Capítulo I: Introdução à Probabilidade

Probabilidades e Aplicações

Licenciatura em Matemática
Licenciatura em Ciências da Computação
Universidade do Minho
Ano Letivo 2025/2026

1. História Breve. Considerações Gerais

A teoria das probabilidades tem como objetivo encontrar modelos matemáticos que permitam descrever *fenómenos aleatórios*, i.e.,

fenómenos cujo futuro não é possível prever, com exatidão, a partir do passado.

Historicamente, esta teoria surge com o objetivo de fornecer uma explicação sobre certos fenómenos que se observavam nos chamados jogos de sorte e azar (jogos com dados, moedas, cartas, etc.). Tais fenómenos surgiam quando se efetuava um grande número de partidas de tais jogos.

Exemplo: No Século XVII, De Méré observou que na experiência que consiste em lançar três vezes consecutivas um dado equilibrado se obtém, após um grande número de repetições, mais vezes o acontecimento “soma 10” do que acontecimento “soma 9”.

Como explicar isto através de um modelo matemático? Na altura, não era claro!

Numa tentativa de explicar tais fenómenos, surge aquela que hoje é conhecida como *teoria elementar de probabilidades* e que assenta no facto de os fenómenos em causa apresentarem as seguintes características:

i) de cada vez que se realiza a experiência (lançar um dado, lançar uma moeda, escolher uma carta num baralho, etc.) obtém-se um resultado individual que não conseguimos prever com exatidão;

ii) repetindo a experiência um grande número de vezes, sempre nas mesmas condições, os resultados apresentam ter uma certa certa regularidade estatística.

Exemplos: (de regularidade estatística)

- Quando lançamos uma moeda equilibrada n vezes, observa-se que a frequência relativa do acontecimento “saiu cara” tende, quando $n \to \infty$, a estabilizar em torno de $1/2$;
- Quando lançamos um dado equilibrado n vezes, observa-se que a frequência relativa do acontecimento “saiu face 1” tende, quando $n \to \infty$, a estabilizar em torno de $1/6$.

Foi precisamente este tipo de regularidade estatística (existência do limite da frequência relativa), que estes jogos de sorte e de azar apresentam, que conduziu à chamada *definição frequencista de probabilidade*.

Definição Frequencista de Probabilidade

A probabilidade de um acontecimento é o valor em torno do qual a sua frequência relativa tende a estabilizar quando repetimos a experiência, nas mesmas condições, um número suficientemente grande de vezes.

Nos dois exemplos atrás mencionados (lançamento de uma moeda equilibrada e lançamento de um dado equilibrado) a experiência realizada é chamada de *experiência aleatória* uma vez que satisfaz as seguintes condições:

i) pode ser repetida uma infinidade de vezes, sempre nas mesmas condições;
ii) conhecemos todos os resultados possíveis da experiência;
iii) de cada vez que a experiência é efetuada não se conhece, com exatidão, qual dos resultados possíveis vai ocorrer.

Quando uma experiência aleatória apresenta ainda as seguintes características:

iv) número de resultados possíveis é finito;

v) os resultados elementares são igualmente possíveis (princípio da equiprobabilidade);

usa-se a definição clássica de probabilidade (também conhecida por definição clássica de Laplace), que é a seguinte:

Definição Clássica de Laplace

A probabilidade de um acontecimento A, decorrente de uma experiência aleatória que satisfaça as condições iv) e v), é igual ao quociente entre o número de resultados elementares favoráveis a A e o número total de resultados elementares da experiência.

Observação: A definição clássica de Laplace não pode ser usada em experiências envolvendo lançamentos de moedas ou dados não equilibrados. Note que, nestes casos, a condição v) não é verificada.

Nos anos 30 do Séc. XX, surge a definição moderna de probabilidade, devida ao matemático russo A.N. Kolmogorov, que assenta em teoria da medida (teoria matemática que, entre outros, permite “medir” conjuntos e funções).

Esta é a definição que vamos utilizar nesta UC e que é habitualmente designada de *definição axiomática de probabilidade*.

Esta definição assenta em conceitos de teoria da medida, pelo que vamos começar pela apresentação de alguns desses conceitos que serão relevantes para esta UC.

Nota: No que se segue, assume-se sempre que Ω é um conjunto não-vazio.

2. Definição Axiomática de Probabilidade

Nota: No que se segue, assume-se sempre que Ω é um conjunto não-vazio.

