Tópico 1: Conceitos de Probabilidade - Experimento Aleatório; Espaço de Probabilidade; Probabilidade Condicional e Independência. Teorema de Bayes

Ben Dêivide

6 de outubro de 2021

Quando desejamos compreender algum fenômeno da natureza, tentamos estudá-lo por meio de um processo de observação chamado experimento. Para o nosso estudo, definimos

Definição 1 (Experimento Aleatório)**.** *Todo experimento cujo resultado não pode ser previsto antes de sua execução, é chamado de experimento aleatório.*

Podemos apresentar alguns exemplos.

Exemplo 1. *Lançar um dado equilibrado e observar o resultado obtido na face superior do dado.*

Exemplo 2. *Observar o número de chamadas telefônicas que chegam a uma central telefônica em um determinado intervalo de tempo.*

Exemplo 3. *Para a escolha ao acaso de uma lâmpada que acabou de sair do processo de fabricação, verificar o tempo de duração da lâmpada em funcionamento.*

Por mais que não seja possível prever o resultado antes de sua execução, sabemos que diversos resultados possíveis podem ocorrer. Assim, definimos,

Definição 2 (Espaço amostral)**.** *O conjunto de todos os resultados possíveis de um experimento, denotado por Ω, é chamado de espaço amostral.*

Cada resultado de Ω é chamado de ponto ou elemento amostral. Denotamos o elemento por w, e expressar $w \in \Omega$, isto é, w pertence a Ω. Cada resultado possível corresponde um, e somente um ponto $\omega \in \Omega$, e resultados distintos correspondem a pontos distintos de $\omega \in \Omega$, isto é, ω representa apenas um único resultado de Ω. Vejamos alguns exemplos.

Exemplo 4. *Com base nos Exemplos anteriores, temos para o Exemplo 1 temos $\Omega = \{1, 2, 3, 4, 5, 6\}$. Para o Exemplo 2 temos $\Omega = \mathbb{N}$. E para o Exemplo 3, temos $\Omega = \mathbb{R}^+$.*

Exemplo 5. *Um experimento lança duas moedas honestas, e deseja-se verificar a face superior dessas moedas. Sabe-se que cada moeda apresenta duas faces: cara (H) e coroa (T). Dessa forma, o espaço amostral é dado por:*

$$\Omega \;=\; \{(H, H), (H, T), (T, H), (T, T)\}.$$

Entretanto, também podemos ter um conjunto qualquer A, que contém parte do elementos de Ω, isto é, $A \subset \Omega$, A passa a ser chamado de subconjunto de Ω.

Definição 3 (Subconjunto)**.** *Se todo elemento do conjunto A é também elemento de Ω, então A é definido como um subconjunto de Ω, sendo representado $A \subset \Omega$ ou $\Omega \supset A$ (lê-se: A está contido em Ω ou Ω contém A).*

Essa definição pode ser aplicada entre subconjuntos de Ω, como no exemplo a seguir.

Exemplo 6. *Sejam o conjunto $\Omega = \{1, 2, 3, 4, 5, 6\}$, e seus subconjuntos,*

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

então A é um subconjunto de B, pois, os elementos que contém em A, também contém em B. Assim, $A \subset B$.

Definição 4 (Evento)**.** *Todo subconjunto do espaço amostral (Ω), representado por letras latinas em maiúsculo, A, B, ..., é chamado de evento.*

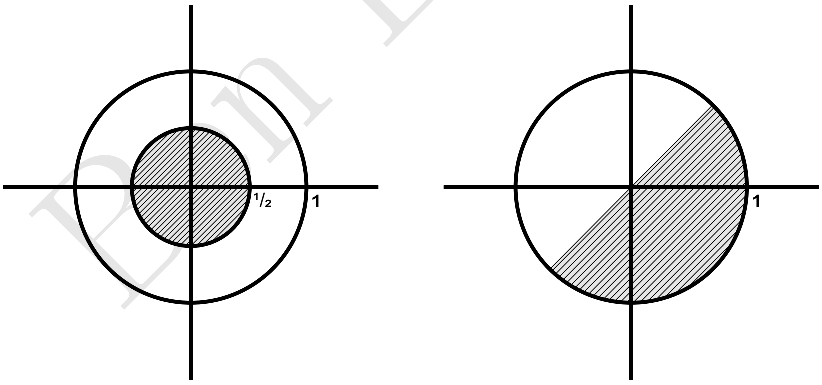
Exemplo 7. *Escolher ao acaso um ponto no círculo de raio 1 centrado na origem. Então*

$$\Omega = \textit{círculo unitário} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$

Vejamos alguns eventos para esse exemplo:

$$\begin{aligned}
A &= \text{“distância entre o ponto escolhido e a origem é”} \leq 1/2 \\
B &= \text{“distância entre o ponto escolhido e a origem é”} \geq 15 \\
C &= \text{“1}^{\underline{a}}\text{ Coordenada do ponto escolhido é maior que a 2}^{\underline{a}}.
\end{aligned}$$

Se $\omega = (x, y)$ for um resultado do experimento, então ω pertencerá a A se, e somente se, $x^2 + y^2 \leq 1/4$. Pertencerá ao evento C se, e somente se, $x > y$. Nenhum ponto ω pertencerá a B.



(a) Evento A (b) Evento C

Figura 1: Escolha do ponto em um círculo unitário.

Logo temos:

$$\begin{aligned}
A &= \{(x, y) \in \Omega : \sqrt{x^2 + y^2} \leq 1/2\}, \\
B &= \emptyset = \textit{conjunto vazio}, \\
A &= \{(x, y) \in \Omega : x > y\}.
\end{aligned}$$

Então, todo experimento associado a este experimento pode ser identificado por um subconjunto do espaço amostral.

Definição 5 (Evento certo, impossível e elementar)**.** *Seja* Ω *o espaço amostral do experimento. Então dizemos que* Ω *é o evento certo, e* $\emptyset$ *é o evento impossível, e o evento* $\{\omega\}$ *é dito elementar.*

Para a compreensão de algumas propriedades, a seguir definimos mais alguns tipos de eventos.

Definição 6 (União de eventos)**.** *Sejam* A *e* B*, dois eventos quaisquer de* Ω*, então o conjunto de todos os elementos que estão em* A *ou* B *ou em ambos, é definido conjunto união de* A *e* B*, denotado por* $A \cup B$*.*

Dessa forma, percebemos que $A \cup B$ ocorre se ao menos um dos eventos A ou B ocorrer.

Exemplo 8. *Sejam os conjuntos:*

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

então

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

Definição 7 (Intersecção de eventos)**.** *Sejam* A *e* B*, dois eventos quaisquer de* Ω*, então o conjunto que contém todos os elementos que estão em* A *e* B*, é definido a intersecção de* A *e* B*, e escrito* $A \cap B$ *ou* AB*.*

Exemplo 9. *Do exemplo anterior, temos que a intersecção de* $AB = \{3\}$*.*

Definição 8 (Eventos Disjuntos ou multuamente esxclusivos)**.** *Sejam* A *e* B*, dois eventos quaisquer de* Ω*, então estes são disjuntos ou mutuamente exclusivos quando não existir elementos em comum entre* A *e* B*, isto é,* $A \cap B = \emptyset$*.*

Teorema 1. *Sejam dois eventos* A *e* B *em* Ω*. Se* $A \cap B = \emptyset$*, então* $A^c \cap B^c \neq \emptyset$*, a menos que* A *e* B *sejam complementares.*

Demonstração. Considere $A \cap B = \emptyset$ e que

$$\begin{aligned} A \cup B &= (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B) \\ &= (A \cap B^c) \cup (A^c \cap B) \\ &\neq \Omega \quad \text{(Pelo fato de A e B não serem complementares).} \end{aligned} \tag{1}$$

Usando a Lei de Morgan $A^c \cap B^c = (A \cup B)^c$, logo percebemos pela expressão (1) que $A^c \cap B^c \neq \emptyset$, o que completa a prova. □

Exemplo 10. *Sejam os eventos* $A = \{1, 2, 3, 4\}$ *e* $B = \{5, 6\}$*, então* $A \cap B = \varnothing$

Uma relação de eventos que será muito importante para o estudo da teoria da probabilidade, é a definição de complemento, abordado a seguir.

Definição 9 (Evento complementar)**.** *Seja* A *um evento de* Ω*. Então o complemento do evento* A *com respeito a* Ω*, denotado por* $\overline{A}$*,* A^c*, ou* $\Omega - A$*, é o subconjunto dos elementos de* Ω *exceto os elementos do evento* A*.*

Observemos o seguinte exemplo.

