Capítulo 1

Teoria das Probabilidades

1.1 Introdução

Apesar da estatística descritiva ser um ramo importante da estatística, muito frequentemente a informação que dispomos existe apenas para um subgrupo de um grande conjunto de items de interesse (uma amostra), significando a necessidade de generalizações para além dos dados. O objectivo da inferência estatística prende-se precisamente com estas generalizações.

Neste processo estão sempre presentes incertezas, quer porque a informação não é completa ou porque é apenas parte de um todo ou ainda porque é de natureza indirecta. Estas incertezas são quantificadas através da **teoria das probabilidades**.

A **teoria das probabilidades** tem assim como objectivo a formulação de modelos de fenómenos naturais em que intervém o acaso. As suas origens remontam aos chamados jogos de azar, como sendo a roleta do casino ou os jogos de cartas ou de dados!

Definição 1.1 *Uma* ***experiência aleatória*** *é uma experiência na qual:*

- *todos os possíveis resultados da experiência são conhecidos à partida;*
- *para qualquer realização da experiência não se sabe, antes desta ocorrer, qual dos seus possíveis resultados vai acontecer;*
- *a experiência pode sempre ser repetida sob idênticas condições.*

Vários são os exemplos do nosso dia-a-dia de experiências aleatórias - lançamento de uma moeda ao ar (assumindo que não "aterra"de lado!), lançamento de um dado, a extracção do totoloto, o tempo de vida de duração de uma lâmpada, o tempo que se demora na fila de espera dos correios, o sorteio dos turnos práticos de IPE!...

O nosso objectivo é então estudar a incerteza associada a estas experiências aleatórias, se possível quantificá-la. Laplace, em 1812, forneceu a primeira definição de probabilidade, dita **definição clássica de probabilidade** ou **Lei de Laplace**:

Definição 1.2 *Se uma experiência aleatória tem a si associado um número finito N de resultados mutuamente exclusivos e igualmente prováveis e se, desses resultados, N_A têm um certo atributo A, então a probabilidade de A, $P(A)$, é dada por:*

$$P(A) = \frac{N_A}{N} = \frac{n^o \text{ de resultados favoráveis}}{n^o \text{ de resultados possíveis}}$$

Esta definição é no entanto restritiva e inadequada em muitas situações, por exemplo se os resultados da experiência aleatória não forem equiprováveis. Surge então o **conceito frequencista de probabilidade**:

Definição 1.3 *A probabilidade de um acontecimento A é avaliada através de informação existente sobre A, sendo dada pela proporção de vezes em que se observou o resultado A, n_A, num número n suficientemente grande de realizações da experiência aleatória:*

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

Este é o conceito de probabilidade que trataremos neste curso. Notamos que esta interpretação não é única. No entanto, a matemática das probabilidades que vamos aprender de seguida é desenvolvida numa base inteiramente **axiomática**, independente da referida interpretação. Deve-se a Kolmogorov, que a apresentou em 1933.

De acordo com o desenvolvimento de Kolmogorov os acontecimentos aleatórios são representados por conjuntos e a probabilidade é uma medida normada definida sobre estes conjuntos.

1.2 Espaço amostral

Definição 1.4 *O **espaço amostral** de uma experiência aleatória é um par $(\Omega, \mathcal{S})$ onde:*

1. *Ω é o conjunto de todos os possíveis resultados da experiência (**espaço de resultados** ou **universo**);*
2. *$\mathcal{S}$ é uma $\sigma-$álgebra,* i.e.*:*
   - *(i) $\varnothing \in \mathcal{S}$;*
   - *(ii) Se $A \in \mathcal{S}$ então $\bar{A} \in \mathcal{S}$, onde $\bar{A}$ é o conjunto complementar de A;*
   - *(iii) Se $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{S}$ então $\bigcup_{i=1}^{\infty} A_i \in \mathcal{S}$.*

Observações:

1. Os pontos em Ω designam-se por **pontos amostrais**.
2. Muito frequentemente $\mathcal{S}$ é o conjunto de todos os subconjuntos de Ω, $\mathcal{S} \equiv \mathcal{P}(\Omega)$. Este conjunto é sempre uma $\sigma-$álgebra.
3. Qualquer conjunto $A \in \mathcal{S}$ é chamado um **acontecimento**. A é um conjunto de pontos amostrais.
4. Qualquer acontecimento A diz-se ter ocorrido se algum ponto de A corresponder ao resultado de uma experiência aleatória.
5. Cada conjunto formado por apenas um ponto amostral é dito um **acontecimento simples** ou **elementar**.
6. Ao conjunto Ω chamamos **acontecimento certo**.
7. Ao conjunto $\varnothing$ chamamos **acontecimento impossível**.