Definição [σ- álgebra e Espaço Mensurável]

Sejam Ω um conjunto e $\mathcal{A}$ uma família (ou *colecção* ou *conjunto*) de subconjuntos de Ω.
$\mathcal{A}$ diz-se uma σ-álgebra sobre Ω se satisfaz as seguintes condições:

i) $\Omega \in \mathcal{A}$;

ii) Se $E \in \mathcal{A}$ então $\overline{E} \in \mathcal{A}$, com $\overline{E} \equiv \{ x \in \Omega : x \notin E\}$;

iii) Se $(E_n)_{n\in\mathbb{N}}$ é uma sucessão de elementos de $\mathcal{A}$ então

$$\left(\bigcup_{n\in\mathbb{N}} E_n \right) \in \mathcal{A}.$$

Se $\mathcal{A}$ é uma σ-álgebra sobre Ω, ao par $(\Omega, \mathcal{A})$ chamamos um espaço mensurável.

Nota: ii) é conhecida por "fecho para a complementaridade" e iii) por "fecho para a união infinita numerável".

Observações: Da definição de σ-álgebra, concluimos facilmente que:

1) $\emptyset \in \mathcal{A}$;
2) Se $(E_n)_{n\in\mathbb{N}}$ é uma sucessão de elementos de $\mathcal{A}$ então
$$\left(\bigcap_{n\in\mathbb{N}} E_n\right) \in \mathcal{A}.$$
3) Se $E_1, E_2, \ldots, E_m$ são m elementos de $\mathcal{A}$, com $m \in \mathbb{N}$ fixo e $m \geq 2$, então
$$\left(\bigcup_{i=1}^{m} E_i\right) \in \mathcal{A} \quad \text{e} \quad \left(\bigcap_{i=1}^{m} E_i\right) \in \mathcal{A}$$
4) $\{\emptyset, \Omega\}$ é uma σ-álgebra sobe Ω. É chamada de *σ-álgebra trival* e é, na verdade, a menor σ-álgebra sobre Ω.
(note que qualquer σ-álgebra sobre Ω contém $\{\emptyset, \Omega\}$)
5) Partes de Ω (denotada por $\mathcal{P}(\Omega)$) é uma σ-álgebra sobre Ω. É, na verdade, a maior σ-álgebra sobre Ω.
(note que qualquer σ-álgebra sobre Ω está contida em $\mathcal{P}(\Omega)$)

Definição [σ-álgebra gerada por uma família de subconjuntos de Ω]

Seja $\mathcal{C}$ uma família de subconjuntos de Ω.
Chama-se σ-álgebra gerada por $\mathcal{C}$, denota-se por $\sigma(\mathcal{C})$, à menor σ-álgebra sobre Ω que contém a família $\mathcal{C}$.

Exemplos:

I) Seja E um subconjunto próprio e não-vazio de Ω.
Facilmente se conclui que

$$\{\emptyset, E, \overline{E}, \Omega\} \tag{1}$$

é uma σ-álgebra sobre Ω.
Na verdade, esta é a σ-álgebra gerada pelo conjunto E, i.e., é a menor σ-álgebra sobre Ω que contém a família $\{E\}$ (observe que qualquer outra σ-álgebra sobre Ω que contenha $\{E\}$ tem necessariamente que conter a família (1)).
Neste exemplo, a família $\mathcal{C}$ inicial é formada apenas pelo elemento E, i.e., $\mathcal{C} = \{E\}$ e $\sigma(\mathcal{C}) = \{\emptyset, E, \overline{E}, \Omega\}$.

Note que, caso $E = \emptyset$ ou $E = \Omega$, ter-se-ia obtido a σ-álgebra trivial $\{\emptyset, \Omega\}$.

Exemplos: (continuação)

II) Considere $\Omega = \{a, b, c, d\}$ e seja $\mathcal{C} = \{\Omega, \{a, b\}, \{b\}\}$.
Observe que $\mathcal{C}$ não é uma σ-álgebra sobre Ω. Vamos então construir $\sigma(\mathcal{C})$:

$$\sigma(\mathcal{C}) = \{\, \Omega, \{a, b\}, \{b\},$$

Exemplos: (continuação)

II) Considere $\Omega = \{a, b, c, d\}$ e seja $\mathcal{C} = \{\Omega, \{a, b\}, \{b\}\}$.
Observe que $\mathcal{C}$ não é uma σ-álgebra sobre Ω. Vamos então construir $\sigma(\mathcal{C})$:

$$\sigma(\mathcal{C}) = \{\, \Omega, \{a, b\}, \{b\}, \emptyset, \{c, d\}, \{a, c, d\},$$

Exemplos: (continuação)

II) Considere $\Omega = \{a, b, c, d\}$ e seja $\mathcal{C} = \{\Omega, \{a, b\}, \{b\}\}$.
Observe que $\mathcal{C}$ não é uma σ-álgebra sobre Ω. Vamos então construir $\sigma(\mathcal{C})$:

$$\sigma(\mathcal{C}) = \{\, \Omega, \{a, b\}, \{b\}, \emptyset, \{c, d\}, \{a, c, d\}, \{b, c, d\},$$

Exemplos: (continuação)

II) Considere $\Omega = \{a, b, c, d\}$ e seja $\mathcal{C} = \{\Omega, \{a, b\}, \{b\}\}$.
Observe que $\mathcal{C}$ não é uma σ-álgebra sobre Ω. Vamos então construir $\sigma(\mathcal{C})$:

$$\sigma(\mathcal{C}) = \{\, \Omega, \{a, b\}, \{b\}, \emptyset, \{c, d\}, \{a, c, d\}, \{b, c, d\}, \{a\} \,\} .$$

Observe ainda que $\sigma(\mathcal{C})$ não coincide com $\mathcal{P}(\Omega)$.

