{cases} \dfrac{e^{-5t}\, (-5t)^x}{x!} & se \quad x = 0,1,2,\ldots \\ 0 & \quad \text{para outros valores de } x \end{cases}.$$

$$\begin{aligned} P(Z=0) &= f(0) \\ &= \frac{(5t)^0\, e^{-5t}}{0!} \\ &= e^{-5t}. \end{aligned}$$

3.2 Distribuições Contínuas

À semelhança do que se apresentou na secção das distribuições discretas, na presente secção vamos introduzir as distribuições contínuas que traduzem os modelos matemáticos adequados ao tratamento de determinado tipo de v.a. contínuas. Apenas iremos referir as mais utilizadas.

3.2.1 Distribuição Uniforme

Quando se pretende representar uma quantidade que varia aleatoriamente num intervalo, $[a, b]$, utiliza-se esta distribuição, assumindo que a probabilidade de obter valores num sub-intervalo de $[a, b]$ é proporcional à amplitude do mesmo. Os números aleatórios são um exemplo de concretização desta distribuição, pois se pretendermos gerar um número aleatório no intervalo $[0, 1]$, o que procuramos é uma concretização da v.a. X com distribuição uniforme no referido intervalo. Também se pode utilizar esta distribuição em situações onde se sabe que uma determinada quantidade varia num dado intervalo conhecendo pouco mais sobre a situação.

Definição 3.2.1 (Distribuição Uniforme) *Uma v.a. contínua X tem distribuição Uniforme no intervalo $[a, b]$, escreve-se $X \sim U(a, b)$, quando a sua função de densidade é dada por*

$$f(x) = \begin{cases} \dfrac{1}{b-a} & se \quad a \leq x \leq b \\ 0 & para\ outros\ valores\ de\ x \end{cases}.$$

e cuja função de distribuição é

$$F(x) = \begin{cases} 0 & se \quad x < a \\ \frac{x-a}{b-a} & se \quad a \leq x \leq b \\ 1 & se \quad x > b \end{cases}.$$

A distribuição tem como parâmetros os valores a e b, com $a < b$, que são os extremos do intervalo.

Teorema 3.2.1 *Se X é uma v.a. contínua com distribuição uniforme tem-se que:*

$$\mu_X = E[X] = \frac{1}{2}\ (a+b)$$

$$\sigma_X^2 = Var(X) = \frac{1}{12}\ (b-a)^2.$$

Exemplos:

1. Considere o processo de seleccionar números reais, ao acaso, no intervalo [5,7]. Determine a probabilidade do número estar compreendido entre:

a) 6.3 e 6.95.
Considerando que
$$X \sim U(5,7)$$
com representação gráfica na Figura 3.9

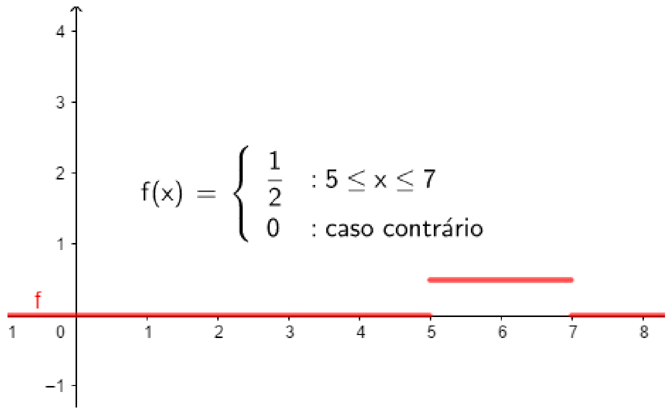

Figura 3.9: Função densidade da uniforme em [5,7].

Logo,
$$\begin{aligned} P(6.3 \leq X \leq 6.95) &= F(6.95) - F(6.3) \\ &= \frac{6.95 - 6.3}{7-5} \cong 0.325 \end{aligned}$$
ou
$$P(6.3 \leq X \leq 6.95) = \int_{6.3}^{6.95} \frac{1}{7-5}\, dx \cong 0.325$$

b) 6.5 e 7.5
$$\begin{aligned} P(6.5 \leq X \leq 7.5) &= F(7.5) - F(6.5) \\ &= 1 - \frac{6.5 - 5}{7-5} \cong 0.25 \end{aligned}$$

2. A dureza H de uma peça de aço pode ser representada por uma v.a. com distribuição uniforme no intervalo $[50, 70]$. Calcule a probabilidade de uma peça ter dureza entre 55 e 60.
Definindo
$$H \sim U(50, 70).$$

A função de densidade (Figura 3.10)é:

$$f(h) = P(H = h) = \begin{cases} \dfrac{1}{20} & se \quad 50 \leq h \leq 70 \\ 0 & \quad \text{para outros valores de } h \end{cases} .$$

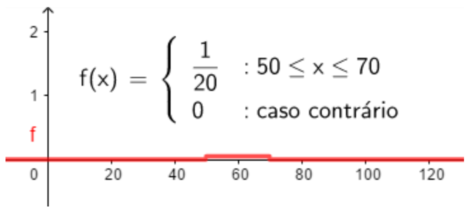

Figura 3.10: Função densidade da uniforme em [50,70].

$$P(55 \leq H \leq 60) \quad = \quad \int_{55}^{60} \frac{1}{20}\, dx \cong 0.25.$$

3.2.2 Distribuição Exponencial

Esta distribuição está, habitualmente, relacionada com o processo de Poisson. Isto, porque demonstra-se que num processo de Poisson o tempo de espera até ao primeiro sucesso, ou entre dois sucessos consecutivos, segue uma determinada distribuição exponencial. Esta distribuição é muito aplicada em estudos de filas de espera e de fiabilidade de sistemas complexos (sendo a fiabilidade num instante t a probabilidade do sistema em causa estar a funcionar). Pode utilizar-se esta distribuição para descrever o tempo de vida de componentes, de organismos, etc.

Definição 3.2.2 (Distribuição Exponencial) *Uma v.a. contínua X tem distribuição exponencial, escreve-se $X \sim E(\alpha)$, quando a sua função de densidade é dada por*

$$f(x) = \begin{cases} \alpha e^{-\alpha\, x} & se \quad x > 0 \wedge \alpha > 0 \\ 0 & se \quad x \leq 0 \end{cases} .$$

e cuja função de distribuição é

$$F(x) = \begin{cases} 0 & se \quad x \leq 0 \\ 1 - e^{-\alpha\, x} & se \quad x > 0 \end{cases} .$$

A distribuição tem como parâmetro α que é um número real positivo.

Teorema 3.2.2 *Se X é uma v.a. contínua com distribuição exponencial tem-se que:*

$$\mu_X = E[X] = \frac{1}{\alpha}$$

$$\sigma_X^2 = Var(X) = \frac{1}{\alpha^2}.$$

Exemplos:

1. O tempo entre duas chamadas consecutivas que chegam a uma central telefónica tem uma distribuição exponencial cuja média é de 2 segundos. Qual a probabilidade da central não receber chamadas nos próximos 5 segundos?
   Sendo

   $$X=\text{"tempo entre duas chamadas consecutivas (segundos)"}$$

   sendo a média de 2 segundos vem $\alpha = \frac{1}{2} = 0,5$ com

   $$X \sim E(\alpha = 0,5).$$

   A função de densidade (Figura 3.11) é

   $$f(x) = \begin{cases} 0.5\ e^{-0,5x} & se \quad x > 0 \\ 0 & se \quad x \leq 0 \end{cases}.$$

   e tem-se

   $$\begin{aligned} P(X > 5) &= 1 - P(X \leq 5) \\ &= 1 - F(5) \\ &= e^{-2.5} \cong 0,082. \end{aligned}$$

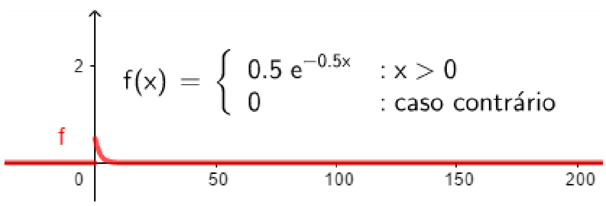

Figura 3.11: Função densidade de X="tempo entre duas chamadas consecutivas (segundos)".

2. O tempo de funcionamento T entre avarias consecutivas, de um determinado modelo de máquinas de empacotar produtos alimentares, é uma variável exponencial com média de uma hora. Considerando uma qualquer máquina, qual a probabilidade de estar a funcionar ao fim de uma hora?
   Sendo

   T="tempo de funcionamento entre duas avarias consecutivas (horas)"

   sendo a média de 1 hora vem $\alpha = 1$ com

   $$T \sim E(\alpha = 1).$$

   A função de densidade (Figura 3.12) é

   $$f(x) = \begin{cases} e^{-x} & se \quad x > 0 \\ 0 & se \quad x \leq 0 \end{cases}.$$

   Logo,

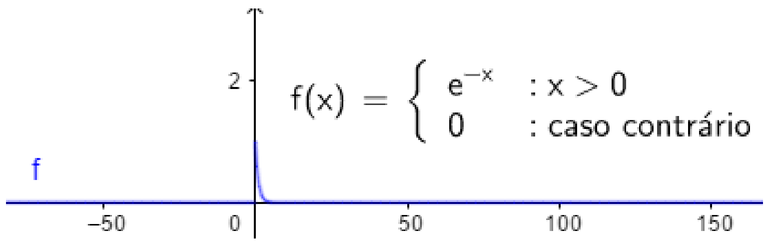

Figura 3.12: Função densidade de T="tempo de funcionamento entre duas avarias consecutivas (horas)".

$$\begin{aligned} P(T > 1) &= 1 - P(T \leq 1) \\ &= 1 - F(1) \\ &= e^{-1} \cong 0,368. \end{aligned}$$

3.2.3 Distribuição Normal ou de Gauss

São inúmeras as v.a. que descrevem fenómenos do dia a dia, nomeadamente características humanas como por exemplo a altura e o peso, que seguem uma distribuição normal. Assim, esta distribuição utiliza-se como modelo para representar quantidades ou características de populações que resultem de medições ou dos respectivos erros. Existem casos em que as v. a. não seguem uma distribuição normal, mas aproximam-se muito dela. Esta é a distribuição contínua mais importante.

