Capítulo 2 - Noções básicas de probabilidade

Conceição Amado, Ana Pires e M. Rosário Oliveira

2.1 Experiências aleatórias. Espaço de resultados. Acontecimentos

"Sempre que aplicamos matemática a fim de estudar alguns fenómenos de observação, devemos essencialmente começar por construir um modelo matemático (determinístico ou não) para esses fenómenos." Neyman, J.

Definição: Um modelo que estipula que as condições sob as quais uma experiência é realizada determinam o resultado dessa experiência denomina-se modelo determinístico.

Exemplos: equação do movimento uniforme: $s(t) = v \times t$; lei de Ohm: $V = R \times I$.

Definição: Denomina-se modelo probabilístico quando a realização de uma dada experiência sob determinadas condições irá ter vários resultados possíveis, aos quais, se possível, vamos associar um número a que chamaremos probabilidade desse acontecimento.

Exemplos: fenómenos meteorológicos; euromilhões; lançamento de uma moeda ...

Definição: Uma experiência aleatória é a realização de um fenómeno aleatório.

Características:

(i) os resultados particulares são imprevisíveis mas é possível des-crever o conjunto dos resultados possíveis;

(ii) apesar dos resultados particulares serem imprevisíveis é possível observar um padrão de regularidade ao fim de um grande número de realizações.

Exemplos:

- jogos de azar:
  - lançamento de uma moeda;
  - lançamento de um dado;
  - escolha de uma carta num baralho.
- energia consumida numa reacção química;
- duração de uma chamada telefónica;
- característica "defeituosa" ou "não defeituosa" de peças produzidas em série.

2.1 (cont.)

Definição: Denomina-se por resultado possível ou elementar a toda e qualquer informação que pode ser registada como resultado de uma experiência aleatória.

Definição: Chama-se espaço de resultados ao conjunto de todos os resultados possíveis de uma experiência aleatória.

Representa-se geralmente por Ω ou S.

- A formulação de um modelo probabilístico associado a uma experiência aleatória inicia-se pela definição do espaço de resultados;
- Cada resultado elementar é representado por um e um só elemento de Ω;
- Os elementos de Ω podem ser números, atributos ou uma combinação de elementos quantitativos e qualitativos;
- Ω pode ser finito, infinito numerável ou infinito não numerável.

2.1 (cont.)

Exemplos:

	Experiência	Ω
E_1	lançamento de moeda	$\Omega_1 = \{\text{cara}, \text{coroa}\}$
E_2	lançamento de dado	$\Omega_2 = \{$ ⚀ , ⚁ , ⚂ , ⚃ , ⚄ , ⚅ $\}$
E_3	duração de chamada telefónica	$\Omega_3 = [0, +\infty[$
E_4	classificação de uma peça	$\Omega_4 = \{\text{defeituosa},\ \text{não defeituosa}\} = \{\text{d},\ \text{n}\}$
E_5	obs. sucessiva de peças até encontrar uma defeituosa	$\Omega_5 = \{\text{d},\ \text{nd},\ \text{nnd},\ \text{nnnd},\ \ldots\}$

2.1 (cont.)

Definição: Um acontecimento é um subconjunto do espaço de resultados de uma experiência aleatória.

- Em geral, os acontecimentos, são representados pelas primeiras letras maiúsculas do alfabeto latino.
- Alguns acontecimentos especiais. Seja $\Omega = \{\omega_1, \omega_2, \cdots, \omega_k, \cdots\}$ um espaço de resultados, então define-se acontecimento:
  - elementar como sendo qualquer conjunto $\{\omega_i\}, \quad i = 1, 2, \cdots;$
  - certo se contém todos os elementos de Ω ;
  - impossível se não contém nenhum elemento de Ω (conjunto $\emptyset$).

2.1 (cont.)