Exemplos: (continuação)

III) Considere-se o caso $\Omega = \mathbb{R}$.

Em teoria de probabilidades, a σ-álgebra sobre $\mathbb{R}$ mais importante é a chamada *σ-álgebra de Borel sobre* $\mathbb{R}$, que é denotada por $\mathcal{B}(\mathbb{R})$ e é dada por

$$\mathcal{B}(\mathbb{R}) = \sigma\left(\{G \subseteq \mathbb{R} : G \text{ é um aberto}\}\right).$$

Observação: Os subconjuntos de $\mathbb{R}$ com que lidamos habitualmente são elementos de $\mathcal{B}(\mathbb{R})$. No entanto, apesar de difícil, é possível mostrar que existem subconjuntos de $\mathbb{R}$ que não pertencem a $\mathcal{B}(\mathbb{R})$. Temos assim a seguinte relação:

$$\mathcal{B}(\mathbb{R}) \subsetneq \mathcal{P}(\mathbb{R}).$$

Definição [Medida e Espaço de Medida]

Seja $(\Omega, \mathcal{A})$ um espaço mensurável.
Uma função $\mu : \mathcal{A} \to [0, +\infty]$ diz-se uma medida sobre $(\Omega, \mathcal{A})$ se satisfaz as seguintes condições:

i) $\mu(\emptyset) = 0$;

ii) Se $(E_n)_{n\in\mathbb{N}}$ é uma sucessão de elementos de $\mathcal{A}$, disjuntos 2 a 2 (i.e., $E_i \cap E_j = \emptyset, i \neq j$), então

$$\mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \sum_{n\in\mathbb{N}} \mu(E_n).$$

Se μ é uma medida sobre $(\Omega, \mathcal{A})$, ao triplo $(\Omega, \mathcal{A}, \mu)$ chamamos espaço de medida.

Nota: A condição ii) da definição é conhecida como σ- *aditividade*.

E estamos em condições de conhecer a definição axiomática de probabilidade.

Definição [Medida de Probabilidade e Espaço de Probabilidade]

Seja $(\Omega, \mathcal{A})$ um espaço mensurável.
Uma função $P : \mathcal{A} \to [0, +\infty]$ diz-se uma medida de probablidade sobre $(\Omega, \mathcal{A})$ se P é uma medida sobre $(\Omega, \mathcal{A})$ e $P(\Omega) = 1$.
Se P é uma medida de probabilidade sobre $(\Omega, \mathcal{A})$, ao triplo $(\Omega, \mathcal{A}, P)$ chamamos espaço de probabilidade.

Observações:

1) Se Ω é finito, a condição ii) da definição de medida reduz-se a:

Se A e B são elementos disjuntos (i.e., $A \cap B = \emptyset$) de $\mathcal{A}$ então

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

2) Num espaço de probabilidade $(\Omega, \mathcal{A}, P)$ em que Ω é infinito não-numerável (por exemplo, $\mathbb{R}$ ou um qualquer intervalo real), a σ-álgebra $\mathcal{A}$ é necessariamente uma parte própria de $\mathcal{P}(\Omega)$, i.e., teremos necessariamente

$$\mathcal{A} \subsetneq \mathcal{P}(\Omega).$$

[Para esta discussão, ver Lopes e Gonçalves, 2000]

Quando Ω é finito ou infinito numerável, podemos ter $\mathcal{A} = \mathcal{P}(\Omega)$ sem qualquer problema.

3. Espaço de Probabilidade de uma Experiência Aleatória

Seja ξ uma qualquer experiência aleatória.

Em cada realização da experiência ξ obtém-se um resultado elementar (ou individual), usualmente denotado por ω. Cada ω pertence a um conjunto, usualmente denotado por Ω, que é formado por todos os resultados elementares da experiência.

Ao conjunto Ω chamamos *espaço de resultados* ou *espaço amostral* da experiência aleatória.

Os *acontecimentos* decorrentes da experiência aleatória (elementares ou não) serão elementos de $\mathcal{P}(\Omega)$, i.e., acontecimentos corresponderão a subconjuntos de Ω.