8. A álgebra assim construída, também designada por **álgebra de acontecimentos**, é "parecida" com a álgebra de conjuntos, "herdando" propriedades desta. Assim salientamos:

   (i) A é **subacontecimento** de B, e escreve-se $A \subset B$, se e só se a realização de A implica a realização de B;

   (ii) Dado o acontecimento A, chama-se **acontecimento complementar de A** e escreve-se $\overline{A}$, ao acontecimento constituído pelos elementos de Ω que não estão em A;

   (iii) Dados os acontecimentos A e B, dá-se o nome de **união de A com B** ao acontecimento que consiste na realização de pelo menos um deles e representa-se por $A \cup B$;

   (iv) **Intersecção de A com B** é o acontecimento que se realiza se e só se realizam em simultâneo os acontecimentos A e B. Representa-se por $A \cap B$;

   (v) A **união de acontecimentos disjuntos A e B** representa-se por $A + B$.

   (vi) Chama-se **diferença dos acontecimentos A e B** ao acontecimento $A - B = A \cap \overline{B}$, ou seja, ao acontecimento que se realiza se e só se A se realiza mas não se realiza B.

9. Dois acontecimentos A e B dizem-se **mutuamente exclusivos** se não têm elementos em comum, ou seja se $A \cap B = \varnothing$.

10. Se Ω contiver apenas um número finito de elementos dizemos que $(\Omega, \mathcal{S})$ é um **espaço amostral finito**. Se Ω for no máximo um conjunto numerável de pontos dizemos que $(\Omega, \mathcal{S})$ é um **espaço amostral discreto**. Se os pontos em Ω não forem contáveis dizemos que $(\Omega, \mathcal{S})$ é um **espaço amostral não contável**. Em particular, se $\Omega = \mathbb{R}^k$, dizemos que temos um **espaço amostral contínuo**.

Exemplo 1.1 *Considere-se a experiência aleatória simples do lançamento ao ar de uma moeda equilibrada. Representando "Ca" o resultado "sair cara" e "Co" o resultado "sair coroa", temos que* $\Omega = \{Ca, Co\}$. *Escolhemos então a seguinte* $\sigma-$*álgebra,* $\mathcal{S} = \mathcal{P}(\Omega) = \{\varnothing, \{Ca\}, \{Co\}, \{Ca, Co\}\}$, *formando o espaço amostral* $(\Omega, \mathcal{S})$.

Consideremos agora a experiência aleatória do lançamento ao ar de duas moedas equilibradas. Temos que $\Omega = \{(Ca, Ca), (Ca, Co), (Co, Ca), (Co, Co)\}$ *e podemos escolher* $\mathcal{S} = \mathcal{P}(\Omega) = \{\varnothing, \{(Ca, Ca)\}, \{(Ca, Co)\}, \{(Co, Ca)\}, \{(Co, Co)\}, \{(Ca, Ca), (Ca, Co)\}, \{(Ca, Ca), (Co, Ca)\}, \{(Ca, Ca), (Co, Co)\}, \{(Ca, Co), (Co, Ca)\}, \{(Ca, Co), (Co, Co)\}, \{(Co, Ca), (Co, Co)\}, \{(Ca, Ca), (Ca, Co), (Co, Ca)\}, \{(Ca, Ca), (Ca, Co), (Co, Co)\}, \{(Ca, Ca), (Co, Ca), (Co, Co)\}, \{(Ca, Co), (Co, Ca), (Co, Co)\}, \Omega\}$, *formando o espaço amostral* $(\Omega, \mathcal{S})$.

□

1.3 Axiomática das probabilidades

Definição 1.5 *Seja* $(\Omega, \mathcal{S})$ *um espaço amostral. Uma função* $P : \mathcal{S} \to [0, 1]$ *diz-se uma* ***probabilidade*** *se satisfaz as seguintes* ***condições*** *ou* ***axiomas:***

1. $P(A) \geq 0, \qquad \forall A \in \mathcal{S}$;

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

3. *Sejam* $\{A_i,\ i = 1, 2, \ldots\}$, $A_i \in \mathcal{S}$, *uma sucessão de conjuntos disjuntos* $(A_j \cap A_k = \varnothing,\ j \neq k)$. *Então:*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \qquad \textbf{Aditividade contável}$$

Nota: Como caso particular do 3º axioma temos a chamada **aditividade finita**, para Ω finito:

Sejam $A, B \in \mathcal{S}:\ A \cap B = \varnothing$. Então $P(A \cup B) = P(A) + P(B)$.