Exemplo 11. *Um experimento lança três moedas honestas, e deseja-se verificar a face superior dessas moedas. Sabe-se que cada moeda apresenta duas faces: cara (H) e coroa (T). Dessa forma, o espaço amostral é dado por:,*

$$\begin{aligned}\Omega &= \{(H,H,H),(H,H,T),(H,T,H),(H,T,T),\\ &\quad (T,H,H),(T,H,T),(T,T,H),(T,T,T)\}.\end{aligned}$$

e um evento de Ω, pode ser dado por

$$A = \{(H,H,H),(H,H,T),(H,T,T)\}.$$

Então o complemento de A será:

$$\overline{A} = \{(T,H,H),(T,H,T),(T,T,T)\}.$$

Seja um evento A contido no espaço amostral Ω. Desejamos associar ao evento A uma medida que assume valores entre 0 e 1, que chamamos de medida de probabilidade de A, denotada por $P(A)$. Assim, diremos que $P(A)$ é a probabilidade de que o evento A ocorra no espaço amostral Ω. Voltando ao Exemplo 1, considerando que esse dado é equilibrado, e o evento $A \subset \Omega$, então poderemos atribuir uma probabilidade para A da seguinte forma:

$$P(A) = \frac{\#A}{6} = \frac{\text{número de resultados favoráveis a A}}{\text{número de resultados possíveis}}.$$

Esta é a definição clássica de probabilidade quando Ω é finito. Entretanto, a probabilidade que o evento A ocorra no espaço amostral nem sempre é possível, devido a complexidade desses eventos. Retornando ao Exemplo 7, podemos interpretar a probabilidade de $A \subset \Omega$ como:

$$P(A) = \frac{\text{área } A}{\text{área } \Omega} = \frac{\text{área } A}{\pi},$$

sendo a área de A bem definida. Segundo um teorema profundo da teoria da medida, não se pode definir $P(A)$ para $A \subset \Omega$ de modo que a área de A não esteja bem definida. A prova disso depende do **Axioma da escolha**. Um exemplo clássico desses eventos são os **conjuntos de Vitali de** $\mathbb{R}$, os quais não podemos atribuir nenhuma medida quando ela generaliza o comprimento de intervalos de $\mathbb{R}$. De fato é impossível atribuir comprimento a todos subconjuntos de $\mathbb{R}$ preservando a aditividade e invariância por translação.

Dessa forma, estaremos apenas interessados em eventos cuja área está bem definida. Assim, definimos

Definição 10 (Evento Aleatório)**.** *Todo evento de Ω que podemos atribuir uma probabilidade, chamamos de evento aleatório.*

Vamos contextualizar algumas definições de Teoria da medida com relação ao conjunto do espaço amostral Ω.

Definição 11 (Classe de um conjunto Ω)**.** *Uma coleção de subconjuntos de um dado conjunto Ω, é chamado de classe.*

Exemplo 12. *Considere $\Omega = \{1,2\}$ e seja $\mathcal{C}_1 = \{\emptyset,\{1\},\{2\}\}$ e $\mathcal{C}_2 = \{\emptyset,\{1\},\{2\},\{1,2\}\}$, então dizemos que $\mathcal{C}_1$ e $\mathcal{C}_2$ são classes de Ω.*

Vamos estar interessados numa classe de eventos aleatórios que atendem algumas propriedades, que serão importantes para a teoria e cálculo de probabilidades. Denotemos por $\mathcal{A}$, uma classe de eventos aleatórios definida da seguinte forma:

Definição 12 (Álgebra)**.** *Seja Ω o espaço amostral, então uma classe de Ω é chamada de álgebra, denotada por $\mathcal{A}$, se satisfaz as seguintes propriedades:*

A1. $\Omega \in \mathcal{A}$*;*

A2. $\forall\ A \in \mathcal{A},\ A^c \in \mathcal{A}$*;*

A3. Se $A \in \mathcal{A}$ *e* $B \in \mathcal{A}$*, então* $A \cup B \in \mathcal{A}$*.*

Como consequência dessas propriedades, apresentamos o seguinte Teorema,

Teorema 2. *Seja $\mathcal{A}$ uma álgebra do espaço amostral Ω. Então valem as seguintes propriedades:*

A4. $\emptyset \in \mathcal{A}$ *e*

A5. $\forall n$*,* $\forall A_1, A_2, \ldots, A_n \in \mathcal{A}$*, temos* $\bigcup_{i=1}^n A_i \in \mathcal{A}$ *e* $\bigcap_{i=1}^n A_i \in \mathcal{A}$*.*

A6. Se $A \in \mathcal{A}$ *e* $B \in \mathcal{A}$*, então* $A - B = A \cap B^c \in \mathcal{A}$*.*

Demonstração. Podemos observar que $A1$ e $A2$ implicam em $A4$. Para $A5$, temos que $A_1 \cup A_2 \in \mathcal{A} \Rightarrow (A_1 \cup A_2) \cup A_3 \in \mathcal{A} \Rightarrow \ldots \Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{A}$, por indução. Usando o fato de que:

$$\bigcap_{i=1}^n A_i = \left(\bigcup_{i=1}^n A_i^c\right)^c.$$

Por $A2$ sabemos que se $A \in \mathcal{A}$, então $A^c \in \mathcal{A}$. Por indução provamos que $\bigcup_{i=1}^n A_i \in \mathcal{A}$, e por consequência de $A3$, também $\bigcup_{i=1}^n A_i^c \in \mathcal{A}$. Portanto, se $\bigcup_{i=1}^n A_i^c \in \mathcal{A}$, por consequência de $A2$, logo $(\bigcup_{i=1}^n A_i^c)^c = \bigcap_{i=1}^n A_i \in \mathcal{A}$. Para provar A6, temos por A2 que se $B \subset \mathcal{A}$, então $B^c \subset \mathcal{A}$. E como A e B^c pertencem a $\mathcal{A}$, então por A5 $A \cap B^c = A - B \in \mathcal{A}$. □

Vamos supor para a Definição 12 que também satisfaça a seguinte propriedade:

A3*. Se $A_k \in \mathcal{A}$, para $k = 1, 2, \ldots$, então $\bigcup_{k=1}^\infty A_k \in \mathcal{A}$.

Dessa forma, definimos

Definição 13 (σ-álgebra)**.** *Uma classe de eventos de Ω, denotado por $\mathcal{F}$, é definido σ-álgebra se satisfizer as seguintes condições:*

A1. $\Omega \in \mathcal{F}$*;*

A2. Se $A \in \mathcal{F}$*, então* $A^c \in \mathcal{F}$*;*

A3. Se uma sequência finita ou infinita contável (enumerável) de eventos* $A_1, A_2, \ldots \in \mathcal{F}$*, então* $\bigcup_{i=1}^\infty A_i \in \mathcal{F}$*.*

Uma σ-álgebra é sempre uma álgebra, pois A3 é consequência de A3*, uma vez que $A \cup B = A \cup B \cup B \cup B \ldots \in \mathcal{F}$, se $\mathcal{F}$ é uma σ-álgebra. Sem perda de generalidade, podemos afirmar que a σ-álgebra é sempre uma álgebra, pelo **Teorema da Extensão de Carathéodory**. Este teorema garante que uma probabilidade definida em uma álgebra, e de acordo com os axiomas usuais, pode ser extendida de uma única maneira para a σ-álgebra gerada pela álgebra.

As duas σ-álgebras canônicas de um conjunto Ω, considerando Ω finito ou enumerável, são definidas a seguir.

Definição 14 (σ-álgebras triviais de Ω)**.** *As duas σ-álgebras triviais de Ω, considerando Ω finito ou enumerável, são:*

a) $\mathcal{C} = \{\emptyset, \Omega\}$*, a menor σ-álgebra de Ω;*

b) $\mathcal{P}(\Omega)$*, é o conjunto de todas as partes de Ω, e representa a maior σ-álgebra de Ω. O número de subconjuntos é 2^n, considerando que Ω tem n elementos.*

Alguns exemplos a seguir, complementam o conceito de σ-álgebra.

Exemplo 13. *Seja $\Omega = \mathbb{R}$, o conjunto dos números reais, e seja $\mathcal{F}$ uma coleção de todos os subconjuntos de $\mathbb{R}$. Então $\mathcal{F}$ é uma σ-álgebra.*

Exemplo 14. *Seja $\Omega = [0, 1]$ e seja $\mathcal{F} = \{\emptyset, \Omega, [0, 1/2], [1/2, 1]\}$. Então $\mathcal{F}$ é uma σ-álgebra.*

Exemplo 15. *Seja $\Omega = \mathbb{N}^*$ o conjunto dos números naturais maiores ou iguais a um, sendo $P = \{x \in \mathbb{N}^* : x \text{ é par}\}$ e $I = \{x \in \mathbb{N}^* : x \text{ é ímpar}\}$. Então $\mathcal{F} = \{\Omega, P, I, \emptyset\}$ é uma σ-álgebra de subconjuntos de Ω.*

Exemplo 16. *Seja Ω um conjunto infinito, e $\mathcal{Z}$ uma coleção de todos os conjuntos finitos de Ω. Então $\mathcal{Z}$ não contém Ω e não é fechado para complementação. Assim, $\mathcal{Z}$ não é σ-álgebra de Ω.*

Para definirmos uma σ-álgebra de $\Omega = \mathbb{R}$, apresentamos alguns teoremas a seguir.

Teorema 3. *Seja Ω um espaço amostral não vazio, e $\mathcal{S} = (\mathcal{F})_{i \in I}$ uma coleção arbitrária não vazia de σ-álgebras de Ω, então*

$$\mathcal{J} = \bigcap_{i \in I} \mathcal{F}_i = \{E \in \mathcal{F}_i,\ \forall i \in I\}$$

é a interseção de todas as σ-álgebras que pertencem a $\mathcal{S}$, que também é uma σ-álgebra de Ω, em que I é um conjunto não-vazio de índices.

Demonstração. Sabemos que $\mathcal{S}$ é uma coleção não vazia de σ-álgebras de Ω, e que $\mathcal{J} = \bigcap_{i \in I} \mathcal{F}_i$ pertence a $\mathcal{S}$.