A distribuição normal é muito utilizada na inferência estatística, pois mesmo quando a distribuição da população não é normal a distribuição da média amostral pode aproximar-se pela distribuição normal. Existem algumas v.a. que podem ser aproximadas por uma v.a. normal, exemplo disto são a Binomial e a Poisson.

Definição 3.2.3 (Distribuição Normal) *Uma v.a. contínua X tem distribuição normal, escreve-se $X \sim N(\mu, \sigma)$, quando a sua função de densidade é dada por*

$$f(x) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R} \tag{3.1}$$

e cuja função de distribuição é

$$F(x) = P(X \leq x) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}\, dt. \tag{3.2}$$

As características da função densidade da normal são:

- Tem a forma de sino e um único máximo em $x = \mu$.
- É simétrica em relação a um eixo vertical $x = \mu$.
- Tem pontos de inflexão em $x = \mu \pm \sigma$ com

$$\lim_{x \to \pm\infty} f(x) = 0.$$

A distribuição normal tem como parâmetros a média μ, uma medida de localização, e o desvio padrão σ, um parâmetro de escala.

Teorema 3.2.3 *Se X é uma v.a. contínua com distribuição normal tem-se que:*

$$\mu_X = E[X] = \mu$$

$$\sigma_X^2 = Var(X) = \sigma^2.$$

Distribuição Normal Estandardizada

Introduzindo uma mudança de variável, associada a $X \sim N(\mu, \sigma)$, tem-se:

$$U = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

onde U tem distribuição normal reduzida ou distribuição normal estandardizada ou normal-padrão. A função de densidade (Figura 3.13) desta nova variável é

$$f(u) = \frac{1}{\sqrt{2\,\pi}}\, e^{-\dfrac{u^2}{2}}, \quad u \in \mathbb{R}$$

que corresponde a fazer $\mu = 0$ e $\sigma = 1$ na função (3.1), e a função de distribuição

$$F(u) = P(U \leq u) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{u} e^{-\dfrac{v^2}{2}}\, dv.$$

que corresponde a fazer $\mu = 0$ e $\sigma = 1$ na função (3.2).

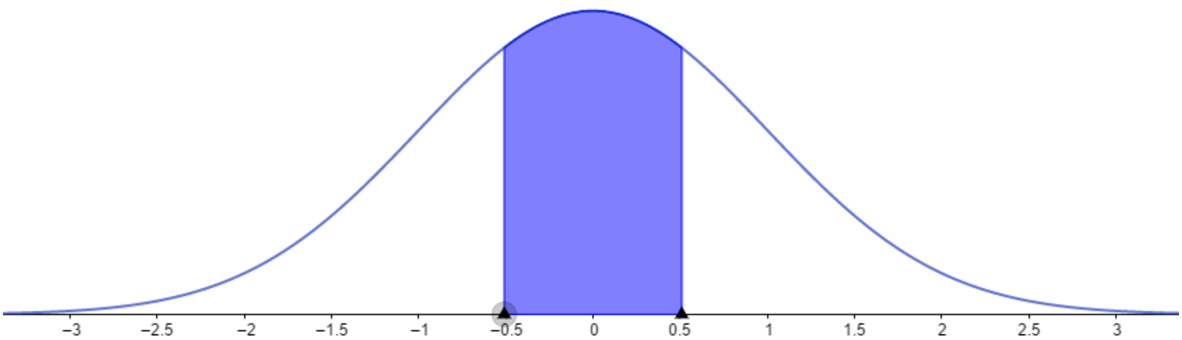

Figura 3.13: Função de densidade da v.a. $U \sim N(0, 1)$.

É usual representar a função de densidade da normal reduzida por ϕ e a função de distribuição por Φ.
A distribuição normal reduzida está tabelada. Para se utilizar a referida tabela para $X \sim N(\mu, \sigma)$ é necessário proceder à mudança de variável e tem-se

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\,\pi\,\sigma^2}}\, e^{-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2} dx$$

fazendo

$$u_1 = \frac{x_1 - \mu}{\sigma}, \; u_2 = \frac{x_2 - \mu}{\sigma}$$

tem-se

$$P(u_1 \leq U \leq u_2) = \int_{u_1}^{u_2} \frac{1}{\sqrt{2\,\pi}}\, e^{-\dfrac{u^2}{2}}\, du = \Phi(u_2) - \Phi(u_1).$$

Para determinar a ordenada da curva normal correspondente a uma abcissa $X = x_0$ tem-se

$$f(x_0) = \frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma}).$$

Para a v.a $U \sim N(0,1)$ tem-se (verifique graficamente):

$$P(U \leq -u) = P(U \geq u),\ P(-1 \leq U \leq 0) = P(0 \leq U \leq 1)$$

e

$$\Phi(u) = 1 - \Phi(-u),\ \phi(u) = \phi(-u).$$

Exemplos:

1. Vejamos a correspondência entre X e U, $X \sim N(20,2)$ (Figura 3.14), considerando os seguintes valores de X: $x_1 = 16, x_2 = 18, x_3 = 20, x_4 = 22$ e $x_5 = 24$ os correspondentes valores em U são:

$$\begin{array}{rcl} u_1 & = & \dfrac{16-20}{2} & = & -2 \\ u_2 & = & \dfrac{18-20}{2} & = & -1 \\ u_3 & = & \dfrac{20-20}{2} & = & 0 \\ u_4 & = & \dfrac{22-20}{2} & = & 1 \\ u_5 & = & \dfrac{24-20}{2} & = & 2 \end{array}$$

   Sendo $U = \dfrac{X-\mu}{\sigma} \Leftrightarrow X = U\,\sigma + \mu$ pode escrever-se

$$\begin{array}{rcl} 16 & = & \mu - 2\sigma \\ 18 & = & \mu - \sigma \\ 20 & = & \mu \\ 22 & = & \mu + \sigma \\ 24 & = & \mu + 2\sigma \end{array}$$

2. Sendo $X \sim N(3,2)$, determine $P(X \leq 2,14)$ (Figura 3.15).

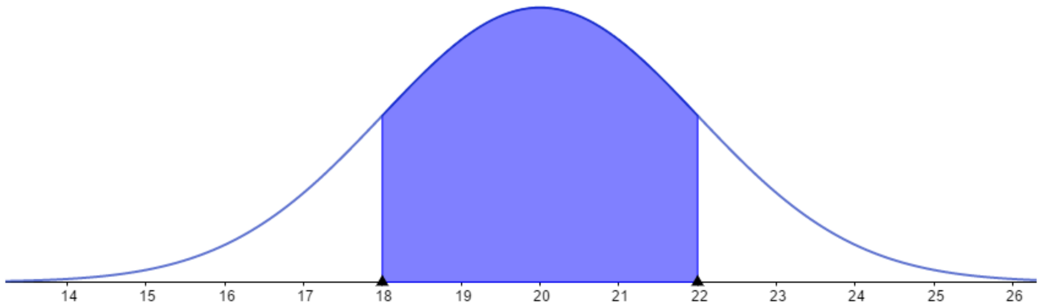

Figura 3.14: Função de densidade da v.a. $X \sim N(20, 2)$.