Definição 1.6 *Chama-se espaço de probabilidades ao triplo* $(\Omega, \mathcal{S}, P)$.

Exemplo 1.2 *Relativamente ao exemplo 1.1, onde* $\Omega = \{Ca, Co\}$ *e* $\mathcal{S} = \mathcal{P}(\Omega)$, *podemos definir a função* P *em* $\mathcal{S}$ *como* $P(\varnothing) = 0$, $P(\{Ca\}) = \frac{1}{2}$, $P(\{Co\}) = \frac{1}{2}$ *e* $P(\Omega) = 1$. *Facilmente se verifica que esta função satisfaz os axiomas acima sendo, por isso, uma probabilidade.*

Já se $\Omega = \{(Ca, Ca), (Ca, Co), (Co, Ca), (Co, Co)\}$ *e* $\mathcal{S} = \mathcal{P}(\Omega)$, *podemos definir a função* P *por:*

$$\begin{aligned}
&P(\varnothing) = 0\\
&P(\{(Ca, Ca)\}) = P(\{(Ca, Co)\}) = P(\{(Co, Ca)\}) = P(\{(Co, Co)\} = \frac{1}{4}\\
&P(\{(Ca, Ca), (Ca, Co)\}) = P(\{(Ca, Ca), (Co, Ca)\}) = P(\{(Ca, Ca), (Co, Co)\}) =\\
&= P(\{(Ca, Co), (Co, Ca)\}) = P(\{(Ca, Co), (Co, Co)\}) = P(\{(Co, Ca), (Co, Co)\}) = \frac{1}{2}\\
&P(\{(Ca, Ca), (Ca, Co), (Co, Ca)\}) = P(\{(Ca, Ca), (Ca, Co), (Co, Co)\}) =\\
&= P(\{(Ca, Ca), (Co, Ca), (Co, Co)\}) = P(\{(Ca, Co), (Co, Ca), (Co, Co)\}) = \frac{3}{4}\\
&P(\Omega) = 1
\end{aligned}$$

Também esta função satisfaz os axiomas acima enunciados sendo, por isso, uma probabilidade.

Notemos que em ambas as situações anteriores, ao definirmos os valores que a função P *deve tomar para os acontecimentos elementares, estes necessariamente implicam os valores que* P *deve assumir para os restantes acontecimentos, de forma a que* P *seja de facto uma probabilidade.*

□

Passam-se a enumerar de seguida algumas consequências da axiomática das probabilidades acima definida, esboçando as suas demonstrações, sem grande detalhe.

Proposição 1.1 $P(\varnothing) = 0$.

Demonstração: $\varnothing \cap \Omega = \varnothing$. Logo $\varnothing$ e Ω são conjuntos disjuntos. Então, pela aditividade e porque $\varnothing \cup \Omega = \Omega$,

$$\begin{aligned}
&P(\varnothing \cup \Omega) = P(\varnothing) + P(\Omega) \Leftrightarrow P(\Omega) = P(\varnothing) + P(\Omega) \Leftrightarrow \qquad \text{(pelo 2º axioma)}\\
&1 = P(\varnothing) + 1 \Leftrightarrow P(\varnothing) = 0
\end{aligned}$$

■

Teorema 1.1 *Se* $A,\ B \in \mathcal{S}$ *e* $A \subseteq B$ *então:*

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

Demonstração: Se $A \subseteq B$, $B = (A \cap B) \cup (B - A) = A \cup (B - A)$. Assim, sendo A e $(B - A)$ disjuntos, temos pelo axioma da aditividade que:

$$\begin{aligned} &P(B) = P(A + (B - A)) = P(A) + P(B - A) \Leftrightarrow \\ &P(B - A) = P(B) - P(A) \end{aligned} \tag{1.3.1}$$

Note-se que de (1.3.1), $P(B) \geq P(A)$, já que $P(B - A) \geq 0$, pelo 1º axioma.

■

Corolário 1.1.1 $\forall A \in \mathcal{S}, \quad 0 \leq P(A) \leq 1$.

Demonstração: Como $\forall A \in \mathcal{S}, \quad A \subseteq \Omega$ e como, pelo axioma 2, $P(\Omega) = 1$, segue o pretendido como consequência do primeiro ponto do teorema anterior.

■

Corolário 1.1.2 $\forall A, B \in \mathcal{S}, \quad P(A - B) = P(A) - P(A \cap B)$.