(i) O conjunto Ω pertence a $\mathcal{J}$, pois Ω pertence a cada σ-álgebra em $\mathcal{S}$;

(ii) Suponha que $A \in \mathcal{J}$. Cada σ-álgebra que pertence a $\mathcal{S}$ contém A e contém A^c. Assim, A^c pertence a interseção $\mathcal{J}$ dessas σ-álgebras;

(iii) Por fim, suponha que $\{A_i\}$ seja uma sequência de conjuntos disjuntos que pertence a $\mathcal{J}$, e então pertence em cada σ-álgebra de $\mathcal{S}$. Assim $\cup A_i \in \mathcal{S}$ que também pertence a $\mathcal{J}$.

Portanto, $\mathcal{J}$ é uma σ-álgebra. $\square$

Contudo, a união de σ-álgebras não necessariamente é σ-álgebra.

Exemplo 17. *Seja o espaço amostral* $\Omega = \{1, 2, 3, 4\}$*, e σ-álgebras de Ω dadas por* $\mathcal{F}_1 = \{\emptyset, \Omega, \{1\}, \{2, 3, 4\}\}$ *e* $\mathcal{F}_2 = \{\emptyset, \Omega, \{4\}, \{1, 2, 3\}\}$*. Então*

$$\mathcal{F}_1 \cup \mathcal{F}_2 = \{\emptyset, \Omega, \{1\}, \{4\}, \{1, 2, 3\}, \{2, 3, 4\}\}\}$$

que não é uma σ-álgebra.

Corolário 1. *Seja ε uma classe de subconjuntos de Ω, sendo Ω um conjunto não vazio, que não necessariamente seja uma σ-álgebra. Existe então, uma σ-álgebra de Ω, denotada por $\mathcal{J}(\varepsilon)$, que contém ε que é a menor σ-álgebra, chamada a σ-álgebra gerada por ε.*

Demonstração. Seja $\mathcal{S}$ uma coleção de todas as σ-álgebras que inclui ε de Ω . Então $\mathcal{S}$ é não vazio, pois contém $\mathcal{P}(\Omega)$ que consiste de todos os subconjuntos de Ω. Pelo Teorema 3, a intersecão das σ-álgebras é uma σ-álgebra, $\mathcal{J}(\varepsilon)$, que inclui ε e está inclusa em todas as σ-álgebras de $\mathcal{S}$, isto é, $\mathcal{J}(\varepsilon)$ está contida em toda σ-álgebra de Ω que contém ε. Portanto, $\mathcal{J}(\varepsilon)$ é a menor σ-álgebra de Ω que contém ε. $\square$

Usaremos o colorário anterior para definir uma importante família das σ-álgebras, a chamada σ-álgebra de Borel.

Definição 15 (σ-álgebra de Borel)**.** *Seja ε uma coleção de subconjuntos abertos de $\mathbb{R}$. Então $\mathcal{J}(\varepsilon)$ é chamado de σ-álgebra de Borel de $\mathbb{R}$, usualmente escrito $\mathcal{B}(\mathbb{R})$. Seus elementos são chamados de conjuntos de Borel ou borelianos. Da mesma forma, definimos $\mathcal{B}(\mathbb{R}^n)$ como a σ-álgebra gerada pelos subconjuntos abertos de $\mathbb{R}^n$.*

Os elementos da σ-álgebra de Borel inclui os conjuntos abertos, conjuntos fechados (os complementares dos conjuntos abertos), interseções enumeráveis de conjuntos abertos (lembrando que uniões enumeráveis e conjuntos abertos já são abertos), uniões enumeráveis de conjuntos fechados (lembrando que interseções enumeráveis de conjuntos fechados já são fechados), etc., como será visto no teorema seguinte.

Teorema 4 (Conjuntos de $\mathcal{B}(\mathbb{R})$)**.** *Os subconjuntos seguintes de $\mathbb{R}$ pertencem a $\mathcal{B}(\mathbb{R})$:*

(i) (a, b) *para qualquer* $a < b$;

(ii) $(-\infty, a)$ *para qualquer* $a \in \mathbb{R}$;

(iii) (a, ∞) *para qualquer* $a \in \mathbb{R}$;

(iv) $[a, b]$ *para qualquer* $a \leq b$;

(v) $(-\infty, a]$ *para qualquer* $a \in \mathbb{R}$;

(vi) $[a, \infty)$ *para qualquer* $a \in \mathbb{R}$;

(vii) $(a, b]$ *para qualquer* $a < b$;

(viii) $[a, b)$ *para qualquer* $a < b$;

(ix) $\{x\} \in (a, b)$ *para qualquer* $a < b$;

(x) qualquer subconjunto fechado de $\mathbb{R}$.

Demonstração. Os itens (i), (ii) e (iii) são conjuntos abertos e portanto pertencem a $\mathcal{B}(\mathbb{R})$ pela própria definição.

(iv) $[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n}\right) \in \mathcal{B}(\mathbb{R})$;

(v) $(-\infty, a] = \bigcap_{n=1}^{\infty} \left(-\infty, a + \frac{1}{n}\right) \in \mathcal{B}(\mathbb{R})$;

(vi) $[a, \infty) = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, \infty\right) \in \mathcal{B}(\mathbb{R})$;

(vii) $(a, b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right) \in \mathcal{B}(\mathbb{R})$;

(viii) $[a, b) = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b\right) \in \mathcal{B}(\mathbb{R})$;

(ix) $\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x + \frac{1}{n}\right) \in \mathcal{B}(\mathbb{R}),\ \forall x \in (a, b)$;

(x) Se B é um subconjunto fechado em $\mathbb{R}$, então B^c é aberto, assim este pertence a $\mathcal{B}(\mathbb{R})$. Mas $B = (B^c)^c \in \mathcal{B}(\mathbb{R})$.

□

De fato, todas essas classes de subconjuntos de $\mathbb{R}$ gera a σ-álgebra $\mathcal{B}(\mathbb{R})$.

Teorema 5. *Seja $\mathcal{B}(\mathbb{R})$ uma σ-álgebra de Borel de $\mathbb{R}$. Então, esta é a menor σ-álgebra de $\mathbb{R}$ que inclui todos os intervalos.*

Demonstração. Seja $\mathcal{B}' \subseteq \mathcal{B}$ uma σ-álgebra tal que $\mathcal{B}'$ contém todos os intervalos. Isso implica que $\mathcal{B}'$ contém todos os intervalos abertos. Consequentemente, $\mathcal{B} \subseteq \mathcal{B}'$ pela Definição 15. Concluímos então que $\mathcal{B}' = \mathcal{B}$. □

Vamos definir formalmente, a probabilidade associada aos eventos aleatórios da σ-álgebra, da qual chamaremos de medida de probabilidade. Inicialmente, definimos

Definição 16 (Medida)**.** *Seja $\mathcal{F}$ uma σ-álgebra e um espaço amostral Ω, então uma medida, denotada por μ, é uma função tal que $\mu : \mathcal{F} \to [0, \infty)$, que satisfaz:*

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

ii) (σ-aditividade) Se $A_1, A_2, \ldots$, é uma sequência disjunta em $\mathcal{F}$, então $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$.

Para o caso em que $\mu : \mathcal{F} \to [0, 1]$, μ é chamado de medida de probabilidade e passa a ser denotado por P.

Exemplo 18. *Seja um espaço amostral Ω qualquer e uma σ-álgebra $\mathcal{F} = \mathcal{P}(\Omega)$, tal que uma medida $\mu : \mathcal{F} \to [0, \infty]$ é dado por:*

$$\mu(A) = \begin{cases} \#A, & \text{se } A \text{ é finito}, \\ \infty, & \text{se } A \text{ é infinito}. \end{cases}$$

Algumas condições são impostas sobre $\mu(A)$:

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

ii) $(A_i)_{i \in \mathbb{N}}$ *uma sequência disjunta em $\mathcal{F}$:*

(a) $\bigcup_{i\in\mathbb{N}} A_i$ *conjunto finito;*

(b) $\bigcup_{i\in\mathbb{N}} A_i$ *conjunto infinito.*

Para (ii).a temos, pela Definição 16.(ii) que

$$\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \#\bigcup_{i\in\mathbb{N}} A_i = \bigcup_{i\in\mathbb{N}} \#A_i$$
$$= \sum_{i=0}^{\infty} \mu(A_i).$$

Para o caso (ii).b, considerando $\bigcup_{i\in\mathbb{N}} A_i$ *um conjunto infinito e* A_i *um conjunto finito não vazio, para todo* $i \in \mathbb{N}$*, temos*

$$\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \infty = \sum_{i=1}^{\infty} \mu(A_i)$$
$$\sum_{i=1}^{\infty} \mu(A_i) = \sum_{i=1}^{\infty} \#A_i = \infty$$

Para o caso (ii).b, considerando $\bigcup_{i\in\mathbb{N}} A_i$ *um conjunto infinito e ao menos um* A_i *seja infinito, para todo* $i \in \mathbb{N}$*, temos*

$$\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \infty$$
$$\sum_{i=1}^{\infty} \mu(A_i) = \infty$$

Definição 17 (Medida de Probabilidade)**.** *Seja* Ω *o espaço amostral e* $\mathcal{F}$ *uma* σ*-álgebra dos eventos de* Ω*. Então, uma função* P *tal que* $P : \mathcal{F} \to [0, 1]$*, é chamada de medida de probabilidade sob os seguintes axiomas de Kolmogorov:*

1. *(Normalidade)* $P(\Omega) = 1$*;*

2. *(Positividade)* $\forall A \in \mathcal{F}$*,* $P(A) \geq 0$*;*

3. *(*σ*-aditividade) Para uma sequência finita ou infinita contável de eventos* $A_1, A_2, \ldots \in \mathcal{F}$ *multuamente exclusivos,*

$$P(U_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

□

Observe a seguinte propriedade:

3*) (Aditividade finita) se $A_1, A_2, \ldots, A_n$ é uma sequência disjunta dois a dois em $\mathcal{F}$, então $P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k)$.