Exemplos:

E_1: $A_1 = \{\text{cara}\}$

E_2: $A_2 = \{\text{n.}^{\text{o}}\text{de pontos inferior a 3}\} = \{$ ⚀ , ⚁ $\}$

(definição em compreensão) (definição em extensão)

E_3: $A_3 = \{\text{duração inferior a 30 unidades}\} = [0, 30[$

E_4: $A_4 = \{\text{d}\}$

Definição: Espaço de acontecimentos de uma experiência aleatória, $\mathscr{A}$, é o conjunto de todos os acontecimentos definidos num espaço de resultados.

Definição: Dada uma experiência aleatória, diz-se que ocorreu o acontecimento A se e só se ao realizar a experiência (uma única vez) o resultado obtido é um elemento de A.

Considere-se E_2,

- se o dado for lançado e sair ⚁ pode dizer-se que ocorreu A_2;
- se o dado for lançado e sair ⚄ pode dizer-se que não ocorreu A_2.

Mais exemplos...

- E6 = Lançamento de dois dados, com faces numeradas de 1 a 6, com o objectivo de registar os números das faces voltadas para cima.
  - $\Omega = \{$(⚀, ⚀), (⚀, ⚁), ..., (⚅, ⚅)$\}$ é o espaço de resultados e $\#\Omega = 36$.
  - O resultado (⚂, ⚀) é um acontecimento elementar/possível.
  - O acontecimento A="ocorrer faces iguais" é representado por $A = \{$(⚀, ⚀), (⚁, ⚁), ..., (⚅, ⚅)$\}$.
- E7 = Lançamento de dois dados, com faces numeradas de 1 a 6, com o objectivo de registar a soma dos números das faces voltadas para cima.
  - O espaço de resultados é $\Omega = \{2, 3, \ldots, 12\}$ e $\#(\Omega) = 11$.
  - O acontecimento A="a soma das faces ser múltiplo de 3" , que corresponde à ocorrência dos pares (⚀,⚁), (⚁,⚀), (⚂,⚂), (⚃,⚁), (⚁,⚃), (⚃,⚄), (⚄,⚃), (⚅,⚅) é representado por $A = \{3, 6, 9, 12\}$. Se lançarmos os dados e sair (⚀,⚁) dizemos que A se realizou.

2.1 (cont.)

Operações com acontecimentos ($\Leftrightarrow$ operações com conjuntos):

– complementação ($\overline{A}$);

– união ($A \cup B$);

– intersecção ($A \cap B$);

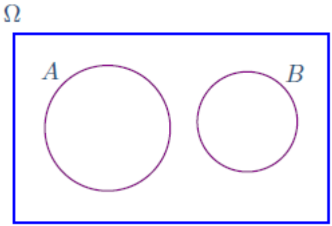
– diferença ($A \backslash B$)

Rever:

- diagramas de Venn;
- propriedades das operações (comutativas, associativas, distributivas, elementos neutros, elementos absorventes, leis de De Morgan, dupla negação).

2.1 (cont.)

Definição: Dois acontecimentos A e B dizem-se mutuamente exclusivos se não puderem ocorrer simultaneamente, ou seja, se $A \cap B = \emptyset$.



A probabilidade é uma medida que pretende quantificar a "possibilidade" de ocorrência de cada acontecimento.

A noção de probabilidade é um conceito complexo, no entanto pode-se adiantar algumas das suas interpretações.

Dado um acontecimento A, pertencente a um determinado Ω, represente-se por $P(A)$ a probabilidade desse acontecimento se realizar, a qual é traduzida por um número real no intervalo $[0, 1]$.

Interpretação/definição clássica ou de Laplace:

Dado um espaço de resultados com N elementos cuja ocorrência (por questões de simetria/indiferença) é igualmente possível, a probabilidade de qualquer acontecimento A é dada por

$$P(A) = \frac{\#A}{N} = \frac{\text{n.}^{\text{o}}\text{de casos favoráveis a } A}{\text{n.}^{\text{o}}\text{de casos possíveis}}$$

$\Longrightarrow$ Rever cálculo combinatório.