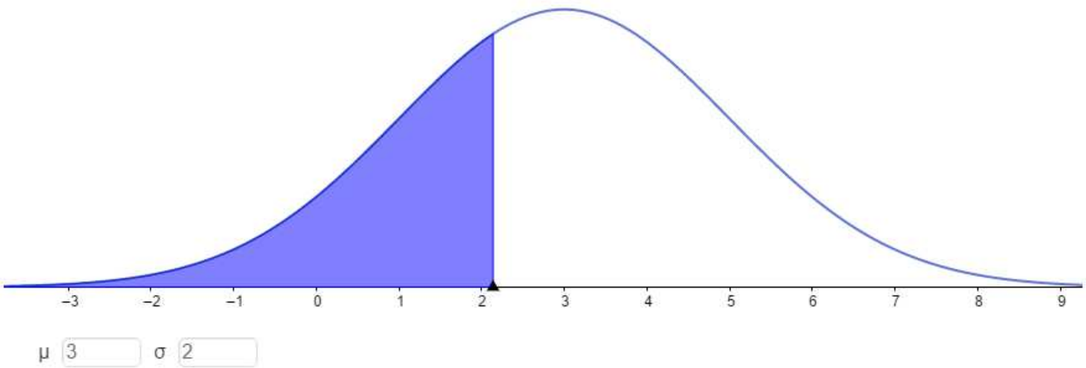

Figura 3.15: $P(X \leq 2,14)$ com $X \sim N(3,2)$.

$$\begin{aligned} P(X \leq 2,14) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{2,14-\mu}{\sigma}\right) \\ &= P\left(\frac{X-3}{2} \leq \frac{2,14-3}{2}\right) \\ &= P(U \leq -0,43) \\ &= \Phi(-0,43) \\ &= 1-\Phi(0,43) = 1-0,6664 \cong 0,3336. \end{aligned}$$

3. Um fabricante de baterias sabe, por experiência passada, que as baterias que produz têm vida média de 600 dias, desvio padrão de 100 dias e que a duração tem distribuição aproximadamente normal. Oferece uma garantia de 312 dias ou seja, troca as baterias que apresentam falhas nesse período. São fabricadas 10000 baterias mensalmente. Quantas deverá trocar, pelo uso da garantia, mensalmente?
Sendo $X =$"duração da bateria"(em dias) com $\mu = 600$ dias, $\sigma = 100$

dias é necessário determinar $P(X < 312)$.

$$\begin{aligned} P(X < 312) &= P\left(\frac{X-600}{100} < \frac{312-600}{100}\right) \\ &= P(U < -2,88) \\ &= \Phi(-2,88) \\ &= 1 - \Phi(2,88) = 1 - 0,998 \cong 0,002. \end{aligned}$$

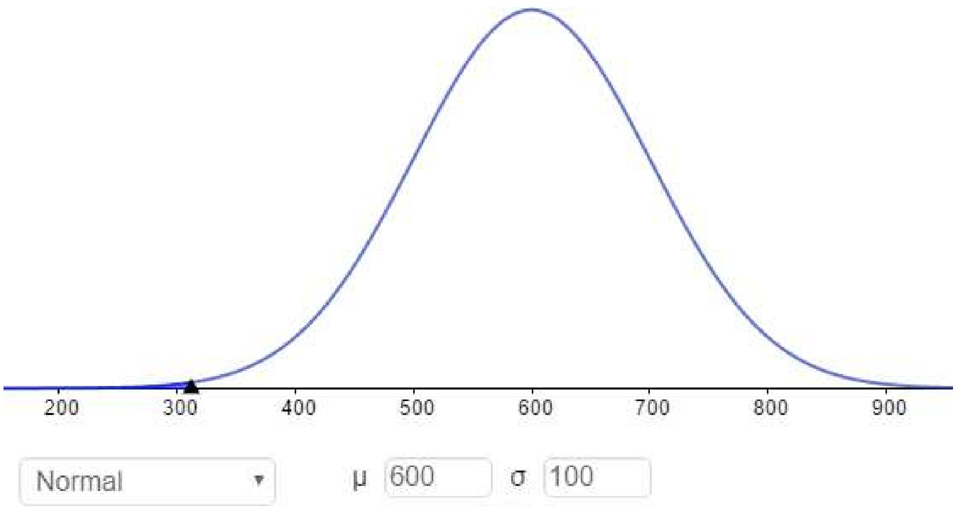

Figura 3.16: $P(X < 312)$ com $X \sim N(600, 100)$.

Assim, mensalmente deverá substituir $10000 \times 0.002 = 20$ baterias.

4. Uma fábrica de automóveis sabe que os motores que fabrica têm duração com distribuição aproximadamente normal e que fazem em média 150000 km, com um desvio padrão de 5000 km. Qual a probabilidade de um automóvel, escolhido ao acaso, dos fabricados por esta empresa ter um motor que dure:

a) menos de 170000 km?

Definindo $X =$"duração do motor"(em km) com $\mu = 150000$ km, $\sigma = 5000$ km tem-se

$$\begin{aligned} P(X < 170000) &= P\left(\tfrac{X-150000}{5000} < \tfrac{170000-150000}{5000}\right) \\ &= P(U < 4) \\ &= \Phi(4) \cong 1. \end{aligned}$$

b) entre 140000 km e 165000 km?

$$\begin{aligned}
P(140000 < X < 165000) &= P\left(\frac{140000-150000}{5000} < \frac{X-\mu}{\sigma} < \frac{165000-150000}{5000}\right) \\
&= P(-2 < U < 3) \\
&= \Phi(3) - \Phi(-2) \\
&= \Phi(3) - 1 + \Phi(2).
\end{aligned}$$

c) Se a fábrica substituir o motor que apresenta duração inferior à garantia, qual deve ser essa garantia, para que a percentagem de motores substituídos seja inferior a 0,2%?

$$\begin{aligned}
P(X \leq X_\alpha) &= 0,002 \Leftrightarrow \\
\Leftrightarrow P\left(\frac{X-150000}{5000} \leq \frac{X_\alpha-150000}{5000}\right) &= 0,002 \Leftrightarrow \\
\Leftrightarrow P\left(U \leq u_\alpha\right) &= 0,002 \Leftrightarrow \\
\Leftrightarrow \Phi(-u_\alpha) &= 0,002 \Leftrightarrow \\
\Leftrightarrow 1 - \Phi(u_\alpha) &= 0,002 \Leftrightarrow \\
\Leftrightarrow \Phi(u_\alpha) &= 0,998
\end{aligned}$$

Pela tabela da normal reduzida tem-se $u_\alpha = 2,88$ de

$$X = U\,\sigma + \mu$$

vem

$$x_\alpha = -2,88 \times 5000 + 150000 \cong 135600$$

podendo concluir-se que a garantia a oferecer deve ser de 135 600 km.

Alguns Resultados Associados à Distribuição Normal

Teorema 3.2.4 (Aditividade da Normal) *Se as v.a. X_i com $i \in \{1, \ldots, n\}$ são independentes então*

$$X_i \sim N(\mu_i, \sigma_i)$$

então

$$\sum_{i=1}^{n} \alpha_i\, X_i \sim N(\sum_{i=1}^{n} \alpha_i\, \mu_i, \sqrt{\sum_{i=1}^{n} (\alpha_i)^2\, (\sigma_i)^2})$$

com $\alpha_i \in \mathbb{R}$

Corolário 3.2.1 *Se as v.a. X_i com $i \in \{1, \ldots, n\}$ são independentes e identicamente distribuídas então*

$$X_i \sim N(\mu, \sigma)$$

então

$$\sum_{i=1}^{n} X_i \sim N(n\,\mu, \sigma\sqrt{n}).$$

Exemplo: O conteúdo de certo tipo de garrafas é aleatório e com distribuição normal de $\mu = 1$ e $\sigma = 0.02$. Se 3 garrafas forem despejadas para um recipiente, qual a probabilidade de este ficar com um volume de líquido superior a 3,1 litros?
Definindo $X_i =$"quantidade de líquido da garrafa i" com $i \in \{1, 2, 3\}$ e sendo as variáveis independentes e identicamente distribuídas com

$$X_i \sim N(\mu = 1, \sigma = 0,02)$$

considere-se $V =$"volume de líquido no recipiente" com $V = X_1 + X_2 + X_3$ e

$$V \sim N(\mu = 3, \sigma = \sqrt{3 \times (0.002)^2}).$$

O que se pretende é

$$\begin{aligned}
P(V > 3,1) &= P\left(U > \tfrac{3,1-3}{0,03464}\right) \\
&= P(U > 2,89) \\
&= 1 - P(U \leq 2,89) \\
&= 1 - \Phi(2.89) = 1 - 0,9981 = 0,0019.
\end{aligned}$$

Teorema 3.2.5 (Limite Central) *Sejam $X_1, X_2, X_3, \ldots$ com $\{X_n\}$ uma sucessão de v.a. independentes, identicamente distribuídas com média μ e variância σ^2 finitas. Seja*

$$Z_n = X_1 + X_2 + \ldots + X_n, \quad n = 1, 2, \ldots$$

a v.a.

$$U = \frac{Z_n - n\mu}{\sigma\sqrt{n}}$$

tem distribuição assimptótica normal reduzida.

Exemplo: Uma pessoa poupa por dia 2,5 euros, 5 euros e 10 euros com probabilidade 0.45, 0.15 e 0.4 respectivamente. Determine a probabilidade da referida pessoa poupar menos de 1000 euros em 180 dias.
Definindo $X_i =$"quantia poupada no dia i" com $i \in \{1, 2, \ldots, 180\}$, assumindo que as variáveis são independentes ou seja, a quantia poupada num dia não é influenciada pela que tenha poupado nos dias anteriores, e que as v.a. são identicamente distribuídas tem-se:

$$\mu = \sum_x x f(x) = 2.5 \times 0.45 + 5 \times 0.15 + 10 \times 0.4 = 5.875$$

e

$$\sigma^2 = \sum_x x^2\, f(x) - \left(\sum_x x\, f(x)\right)^2$$

que efectuando os cálculos vem

$$\sigma^2 = 46.5625 - 34.5156 = 12.0469.$$

O que se pretende é

$$\begin{aligned} P\left(\sum_{i=1}^{180} X_i < 1000\right) &= P\left(U < \tfrac{1000-180\times 5.875}{\sqrt{12.0469}\sqrt{180}}\right) \\ &= P(U < 1.26) \\ &= \Phi(1.26) = 0.8962. \end{aligned}$$

3.2.4 Distribuição do Qui-Quadrado

Esta distribuição é importante na inferência estatística, pois permite fazer inferência sobre a variância de uma população a partir de uma amostra, sendo também utilizada nas situações em que se considera o quadrado de uma v.a. com distribuição normal estandardizada.
Considerando um conjunto de n v.a. Z_i com $i \in \{1, \ldots, n\}$ que verificam:

i) cada variável Z_i segue uma distribuição normal estandardizada $Z_i \sim N(0, 1)$,

ii) as variáveis Z_i são mutuamente independentes,

a v.a.

$$X = \sum_{i=1}^{n} Z_i^2 = Z_1^2 + Z_2^2 + \ldots + Z_n^2$$

segue uma distribuição χ_n^2, onde n é o número de graus de liberdade (número de v.a. envolvidas na definição da v.a. X). Na figura 3.17 representa-se a função de densidade de uma v.a. $X \sim \chi_6^2$.