Teorema 6. *O Axioma* 3 *implica o Axioma* 3^**, isto é, se* P *é* σ*-aditiva, então é finitamente aditiva.*

Demonstração. Supondo satisfeito o Axioma (3), e sejam $A_1, A_2, \ldots, A_n \in \mathcal{F}$. Considerando $P(\emptyset) = 0$, uma vez que

$$P(\Omega) = P(\Omega \cup \emptyset \cup \emptyset \cup \ldots) = P(\Omega) + P(\emptyset) + P(\emptyset) + \ldots.$$

Definimos $A_k = \emptyset$, para $k = n+1, n+2, \ldots$. Como $A_1, A_2, \ldots$ são disjuntos, então

$$\begin{aligned}
P\left(\bigcup_{k=1}^{n} A_k\right) &= P\left(\bigcup_{k=1}^{\infty} A_k\right)\\
&= \sum_{k=1}^{\infty} P(A_k)\\
&= \sum_{k=1}^{n} P(A_k) + P(\emptyset) + P(\emptyset) + \ldots\\
&= \sum_{k=1}^{n} P(A_k).
\end{aligned}$$

$\square$

Um quarto Axioma, pode ser complementado sobre a medida de probabilidade, que segue:

Axioma 4. ("Continuidade do vazio") Se a sequência $\{A_k\}_{k\geq 1}$, em que $A_k \in \mathcal{F}\ \forall k$, decrescer para o vazio, então $\lim_{k\to\infty} P(A_k) \to 0$.

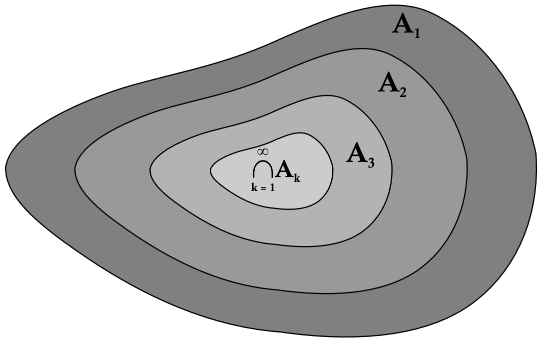
Este axioma indica que se $(A_k)_{k\geq 1}$ decrescer para o vazio, $A_k \downarrow \emptyset$, significa que $A_k \supset A_{k+1}\ \forall k$, ou seja, $(A_k)_{k\geq 1}$ decresce, e $\bigcap_{k\geq 1} A_k = \emptyset$.

Teorema 7 (Equivalência dos Axiomas 4 e 3*)**.** *Dados os axiomas de Kolmogorov, o Axioma 4 é equivalente ao Axioma 3*, isto é, uma probabilidade finitamente aditiva é uma probabilidade se, e somente se, é contínua no vazio.*

Demonstração. (i) Suponhamos o Axioma 4. Sejam $A_1, A_2, \ldots \in \mathcal{F}$ tais que $A_k \downarrow \emptyset$. Vamos provar que $P(A_k) \to 0$. Considere

$$A_1 = (A_1 - A_2) \cup (A_2 - A_3) \cup \ldots = \bigcup_{k=1}^{\infty} (A_k - A_{k+1}),$$

pelo diagrama:



As regiões $A_k - A_{k+1}$ são disjuntas, uma vez que a sequência é decrescente e $\mathcal{F}$ é fechada para diferenças. Pelo Axioma 3*,

$$P(A_1) = P\left(\bigcup_{k=1}^{\infty}(A_k - A_{k+1})\right) = \sum_{k=1}^{\infty} P(A_k - A_{k+1}),$$

portanto a série é convergente e

$$\begin{aligned}\sum_{k=1}^{\infty} P(A_k - A_{k+1}) &= \sum_{k=1}^{n-1} P(A_k - A_{k+1}) + P(\emptyset) + P(\emptyset) + P(\emptyset) + \ldots \\ &= \sum_{k=1}^{n-1} P(A_k - A_{k+1}) \underset{n\to\infty}{\to} P(A_1).\end{aligned}$$

Pela aditividade finita,

$$P(A_k - A_{k+1}) = P(A_k) - P(A_{k+1}),$$

logo

$$\begin{aligned}P(A_1) &= \lim_{n\to\infty} \sum_{k=1}^{n-1}[P(A_k) - P(A_{k+1})] \\ &= \lim_{n\to\infty} \{[P(A_1) - P(A_2)] + [P(A_2) - P(A_3)] + \ldots + [P(A_{n-1}) - P(A_n)]\} \\ &= \lim_{n\to\infty} \{P(A_1) - P(\cancel{A_2}) + P(\cancel{A_2}) - P(\cancel{A_3}) + \ldots + P(\cancel{A_{n-1}}) - P(A_n)\} \\ &= \lim_{n\to\infty} [P(A_1) - P(A_n)],\end{aligned}$$

e então $P(A_n) \to 0$.

(ii) Suponhamos o Axioma 4 e sejam $A_1, A_2, \ldots \in \mathcal{F}$ disjuntos. Vamos provar que $P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$. Seja $A = \bigcup_{k=1}^{\infty} A_k$, então

$$A = \left(\bigcup_{k=1}^{n} A_k\right) \cup \left(\bigcup_{k=n+1}^{\infty} A_k\right)$$

e pela aditividade finita,

$$P(A) = \sum_{k=1}^{n} P(A_k) + P\left(\bigcup_{k=n+1}^{\infty} A_k\right).$$

Seja $B_k = \bigcup_{k=n+1}^{\infty} A_k$, então $B_k \downarrow \emptyset$ e portanto $P(B_k) \to 0$ (Pelo Axioma 4). Logo

$$\sum_{k=1}^{n} P(A_k) \underset{n\to\infty}{\to} P(A),$$

isto é, $P(A) = \sum_{k=1}^{\infty} P(A_k)$. □

Corolário 2. *Os dois seguintes sistemas de axiomas são equivalentes:*

*Sistema I: Axiomas 1, 2 e 3**
Sistema II: Axiomas 1, 2, 3 e 4.

Demonstração. O sistema I é equivalente aos Axiomas 1, 2, 3 e 3^*, pois já vimos que o Axioma 3^* implica no Axioma 3. Agora, usando o Teorema 7, provamos que o Axioma 3 implica no Axioma 4, e a prova é concluída. □

Então para verificar se P é uma probabilidade em $\mathcal{F}$, basta verificar os axiomas do sistema I ou os axiomas do sistema II.

Vejamos algumas propriedades de uma medida de probabilidade,

Teorema 8 (Propriedades de P)**.** *Seja P uma medida de probabilidade associada a sequênia de eventos aleatórios $A_i \in \mathcal{F}$ $\forall i \in \mathbb{N}$ e o evento $B \in \mathcal{F}$, sendo $\mathcal{F}$ uma σ-álgebra, e $A \in \mathcal{F}$. Então são válidas as seguintes propriedades:*

i) (Complemento) $P(A) = 1 - P(A^c)$;

ii) $P(\emptyset) = 0$;

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

iv) (Monotonicidade) Se $A \subset B$, *então* $P(A) \leq P(B)$;

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

vi) (Limitante Superior) $0 \leq P(A) \leq 1$;

vii) (Desigualdade de Boole) $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$;

viii) (subaditividade) $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$;

ix) (Desigualdade de Bonferroni) $P(\bigcap_{i=1}^{n} A_i) \geq 1 - \sum_{i=1}^{n} P(A_i^c)$

x) (Continuidade da probabilidade) Se $A_i \downarrow A$, *então* $P(A_i) \downarrow P(A)$. *Se* $A_i \uparrow A$, *então* $P(A_i) \uparrow P(A)$.

xi) (Inclusão-Exclusão) $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j}^{n} P(A_i \cap A_j) + \sum_{i<j<k}^{n} P(A_i \cap A_j \cap A_k) + (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n)$.