Demonstração:

$$\begin{aligned} P(A - B) &= P(A - (A \cap B)) = \quad \text{(Pelo 2º ponto do teorema anterior e porque } (A \cap B) \subseteq A) \\ &= P(A) - P(A \cap B) \end{aligned}$$

■

Teorema 1.2 *(***Regra da adição***) Para* $A, B \in \mathcal{S}$*,* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$*.*

Demonstração:

1. $A \cup B = (A - B) + (B - A) + (A \cap B)$;
2. $A = (A \cap B) + (A - B) \Leftrightarrow (A - B) = A - (A \cap B)$
3. $B = (A \cap B) + (B - A) \Leftrightarrow (B - A) = B - (A \cap B)$

Assim, pelo axioma da aditividade:

$$\begin{aligned} P(A \cup B) &= P(A - B) + P(B - A) + P(A \cap B) = \\ &= P(A - (A \cap B)) + P(B - (A \cap B)) + P(A \cap B) = \qquad \text{(Pelo corolário (1.1.2))} \\ &= P(A) - P(A \cap B) + P(B) - P(A \cap B)) + P(A \cap B) = \\ &= P(A) + P(B) - P(A \cap B)) \end{aligned}$$

■

Corolário 1.2.1 *$\forall A \in \mathcal{S}$, $P(\bar{A}) = 1 - P(A)$.*

Demonstração: Tome-se na regra da adição anteriormente enunciada $B = \bar{A}$:

$$
\begin{aligned}
&P(A \cup \bar{A}) = P(A) + P(\bar{A}) - P(A \cap \bar{A}) \Leftrightarrow \\
&P(\Omega) = P(A) + P(\bar{A}) + P(\varnothing) \Leftrightarrow \qquad \text{(Pelo 2º axioma e prop. (1.1))} \\
&1 = P(A) + P(\bar{A}) + 0 \Leftrightarrow \\
&P(\bar{A}) = 1 - P(A)
\end{aligned}
$$

■

Corolário 1.2.2 *Para $A_i \in \mathcal{S}$, $i = 1, \ldots, n$,*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) + \sum_{i \neq j \neq k} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

Observações:

1. Dois **acontecimentos** A e B dizem-se **incompatíveis** se $P(A \cap B) = 0$.
2. Se temos um acontecimento $A \neq \Omega$ mas tal que $P(A) = 1$ dizemos que A é um **acontecimento quase certo**.
3. Se temos um acontecimento $B \neq \varnothing$ mas tal que $P(B) = 0$ dizemos que B é um **acontecimento quase impossível**.

Exemplo 1.3 *Relativamente ao exemplo 1.1, continuado em 1.2, relativamente à experiência aleatória do lançamento de 2 moedas equilibradas, definamos os seguintes acontecimentos:*
A-"Sair pelo menos uma cara"
B-"Sair pelo menos uma coroa"
Temos que:

$$
\begin{aligned}
&A = \{(Ca, Co), (Co, Ca), (Ca, Ca)\} \qquad P(A) = \frac{3}{4} \\
&B = \{(Ca, Co), (Co, Ca), (Co, Co)\} \qquad P(B) = \frac{3}{4}
\end{aligned}
$$

A partir daqui consideremos os seguintes acontecimentos:

Ocorrerem os dois acontecimentos simultaneamente:

$$A \cap B = \{(Ca, Co), (Co, Ca)\} \qquad P(A \cap B) = \frac{1}{2}$$

Ocorrer A mas não B:

$$A - B = \{(Ca, Ca)\} \qquad P(A - B) = P(A) - P(A \cap B) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$$

Ocorrer pelo menos um dos acontecimentos:

$$A \cup B = \Omega \qquad P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{4} + \frac{3}{4} - \frac{1}{2} = 1$$

Não ocorrer B:

$$\bar{B} = \{(Ca, Ca)\} \qquad P(\bar{B}) = 1 - P(B) = 1 - \frac{3}{4} = \frac{1}{4}$$

□

1.4 Probabilidade condicionada e Teorema de Bayes

O cálculo de probabilidades de acontecimentos associados a uma experiência aleatória pode ser alterado quando existe informação disponível para além do espaço amostral da experiência em questão. Vejamos o seguinte exemplo:

Exemplo 1.4 *Em determinada aldeia apareceu um surto de cólera, que se pensa estar associado ao consumo de água de um determinado poço. São conhecidas as seguintes proporções relativas à quantidade de pessoas que desenvolveram a doença (representando esse acontecimento pela letra D) e às pessoas que beberam água do referido poço (representando esse acontecimento pela letra B):*

	B	$\overline{B}$	*Total*
D	*0.18*	*0.02*	*0.20*
$\overline{D}$	*0.01*	*0.79*	*0.80*
Total	*0.19*	*0.81*	*1.00*