Exemplos:

1) ξ_1: "lançamento de um dado"
O espaço amostral é $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$.
Os acontecimentos

A_1: "saiu um ás",
B_1: "saiu uma face par",
C_1: "saiu uma face numerada com um múltiplo de 3",

correspondem aos seguintes elementos de $\mathcal{P}(\Omega_1)$:

$$A_1 = \{1\}, \;\; B_1 = \{2, 4, 6\}, \;\; C_1 = \{3, 6\}$$

Exemplos: (cont.)

2) ξ_2: "número de chamadas recepcionadas, num determinado intervalo de tempo, pela linha do cliente de uma certa empresa"
O espaço amostral é $\Omega_2 = \{0, 1, 2, \ldots\} = \mathbb{N}_0$.
O acontecimento

A_2: "número de chamadas foi inferior a 6"

corresponde ao seguinte elemento de $\mathcal{P}(\Omega_2)$:

$$A_2 = \{0, 1, 2, 3, 4, 5\}$$

3. Espaço de Probabilidade de uma Experiência Aleatória

Exemplos: (cont.)

3) ξ_3: "duração, em horas, de uma chamada"
O espaço amostral é $\Omega_3 =]0, +\infty[$
O acontecimento

A_3: "chamada teve a duração de pelo menos hora e meia"

corresponde ao seguinte elemento de $\mathcal{P}(\Omega_3)$:

$$A_3 = [1.5, +\infty[$$

3. Espaço de Probabilidade de uma Experiência Aleatória

Definições: Seja ξ uma experiência aleatória com espaço amostral Ω.

1) Aos acontecimentos que correspondem a subconjuntos singulares de Ω chamamos de *acontecimentos elementares*.

2) Ω é designado de *acontecimento universal* ou *acontecimento certo*.
$\emptyset$ é designado de *acontecimento impossível*.

3) Dois acontecimentos A e B dizem-se *disjuntos* (ou *incompatíveis* ou *mutuamente exclusivos*) se não podem ocorrer em simultâneo numa única realização de ξ, i.e., se $A \cap B = \emptyset$.

4) Dois acontecimentos dizem-se *equivalentes* se correspondem ao mesmo subconjunto de Ω.

Exemplo: Seja ξ: "dois lançamentos consecutivos de uma moeda".
Tem-se $\Omega = \{(Ca, Ca), (Ca, Co), (Co, Ca), (Co, Co)\}$ e os acontecimentos

A: "saiu exatamente uma cara" e B: "saiu exatamente uma coroa"

são equivalentes, uma vez que correspondem ao mesmo subconjunto de Ω

$$A = B = \{(Ca, Co), (Co, Ca)\}.$$

3. Espaço de Probabilidade de uma Experiência Aleatória

Uma vez determinado o espaço de resultados, Ω, é necessário definir a σ-álgebra sobre Ω, digamos $\mathcal{F}$, que contenha todos os acontecimentos decorrentes de ξ e que queremos probabilizar.

Na definição de $\mathcal{F}$ temos então duas situações distintas, dependendo do cardinal de Ω:

• Quando Ω é finito ou infinito numerável (exemplos das experiências ξ_1 e ξ_2 vistas atrás), a σ-álgebra indicada é $\mathcal{P}(\Omega)$.

• Quando Ω é infinito não numerável (exemplo da experiência ξ_3), a σ-álgebra $\mathcal{F}$ terá que ser uma parte própria de $\mathcal{P}(\Omega)$, i.e., teremos

$$\mathcal{F} \subsetneq \mathcal{P}(\Omega).$$

3. Espaço de Probabilidade de uma Experiência Aleatória

O modelo matemático, que irá descrever a experiência aleatória, fica depois completo com a indicação da função $P : \mathcal{F} \to [0, +\infty]$, a medida de probabilidade sobre $(\Omega, \mathcal{F})$. A experiência aleatória será então modelada pelo espaço de probabilidade

$$(\Omega, \mathcal{F}, P).$$

Nota Importante:
Quando Ω **é finito** e os **acontecimentos elementares são equiprováveis**, o espaço de probabilidade que modela a experiência aleatória é $(\Omega, \mathcal{P}(\Omega), P)$, em que P é a conhecida medida de probabilidade de Laplace

$$\begin{aligned} P : \mathcal{P}(\Omega) &\longrightarrow [0, 1] \\ A &\longrightarrow P(A) = \frac{\#A}{\#\Omega}. \end{aligned}$$

Observação: Na Folha Prática 2, prova-se que a função P assim definida é uma medida de probabilidade sobre $(\Omega, \mathcal{P}(\Omega))$.