□

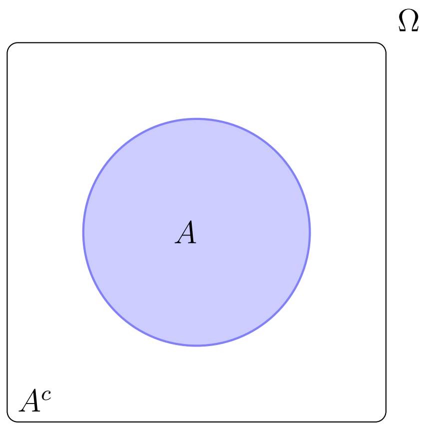
Demonstração. (Item i). Sabemos que $A \cup A^c = \Omega$ e que estes eventos são disjuntos. Veja o diagrama de Venn,

Assim,

$$\begin{aligned} P(A \cup A^c) &= P(A) + P(A^c) \\ 1 &= P(A) + P(A^c) \\ P(A) &= 1 - P(A^c). \end{aligned}$$

(Item ii). Sabemos que Ω e $\emptyset$ são eventos disjuntos. Assim,

$$\begin{aligned} P(\Omega \cup \emptyset) &= P(\Omega) + P(\emptyset) \\ 1 &= 1 + P(\emptyset) \\ P(\emptyset) &= 0. \end{aligned}$$



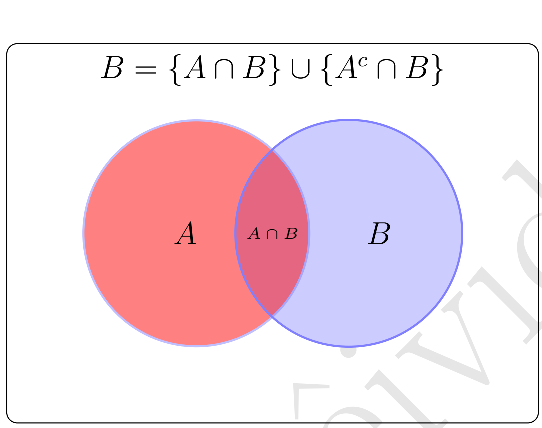


Figura 2: Diagrama de Venn para o evento $B = \{A \cap B\} \cup \{A^c \cap B\}$.

(Item iii). Podemos observar que $B = \{A \cap B\} \cup \{A^c \cap B\}$, e que $\{A \cap B\}$ e $\{A^c \cap B\}$ são disjuntos. Veja o diagrama de Venn na Figura 2.

Portanto,

$$\begin{aligned} P(B) &= P(\{A \cap B\} \cup \{A^c \cap B\}) \\ &= P(A \cap B) + P(A^c \cap B). \end{aligned}$$

(Item iv) Se $A \subset B$, então $A \cap B = A$. Usando (i), temos

$$\begin{aligned} P(B) &= P(A \cap B) + P(A^c \cap B) \\ &= P(A) + P(A^c \cap B) \geq P(A). \end{aligned}$$

. (Item v). Observando a seguinte identidade,

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

união de eventos disjuntos, que pode ser observado pela Figura 3, então

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

Por (iii) sabemos que

$$\begin{aligned} P(A) = P(B \cap A) + P(B^c \cap A) \Rightarrow \quad & P(B^c \cap A) = P(A) - P(B \cap A), \\ P(B) = P(A \cap B) + P(A^c \cap B) \Rightarrow \quad & P(A^c \cap B) = P(B) - P(A \cap B). \end{aligned}$$

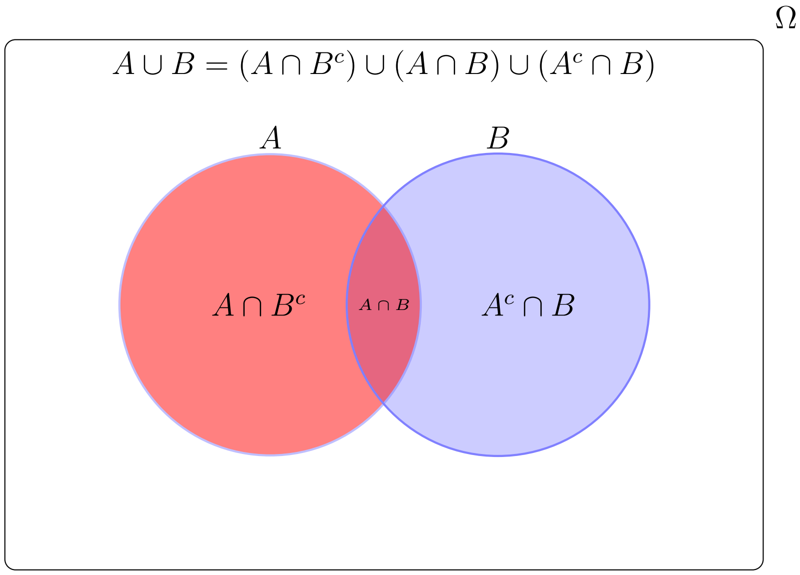


Figura 3: Diagrama de Venn para o evento $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$.

Logo

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

(Item vi). Pela Definição da medida de probabilidade P ser uma função $P : \mathcal{F} \to [0,1]$, logo se $A \in \mathcal{A}$, então $0 \leq P(A) \leq 1$.
(Item vii).Vamos inicialmente criar uma sequência disjunta $A_1^*, A_2^*, \ldots$, com a propriedade $\bigcup_{i=1}^{\infty} A_i^* = \bigcup_{i=1}^{\infty} A_i$. Definimos

$$A_1^* = A_1, \qquad A_i^* = A_i \cap \left(\bigcup_{j=1}^{i-1} A_j\right)^c, \quad i = 2, 3, \ldots.$$

É fácil perceber que

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i^*\right) = \sum_{i=1}^{\infty} P(A_i^*),$$

onde a última igualdade segue, pois A_i^* são disjuntos. Observemos que pela construção, $A_i^* \subseteq A_i$, portanto, pela propriedade (iv), $P(A_i^*) \leq P(A_i)$, logo

$$\sum_{i=1}^{\infty} P(A_i^*) \leq \sum_{i=1}^{\infty} P(A_i).$$

Concluindo a prova,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i^*) \;\leq\; \sum_{i=1}^{\infty} P(A_i).$$

(item $viii$). Como vimos pela Definição 17 que a σ-aditividade implica na aditividade finita, logo o item (vii) implica em $(viii)$.
(item ix).

(item x). Vamos supor que $A_i \downarrow A$, isto é, $A_i \supset A_{i+1}$ e $\bigcap_{i\geq 1} A_i = A$. Então $P(A_i) \geq P(A_{i+1})$ pelo item (iv), e $(A_i - A) \downarrow \emptyset \Rightarrow P(A_i - A) \to 0$ pela continuidade do vazio. Pela aditividade finita $P(A_i - A) = P(A_i) - P(A)$, e como $\{P(A_i)\}_{i\in\mathbb{N}}$ é decrescente, logo $P(A_i) \downarrow P(A)$. Agora, se $A_i \uparrow A$, isto é, $A_i \subset A_{i+1}$ e $\bigcup_{i\geq 1} A_i = A$, então $A_i^c \downarrow A^c$, logo $P(A_i^c) \downarrow P(A^c) \Rightarrow 1 - P(A_i) \downarrow 1 - P(A)$. Portanto, $P(A_i) \uparrow P(A)$.
(item xi). Associar a prova aos diagramas de Venn abaixo. Vamos apresentar duas provas.

- Primeiro para $n = 2$, temos a propriedade (iv) do Teorema 8, isto é,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \tag{2}$$

- Para $n = 3$, vamos considerar $A = \{A_1 \cup A_2\}$, e que

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A \cup A_3) \\ &= P(A) + P(A_3) - P(A \cap A_3). \end{aligned} \tag{3}$$

  Segue que $P(A) = P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$ e que

$$\begin{aligned} P(A \cap A_3) &= P[(A_1 \cup A_2) \cap A_3] \\ &= P[(A_1 \cap A_3) \cup (A_2 \cap A_3)] \\ &= P(A_1 \cap A_3) + P(A_2 \cap A_3) - P(A_1 \cap A_2 \cap A_3) \end{aligned} \tag{4}$$

  Substituindo as expressões (2) e (4) em (3), logo

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + \\ + P(A_1 \cap A_2 \cap A_3). \end{aligned}$$

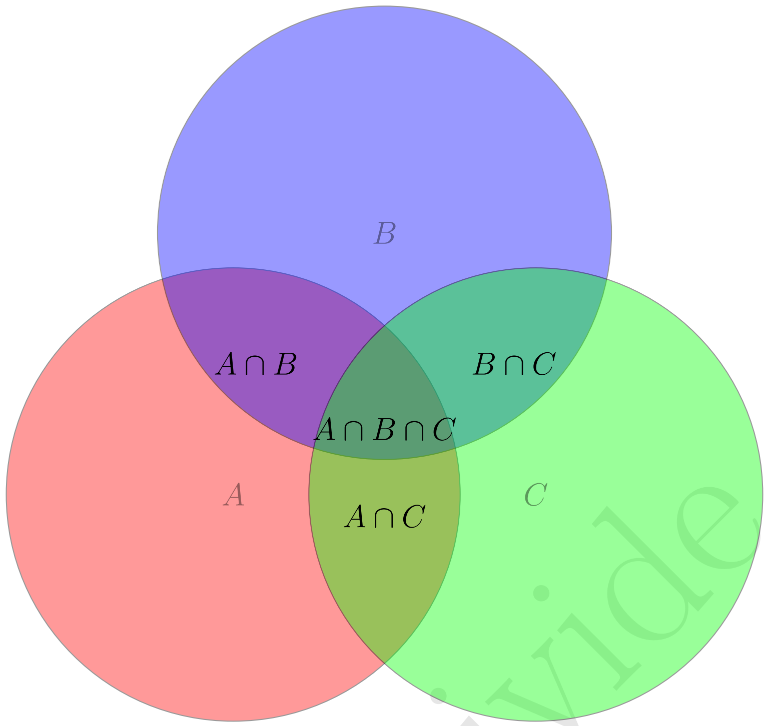
- Para $n = 4$, já apresentando o resultado direto, temos

