

Notas de Aulas - MI401

Probabilidade

Caio Gomes Alves

03/2025

Conteúdos

1	Prefácio	2
2	Definições Básicas	3
2.1	Modelo Probabilístico	3
2.2	Álgebras de Conjuntos	4
2.3	Axiomas de Kolmogorov	6
2.4	Propriedades da medida de probabilidade	6
2.5	Propriedades de probabilidade	8
2.6	Probabilidade Condicional e Independência	8
2.7	Fórmula de Poincaré e Teorema de Bayes	10
2.8	Exercícios	12
3	Variáveis Aleatórias	29
3.1	Variáveis aleatórias e funções de distribuição	29
3.2	Natureza das variáveis aleatórias	31
3.3	Variáveis aleatórias e σ -álgebra de Borel	33
3.4	Variáveis contínuas	34
3.5	Variáveis aleatórias multidimensionais	35
3.6	Independência	37
3.7	Distribuições de funções de vetores	39

1 Prefácio

Este “livro” consiste de notas de aulas da matéria MI401 - Probabilidade, do Programa de Mestrado em Estatística, do Instituto de Matemática, Estatística e Computação Científica (IMECC), da Universidade Estadual de Campinas - UNICAMP.

Os seus conteúdos são baseados nas notas tomadas durante as aulas, do livro “Probabilidade e Variáveis Aleatórias”, do autor Marcos Nascimento Magalhães, 3ª edição, e do livro “Probabilidade: um curso em nível intermediário”, do autor Barry R. James, 5ª edição.

2 Definições Básicas

2.1 Modelo Probabilístico

Suponha que é realizado um experimento “sob certas condições”, sendo Ω o conjunto de resultados possíveis do experimento (também chamado de resultados elementares). Chamamos Ω de **espaço amostral do experimento**, com a representação axiomática sendo dada por: $\Omega = \{\omega : \omega \in \Omega\}$.

Example 2.1. Considere o lançamento de um dado honesto. Nesse caso, temos que $\Omega = \{1, 2, 3, 4, 5, 6\}$, em que cada $\{i\}$ é um evento elementar, sendo eles $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$ e $\{6\}$.

Temos então que eventos são coleções de pontos em Ω , por exemplo um evento $A = \{2, 4, 6\}$ (números pares no lançamento de um dado honesto). Assim, temos as seguintes suposições para eventos:

1. Todo resultado possível no experimento corresponde a um e somente um $\omega \in \Omega$;
2. Resultados diferentes correspondem a elementos diferentes em Ω .

Definition 2.1. Seja um espaço amostral Ω de um experimento. Todo subconjunto $A \subset \Omega$ é um evento. Ω é o evento certo e \emptyset é o evento impossível. Além disso, $\omega \in \Omega \rightarrow \{\omega\}$ é um evento elementar.

Note-se que, dados A e B eventos, tais que $A \subset \Omega$ e $B \subset \Omega$, temos que:

- $A \cup B \rightarrow (\omega \in A \text{ e } \omega \notin B) \text{ ou } (\omega \notin A \text{ e } \omega \in B) \text{ ou } (\omega \in A \text{ e } \omega \in B)$;
- $A \cap B \rightarrow (\omega \in A \cup \omega \in B)$;
- $A^c \rightarrow (\omega \notin A)$;
- $A \subset B \rightarrow$ a ocorrência de A implica a ocorrência de B ;
- $A \cap B = \emptyset \rightarrow$ os eventos A e B são mutuamente exclusivos.

No campo probabilístico, pensamos em atribuir probabilidades (leia-se chances) a eventos em Ω .

Definition 2.2 (Clássica). A probabilidade de ocorrência de um evento A , denotada por $P(A)$ é dada por:

$$P(A) = \frac{\#(A)}{\#(\Omega)} = \frac{\text{nº de resultados favoráveis a } A}{\text{nº de resultados possíveis em } \Omega}$$

Onde $\#$ indica a cardinalidade de um conjunto (quantidade de elementos no conjunto).

Example 2.2. Seja $A = \{2, 4, 6\}$, os lançamentos pares em um dado honesto. Como $\Omega = \{1, \dots, 6\}$, temos que:

$$P(A) = \frac{3}{6} = \frac{1}{2}$$

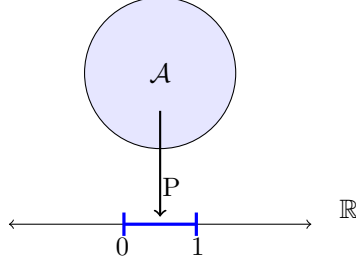
Note que o conjunto A pode ser descrito como a união dos eventos elementares, tais que $A = \{2\} \cup \{4\} \cup \{6\}$. Nesse caso, podemos ver que a probabilidade de A não muda, pois:

$$\begin{aligned} P(\{i\}) &= \frac{\#(\{i\})}{\#(\Omega)} = \frac{1}{6} \\ P(A) &= \frac{\#(\{2\}) + \#(\{4\}) + \#(\{6\})}{\#(\Omega)} = \frac{1 + 1 + 1}{6} = \frac{1}{2} \end{aligned}$$

Definition 2.3. Um evento A ao qual atribuímos uma probabilidade é um evento aleatório.