4. Propriedades de uma Medida de Probabilidade

Seja $(\Omega, \mathcal{F}, P)$ é um espaço de probabilidade. A medida de probabilidade, P, tem as seguintes propriedades:

I) Se A e B são elementos de $\mathcal{F}$ tais que $A \subseteq B$ então

$$P(A) \leq P(B) \text{ e } P(B \cap \overline{A}) = P(B) - P(A),$$

onde $\overline{A} \equiv \Omega \setminus A = \{x \in \Omega : x \notin A\}$.
<u>Nota</u>: Desta propriedade deduz-se que, para todo o $A \in \mathcal{F}$,

$$0 \leq P(A) \leq 1 \text{ e } P(\overline{A}) = 1 - P(A).$$

II) Sejam A e B quaisquer elementos de $\mathcal{F}$. Tem-se que

$$P(B \cap \overline{A}) = P(B) - P(B \cap A).$$

III) Sejam A e B quaisquer elementos de $\mathcal{F}$. Tem-se

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

[TPC] Demonstração destas 3 propriedades. Faz uso das propriedades de operações entre conjuntos e da definição axiomática de probabilidade.

4. Propriedades de uma Medida de Probabilidade

IV) [Fórmula de Poincaré]
Sejam $A_1, A_2, \ldots, A_k$ quaisquer k elementos de $\mathcal{F}$, com $k \in \mathbb{N}$ (finito) e $k \geq 2$. Então:

$$\begin{aligned} P\left(\bigcup_{i=1}^{k} A_i\right) &= \sum_{i=1}^{k} P(A_i) - \sum_{i=1}^{k}\sum_{j=i+1}^{k} P(A_i \cap A_j) \\ &+ \sum_{i=1}^{k}\sum_{j=i+1}^{k}\sum_{l=j+1}^{k} P(A_i \cap A_j \cap A_l) \\ &- \sum_{i=1}^{k}\sum_{j=i+1}^{k}\sum_{l=j+1}^{k}\sum_{m=l+1}^{k} P(A_i \cap A_j \cap A_l \cap A_m) \\ &+ \ldots + (-1)^{k+1} P\left(\bigcap_{i=1}^{k} A_i\right) \end{aligned}$$

[TPC] Demonstração por indução sobre k.

4. Propriedades de uma Medida de Probabilidade

V) [Sub-σ-aditividade]
Seja $(A_n)_{n\in\mathbb{N}}$ uma qualquer sucessão de elementos de $\mathcal{F}$. Tem-se que

$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} P(A_n).$$

VI) Seja $(A_n)_{n\in\mathbb{N}}$ uma sucessão crescente de elementos de $\mathcal{F}$, i.e., tal que $A_n \subseteq A_{n+1}, \forall n \in \mathbb{N}$. Então a sucessão de números reais $(P(A_n))_{n\in\mathbb{N}}$ é monótona não-decrescente e convergente. Mais,

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n\in\mathbb{N}} A_n\right).$$

VII) Seja $(A_n)_{n\in\mathbb{N}}$ uma sucessão decrescente de elementos de $\mathcal{F}$, i.e., tal que $A_n \supseteq A_{n+1}, \forall n \in \mathbb{N}$. Então a sucessão de números reais $(P(A_n))_{n\in\mathbb{N}}$ é monótona não-crescente e convergente. Mais,

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n\in\mathbb{N}} A_n\right).$$

Demonstração - Ver detalhes em Lopes e Gonçalves, 2000.

4. Propriedades de uma Medida de Probabilidade

Nota: A demonstração da propriedade V) faz uso da seguinte sucessão $(B_n)_{n\in\mathbb{N}}$ de elementos de $\mathcal{F}$, disjuntos 2 a 2 :

$$\begin{aligned} B_1 &= A_1, \\ B_2 &= A_2 \cap \overline{A_1}, \\ B_3 &= A_3 \cap \overline{A_2} \cap \overline{A_1}, \\ &\vdots \\ B_n &= A_n \cap \overline{A_{n-1}} \cap \ldots \overline{A_1},\ n \geq 2. \end{aligned}$$

Justifique que $(B_n)_{n\in\mathbb{N}}$ é efetivamente uma sucessão de elementos de $\mathcal{F}$, disjuntos 2 a 2, e observe que

$$\underset{n\in\mathbb{N}}{\cup} A_n = \underset{n\in\mathbb{N}}{\cup} B_n,$$

permitindo decompor o elemento de interesse, $\underset{n\in\mathbb{N}}{\cup} A_n$, na união de elementos disjuntos de $\mathcal{F}$.

Observe ainda que esta decomposição é válida em condições muito gerais, i.e., é válida para qualquer sucessão $(A_n)_{n\in\mathbb{N}}$ de elementos de $\mathcal{F}$.

5. Probabilidade Condicionada

Considere uma experiência aleatória, modelada por um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, e seja $B \in \mathcal{F}$ tal que $P(B) > 0$, i.e., B é um acontecimento decorrente da experiência e que tem probabilidade estritamente positiva de ocorrer.

Como será que se alteram as probabilidades dos acontecimentos quando sabemos que o acontecimento B ocorreu?