Definição 3.2.4 (Distribuição do Qui-Quadrado) *Uma v.a. contínua X tem distribuição do qui-quadrado com n graus de liberdade, escreve-se $X \sim \chi_n^2$, quando a sua função de densidade é dada por*

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}} G(\frac{n}{2})}\, e^{-\frac{x}{2}}\, x^{\frac{n}{2}-1} & se \quad x > 0 \\ 0 & se \quad x \leq 0 \end{cases}$$

e se n for um número par positivo a função de distribuição é

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{2}} \displaystyle\sum_{i=0}^{\frac{n}{2}-1} \frac{1}{i!} \left(\frac{x}{2}\right)^i & se \quad x > 0 \\ 0 & se \quad x \leq 0 \end{cases}.$$

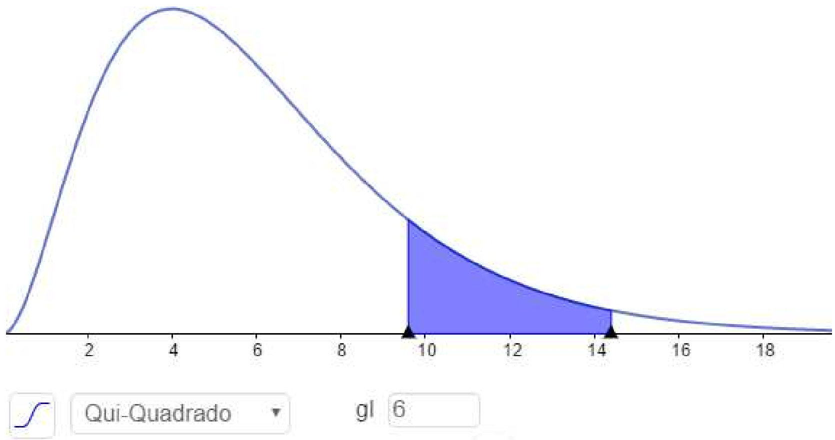

Figura 3.17: $X \sim \chi^2_6$.

A função $G(u)$, que aparece na definição anterior, denomina-se função gama e representa-se por

$$G(u) = \int_0^{+\infty} x^{u-1}\, e^{-x}\, dx \quad com \quad u > 0$$

sendo tal que

$$G\left(\frac{n}{2}\right) = \begin{cases} \dfrac{1 \times 3 \times 5 \times \ldots \times (n-2)}{2^{\frac{n-1}{2}}} \sqrt{\pi} & se \quad n > 1 \ e \ impar \\ \left(\frac{n}{2} - 1\right)! & se \quad n \ par \end{cases} .$$

A distribuição qui-quadrado tem como parâmetro o número de graus de liberdade que é um número inteiro positivo. A função de densidade é uma função positiva e não simétrica e o seu aspecto gráfico depende do parâmetro n.

Teorema 3.2.6 *Se X é uma v.a. contínua com distribuição qui-quadrado com n graus de liberdade tem-se que:*

$$\mu_X = E[X] = n$$

$$\sigma^2_X = Var(X) = 2n.$$

Se $X \sim \chi^2_n$ então, quando n aumenta, prova-se que a função de distribuição da v.a.

$$Z = \sqrt{2X} - \sqrt{2n-1}$$

aproxima-se da função de distribuição normal estandardizada. Este resultado é muito útil quando não existem valores tabelados para o número de graus de liberdade que se pretende, por ser muito grande, e não se tem um computador para determinar o valor pretendido. Neste caso, o que se faz para determinar $P(X \leq x)$ é

$$P(X \leq x) \approx P(Z \leq \sqrt{2x} - \sqrt{2n-1}), \quad com \ \ Z \sim N(0,1).$$

A distribuição em causa está tabelada e na tabela que vamos utilizar têm-se valores de χ_n^2 para $n = 1, 2, \ldots, 29, 30, 40, 50, \ldots, 100$ e os valores da tabela correspondem às áreas

$$\alpha = 0.995, 0.99, 0.975, 0.95, 0.9, 0.75, 0.5, 0.25, 0.1, 0.05, 0.025, 0.01, 0.005.$$

Sendo

$$\alpha = \int_{\chi_n^2(\alpha)}^{+\infty} f(u)du = P(X \geq \chi_n^2(\alpha))$$

o valor $\chi_n^2(\alpha)$ é um valor da v.a. X (será um valor do interior da tabela). Assim, pode dizer-se que na tabela a pares (n, α) adequados se faz corresponder $x = \chi_n^2(\alpha)$ tal que

$$P(X \geq x) = P\left(X \geq \chi_n^2(\alpha)\right) = \alpha$$

que é equivalente a

$$P(X \leq x) = P\left(X \leq \chi_n^2(\alpha)\right) = 1 - \alpha$$

(visualizar graficamente).

Exemplos:

1. Considere um circuito onde é induzida uma corrente eléctrica cuja intensidade I (ampere) tem uma distribuição normal estandardizada. Sabe-se que a potência P (watt) dissipada numa resistência é $P = RI^2$, em que R é o valor da resistência (ohm).

   a) Identifique a distribuição de P, supondo a resistência constante em 1 ohm.
   Sendo $P = RI^2 = I^2$ tem-se $P \sim \chi_1^2$.

   b) Determine os percentis 0.5, 1 e 10 de P.
   Para identificar o valor α associado a cada percentil basta considerar, por exemplo para 0.5

   $$0.5 = 100(1-p) \Leftrightarrow \alpha = 0.995$$

   de igual forma calculamos os restantes valores. Sendo a resposta:
   Percentil 0.5 $\chi_1^2(\alpha = 0.995) = 0.000$
   Percentil 1 $\chi_1^2(\alpha = 0.99) = 0.000$
   Percentil 10 $\chi_1^2(\alpha = 0.9) = 0.015$.

2. A posição de uma dada partícula no espaço é dada pelas suas coordenadas cartesianas (X, Y, Z) que se supõe serem v.a. independentes e normais estandardizadas. O quadrado da distância à origem é $D^2 = X^2 + Y^2 + Z^2$.

   a) Identifique a distribuição de D^2 e determine a sua média e variância.
   Sendo
   $$D^2 = X^2 + Y^2 + Z^2$$
   tem-se
   $$D^2 \sim \chi^2_3,$$
   com
   $$E(D^2) = 3$$
   e
   $$Var(D^2) = 2 \times 3 = 6.$$

   b) Determine $P(D^2 \leq 0.584)$.
   Da tabela do qui-quadrado
   $$P\left(D^2 \geq 0.584\right) = 0.9$$
   logo
   $$P\left(D^2 \leq 0.584\right) = 1 - 0.9 = 0.1.$$

3.2.5 Distribuição do T-Student

Esta distribuição é importante na inferência estatística, pois permite fazer inferências sobre a média de uma população a partir de uma amostra.

Considerando uma v.a. Z com distribuição normal estandardizada e Y uma v.a. com $Y \sim \chi^2_n$ e se X, Y são v.a. independentes então a v.a.

$$X = \frac{Z}{\sqrt{\frac{Y}{n}}}$$

segue uma distribuição T_n, onde n é o número de graus de liberdade.

Definição 3.2.5 (Distribuição t-Student) *Uma v.a. contínua X tem distribuição do t-student com n graus de liberdade, escreve-se $X \sim T_n$, quando a sua função de densidade é dada por*

$$f(x) = \frac{G\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\, G\left(\frac{n}{2}\right)} \left(\frac{1+x^2}{n}\right)^{-\frac{n+1}{2}} \quad , \quad x \in \mathbb{R}\,, \quad n > 0.$$

A distribuição t-student tem como parâmetro o número de graus de liberdade que é um número inteiro positivo. A função de densidade é uma função simétrica em relação ao eixo $x = 0$, tal como a normal estandardizada, tem a forma de sino e tem um único máximo ($x = 0$). O seu aspecto gráfico depende do número de graus de liberdade, na Figura 3.18 representa-se o caso particular da função de densidade de uma v.a. $X \sim T(10)$. À medida que o número de graus de liberdade aumenta (na prática $n \geq 30$) a curva aproxima-se da normal estandardizada.