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup A_4) = \sum_{i=1}^{4} P(A_i) - \sum_{i<j}^{4} P(A_i \cap A_j) + \sum_{i<j<k}^{4} P(A_i \cap A_j \cap A_k) - \\ - P(\bigcap_{i=1}^{4} A_i). \end{aligned} \tag{5}$$

- Para $n = 5$, já apresentando o resultado direto, temos

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5) = \sum_{i=1}^{5} P(A_i) - \sum_{i<j}^{5} P(A_i \cap A_j) + \sum_{i<j<k}^{5} P(A_i \cap A_j \cap A_k) - \\ - \sum_{i<j<k<l}^{5} P(A_i \cap A_j \cap A_k \cap A_l) + P(\bigcap_{i=1}^{5} A_i). \end{aligned} \tag{6}$$

  Observe que a última expressão para a união das probabilidades alterna o sinal à medida que n aumenta, isto é, quando n é positivo, a interseção de todos os eventos



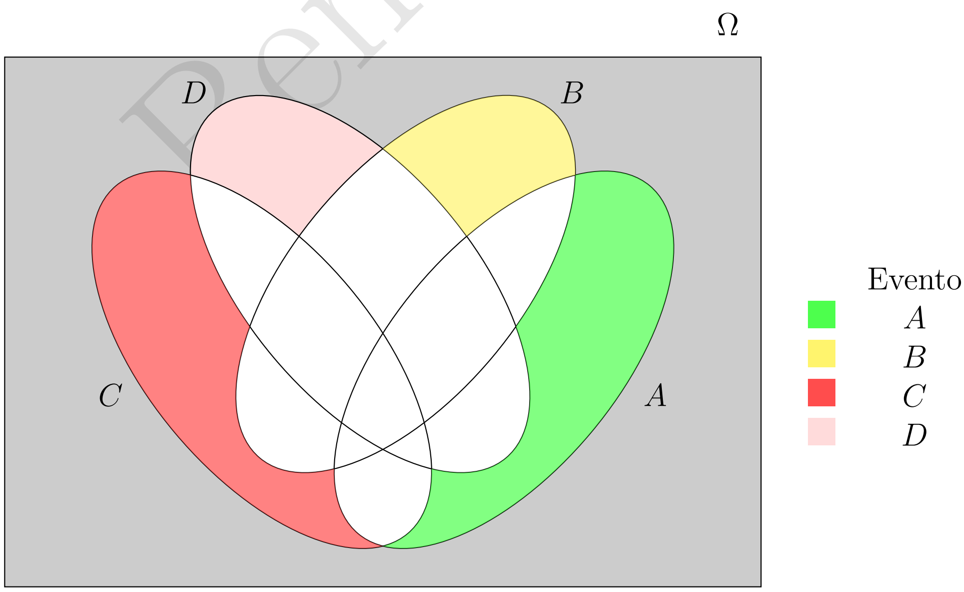


Figura 4: Diagrama de Venn 4 eventos.

é eliminada. Quando n é ímpar, a interseção de todos os eventos é somada. Assim, para um n qualquer induzindo pela expressão (6), temos

$$
\begin{aligned}
P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j}^{n} P(A_i \cap A_j) + \sum_{i<j<k}^{n} P(A_i \cap A_j \cap A_k) - \ldots \\
\ldots (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n).
\end{aligned}
$$

Para uma segunda demonstração, vamos provar para $n-1$ eventos e induzir para n. Considere a probabilidade da união de eventos $\{A_i\}_{i=1}^{n}$, isto é,

$$
\begin{aligned}
P(A_1 \cup A_2 \cup \ldots \cup A_n) &= P\left(\bigcup_{i=1}^{n-1} A_i \cup A_n\right) \\
&= P\left(\bigcup_{i=1}^{n-1} A_i\right) + P(A_n) - P\left(\bigcup_{i=1}^{n-1} A_i \cap A_n\right) \\
&= \left[\sum_{i=1}^{n-1} P(A_i) - \sum_{i<j}^{n-1} P(A_i \cap A_j) + \sum_{i<j<k}^{n-1} P(A_i \cap A_j \cap A_k) - \ldots \right. \\
&\quad \left. \ldots + (-1)^{(n-1)+1} P\left(\bigcap_{i=1}^{n} A_i\right)\right] + P(A_n) - P\left(\bigcup_{i=1}^{n-1} A_i \cap A_n\right).
\end{aligned}
\tag{7}
$$

Observe que

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{n-1} A_i \cap A_n\right) &= \sum_{i=1}^{n-1} P(A_i \cap A_n) - \sum_{i<j}^{n-1} P[(A_i \cap A_n) \cap (A_j \cap A_n)] + \\
&\quad + \sum_{i<j<k}^{n-1} P[(A_i \cap A_n) \cap (A_j \cap A_n) \cap (A_k \cap A_n)] - \ldots \\
&\quad \ldots + (-1)^{(n-1)+1} P\left(\bigcap_{i=1}^{n-1} (A_i \cap A_n)\right) \\
&= \sum_{i=1}^{n-1} P(A_i \cap A_n) - \sum_{i<j}^{n-1} P(A_i \cap A_j \cap A_n) + \\
&\quad + \sum_{i<j<k}^{n-1} P(A_i \cap A_j \cap A_k \cap A_n) - \ldots + \\
&\quad (-1)^n \sum_{i_1<i_2<\ldots<i_{n-2}}^{n-1} P(A_{i_1} \cap \ldots \cap A_{i_{n-2}} \cap A_n) \\
&\quad \ldots + (-1)^{(n-1)+1} P\left(\bigcap_{i=1}^{n} A_i\right).
\end{aligned}
\tag{8}
$$

Observe a expressão $P\left(\bigcup_{i=1}^{n-1} A_i \cap A_n\right)$ recebe o sinal negativo, portanto iremos (8) em (7), levando em consideração a substituição do sinal, que segue

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = \left[\sum_{i=1}^{n-1} P(A_i) - \sum_{i<j}^{n-1} P(A_i \cap A_j) + \sum_{i<j<k}^{n-1} P(A_i \cap A_j \cap A_k) - \ldots \right.$$
$$\left. \ldots + (-1)^{(n-1)+1} P\left(\bigcap_{i=1}^{n} A_i\right)\right] + P(A_n) - [\,]. \qquad (9)$$

□

Vejamos mais algumas propriedades da probabilidade em uma sequência de eventos. (ACRESCENTAR)

Agora diante dessas definições, definimos

Definição 18 (Espaço de probabilidade)**.** *Seja Ω o espaço amostral e $\mathcal{F}$ uma coleção de subconjuntos de Ω que pode ser atribuído probabilidade. Se P é uma função que mede a probabilidade dos eventos em $\mathcal{F}$, então a tripla $(\Omega, \mathcal{F}, P)$ é chamada de espaço de probabilidade.* □

Dizemos também que $(\Omega, \mathcal{F})$ é o espaço mensurável. As próximas Definições e Teoremas iremos apresentar com base em uma motivação:

Motivação 1. *Paulo é um jovem empreendedor e quer abrir seu próprio negócio. Ele observou que o mercado de sandálias era lucrativo. Então resolveu abrir uma fábrica de sandálias. Devido a dificuldade financeira, resolveu comprar três máquinas de sandálias usadas. As informações anteriores sobre estas máquinas dadas pelo proprietário foram:*

Máquina	**Produto**	**Total da produção**	**Produto com defeito**
M1	*Pantufas*	*50%*	*1%*
M2	*Sandálias baixas*	*40%*	*2%*
M3	*Sandálias de couro*	*10%*	*3%*

Surgiu as seguintes indagações:

- *Do total de sandálias produzidas, qual a probabilidade de Paulo produzir uma sandália com defeito?*
- *Pensando em aumentar o lucro da fábrica, Paulo pensa e substituir uma das máquinas, qual seria sua decisão?*
  - *Será que a máquina M1 que produz mais sandálias e consequentemente tem maior desgaste, deve ser trocada primeiro?*
  - *Ou será que apesar da máquina M3 ter menor produção, é a que gera mais defeito por sandália, e portanto, deve ser trocada primeiro?*

Muitas vezes nos deparamos com situações em que antes da realização de algum experimento descrito em $(\Omega, \mathcal{F}, P)$, temos alguma informação adicional. Queremos saber o quanto que essa informação pode afetar a medida de probabilidade dos eventos de $\mathcal{F}$.

Definição 19 (Probabilidade condicional)**.** *Seja um espaço de probabilidade $(\Omega, \mathcal{F}, P)$. Dados dois eventos A e B definidos em $\mathcal{F}$, então a probabilidade condicional do evento A dado que ocorreu o evento B, denotado por $P(A|B)$, é definida por:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad (10)$$

para $P(B) > 0$. □

Com base no problema de Paulo, denote o evento D as sandálias produzidas com defeitos, M_1 o evento que representa as sandálias produzidas pela máquina M1, M_2 o evento que representa as sandálias produzidas pela máquina M2 e M_3 o evento que representa as sandálias produzidas pela máquina M3. Assim,

$$P(D|M_1) = 0,01,$$
$$P(D|M_2) = 0,02,$$
$$P(D|M_3) = 0,03.$$

Teorema 9 ($P(A|B)$ é uma medida de probabilidade)**.** *Sejam A e B eventos no espaço de probabilidade $(\Omega, \mathcal{F}, P)$, tal que $P(B) > 0$, e considere $Q : \mathcal{F} \to [0,1]$ definida por*

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

que é a probabilidade condicional de A dado B. Então $(\Omega, \mathcal{F}, Q)$ é também um espaço de probabilidade. □

Demonstração. Verificar que $(\Omega, \mathcal{F}, Q)$ é um espaço de probabilidade, é o mesmo que dizer que Q é uma medida de probabilidade sobre $(\Omega, \mathcal{F})$, isto é, Q satisfaz os axiomas de Kolmogorov.