Com a informação de que B ocorreu, podemos (aliás, devemos) atribuir uma "nova" probabilidade aos diferentes acontecimentos.

Exemplo: No lançamento de um dado equilibrado tem-se

$$\Omega = \{1, 2, 3, 4, 5, 6\} \text{ e } P(A_i) = \frac{1}{6},$$

onde A_i: "saiu a face com o número i", $i = 1, \ldots, 6$.

Se soubermos que no lançamento do dado saiu uma face par, como será que se alteram as probabilidades dos acontecimentos A_i?

Exemplo: No lançamento de um dado equilibrado tem-se

$$\Omega = \{1, 2, 3, 4, 5, 6\} \text{ e } P(A_i) = \frac{1}{6},$$

onde A_i: "saiu a face com o número i", $i = 1, \ldots, 6$.
Se soubermos que no lançamento do dado saiu uma face par, como será que se alteram as probabilidades dos acontecimentos A_i?

Ora, se soubermos que ocorreu o seguinte acontecimento B

B: "saiu uma face par",

os acontecimentos A_1, A_3 e A_5 passam a ter probabilidade nula! E os acontecimentos A_2, A_4 e A_6 passam agora a ter probabilidade igual a $\frac{1}{3}$.
Note-se que também o acontecimento B tem uma "nova" probabilidade: era $\frac{1}{2}$ e passou a ser 1.

5. Probabilidade Condicionada

Será então natural pensar que a "nova" probabilidade de um qualquer acontecimento, $A \in \mathcal{F}$, irá depender do que existir em comum entre A e B, i.e., irá depender da $A \cap B$. Mais, esta"nova" probabilidade irá atribuir a B o valor 1.

Definição [Probabilidade Condicionada por um Acontecimento B]

Sejam $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade e $B \in \mathcal{F}$ tal que $P(B) > 0$. Chama-se probabilidade condicionada por B à função P_B dada por:

$$\begin{aligned} P_B : \mathcal{F} &\longrightarrow [0, 1] \\ A &\longrightarrow P_B(A) = \frac{P(A \cap B)}{P(B)}. \end{aligned}$$

Ao valor $P_B(A)$ chamamos "probabilidade de A condicionada por B" ou ainda "probabilidade de A sabendo que B ocorreu" ou "probabilidade de A dado B".

<u>Nota</u>: Uma notação alternativa a $P_B(A)$ é $P(A|B)$. Esta última notação é muito usada, mas requer cuidado com o argumento da função (à esquerda da barra).

Propriedades de uma probabilidade condicionada:

1) P_B é uma medida de probabilidade sobre $(\Omega, \mathcal{F})$.
   Demonstração: Ver Ex. 1 da Folha Prática 2.

2) [Regra da Multiplicação]
   Se $A_1, A_2, \ldots, A_{k-1}, A_k$ são k acontecimentos, com $k \in \mathbb{N}$ fixo e $k \geq 2$, e tais que

$$P\left(\bigcap_{i=1}^{k-1} A_i\right) > 0,$$

   então

$$P\left(\bigcap_{i=1}^{k} A_i\right) = P(A_1)\,P(A_2|A_1)\,P(A_3|A_1 \cap A_2)\ldots P\left(A_k \;\middle|\; \bigcap_{i=1}^{k-1} A_i\right).$$

   Demonstração: TPC [Deve verificar que todas as probabilidades condicionadas aqui utilizadas estão todas definidas. É verdade? Porquê?]

Propriedades de uma probabilidade condicionada: (continuação)

3) [Teorema da Probabilidade Total (TPT)] Seja $(E_n)_{n\in\mathbb{N}}$ uma sucessão de elementos de $\mathcal{F}$, disjuntos 2 a 2, e tais que $P(E_n) > 0$, $\forall\, n \in \mathbb{N}$.
Se $A \in \mathcal{F}$ é tal que $A \subseteq \underset{n\in\mathbb{N}}{\cup} E_n$ então

$$P(A) = \sum_{n\in\mathbb{N}} P(A|E_n)\, P(E_n).$$

Demonstração: Basta observar que

$$A = A \cap \left(\underset{n\in\mathbb{N}}{\cup} E_n \right) = \underset{n\in\mathbb{N}}{\cup} (A \cap E_n).$$

Como $(A \cap E_n)_{n\in\mathbb{N}}$ é uma sucessão de elementos de $\mathcal{F}$, disjuntos 2 a 2, usando a σ-aditividade de P, tem-se

$$P(A) = \sum_{n\in\mathbb{N}} P(A \cap E_n) = \sum_{n\in\mathbb{N}} P(A|E_n) P(E_n).$$

Note que as probabilidades condicionadas aqui utilizadas estão definidas.