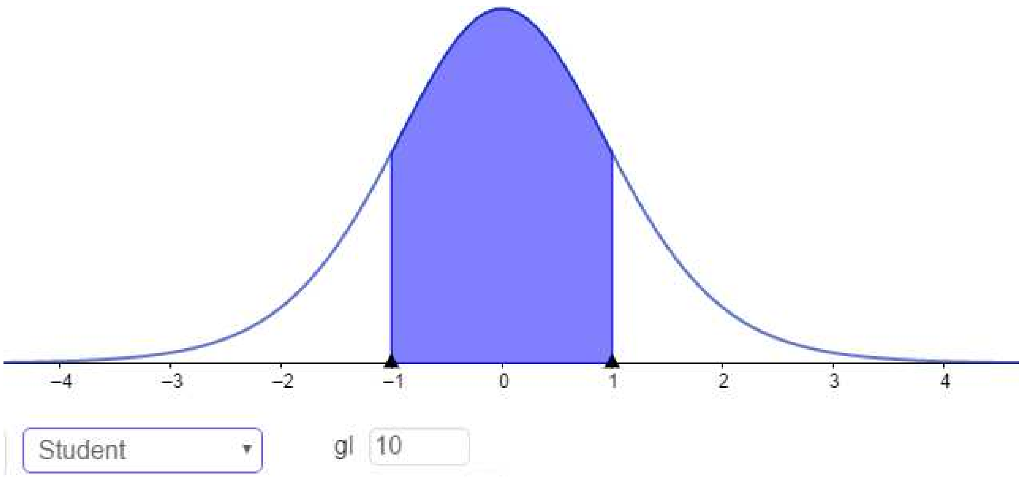

Figura 3.18: $X \sim T(10)$.

Teorema 3.2.7 *Se X é uma v.a. contínua com distribuição t-student com n graus de liberdade tem-se que:*

$$\mu_X = E[X] = 0, \quad se \ \ n > 1$$

$$\sigma_X^2 = Var(X) = \frac{n}{n-2} \quad se \ \ n > 2.$$

A distribuição em causa está tabelada e na tabela que vamos utilizar têm-se valores de T_n para $n = 1, 2, \ldots, 29, 30, 40, 50, \ldots, 100, 120, \infty$ e os valores da tabela correspondem às áreas

$$\alpha = 0.25, 0.2, 0.1, 0.15, 0.05, 0.025, 0.01, 0.005, 0.0005.$$

Sendo

$$\alpha = \int_{T_n(\alpha)}^{+\infty} f(u)du = P(X \geq T_n(\alpha))$$

o valor $T_n(\alpha)$ é um valor da v.a. X (será um valor do interior da tabela). Assim, pode dizer-se que na tabela a pares (n,α) adequados faz corresponder $x = T_n(\alpha)$ tal que

$$P(X \geq x) = P(X \geq T_n(\alpha)) = \alpha$$

que, dada a simetria da distribuição, é equivalente a

$$P(X \leq -x)$$

(visualizar graficamente).
Exemplo:
Seja $X \sim N(5,2)$ e $Y \sim \chi^2_{10}$ calcule a probabilidade da v.a.

$$Z = \frac{\frac{X-5}{2}}{\sqrt{\frac{Y}{10}}}$$

ser superior a 2.2.
Se $X \sim N(5,2)$ então $\dfrac{X-5}{2} \sim N(0,1)$ pelo que $Z \sim T_{10}$. O que se pretende é $P(Z > 2.2) = 0.025$, dado pela tabela.

3.2.6 Distribuição do F-Snedecor

Esta distribuição permite comparar variâncias amostrais e fazer inferências sobre as variâncias das populações.

Considerando as v.a. $X_1 \sim \chi^2_{n_1}$ e $X_2 \sim \chi^2_{n_2}$ independentes então a v.a.

$$X = \frac{\frac{X_1}{n_1}}{\frac{X_2}{n_2}}$$

segue uma distribuição F_{n_1,n_2}, com n_1 e n_2 graus de liberdade.

Definição 3.2.6 (Distribuição F-Snedecor) *Uma v.a. contínua X tem distribuição do F-Snedecor com n_1 e n_2 graus de liberdade, escreve-se*

$$X \sim F_{n_1,n_2},$$

quando a sua função de densidade é dada por

$$f(x) = \frac{G\left(\frac{n_1+n_2}{2}\right)}{G\left(\frac{n_1}{2}\right)\ G\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{-\frac{n_1}{2}} x^{\frac{n_1}{2}-1} \left(1+\frac{n_1}{n_2}x\right)^{-(n_1+n_2)/2},\ x>0,\ n_1,n_2>0.$$

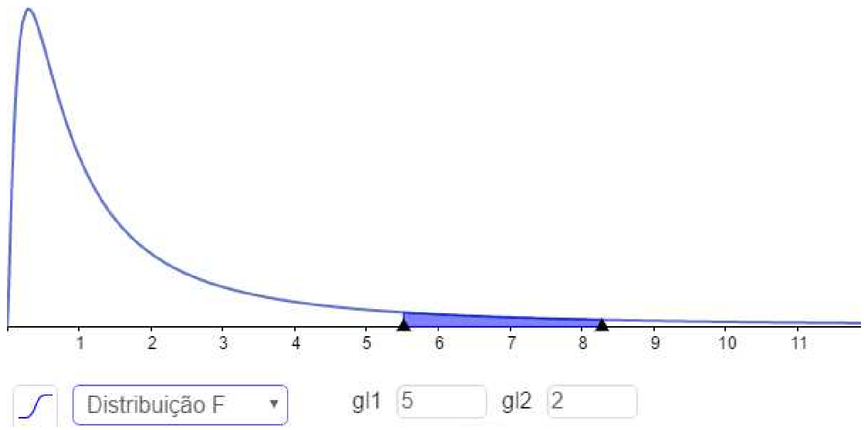

Figura 3.19: $X \sim F_{5,2}$.

A distribuição F-Snedecor tem como parâmetros os números de graus de liberdade n_1 e n_2 que são inteiros positivos, e a distribuição depende da ordem dos parâmetros. A função de densidade é uma função positiva e não simétrica. O seu aspecto gráfico depende dos números de graus de liberdade. Na Figura 3.19 representa-se a função de densidade de uma v.a. $X \sim F_{5,2}$.

Teorema 3.2.8 *Se X é uma v.a. contínua com distribuição F-Snedecor com n_1 e n_2 graus de liberdade tem-se que:*

$$\mu_X = E[X] = \frac{n_2}{n_2 - 2}, \quad se \ \ n_2 > 2$$

$$\sigma_X^2 = Var(X) = \frac{2n_2^2\,(n_1 + n_2 - 2)}{n_1\,(n_2 - 2)^2\ (n_2 - 4)} \quad se \ \ n_2 > 4.$$

A distribuição em causa está tabelada e na tabela que vamos utilizar têm-se valores de F_{n_1,n_2} correspondentes às áreas $\alpha = 0.01, 0.25, 0.05, 0.1$. Sendo

$$\alpha = \int_{F_{n_1,n_2}(\alpha)}^{+\infty} f(u)du = P(X \geq F_{n_1,n_2}(\alpha))$$

o valor $F_{n_1,n_2}(\alpha)$ é um valor da v.a. X (será um valor do interior da tabela). Assim, pode dizer-se que na tabela temos que identificar (n_1, n_2, α) tais que $x = F_{n_1,n_2}(\alpha)$ sendo

$$P(X \geq x) = P(X \geq F_{n_1,n_2}(\alpha)) = \alpha$$

(visualizar graficamente).
Pode estabelecer-se a seguinte relação:

$$F_{n_1,n_2}(1-\alpha) = \frac{1}{F_{n_2,n_1}(\alpha)},$$

note-se que a ordem do número de graus de liberdade trocou.
Exemplo:
Sendo $X \sim F_{8,4}$ determine o valor esperado, a variância de X e os valores $F_{8,4}(0.05)$ e $F_{8,4}(0.95)$.
Tem-se

$$E(X) = \frac{n_2}{n_2 - 2} = \frac{4}{4-2} = 2$$

e

$$Var(X) = \frac{2n_2^2\,(2 + n_1 + n_2)}{n_1\,(n_2 - 2)^2\ (n_2 - 4)} \quad se \;\; n_2 > 4$$

logo a variância da distribuição não está definida (não toma um valor finito).
Consultando a tabela tem-se:

$$F_{8,4}(0.05) = 6.04$$

e

$$F_{8,4}(0.95) = \frac{1}{F_{4,8}(0.05)} = \frac{1}{3.84} \simeq 0.26.$$

3.3 Aproximação da Distribuição Normal para as Distribuições Binomial e de Poisson

A distribuição normal, apesar de ser uma distribuição contínua, é utilizada com frequência como aproximação de uma distribuição discreta quando esta tem a forma de sino e é simétrica. No caso particular das distribuições binomial e Poisson existe convergência para a distribuição normal quando os parâmetros que as identificam tendem para determinados valores limite. Utilizara a distribuição normal para aproximar estas distribuições é muito prático dado que temos valores tabelados para a distribuição normal reduzida.

Teorema 3.3.1 (Moivre-Laplace) *A sucessão de v.a.*

$$\{Z_n\} = \left\{ \frac{X_n - np}{\sqrt{npq}} \right\}, \quad n = 1, 2, \ldots$$

onde $X_n \sim B(n,p)$ converge em distribuição para uma v.a. normal reduzida ou seja, $\{Z_n\}$ tem distribuição assimptótica normal reduzida.

Teorema 3.3.2 *Seja $X \sim P(\lambda)$ (sendo $\mu = \lambda = \sigma^2$) então quando $\lambda \to +\infty$ a função de distribuição da v.a. Z*

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

converge para a função de distribuição da normal reduzida.