- Axioma 1:

$$\begin{aligned} Q(\Omega) &= \frac{P(\Omega \cap B)}{P(B)} \\ &= \frac{P(B)}{P(B)} \\ &= 1. \end{aligned}$$

- Axioma 2:
  Como P é uma medida de probabilidade, então

$$\forall A \in \Omega : \quad Q(A) \geq 0.$$

- Axioma 3: Seja uma sequência A_1, A_2, $\ldots$, disjunta, então

$$\begin{aligned} Q(\bigcup_{i=1}^{\infty} A_i) &= \frac{P(\bigcup_{i=1}^{\infty} A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} Q(A_i). \end{aligned}$$

  Prova concluída.

$\square$

Teorema 10 (Regra do produto de probabilidade)**.** *Seja os eventos $A_1, A_2, \ldots, A_n$ em $(\Omega, \mathcal{F}, P)$, com $P(\bigcap_{i=1}^{n} A_i) > 0$, então a probabilidade do produto desses eventos é dado por*

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

Demonstração. Por indução, consideremos $n = 2$. Assim, pela Definição 19, temos

$$\begin{aligned} P(A_2|A_1) &= \frac{P(A_1 \cap A_2)}{P(A_1)} \\ P(A_1 \cap A_2) &= P(A_1)P(A_2|A_1), \end{aligned}$$

pois $P(A_1) > 0$. Agora para $n = k$, generalizamos a indução,

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) \;\; = \;\; P[(A_1A_2 \ldots A_{k-1}) \cap A_k],$$

pela definição 19, temos

$$P[(A_1A_2 \ldots A_{k-1}) \cap A_k] \;\; = \;\; P[(A_1A_2 \ldots A_{k-1})]P(A_k|A_1A_2 \ldots A_{k-1})$$

podendo ser reescrito como

$$\begin{aligned} P[(A_1A_2 \ldots A_{k-1}) \cap A_k] \;\; = \;\; & \underbrace{P(A_1A_2 \ldots A_{k-2})P(A_{k-1}|A_1A_2 \ldots A_{k-2})}_{P(A_1A_2 \ldots A_{k-1})} \times \\ & \times P(A_k|A_1A_2 \ldots A_{k-1}). \end{aligned}$$

Assim, usando a indução sucessivas vezes, chegaremos a expressão

$$\begin{aligned} P(A_1A_2 \ldots \cap A_k) \;\; = \;\; & P(A_1)P(A_2|A_1)\ldots P(A_{k-1}|A_1A_2 \ldots A_{k-2}) \times \\ & \times P(A_k|A_1A_2 \ldots A_{k-1}). \end{aligned}$$

Observe que por hipótese, todos os condicionamentos da expressão do lado direito, têm probabilidades positivas, pois contém $\bigcap_{i=1}^{n} A_i$. Portanto, teorema provado. $\square$

Antes de falarmos sobre o teorema da lei da probabilidade total, será interessante fazer a definição sobre a partição de Ω.

Definição 20 (Partição de Ω)**.** *Se a sequência $A_1, A_2, \ldots,$ são disjuntos dois a dois e $\bigcup_{i=1}^{\infty} A_i = \Omega$, então dizemos que essa sequência forma uma partição de Ω.*

Entretanto, para calcular a probabilidade de uma sandália está com defeito, isto é $P(D)$, apresentamos o seguinte Teorema,

Teorema 11 (Teorema da probabilidade total)**.** *Seja o espaço de probabilidade $(\Omega, \mathcal{F}, P)$. Considere uma sequência de eventos $A_1, A_2, \ldots, A_n$ de $\mathcal{F}$, disjuntos, tal que $\bigcup_{i=1}^{n} A_i = \Omega$, e B um evento de $\mathcal{F}$, então a probabilidade de B é dada por:*

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i), \tag{11}$$

para $P(A_i) > 0$, sendo $i = 1, 2, \ldots, n$. $\square$

Demonstração. Sabendo que $A_i \cap_{i \neq j} A_j$ e que $\bigcup_{i=1}^n B \cap A_i = \Omega$, então $\bigcup_{i=1}^n B \cap A_i = B$ e os $B \cap A_i$ são também disjuntos. Dessa forma,

$$\begin{aligned} P(B) =& P\left(\bigcup_{i=1}^n B \cap A_i\right) \\ =& \sum_{i=1}^n P(B \cap A_i) \qquad (B \cap A_i \text{ disjuntos}) \\ =& \sum_{i=1}^n P(B|A_i)P(A_i) \qquad (\text{Teorema 10}) \end{aligned}$$

como queríamos provar. □

Exemplo 19. *Voltando ao problema de Paulo, como* $P(M_1) = 0,50$*,* $P(M_2) = 0,40)$ *e* $P(M_3) = 0,10$*, então a probabilidade de uma sandália ter defeito é*

$$\begin{aligned} P(D) =& \sum_{i=1}^3 P(D|M_i)P(M_i) \\ =& P(D|M_1)P(M_1) + P(D|M_2)P(M_2) + P(D|M_3)P(M_3) \\ =& 0,01 \times 0,50 + 0,02 \times 0,40 + 0,03 \times 0,10 \\ =& 0,016. \end{aligned}$$

Uma outra Definição interessante é a independência de eventos, apresentada a seguir.

Definição 21 (Independência de dois eventos)**.** *Seja um espaço de probabilidade* $(\Omega, \mathcal{F}, P)$*. Dois eventos* A *e* B *de* $\mathcal{F}$ *são independentes se satisfaz ao menos uma das seguintes condições:*

I) $P(A \cap B) = P(A)P(B)$*;*

II) $P(A|B) = P(A)$*, para* $P(B) > 0$*;*

III) $P(B|A) = P(B)$*, para* $P(A) > 0$*.*

□

É fácil mostrar que (I) implica em (II), (II) implica em (III), e (III) implica em (I). $(i) \to (ii)$: Se $P(AB) = P(A)P(B)$, então

$$P(A|B) = \frac{P(AB)}{P(A)} = \frac{P(A)P(B)}{P(B)} = P(A), \quad \text{para } P(B) > 0;$$

$(ii) \to (iii)$: Se $P(A|B) = P(A)$, então

$$P(B|A) = \frac{P(BA)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = P(B), \quad \text{para } P(A) > 0;$$

$(iii) \to (i)$: Se $P(B|A) = P(B)$, então

$$P(AB) = P(B|A)P(A) = P(B)P(A), \text{para } P(A) > 0.$$

A intuição para independência na Definição 21 fica justificada pelo fato de que A é independente de B tanto na ocorrência quanto a não ocorrência de B e isso não muda em nada a probabilidade da ocorrência de A, isto é, $P(A|B) = P(A)$ e $P(A|B^c) = P(A)$. Essas duas expressões significam que

$$\begin{aligned} P(A \cap B) &= P(B)P(A|B) = P(B)P(A) \\ P(A \cap B^c) &= P(B^c)P(A|B) = P(B^c)P(A). \end{aligned}$$

Entretanto, a independência entre dois eventos não implica em independência coletiva. Vejamos,

Exemplo 20. *Seja um experimento cujo objetivo é verificar a face superior de um tetraedro, isto é,* $\Omega = \{1, 2, 3, 4\}$. *Sejam os eventos em* Ω, $A = \{1, 4\}$, $B = \{2, 4\}$ *e* $C = \{3, 4\}$. *Considerando o tetraedro honesto e que cada valor é equiprovável, assim* $P(A) = P(B) = P(C) = 1/2$. *Observamos que estes eventos são independentes dois a dois, isto é,* $P(A \cap B) = 1/4 = P(A)P(B)$, $P(A \cap C) = 1/4 = P(A)P(C)$ *e* $P(B \cap C) = 1/4 = P(B)P(C)$. *Porém,* $P(A \cap B \cap C) = 1/4 \neq P(A)P(B)P(C)$. *Logo, os eventos* A, B *e* C *não são independentes três a três.*

Para uma definição mais geral, temos

Definição 22 (Independência de eventos)**.** *Seja um espaço de probabilidade* $(\Omega, \mathcal{F}, P)$. *Uma sequência de eventos* $A_1, A_2, \ldots, A_n$ *de* $\mathcal{F}$ *são independentes se e somente se:*

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j), \quad \textit{para } i \neq j; \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k), \quad \textit{para } i \neq j \neq k; \\ &\vdots \\ P(\cap_{i=1}^n A_i) &= \prod_{i=1}^n P(A_i). \end{aligned} \tag{12}$$

□

Devemos deixar claro que a Definição 22 implica na Definição 21, mas não o contrário, isto é, a independência a pares não implica em independência coletiva. Vejamos o exemplo.