Limitação

A definição clássica não pode ser aplicada quando:

- o espaço de resultados tem um número infinito de elementos;
- os elementos não são igualmente possíveis.

$\Longrightarrow$ São necessárias outras interpretações de probabilidade!

Definição: Dada uma experiência aleatória que se realizou n vezes, e um acontecimento A, chama-se frequência relativa do acontecimento A, ao quociente

$$f_n(A) = \frac{n(A)}{n}$$

onde $n(A)$ representa o número de vezes que se observou o acontecimento A (ou seja, é a frequência absoluta de A).

2.2 (cont.)

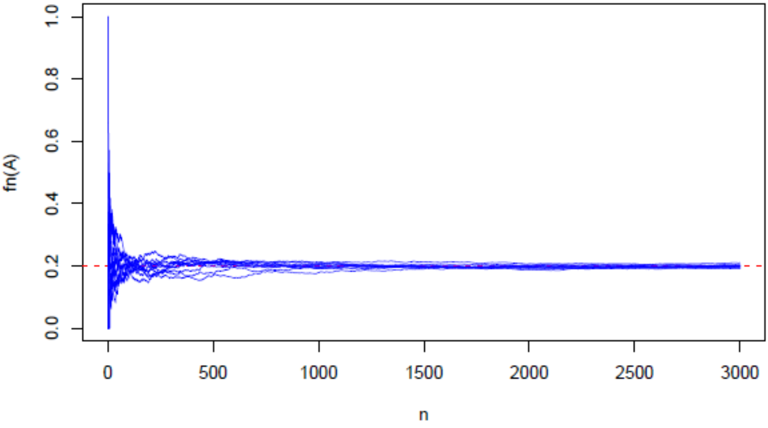
Exemplo: Imagine-se uma experiência aleatória só com dois resultados possíveis mas que não sejam necessariamente igualmente possíveis. Pode ser o caso do lançamento de uma moeda em que não se tem a certeza de que a moeda é equilibrada ou a experiência E_4 da Secção 2.1 (classificação de peças em defeituosas ou não defeituosas).

Repetindo a experiência um número muito elevado de vezes observa-se que a frequência relativa dos acontecimentos elementares (que são só dois, neste caso), tende a estabilizar à medida que o número de repetições cresce (embora a sequência particular de valores seja imprevisível).

Nota 1: Programa (*copy*/*paste* no R)

```
n<-3000 ## n = número de repetições
p<-0.2 ## p = verdadeiro valor de P(A)
k<-10 ## k = número de sequências
aux<-rbinom(n,1,p)
plot(cumsum(aux)/(1:n),type="l",ylim=c(0,1),xlab="n",ylab="fn(A)",col="blue")
abline(h=p,lty=2,col=2)
for (i in 1:10){
aux<-rbinom(n,1,p)
if (interactive()) {cat("\nCarregue em  <Return> para continuar: ")
readline()}
lines(cumsum(aux)/(1:n),col="blue")
}
```

Exemplo (cont.): O gráfico seguinte mostra 10 sequências de frequências relativas (fictícias) cada uma correspondendo a 3000 realizações de uma experiência aleatória com o verdadeiro valor de $P(A) = 0.2$. [1]



[1] Simulação com números pseudo-aleatórios. Código na Nota 1.

Interpretação/definição frequencista:

A probabilidade do acontecimento A, $P(A)$, é o "limite" para o qual tende a frequência relativa, $f_n(A)$, quando $n \to \infty$.

Limitação: A definição frequencista não pode ser aplicada quando

- não é possível repetir a experiência um número muito elevado de vezes;
- não é possível repetir a experiência exactamente nas mesmas condições.

Exemplo:

Qual a probabilidade de ganhar o Totobola com uma única aposta?