Estes resultados advêm do teorema do limite central, apresentado anteriormente. Embora as aproximações apresentadas só sejam exactas no limite, em que

$$na\ \ binomial\ \ n \to +\infty$$

$$na\ \ Poisson\ \ \lambda \to +\infty$$

podem utilizar-se e esperar boas aproximações quando:

- na binomial: n seja grande (sendo usual considerar-se grande quando $np \geq 5$ e $n(1-p) \geq 5$),
- na Poisson: λ grande (sendo usual considerar-se grande quando $\lambda \geq 5$).

Convém ter presente que estamos a aproximar uma distribuição discreta utilizando uma contínua. Assim, a ocorrência $X = x$ tem de ser redefinida para $x - \delta \leq X \leq x + \delta$, dado que numa v.a. contínua $P(X = x) = 0$. A este procedimento chama-se correcção de continuidade, sendo o seu objectivo reduzir o erro que se comete na aproximação que se faz.

Definição 3.3.1 (Correcção de Continuidade) *Seja X uma v.a. discreta. A correcção de continuidade consiste em converter X numa v.a. contínua, reescrevendo o evento $x_1 \leq X \leq x_2$ como $x_1 - 0.5 \leq X \leq x_2 + 0.5$ e calculando*

$$P(x_1 \leq X \leq x_2) \cong P(x_1 - 0.5 \leq X \leq x_2 + 0.5)$$

a partir da função de densidade de probabilidade da v.a. contínua.

Refira-se que se $X \sim B(n,p)$ então para utilizar um a aproximação pela normal tem-se

$$P(X = x) \cong P(x - 0.5 \leq X \leq x + 0.5)$$

assumindo

$$X \sim N(\mu = np, \sigma = \sqrt{npq})$$

pode-se calcular a probabilidade pretendida recorrendo à tabela da normal reduzida

$$P\left(\frac{(x-0.5)-np}{\sqrt{npq}} \leq X \leq \frac{(x+0.5)-np}{\sqrt{npq}}\right) =$$

$$= \Phi\left(\frac{(x+0.5)-np}{\sqrt{npq}}\right) - \Phi\left(\frac{(x-0.5)-np}{\sqrt{npq}}\right).$$

Também no caso dos valores da função de distribuição a correcção deve ser feita sendo

$$P(X \leq x_1) \cong P(X \leq x_1 + 0.5).$$

Exemplos:

1. Uma empresa comercializa garrafas de vinho do Porto de 1 litro. Supõe-se que 40% dessas garrafas contêm realmente uma menor quantidade de líquido do que o indicado no rótulo. Em 100 garrafas existentes numa loja, qual a probabilidade de haver 30 com menos de 1 litro?
Definindo

$$X = \text{"número de garrafas, em 100, com menos de 1 litro"}$$

com $X \sim B(n = 100, p = 0.4)$, pretende-se $P(X = 30)$ cujo valor exacto é

$$P(X = 30) \;\; = \;\; C_{100}^{30} (0.4)^{30} \; (0.6)^{70},$$

mas na tabela da distribuição binomial não existe n=100, pelo que vamos obter um valor aproximado usando a distribuição normal. Sendo $X \sim N(\mu = np = 40, \sigma = \sqrt{npq} = \sqrt{100 \times 0.4 \times 0.6})$ e aplicando a correcção de continuidade tem-se

$$\begin{aligned}
P(X = 30) &= P(30 - 0.5 \leq X \leq 30 + 0.5) \\
&= P(29.5 \leq X \leq 30.5) \\
&= P\left(\frac{29.5 - 40}{4.9} \leq X \leq \frac{30.5 - 40}{4.9}\right) \\
&= P(-2.14 \leq U \leq -1.94) \\
&= 1 - \Phi(1.94) - 1 + \Phi(2.14) \cong 0.01.
\end{aligned}$$

2. O número de chamadas que chegam num período de 5 minutos à central telefónica de uma empresa é uma v.a. com distribuição de Poisson de parâmetro $\lambda = 10$. Calcule a probabilidade de em meia hora chegarem 65 chamadas?
Definindo

$$X = \text{"número de chamadas que chegam em 5 minutos"}$$

com $X \sim P(\lambda = 10)$, pretende-se trabalhar com

$$X' = \text{"número de chamadas que chegam em 30 minutos"}$$

com $X' \sim P(\lambda^{'} = 6 \times 10 = 60)$ sendo a probabilidade pretendida $P(X' = 65)$. Dado que o valor $\lambda^{'} = 60$ não existe na tabela da distribuição de Poisson vamos obter um valor aproximado usando a distribuição normal. Sendo $X' \sim N(\mu = 60, \sigma = \sqrt{60} = 7.75)$ e aplicando a correcção de continuidade tem-se

$$\begin{aligned}
P(X' = 65) &= P(65 - 0.5 \leq X \leq 65 + 0.5) \\
&= P(64.5 \leq X \leq 65.5) \\
&= P\left(\frac{64.5 - 60}{7.75} \leq X \leq \frac{65.5 - 60}{7.75}\right) \\
&= P(0.58 \leq U \leq 0.71) \\
&= \Phi(0.71) - \Phi(0.58) \cong 0.0421.
\end{aligned}$$

Capítulo 4

Teoria da Amostragem

Na teoria das probabilidades partimos de determinado modelo com o qual determinamos a probabilidade de ocorrência de resultados. Neste capítulo, vamos trabalhar no âmbito da estatística, em que para um dado conjunto de observações se pretendem obter conclusões sobre o modelo que melhor as traduz.

4.1 Noções de população amostra e estatística

O conjunto de observações ou dados a tratar resultam da realização de experiências ou de inquéritos sobre um conjunto de elementos, a que se chama **amostra**, mas pretendem-se conclusões para o conjunto mais alargado onde a amostra se insere, este conjunto denomina-se de **população**. Assim, sendo o objectivo inferir sobre características da população com base nos dados recolhidos na amostra, este é um processo de inferência estatística.

É essencial obter amostras representativas da população, para que as conclusões que se possam retirar sejam aceitáveis. A técnica ou método de obtenção das amostras representativas da população denomina-se de **amostragem**. As dificuldades inerentes ao processo de amostragem são minimizadas pela aplicação da **amostragem casual**. Neste caso, as n variáveis observadas que constituem a amostra,X_i com $i \in \{1, 2, \ldots, n\}$, são independentes e identicamente distribuídas.

Podemos ter um processo de **amostragem com reposição**, em que um elemento da população depois de seleccionado é reposto na população, podendo ser seleccionado novamente. Neste caso temos independência entre os elementos seleccionados.

Podemos ter um processo de **amostragem sem reposição**, em que um elemento da população depois de seleccionado não é reposto na população, pelo que apenas será seleccionado uma vez. Neste caso, não existe independência entre os elementos seleccionados.

Exemplo: Supondo que numa fábrica que produz componentes para automóveis se pretende avaliar as peças defeituosas produzidas durante 10 dias, mas recolhendo uma amostra de 50 peças em cada dia, em diferentes períodos do dia, do total de peças produzidas pela fábrica nesse dia. A população em causa é o universo das peças produzidas pela fábrica em 10 dias e a amostra é constituída pelas 500 peças seleccionadas.

A população relativamente à qual se pretende fazer a inferência pode ser finita ou infinita. O número de elementos da população denomina-se de dimensão (ou tamanho) da população (iremos representar por N) e o número de elementos da amostra de dimensão (ou tamanho) da amostra (iremos representar por n).

Quando a dimensão da população é grande quando comparada com a dimensão da amostra as diferenças entre amostragem com reposição e sem reposição tornam-se menos relevantes.Por outro lado, quando uma população finita é submetida a um processo de amostragem com reposição pode considerar-se *infinita* (em termos teóricos).

O ponto de partida para os problemas de inferência que vamos tratar é sempre uma amostra aleatória (obtida por processos aleatórios em que cada elemento da população tem a mesma probabilidade de ser seleccionado para a amostra) $X_1, X_2, \ldots, X_n$ da população. Só depois, com base na amostra, se utiliza uma função $g(X_1, X_2, \ldots, X_n)$, chamada estatística, para estimar os parâmetros pretendidos. Sendo $\widehat{\theta} = g(X_1, X_2, \ldots, X_n)$ um parâmetro a estimar que, como depende das v.a. $X_1, X_2, \ldots, X_n$, é uma v.a. que terá uma determinada distribuição de probabilidade. Assim o que se segue é o estudo destas distribuições de probabilidade, dado que são essenciais para se avaliarem estatísticas e para as comparar com outras que se possam utilizar para estimar os mesmos parâmetros. Vamos supor que a amostra aleatória com que trabalhamos é independente e identicamente distribuída o que nos permite definir uma distribuição de probabilidade conjunta:

$$f(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

Em termos práticos, esta suposição não causa problemas pois ter uma amostra aleatória que é i.i.d é o mesmo que recolher uma amostra aleatória de uma população infinita ou, numa população finita, recolher uma amostra aleatória com reposição. Se for uma população finita sem reposição não teremos grandes problemas se o N for grande quando comparado com o n.

Definição 4.1.1 (Estatística) *Uma estatística é uma função real das v.a. que constituem a amostra*

$$\widehat{\theta} = g(X_1, X_2, \ldots, X_n),$$

sendo também uma v.a..

4.2 Estatísticas Amostrais mais Usuais

Existem várias estatísticas que indicam o centro de um dado conjunto de dados (chamadas medidas de localização central),mas apenas vamos tratar a mais importante que é a média amostral. Esta estima a constante $E(X)$.