Exemplo 21. *Considere um experimento de um arremesso de um tetraedro honesto, cuja suas faces resultam nos números 1, 2, 3 e 4. Sejam os eventos* $A = \{1, 4\}$, $B = 2, 4$ *e* $C = \{3, 4\}$. *Considerando equiprovável o valor de se obter um dos números do tetaedro na face superior, temos que qualquer valor terá probablidade 1/4. Dessa forma, a probabilidade* $P(A) = P(B) = P(C) = 1/2$ *e* $P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/4$. *Logo,* A, B *e* C *são independentes dois a dois. Mas,*

$$P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C).$$

Exemplo 22. *No lançamento de dois dados honestos, sejam os eventos* A = *"face ímpar no primeiro dado",* B = *"face ímpar no segundo dado" e* C = *"'soma ímpar das duas faces". É fácil ver que* A, B *e* C, *têm, cada um, probabilidade 1/2 e são independentes dois a dois. Mas, eles não podem ser simultaneamente, de modo que* $A \cap B \cap C = \emptyset$ *e*

$$P(A \cap B \cap C) = 0 \neq \frac{1}{8} = P(A)P(B)P(C).$$

Retornando a Motivação 1, Paulo poderia indagar, os M_i e D são independentes ou dependentes? Pela Definição 21, temos que

$$P(D|M_i) \neq P(D) = 0,016 \Rightarrow D \text{ e } M_i \text{ não são independentes, para } i = 1, 2, 3.$$

A grande questão agora é qual a máquina que Paulo deveria substituir com o propósito de aumentar seu lucro na empresa. A ideia será calcular $P(M_i|D)$, isto é, dado um defeito na sandália qual a probabilidade de vindo da máquina i? A maior probabilidade será a máquina substituída. Entretanto, ainda não temos ferramenta para resolver essa resposta. Para isso, apresentamos o seguinte Teorema:

Teorema 12 (Teorema de Bayes)**.** *Seja o espaço de probabilidade* $(\Omega, \mathcal{F}, P)$*. Considere uma sequência de eventos* $A_1, A_2, \ldots, A_n$ *de* $\mathcal{F}$*, disjuntos, tal que* $\bigcup_{i=1}^n A_i = \Omega$*, e* B *um evento de* $\mathcal{F}$*, então a probabilidade de* A_k*, para* $k = 1, 2, \ldots, n$*, dado que ocorreu o evento* B*, denotado por* $P(A_k|B)$*, é dado por:*

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^n P(B|A_i)P(A_i)}, \qquad k = 1, 2, \ldots, n, \tag{13}$$

para $P(A_k) > 0$ *e* $P(A_i) > 0$*, sendo* $i = 1, 2, \ldots, n$*.* □

Demonstração. Para um i qualquer, temos

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

ou

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

Isto implica que

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

Pela lei da probabilidade total, a probabilidade de B pode ser dada por $P(B) = \sum_{i=1}^{\infty} P(A_i)\ P(B|A_i)$. Portanto,

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{\infty} P(A_i)P(B|A_i)},$$

prova concluída. □

Tal é a sua importância, que um dos ramos de estudo da inferência estatística é baseado nesse teorema. O Teorema de Bayes fornece uma atualização do conhecimento já existente $P(A_k)$, conhecido como "*a priori*", por meio da ocorrência do evento B. Essa atualização é a probabilidade "*a posteriori*" $P(A_k|B)$.

Com esse resultado, Paulo agora pode tomar uma decisão mais plausível, isto é, dado um defeito numa determinada sandália produzida na fábrica, qual a probabilidade desta ter sido produzida em cada uma das máquinas?

$$P(M_1|D) = \frac{0,01 \times 0,50}{0,016} = 0,3125$$
$$P(M_2|D) = \frac{0,02 \times 0,40}{0,016} = 0,5000$$
$$P(M_3|D) = \frac{0,03 \times 0,10}{0,016} = 0,1875$$

A tomada de decisão será substituir a máquina M2. Poderíamos ter tomado uma decisão equivocada se não fosse o teorema de Bayes.

Devemos abrir uma discussão para que ocorre muito frequente entre as Definições 8 e 22, isto é, eventos disjuntos e independência. Nas próprias definições percebemos a distinção clara entre as características. A primeira se remete a eventos (conjuntos), e a segunda é uma condição probabilística dos eventos. Contudo, em determinados problemas ainda há muita confusão ao tentar resolvê-los. Assim, apresentemos os seguintes teoremas,

Teorema 13 (Eventos disjuntos e independentes)**.** *Considere A e B, dois eventos no espaço de probabilidade $(\Omega, \mathcal{F}, P)$. Se $A \cap B = \emptyset$ (eventos disjuntos), então A e B são independentes apenas, se e somente se, um dos eventos tiver probabilidade 0.*

Demonstração. Considerando que o evento A tenha probabilidade 0, isto é, $P(A) = 0$, implica que $A = \emptyset$, pelo Teorema 8. Assim, $P(A \cap B) = P(\emptyset \cap B) = P(\emptyset) = 0$. A condição de independência entre os dois eventos existe se $P(A)P(B) = P(A \cap B)$, e isso ocorre de fato, $P(A)P(B) = P(\emptyset)P(B) = 0 \times P(B) = 0 = P(A \cap B)$, o que completa a prova. □

Caso esses eventos não tenham probabilidade 0, a condição $A \cap B = \emptyset$ não implica em independência entre os eventos. Vejamos o seguinte exemplo.

Exemplo 23 (Eventos disjuntos não implica em independência de eventos)**.** *A tabela abaixo dá a distribuição das probabilidade dos quatro tipos sanguíneos, numa certa comunidade.*

Tipo sanguíneo	*A*	*B*	*AB*	*O*
Probabilidade de ter o tipo especificado	*0,2*			
Probabilidade de não ter o tipo especificado		*0,9*	*0,95*	

Calcular a probabilidade de que:

a) um indivíduo, sorteado ao acaso nessa comunidade, tenha o tipo O;

b) dois indivíduos, sorteados ao acaso nessa comunidade, tenham tipo A e tipo B, nessa ordem;

c) um indivíduo, sorteado ao acaso nessa comunidade, não tenha o tipo B ou não tenha o tipo AB.

Vejamos que os tipos sanguíneos são multuamente exclusivos e formam a partição do espaço amostral, uma vez que não existe outro tipo sanguineo além dos informados e que não há indivíduo com dois tipos sanguíneos. Assim,

a) $P(\Omega) = P(A) + P(B) + P(AB) + P(O) \Rightarrow 1 = 0,2000 + 0,1000 + 0,0500 + P(O) \Rightarrow P(O) = 0,6500.$

b) Como os eventos A e B são independentes, então $P(A \cap B) = P(A) \times P(B) = 0,2000 \times 0,1000 = 0,0200.$

c) Agora os eventos "não ter o tipo sanguíneo especificado" não implica que os eventos sejam multuamente exclusivos pelo fato dos eventos "ter o tipo sanguíneo especificado" terem sido disjuntos. Veja, o evento não ter o tipo sanguínio AB e o evento não ter o tipo sanguíneo B, pode existir indivíduos comum a estes dois eventos, por exemplo, um indivíduo do tipo sanguíneo A ou O, e a probabilidade destes não é zero, logo, os eventos não ter o tipo sanguínio AB e não ter o tipo sanguíneo B não são disjuntos. Entretanto, esses eventos são independentes, pois a probabilidade de um evento não influencia na probabilidade do outro. Assim,

$$
\begin{aligned}
P[(AB)^c \cup B^c] &= P[(AB)^c] + P(B^c) - P[(AB)^c]P(B^c) \quad \textit{(Independência)} \\
&= 0,9000 + 0,9500 - 0,9000 \times 0,9500 = 0,9950 \approx 1.
\end{aligned}
$$

Ao final, temos a tabela completada da seguinte forma:

Tipo Sanguíneo	*A*	*B*	*AB*	*O*
Prob. tipo esp.	*0,20*	*0,10*	*0,05*	*0,65*
Prob. não tipo esp	*0,80*	*0,90*	*0,95*	*0,35*

Podemos ainda expressar mais três teoremas para complementar as afirmações feitas no Exemplo 23.

Teorema 14. *Se A e B são eventos independentes no espaço de probabilidade* $(\Omega, \mathcal{F}, P)$*, então*

a) A *e* B^c *também são independentes;*

b) A^c *e* B *também são independentes;*

c) A^c *e* B^c *também são independentes.*

Demonstração. Usando as seguintes equivalências:

$$P(A) = P(A \cap B) + P(A \cap B^c) \tag{14}$$
$$P(A^c) = P(A^c \cap B) + P(A^c \cap B^c) \tag{15}$$
$$P(B) = P(B \cap A) + P(B \cap A^c) \tag{16}$$
$$P(B^c) = P(B^c \cap A) + P(B^c \cap A^c) \tag{17}$$

e a condição de que $P(A \cap B) = P(A)P(B)$ (independentes), então usando (14) temos

$$
\begin{aligned}
P(A \cap B^c) &= P(A) - P(A)P(B) \quad \text{(Independência)} \\
&= P(A)[1 - P(B)] = P(A)P(B^c),
\end{aligned}
$$

o que prova o ítem (a). Usando (16) pelo mesmo raciocínio, provamos o ítem (b). Usando o resultado do ítem (a), já provado, e a condição de independência na expressão (17), temos

$$
\begin{aligned}
P(A^c \cap B^c) &= P(B^c) - P(B^c)P(A) \\
&= P(B^c)[1 - P(A)] = P(B^c)P(A^c),
\end{aligned}
$$

o que prova o ítem (c), concluindo assim, a prova do teorema. □