Qual a probabilidade de uma pessoa escolhida ao acaso na população da aldeia ter contraído cólera?

$$P(D) = 0.2$$

De entre as pessoas que beberam água do poço, qual a probabilidade de se escolher ao acaso uma pessoa que contraiu cólera? Agora só estamos interessados no universo das pessoas que beberam água do poço, B, pelo que a probabilidade pretendida é:

$$\frac{0.18}{0.19} \simeq 0.95$$

Repare-se que por sabermos que a pessoa bebeu água do poço tal altera o valor da probabilidade do acontecimento "contrair cólera"de 0.2 para 0.95! O espaço de resultados foi encolhido de toda a população da aldeia, Ω, para apenas os que consumiram água do tal poço, B. Isto reflecte-se na forma como o novo acontecimento "contrair cólera sabendo (ou condicionado a que) que bebeu água do poço" passa a ser designado: $D|B$. A sua probabilidade é dada por:

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

□

Definição 1.7 *Seja* $(\Omega, \mathcal{S}, \mathcal{P})$ *um espaço de probabilidades e seja* $B \in \mathcal{S} : \ P(B) > 0$*. Para* $\forall A \in \mathcal{S}$ *definimos a* ***probabilidade condicionada de*** A ***dado*** B *por:*

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

Teorema 1.3 (Teorema da probabilidade total) *Seja* $(\Omega, \mathcal{S}, \mathcal{P})$ *um espaço de probabilidades e formem* $\{E_1, \ldots, E_n\}$ *uma partição do espaço de resultados* Ω[1]*, com* $P(E_i) > 0, \ \forall i$*. Dado um qualquer acontecimento* $A \in \mathcal{S}$*, tem-se*

$$P(A) = P(A|E_1)P(E_1) + \ldots + P(A|E_n)P(E_n)$$

■

Teorema 1.4 (Teorema de Bayes) *Seja* $(\Omega, \mathcal{S}, \mathcal{P})$ *um espaço de probabilidades e* $\{E_1, \ldots, E_n\}$ *uma partição do espaço de resultados* Ω*, com* $P(E_i) > 0, \ \forall i$*. Dado um qualquer acontecimento* $A \in \mathcal{S}$*, com* $P(A) > 0$*, tem-se*

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{\sum_{i=1}^{n} P(A|E_i)P(E_i)}.$$

Exemplo 1.5 *Suponha que existe um teste para diagnosticar uma certa doença, mas que esse teste é falível. Assim sabe-se que, para um indivíduo portador da doença* (D)*, a probabilidade de o teste dar positivo* (T) *é de 0.98 e que, para um indivíduo são* $(\overline{D})$*, a probabilidade de o teste dar negativo* $(\overline{T})$ *é 0.99. Sabe-se ainda que na população* 10% *são portadores da doença. Assim:*

$$P(T|D) = 0.98 \qquad P(\overline{T}|\overline{D}) = 0.99 \qquad P(D) = 0.10$$

A probabilidade de um indivíduo não ter a doença sabendo que o teste deu positivo é de:

$$P(\overline{D}|T) = \frac{P(T|\overline{D})P(\overline{D})}{P(T|D)P(D) + P(T|\overline{D})P(\overline{D})} = \frac{0.01 \times 0.90}{0.98 \times 0.10 + (1 - 0.99) \times (1 - 0.10)} \simeq 0.084$$

e a probabilidade de um indivíduo ter a doença se o teste deu negativo é de:

$$P(D|\overline{T}) = \frac{P(\overline{T}|D)P(D)}{P(\overline{T}|D)P(D) + P(\overline{T}|\overline{D})P(\overline{D})} = \frac{0.02 \times 0.10}{(1 - 0.98) \times 0.10 + 0.99 \times (1 - 0.10)} \simeq 0.002$$

□

[1] Ou seja, $E_1 \cup \ldots \cup E_n = \Omega$ e $E_i \cap E_j = \varnothing$, $\forall i \neq j$.

1.5 Independência entre acontecimentos

Definição 1.8 *Dois acontecimentos A e B dizem-se* ***independentes*** *se e só se*

$$P(A \cap B) = P(A)\, P(B)$$

Teorema 1.5 *Se A e B são acontecimentos independentes, então:*

$$P(A|B) = P(A) \qquad \textit{se} \quad P(B) > 0$$

e

$$P(B|A) = P(B) \qquad \textit{se} \quad P(A) > 0.$$

Portanto, se dois acontecimentos são independentes, o conhecimento de um deles em nada influencia a probabilidade de ocorrência do outro.