Definição 4.2.1 (Média Amostral) *Dada uma amostra $X_1, X_2, \ldots, X_n$ de dimensão n, a sua média amostral é*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Quando se determina a média amostral num amostra concreta, denota-se o valor por $\overline{x}$. Para descrever adequadamente os dados, além de ser necessário conhecer as medidas de localização, também é importante saber como se distribuem as observações em torno da média. Para tal determinam-se as medidas de variabilidade da amostra: variância amostral e desvio padrão amostral.

Definição 4.2.2 (Variância e Desvio Padrão Amostrais) *Dada uma amostra $X_1, X_2, \ldots, X_n$ de dimensão n, a variância amostral é dada por*

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$$

e o desvio padrão amostral é

$$S = \sqrt{S^2}.$$

Quando para uma concretização da amostra se determinam a variância e o desvio padrão amostrais são denotados por s^2 e s, respectivamente. Se a dimensão da amostra é inferior ou igual a 30 é conveniente utilizar a variância e o desvio padrão amostrais corrigidos, por darem uma melhor estimativa.

Definição 4.2.3 (Variância e Desvio Padrão Amostrais Corrigidos) *Dada uma amostra $X_1, X_2, \ldots, X_n$ de dimensão n, a variância amostral corrigida é dada por*

$$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right]$$

e o desvio padrão amostral corrigido é

$$S' = \sqrt{S'^2}.$$

É fácil verificar que

$$S'^2 = \frac{n}{n-1}S^2.$$

Exemplo: Para estudar a evolução dos preços do café numa determinada cidade escolheram-se, aleatoriamente, 4 estabelecimentos, onde se observaram os seguintes aumentos de preços (em euros) 0.12, 0.15, 0.17 e 0.20 num pacote de 250 gramas. Determine a média e a variância desta amostra aleatória.

$$\overline{x} = \frac{1}{4}(0.12 + 0.15 + 0.17 + 0.2) = 0.16 \ (euros)$$

e

$$s'^2 = \frac{1}{3}\left[(0.12-0.16)^2 + (0.15-0.16)^2 + (0.17-0.16)^2 + (0.2-0.16)^2\right] = 0.11.$$

Outra estatística importante é a proporção amostral $\widehat{P}$, que permite estimar uma proporção p de elementos de uma população que têm determinada característica. Esta estimativa exige que se tenha uma sequência de observações $X_1, X_2, \ldots, X_n$ que sejam independentes e que cada uma delas esteja associada a um sucesso ou a um insucesso (temos um processo de Bernoulli). Assim, o número de sucessos na amostra representa-se pela v.a.

$$X = \sum_{i=1}^{n} X_i$$

sendo a proporção amostral

$$\widehat{P} = \frac{X}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Definição 4.2.4 (Proporção Amostral) *Considere uma amostra $X_1, X_2, \ldots, X_n$ de dimensão n, em que as observações são independentes com distribuição de*

Bernoulli e $X_i = 1$ *ou* $X_i = 0$ *consoante o elemento observado tem ou não a característica pretendida (o sucesso). A proporção amostral é dada por*

$$\widehat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

que indica a proporção de sucessos da amostra.

Exemplo: Supondo que se colocou a um conjunto de 100 indivíduos, de uma dada cidade, a seguinte questão: "É vegetariano?". Tendo-se escolhido aleatoriamente os indivíduos a inquirir, concluiu-se que apenas 5 disseram ser vegetarianos. Para determinar a proporção amostral tem-se:

$$\widehat{p} = \frac{1}{100}\sum_{i=1}^{n} x_i = \frac{5}{100} = 0.05,$$

com base neste valor amostral pode fazer-se inferência sobre o valor da proporção p na população.
Se os erros cometidos na obtenção dos dados fossem desprezáveis(muitos deles compensam-se) e se fosse razoável obter todas as amostras possíveis, que permitiriam obter as estimativas amostrais, seria possível definir a distribuição teórica $f_{\widehat{\theta}}$ da estatística $\widehat{\theta}$, que se designa por distribuição amostral. Contudo, apesar de não ser viável obter todas as amostras possíveis a teoria existente permite estimar propriedades das distribuições das estatísticas amostrais. A distribuição de uma estatística $\widehat{\theta}$, habitualmente, depende da distribuição da população e da dimensão da amostra e permite avaliar e controlar a margem de erro devido à amostragem.

Definição 4.2.5 (Distribuição Amostral) *A distribuição amostral é a distribuição de*

$$\widehat{\theta} = g(X_1, X_2, \ldots, X_n),$$

uma estatística amostral.

4.3 Distribuição de Amostragem da Média

Retomando a v.a. média amostral $\overline{X}$ apresentada na secção anterior, vamos tratar da sua distribuição. Vamos começar por estudar algumas das suas propriedades nomeadamente a sua média e variância.

Teorema 4.3.1 (Média e Variância da Estatística $\overline{X}$) *Considere-se uma v.a. X de uma população infinita (ou finita com selecção com reposição), com uma distribuição f_X com média μ_X e variância σ_X^2 finitas, e uma amostra $X_1, X_2, \ldots, X_n$ de dimensão n independente e identicamente distribuída(i.i.d.) , as observações X_i são independentes e todas têm a mesma distribuição f_X. Nestas condições, a média amostral $\overline{X}$ tem uma distribuição cuja média e a variância são:*

$$\mu_{\overline{X}} = E[\overline{X}] = \mu_X$$

e

$$\sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n}. \tag{4.1}$$

No caso de se alterarem as condições da população e da amostragem surgem alterações na variância da estatística $\overline{X}$.

Teorema 4.3.2 (Média e Variância da Estatística $\overline{X}$) *Considere-se uma v.a. X de uma população de dimensão N, com uma distribuição f_X com média μ_X e variância σ_X^2 finitas, e uma amostra $X_1, X_2, \ldots, X_n$ de dimensão n recolhida sem reposição, sendo $n \leq N$. Nestas condições, a média amostral $\overline{X}$ tem uma distribuição cuja média e a variância são:*

$$\mu_{\overline{X}} = E[\overline{X}] = \mu_X$$

e

$$\sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n} \times \frac{N-n}{N-1}. \tag{4.2}$$

Quando a dimensão da população tende para infinito tem-se (4.2) a tender para (4.1). Definidas as propriedades anteriores seguem-se os resultados relacionados com a distribuição de $\overline{X}$.

Teorema 4.3.3 *Considere-se uma v.a. X de uma população com distribuição normal de média μ_X e variância σ_X^2 finitas, então a média amostral $\overline{X}$ tem uma distribuição normal média μ_X e variância $\frac{\sigma_X^2}{n}$, sendo equivalente a dizer que a v.a.*

$$U = \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}}$$

tem distribuição assimptótica normal reduzida.

O resultado que se segue surgem como consequência do Teorema do limite central.

Teorema 4.3.4 *Considere-se uma v.a. X de uma população infinita (ou finita com selecção com reposição) com média μ_X e variância σ_X^2 finitas , então a média amostral $\overline{X}$ tem uma distribuição aproximadamente normal com média μ_X e variância $\frac{\sigma_X^2}{n}$, sendo equivalente a dizer que a v.a.*

$$U = \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \tag{4.3}$$

tem distribuição assimptótica normal reduzida (para uma dimensão de amostra $n \geq 30$).

Este resultado pode aplicar-se no caso de uma população finita com dimensão N e de amostragem sem reposição se substituirmos (4.3) por

$$U = \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}} \times \frac{N-n}{N-1}}. \tag{4.4}$$

Os resultados apresentados são aplicáveis quando se conhece a variância da população, mas caso σ_X^2 seja desconhecida e se verifiquem determinadas condições, o resultado seguinte permite definir uma v.a. que envolve a média amostral e para a qual se conhece a distribuição.

Teorema 4.3.5 *Considere-se uma v.a. X de uma população com distribuição normal com média μ_X finita e variância σ_X^2 desconhecida, recolhida uma amostra que é i.i.d. de dimensão n, então a v.a. T*

$$T = \frac{\overline{X} - \mu_X}{\frac{S}{\sqrt{n}}} \sim T_{n-1}, \tag{4.5}$$

que envolve a média amostral $\overline{X}$, tem uma distribuição T-student com $n-1$ graus de liberdade.