- Há $3^{13} = 1594323$ casos possíveis, mas não são igualmente possíveis! Inviabiliza a interpretação de Laplace
- Os jogos entre as mesmas equipas não correspondem a repetições nas mesmas condições do próximo jogo, nem são em número suficientemente elevado! Inviabiliza a interpretação frequencista

$\Longrightarrow$ São necessárias outras interpretações de probabilidade!

Interpretação/definição subjectivista ou subjectiva:

Admite-se que cada pessoa pode atribuir a cada acontecimento um número — a que chama "probabilidade do acontecimento" — e que expressa o seu grau de credibilidade pessoal em relação à ocorrência do acontecimento.

A probabilidade subjectiva de um dado acontecimento pode variar de indivíduo para indivíduo, mas deve ser coerente para o mesmo indivíduo.

A coerência é garantida pela **definição axiomática**.

- Os axiomas são inspirados em propriedades verificadas pelas interpretações anteriores (clássica e frequencista) e a sua verificação é exigida no caso da interpretação sujectivista;
- Consoante a situação, é razoável admitir qualquer uma das interpretações.

Definição axiomática (axiomática de Kolmogorov):

As probabilidades dos acontecimentos pertencentes ao conjunto dos acontecimentos definidos em Ω, designado por $\mathscr{A}$ (ver Obs. 1), é um número satisfazendo os três **Axiomas** seguintes:

Axioma 1 (*não negatividade*) – $P(A) \geq 0, \quad \forall_A \in \mathscr{A}$

Axioma 2 (*normalização*) – $P(\Omega) = 1$

Axioma 3 (*sigma-aditividade*) – $P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i),$

$\forall_{A_1, A_2, \ldots \in \mathscr{A}:\ A_i \cap A_j = \emptyset,\ (i \neq j)}$ (ver Obs. 2)

Detalhes técnicos:

Obs. 1 Se Ω for discreto $\mathscr{A}$ pode conter todos os subconjuntos de Ω, caso contrário é necessário impor restrições;

Obs. 2 Para mais detalhes sobre sigma-aditividade ver, por exemplo, Bauer, H. (2001), *Measure and Integration Theory*, Berlin: de Gruyter

2.2 (cont.)

Teoremas (Resultados) decorrentes dos Axiomas:

Os axiomas permitem estabelecer um conjunto de resultados para determinar probabilidades de acontecimentos resultantes de operações entre acontecimentos (mas é sempre necessária uma "base" de partida, obtida através de uma das interpretações anteriores).

Resultado 1: $P(\overline{A}) = 1 - P(A)$

Demonstração: $A \cup \overline{A} = \Omega, \quad A \cap \overline{A} = \emptyset$

$$\left.\begin{array}{ll} \text{pelo Axioma 3,} & P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A}) \\ \text{pelo Axioma 2,} & P(\Omega) = 1 \end{array}\right\} \Leftrightarrow P(A)+P(\overline{A}) = 1$$

Resultado 2: $P(\emptyset) = 0$

Demonstração: consequência do Resultado 1 e do Axioma 1, pois $\emptyset = \overline{\Omega}$

Resultado 3: Se $A \subset B$ então $P(A) \leq P(B)$

Demonstração: Se $A \subset B$, pode escrever-se

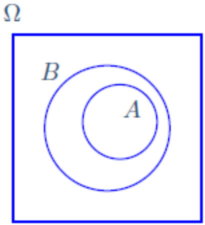
$B = A \cup (B \backslash A)$ e

$A \cap (B \backslash A) = \emptyset$

pelo Axioma 3, $P(B) = P(A) + P(B \backslash A)$

pelo Axioma 2, $P(B \backslash A) \geq 0$

logo $P(B) \geq P(A)$



Obs.: Notar que para A e B genéricos (isto é, A não necessariamente contido em B), se tem $P(B \backslash A) = P(B) - P(A \cap B)$

(Exercício: demonstrar!)