5. Probabilidade Condicionada

Propriedades de uma probabilidade condicionada: (continuação)

4) [Fórmula de Bayes] Nas condições do TPT e desde que $P(A) > 0$, tem-se, para todo o $j \in \mathbb{N}$,

$$P(E_j|A) = \frac{P(A|E_j)P(E_j)}{\sum\limits_{n\in\mathbb{N}} P(A|E_n)P(E_n)}.$$

Demonstração: TPC [Basta usar definição de probabilidade condicionada e o TPT.]

Observações:

1) É usual enunciar o TPT e a Fórmula Bayes com $(E_n)_{n\in\mathbb{N}}$ **uma partição de** Ω, i.e., com $(E_n)_{n\in\mathbb{N}}$ uma sucessão de elementos de $\mathcal{F}$, disjuntos 2 a 2, e tal que

$$\Omega = \bigcup_{n\in\mathbb{N}} E_n.$$

2) O conceito de probabilidade condicionada é de extrema importância em contextos médicos. Por exemplo: diferente prevalência de uma doença em diferentes estratos da população; probabilidades de acerto de testes/exames de diagnóstico (falsos positivos e falsos negativos), etc.

6. Independência de Acontecimentos

Considere uma experiência aleatória, modelada por um espaço de probabilidade $(\Omega, \mathcal{F}, P)$.

Intuitivamente, vamos querer dizer que dois acontecimentos, A e B, serão independentes se a ocorrência de um deles não alterar a probabilidade de ocorrência do outro, i.e.,

$$P(A|B) = P(A) \text{ e } P(B|A) = P(B),$$

desde que $P(A)P(B) > 0$).

A definição de independência é, no entanto, a seguinte:

Definição [Independência de Dois Acontecimentos]

Sejam A e B dois quaisquer elementos de $\mathcal{F}$, i.e., dois quaisquer acontecimentos. A e B dizem-se independentes se

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

Observações:

1) Esta definição de independência capta a **ideia intuitiva** mencionada acima. De facto, se A e B são independentes e desde que

$$P(A)P(B) > 0,$$

teremos:

$$P(A|B) = P(A) \text{ e } P(B|A) = P(B).$$

2) Não confundir acontecimentos disjuntos (ou incompatíveis) com acontecimentos independentes!
A noção de incompatibilidade é intrínseca aos acontecimentos, i.e., não depende da medida de probabilidade (recorde que se exige apenas $A \cap B = \emptyset$). Já a noção de independência faz uso da medida de probabilidade utilizada, P.
Vamos, de seguida, ver um exemplo que ilustra bem esta ideia.

Exemplo: Considere a experiência aleatória que consiste em escolher, ao acaso, uma carta de um baralho com 52 cartas. O espaço de probabilidade é $(\Omega, \mathcal{P}(\Omega), P)$, em que $\Omega = \{1, \ldots, 52\}$ e P é a medida de probabilidade de Laplace

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{52}, \quad A \in \mathcal{P}(\Omega).$$

Os seguintes acontecimentos

C: "saiu uma carta de copas" e F: "saiu uma figura"

são independentes, relativamente à medida P. De facto: $P(C) = \frac{13}{52}$, $P(F) = \frac{12}{52}$ e tem-se que

$$\mathbf{P}(\mathbf{C} \cap \mathbf{F}) = \frac{3}{52} = \mathbf{P}(\mathbf{C})\mathbf{P}(\mathbf{F}).$$

No entanto, C e F não são independentes relativamente à medida de probabilidade P_D, em que D é o acontecimento

D: "saiu uma espada ou uma figura".

De facto: $P(D) = \frac{22}{52}$, $P_D(C) = \frac{3/52}{22/52} = \frac{3}{22}$ e $P_D(F) = \frac{12/52}{22/52} = \frac{12}{22}$, mas

$$\mathbf{P_D}(\mathbf{C} \cap \mathbf{F}) = \frac{P(C \cap F \cap D)}{P(D)} = \frac{3/52}{22/52} = \frac{3}{22} \neq \mathbf{P_D}(\mathbf{C})\,\mathbf{P_D}(\mathbf{F}).$$

[Obs: Num caso destes, é usual dizer que, condicionalmente a D, C e F não são independentes.]

Definição [Família Finita de Acontecimentos Independentes]

Sejam $A_1, A_2, \ldots, A_n$ quaisquer n acontecimentos, com $n \in \mathbb{N}$ fixo (finito) e $n \geq 2$. Diz-se que $A_1, A_2, \ldots, A_n$ formam uma família de n acontecimentos independentes se

$$\underset{\{k_1,k_2,\ldots,k_r\}\subseteq\{1,2,\ldots,n\}}{\forall}, \quad P\left(\bigcap_{i=1}^{r} A_{k_i}\right) = \prod_{i=1}^{r} P(A_{k_i}).$$

Exemplo: ($n = 3$)

Os acontecimentos A, B e C formam uma família de 3 acontecimentos independentes se são satisfeitas as seguintes 4 condições:

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

ii) $P(A \cap C) = P(A)\, P(C)$

iii) $P(B \cap C) = P(B)\, P(C)$

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

7. Considerações Finais

As σ-álgebras são estruturas difíceis de lidar. Por vezes, é suficiente lidar com estrututas mais simples, como os π-sistemas.