Exemplos:

1. O tempo de execução(em segundos) de uma dada rotina num programa informático é uma v.a. exponencial com parâmetro $\alpha = 5$ (sendo $\mu = 5$ e $\sigma = 5$ segundos). Determine a probabilidade de uma amostra aleatória, de 25 desses tempos, ter uma média inferior a 6 segundos.
   Sendo $X_i \sim E(5)$ o tempo da i^{sima} execução da rotina, o que se pretende é $P(\overline{X} < 6)$.

$$\begin{aligned} P(\overline{X} < 6) &\approx P\left(\frac{\overline{X}-5}{\sqrt{\frac{5^2}{25}}} < \frac{6-5}{\sqrt{\frac{5^2}{25}}}\right) \\ &= P(U < 1) \\ &= \Phi(1) = 0,841. \end{aligned}$$

2. A altura dos indivíduos de uma determinada região é uma v.a. com distribuição normal com média 165 cm e desvio padrão 5 cm. Seleccionaram-se aleatoriamente 10 indivíduos dessa população. Qual a probabilidade da média das alturas da amostra se situar entre 1,60 metros e 1,62 metros?
   Sendo X="altura dos indivíduos da população"tem-se $X \sim N(165, 5)$ e o que se pretende é $P(160 < \overline{X} < 162)$. Pelo teorema dado tem-se $\overline{X}$ tem distribuição aproximadamente
   $$N(\mu_{\overline{X}} = 165, \sigma_{\overline{X}} = \frac{5}{\sqrt{10}})$$
   sendo
   $$\begin{aligned} P(160 < \overline{X} < 162) &\approx P\left(\tfrac{160-165}{\frac{5}{\sqrt{10}}} < \tfrac{\overline{X}-165}{\frac{5}{\sqrt{10}}} < \tfrac{162-165}{\frac{5}{\sqrt{10}}}\right) \\ &\approx P(-3,16 < U < -1,9) \\ &= \Phi(-1,9) - \Phi(-3,16) \\ &= -0,9713 + 0,9992 = 0,0279. \end{aligned}$$

3. Numa maternidade, 500 recém-nascidos têm peso médio de 5,02 kg e desvio padrão 0.3 kg. Determine a probabilidade de 100 recém-nascidos, seleccionados sem reposição, terem um peso total entre 496 kg e 500 kg?
   O peso total da amostra de 100 recém-nascidos estará entre 496 kg e 500 kg se o peso médio da amostra estiver entre 4.96 kg e 5,00 kg. Sendo X="peso dos recém-nascidos"com $\mu_X = 5,02$ e $\sigma_X = 0,3$ sendo a população finita e um processo de amostragem sem reposição o resultado dado para esta situação permite-nos dizer que
   $$U = \frac{\overline{X} - 5,02}{\frac{0,3}{\sqrt{100}} \times \sqrt{\frac{500-100}{500-1}}}$$
   tem distribuição assimptótica normal reduzida.
   $$\begin{aligned} P(4,96 < \overline{X} < 5,0) &\approx P\left(\tfrac{4,96-5,02}{0,027} < \tfrac{\overline{X}-5,02}{0,027} < \tfrac{5,0-5,02}{0,027}\right) \\ &\approx P(-2,22 < U < -0,74) \\ &= \Phi(-0,74) - \Phi(-2,22) \\ &= -0,7704 + 0,9868 = 0,2164. \end{aligned}$$

4.4 Distribuição de Amostragem para a Variância

A variância amostral é uma estatística utilizada para medir a variabilidade da amostra e estimar a variância da população. Tal como as restantes estatísticas é uma v.a. (resultante de uma função de v.a.) que também tem uma distribuição.

Considerando uma população normal da qual se obteve uma amostra aleatória de dimensão n, $(X_1, X_2, \ldots, X_n)$, em que

$$X_i \sim N(\mu, \sigma),$$

sendo as n observações independentes dado que o valor de cada uma é independente do valor das restantes, então tem-se

$$U_i = \frac{X_i - \mu}{\sigma} \sim N(0,1)$$

e pela distribuição do χ^2 tem-se

$$\sum_{i=1}^{n} U_i^2 \sim \chi^2_n,$$

Quando a média μ é desconhecida, sendo por isso necessário recorrer ao seu estimador $\overline{X}$, perde-se um grau de liberdade e tem-se:

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}, \tag{4.6}$$

porque quando utilizamos a média amostral as n observações passam a constituir apenas $n-1$ independentes, dado que qualquer observação fica completamente determinada pela média e pelas restantes $n-1$ observações.

De (4.6) temos

$$\begin{aligned} \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} &= \sum_{i=1}^{n} \frac{n(X_i - \overline{X})^2}{n\sigma^2} \\ &= \tfrac{n}{\sigma^2} \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n} \\ &= \tfrac{n\, S^2}{\sigma^2} \sim \chi^2_{n-1}. \end{aligned} \tag{4.7}$$

Por outro lado, dado que

$$S'^2 = \frac{n}{n-1} S^2$$

e pelo resultado (4.7) tem-se

$$\frac{(n-1)\ S'^2}{\sigma^2} \sim \chi^2_{n-1},$$

para as situações em que se utilizar a variância amostral corrigida.

Teorema 4.4.1 *Se S^2 for a variância de uma amostra aleatória i.i.d. de dimensão n, recolhida de uma população normal com variância σ^2, então a v.a.*

$$V = \frac{n\ S^2}{\sigma^2} \sim \chi^2_{n-1}, \tag{4.8}$$

isto é, tem distribuição do qui-quadrado com $n-1$ graus de liberdade.

Exemplo:A vida de certo fármaco segue uma distribuição normal com média 2000 horas e desvio padrão 51.94 horas. Num lote de 10 medicamentos, qual a probabilidade do desvio padrão amostral não exceder 50 horas?
De (4.8) tem-se

$$V = \frac{9\ S'^2}{(51,94)^2} \sim \chi^2_9$$

sendo

$$\begin{aligned} P(S' \leq 50) &= P\left(S'^2 \leq 2500\right) \\ &= P\left(V \leq \tfrac{9\times 2500}{(51,94)^2}\right) \\ &\approx P(V \leq 8,34) = 0,5. \end{aligned}$$

4.5 Distribuição de Amostragem de uma Proporção

Sendo

$$\widehat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

a proporção amostral, que indica a proporção de sucessos (elementos que verificam a característica que se pretende estudar) numa amostra $X_1, X_2, \ldots, X_n$ que são n v.a. de Bernoulli independentes em que $X_i = 1$ (sucesso) e $X_i = 0$ (insucesso), consoante o elemento i tem ou não a característica que corresponde a sucesso. A v.a.

$$X = \sum_{i=1}^{n} X_i$$

representa o número de sucessos na amostra tem uma distribuição Binomial com parâmetros n e p, isto é,

$$X = n\widehat{P} \sim B(n,p) \tag{4.9}$$

note-se que p é a probabilidade de sucesso. De (4.9) tem-se

$$E[X] = E[n\widehat{P}] = np \tag{4.10}$$

e

$$V[X] = V[n\widehat{P}] = np(1-p) \tag{4.11}$$

atendendo a (4.9), (4.10) e (4.11) tem-se

$$E[\widehat{P}] = \mu_{\widehat{P}} = p \tag{4.12}$$

e

$$V[\widehat{P}] = \sigma^2_{\widehat{P}} = \frac{p(1-p)}{n}. \tag{4.13}$$

Sendo razoável aproximar a binomial através da normal para $np \geq 5$ e $n(1-p) \geq 5$. Esta aproximação justifica-se com o auxilio do teorema do limite central, que aplicado à proporção amostral permite obter o resultado que se segue.

Teorema 4.5.1 *Considere-se uma amostra $X_1, X_2, \ldots, X_n$ aleatória i.i.d. de uma população de Bernoulli, em que a probabilidade do sucesso é p, com normais com $X_i = 1$ (sucesso) e $X_i = 0$ (insucesso) consoante o elemento i tem ou não a característica que corresponde a sucesso. Seja*

$$\widehat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

a proporção de sucessos da amostra. Então a v.a.

$$U = \frac{\widehat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \tag{4.14}$$

tem distribuição assimptótica normal reduzida.

Exemplo: Suponha que no Pais da Brincadeira existe um jornal diário, o Escândalo, cujos director financeiro diz que é lido por 10 % dos habitantes. Apesar de o director ter razão, a empresa Tudo Grátis, que plubicita no jornal, decidiu averiguar a veracidade desta afirmação e procedeu a uma estimativa da proporção de habitantes que leêm o jornal. Para tal recolheu uma amostra de 100 habitantes. A empresa Tudo Grátis decidiu que apenas continuará a publicitar no jornal se a proporção amostral for superior a 8,5%. Determine a probabilidade da empresa continuar a plublicitar no referido jornal.

Pelo teorema anterior

$$U = \frac{\widehat{P} - 0,1}{\sqrt{\frac{0,1\times 0,9}{100}}} \sim N(0,1)$$

ou

$$\widehat{P} \sim N(0,1;0,03)$$

pelo que

$$\begin{aligned} P\left(\widehat{P} > 0,085\right) &\approx P\left(U > \tfrac{0,085-0,1}{0,03}\right) \\ &= P(U > -0,5) = 0,691. \end{aligned} \tag{4.15}$$

Conteúdo

Bibliografia

[1] F. Galvão de Mello. *Probabilidades e Estatística, Conceitos e Métodos Fundamentais*, volume I e II. Escolar Editora, 2000.

[2] Jaime Fonseca. *Estatística Matemática*, volume I e II. Ed. Sílabo, 2001.

[3] Lopes M. N. Gonçalves E. *Probabilidades Princípios Teóricos*. Escolar Editora, 2000.

[4] J. Guimarães, R. Campos e Cabral. *Estatística*. McGraw-Hill, 2010.

[5] Paul L. Meyer. *Probabilidade - Aplicações À Estatística*. Livros Técnicos e Científicos, 1983.

[6] M. Murteira B. e Antunes. *Probabilidades e Estatística*, volume I. Escolar Editora, 2012.

[7] M. Murteira B. e Antunes. *Probabilidades e Estatística*, volume II. Escolar Editora, 2013.

[8] Sílvio Pestana, Dinis; Velosa. *Introdução à probabilidade e Ã estatística*, volume I. Fundacão Calouste Gulbenkian, 2008.

[9] Reis E. ; Melo P. ; Andrade R. *Estatística Aplicada*, volume II. Ed. Sílabo, 2005.

[10] Reis E. ; Melo P. ; Andrade R. *Estatística Aplicada*, volume I e II. Ed. Sílabo, 2015.