2.2 (cont.)

Resultado 4: $\forall_{A,B} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$

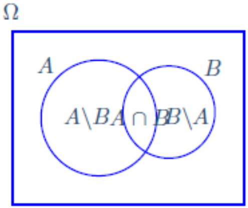
Demonstração: Pode escrever-se $A \cup B = A \cup B \backslash A$ e como $A \cap B \backslash A = \emptyset$

Usando:

Axioma 3: $P(A \cup B) = P(A) + P(B \backslash A)$

Resultado anterior: $P(B \backslash A) = P(B) - P(A \cap B)$

logo dá o resultado pretendido, pois



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

Resultado 5: dados três acontecimentos A, B e C, quaisquer

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - \\ &\quad -P(A \cap B) - P(A \cap C) - P(B \cap C) + \\ &\quad +P(A \cap B \cap C) \end{aligned}$$

Demonstração: Escrever $A \cup B \cup C = (A \cup B) \cup C$ e aplicar o resultado anterior.

Resultado 6: Dados k acontecimentos, A_1, A_2, ..., A_k, quaisquer

$$\begin{aligned} P\left(A_1 \cup \cdots \cup A_k\right) &= P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P\left(A_i\right) - \sum_{i<j=1}^{k} P\left(A_i \cap A_j\right) + \cdots \\ &\quad \cdots + (-1)^{k+1} P\left(A_1 \cap \cdots \cap A_k\right) \end{aligned}$$

Demonstração: Faz-se por indução, aplicando o Resultado 4.

2.3 Probabilidade condicionada

Cálculo de probabilidades quando há alguma informação adicional sobre o resultado de uma experiência.

É importante porque em muitos casos é mais fácil calcular probabilidades condicionadas do que não condicionadas

Exemplo: Considere-se o lançamento de um dado equilibrado com 6 faces ($\Omega = \{1, 2, 3, 4, 5, 6\}$), e os acontecimentos

A - sai a face 2, $\quad A = \{2\}$

B - sai a face 1, $\quad B = \{1\}$

C - sai face com número ≤ 2, $\quad C = \{1, 2\}$

dado equilibrado $\Rightarrow P(A) = P(B) = \frac{1}{6}$, $P(C) = \frac{1}{3}$.

Informação adicional: saiu face par (acontecimento D)

$$D = \{2, 4, 6\} \quad \longrightarrow \quad \text{espaço de resultados reduzido}$$

Exemplo (cont.):
As probabilidades dos acontecimentos A, B e C, são alteradas em face da informação adicional de que ocorreu o acontecimento D!

Notação: representa-se por $P(A|D)$ a probabilidade de A ocorrer sabendo que D ocorreu (pode ler-se *probabilidade condicionada de A dado D*).

É simples verificar que

no espaço reduzido (D)	no espaço original (Ω)
$P(A\|D) = \frac{1}{3}$	$= \dfrac{P(A \cap D)}{P(D)}$
$P(B\|D) = 0$	$= \dfrac{P(B \cap D)}{P(D)}$
$P(C\|D) = \frac{1}{3}$	$= \dfrac{P(C \cap D)}{P(D)}$

Definição: A probabilidade condicionada do acontecimento A sabendo que ocorreu B (tal que $P(B) > 0$) é

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

A probabilidade condicionada (sendo fixo o acontecimento condicionante, D, com $P(D) > 0$) é uma nova medida de probabilidade que verifica os Axiomas e os Resultados decorrentes destes.