Definição [π-sistema]

Seja Ω um conjunto e $\mathcal{S}$ uma família de subconjuntos de Ω. $\mathcal{S}$ diz-se um π-sistema sobre Ω se verifica a seguinte condição:

$$F_1, F_2 \in \mathcal{S} \implies F_1 \cap F_2 \in \mathcal{S}.$$

Exemplo: A família de subconjuntos de $\mathbb{R}$ dada por

$$\pi(\mathbb{R}) = \{\,]-\infty, c] : c \in \mathbb{R}\,\}$$

é um π-sistema sobre $\mathbb{R}$.
É, na verdade, um π-sistema muito importante em probabilidades, pois é possível mostrar que

$$\mathcal{B}(\mathbb{R}) = \sigma(\pi(\mathbb{R})).$$

Em muitas situações, esta igualdade é tudo o que precisamos saber ou usar quando lidamos com a σ-álgebra $\mathcal{B}(\mathbb{R})$.

Lema

Seja Ω um conjunto, $\mathcal{S}$ um π-sistema sobre Ω e $\mathcal{A}$ uma σ-álgebra sobre Ω e tal que $\sigma(\mathcal{S}) = \mathcal{A}$.
Se P_1 e P_2 são duas medidas de probabilidade sobre $(\Omega, \mathcal{A})$ que coincidem em $\mathcal{S}$

(i.e., P_1 e P_2 são tais que $P_1(E) = P_2(E), \forall\, E \in \mathcal{S}$)

então P_1 e P_2 coincidem em toda a σ- álgebra $\mathcal{A}$

(i.e., P_1 e P_2 serão tais que $P_1(E) = P_2(E), \forall\, E \in \mathcal{A}$).

7. Considerações Finais

Nota Importante

O Lema anterior conjugado com o facto de

$$\sigma(\pi(\mathbb{R})) = \mathcal{B}(\mathbb{R}),$$

com $\pi(\mathbb{R}) = \{\,]-\infty, c] : c \in \mathbb{R} \,\}$, permite-nos concluir que, se duas medidas de probabilidade sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ coincidem no π-sistema $\pi(\mathbb{R})$ (i.e., atribuem o mesmo valor aos subconjuntos reais da forma $]-\infty, c],\ c \in \mathbb{R}$) então as duas medidas coincidem em toda a $\mathcal{B}(\mathbb{R})$.

Este resultado/nota vai ser muito importante mais à frente, sobretudo no que respeita à função de distribuição de uma variável aleatória.

7. Considerações Finais

Para terminar este capítulo, vamos demonstrar que

$$\boldsymbol{\sigma}(\boldsymbol{\pi}(\mathbb{R})) = \boldsymbol{\mathcal{B}}(\mathbb{R}), \tag{2}$$

que é um bom exercício sobre propriedades de uma σ-álgebra.
Adicionalmente, é necessário recorrer ao seguinte resultado sobre $\mathbb{R}$:
todo o subconjunto aberto de $\mathbb{R}$ pode ser escrito como uma reunião numerável de intervalos abertos.

Demonstração (esboço): A igualdade (2) prova-se por dupla inclusão.

- $\boldsymbol{\sigma}(\boldsymbol{\pi}(\mathbb{R})) \subseteq \boldsymbol{\mathcal{B}}(\mathbb{R})$: Basta provar a inclusão $\pi(\mathbb{R}) \subseteq \mathcal{B}(\mathbb{R})$ (simples).

- $\boldsymbol{\mathcal{B}}(\mathbb{R}) \subseteq \boldsymbol{\sigma}(\boldsymbol{\pi}(\mathbb{R}))$: Basta provar a inclusão

$$\{G \subseteq \mathbb{R} : G \text{ é um aberto}\} \subseteq \sigma(\pi(\mathbb{R})).$$

Para provar esta inclusão (+ difícil), para além de usar o resultado já referido, observe que todo o intervalo aberto $]a, b[$, com $a < b$ e a e b números reais, se pode escrever na seguinte forma:

$$]\,a, b\,[= \bigcup_{n \in \mathbb{N}} \left] a, b - \frac{\epsilon}{n} \right]$$

com $\epsilon = (b - a)/2$. Bastará depois mostrar que todo o intervalo da forma $]\,a, u\,]$, com $a < u$ e a e u números reais, é um elemento de $\sigma(\pi(\mathbb{R}))$.