A1. $P(A|D) \geq 0$ **A2.** $P(\Omega|D) = 1$

A3. $P\left(\bigcup_{i=1}^{+\infty} A_i|D\right) = \sum_{i=1}^{+\infty} P(A_i|D), \quad \forall_{A_1, A_2, \ldots \in \mathscr{A}:\, A_i \cap A_j = \emptyset,\, (i \neq j)}$

R1. $P(\overline{A}|D) = 1 - P(A|D)$

R4. $P\left(A \cup B|D\right) = P(A|D) + P(B|D) - P(A \cap B|D)$ **etc.**

2.4 Teoremas da probabilidade composta e da probabilidade total. Teorema de Bayes)

Muitas vezes $P(A|B)$ é conhecida ou fácil de obter recorrendo ao espaço de resultados reduzido e a definição de probabilidade condicionada tem também grande aplicação no cálculo de probabilidades de intersecções, como é fácil de verificar se observarmos que:

$$\Downarrow$$

Lei das probabilidades compostas

(ou regra da multiplicação)

dados 2 acontecimentos tais que $P(A) > 0$ e $P(B) > 0$

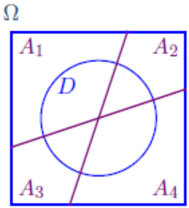
$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

Estas relações podem ser generalizadas e apresentadas no teorema seguinte.

Lei das probabilidades compostas (geral)

dados n acontecimentos tais que $P(\cap_{i=1}^{n-1} A_i) > 0$

$$\begin{aligned} P(A_1 \cap A_2 \cap \cdots \cap A_n) &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots \\ &\cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \end{aligned}$$



Definição: Os subconjuntos não vazios A_1, A_2, ..., A_m formam uma partição de Ω se

$$A_1 \cup A_2 \cup \cdots \cup A_m = \Omega \quad \text{e} \quad A_i \cap A_j = \emptyset, \ \forall_{i \neq j = 1, \ldots m}$$

(ou seja são exaustivos e mutuamente exclusivos)

Teorema da probabilidade total

Se A_1, A_2, ..., A_m é uma partição de Ω tal que $P(A_i) > 0$, $\forall_i$, então

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_m)P(A_m)$$

Teorema de Bayes

Se A_1, A_2, ..., A_m é uma partição de Ω tal que $P(A_i) > 0$, $\forall_i$, então para qualquer acontecimento B tal que $P(B) > 0$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \cdots + P(B|A_m)P(A_m)}$$

Exercício (aula)

Definição: Dois acontecimentos A e B são independentes ($A \perp\!\!\!\perp B$) se e só se
$P(A \cap B) = P(A)P(B)$.

Observações:

- Esta definição é sempre válida.
- Se A é tal que $P(A) = 0$, então A é independente de qualquer outro acontecimento;
- Todo o acontecimento A é independente dos acontecimentos $\emptyset$ e de Ω.
- se $P(A) > 0$ e $P(B) > 0$ e $A \cap B = \emptyset$ (A e B mutuamente exclusivos), então A e B <u>não são</u> independentes;
- Se $A \perp\!\!\!\perp B$ então $P(A|B) = P(A)$ e $P(B|A) = P(B)$, para A e B tais que $P(A) > 0$ e $P(B) > 0$.
- Se $A \perp\!\!\!\perp B$ então $\overline{A} \perp\!\!\!\perp B$, $A \perp\!\!\!\perp \overline{B}$ e $\overline{A} \perp\!\!\!\perp \overline{B}$.

Exercício: demonstrar as afirmações anteriores.

Independência de mais do que dois acontecimentos

Definição: A_1, A_2, ..., A_n são acontecimentos mutuamente ou completamente independentes se para qualquer número inteiro $2 \leq r \leq n$ e qualquer grupo de r acontecimentos
$A_{i_1}, A_{i_2}, \ldots, A_{i_r}$

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_r})$$

Por exemplo:
A, B e C são (mutuamente) independentes se

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

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

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

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

2.5 (cont.)

Definição: Dois acontecimentos são A e B são condicionalmente independentes em relação a um acontecimento C se

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

Observar que:

Independência entre A, B e C implica independência condicional mas não o contrário, i.e, a independência condicional não implica a independência no sentido corrente, a não ser quando $C = \Omega$.

Exercícios (na aula)