2.2 Álgebras de Conjuntos

Considere o conjunto de eventos em uma família \mathcal{A} (subconjuntos de Ω), de tal modo que $P : \mathcal{A} \rightarrow [0, 1]$. Uma representação gráfica da relação P pode ser dada por:



Definition 2.4. Seja Ω um conjunto não-vazio. Seja \mathcal{A} uma classe de subconjuntos de Ω , ela será chamada de “Álgebra de subconjuntos de Ω ”, caso respeite os seguintes axiomas:

- Ax_1 : $\Omega \in \mathcal{A}$, e definimos $P(\Omega) = 1$;
- Ax_2 : Se $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$, e definimos $P(A^c) = 1 - P(A)$;
- Ax_3 : Se $A \in \mathcal{A}, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$.

E por consequência desses axiomas, temos as seguintes extensões:

- Ax_4 : $\emptyset \in \mathcal{A}$;
- Ax_5 : Sejam $A_1, A_2, \dots, A_n : A_i \in \mathcal{A} \forall i \Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{A}$ e $\bigcap_{i=1}^n A_i \in \mathcal{A}$.

É fácil verificar a extensão de Ax_4 a partir de Ax_1 e Ax_2 : Ax_1 define que $\Omega \in \mathcal{A}$, e por Ax_2 temos que $\Omega^c \in \mathcal{A}$, e por definição temos que $\Omega^c = \emptyset$, logo $\emptyset \in \mathcal{A}$. Também é interessante notar que, ainda por Ax_2 , temos que $P(\emptyset) = 1 - P(\Omega)$, e por Ax_1 temos que $P(\Omega) = 1$, portanto $P(\emptyset) = 1 - 1 = 0$.

A extensão de Ax_5 é dada por indução e pelas Leis de De Morgan: Sejam $A_1, A_2 \in \mathcal{A}$. Temos pelo axioma Ax_3 , que $A_1 \cup A_2 \in \mathcal{A}$, podendo assim definir o conjunto $B = A_1 \cup A_2$, sendo possível ver que $B \in \mathcal{A}$. Sejam ainda um conjunto $A_3 \in \mathcal{A}$, podemos ver que $B \cup A_3 \in \mathcal{A}$, e como $B = A_1 \cup A_2$, temos que $(A_1 \cup A_2) \cup A_3 \in \mathcal{A}$. Podemos proceder dessa forma para qualquer quantidade (enumerável) de conjuntos, de modo que $\bigcup_{i=1}^n A_i \in \mathcal{A}$. Pelas Leis de De Morgan, sabemos que:

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

E pela extensão indutiva em n do axioma Ax_2 , temos que se $A_i^c \in \mathcal{A}, \forall i$, então $\bigcup_{i=1}^n A_i^c \in \mathcal{A}$. E como, se um conjunto pertence a \mathcal{A} seu complementar deve pertencer também, e pelo resultado em (1), temos então que:

$$\left(\bigcup_{i=1}^n A_i^c \right)^c = \left(\bigcap_{i=1}^n A_i \right) \in \mathcal{A} \quad (2)$$

Assim provamos o axioma A_5 como extensão indutiva dos axiomas anteriores, indicando que tanto a união quanto a interseção dos A_i pertencem à \mathcal{A} . Podemos também mostrar que a álgebra \mathcal{A} é fechada também para a operação de diferença entre conjuntos: $A \in \mathcal{A}, B \in \mathcal{A}, A - B = A \cap B^c \in \mathcal{A}$.

Proof. Considerando que os conjuntos A e B pertencem à \mathcal{A} , podemos utilizar o axioma Ax_2 para mostrar que $A^c \in \mathcal{A}$ e $B^c \in \mathcal{A}$. A partir disso, por meio do axioma Ax_5 temos que os seguintes conjuntos também pertencem à \mathcal{A} : $A \cup B, A \cup B^c, A^c \cup B, A^c \cup B^c, A \cap B, A \cap B^c, A^c \cap B, A^c \cap B^c$. E como temos que $A \cap B^c = A - B$, temos a prova de que $A - B \in \mathcal{A}$. Além disso, essa prova mostra que a diferença contrária ($B - A = A^c \cap B$) também pertence à álgebra \mathcal{A} . \square

Ainda considerando os conjuntos A e B , existem cinco maneiras como esses conjuntos podem “interagir”, e podemos mostrar que em todos os casos a diferença $A - B \in \mathcal{A}$:

- $A \not\subset B$ e $A \not\supset B$ e $A \cap B \neq \emptyset \Rightarrow A - B = A \cap B^c \in \mathcal{A}$;
- $A \not\subset B$ e $A \not\supset B$ e $A \cap B = \emptyset \Rightarrow A - B = A \in \mathcal{A}$;
- $A \supset B \Rightarrow A - B = A \cap B^c \in \mathcal{A}$;
- $A \subset B \Rightarrow A - B = \emptyset \in \mathcal{A}$;
- $A = B \Rightarrow A - B = \emptyset \in \mathcal{A}$.

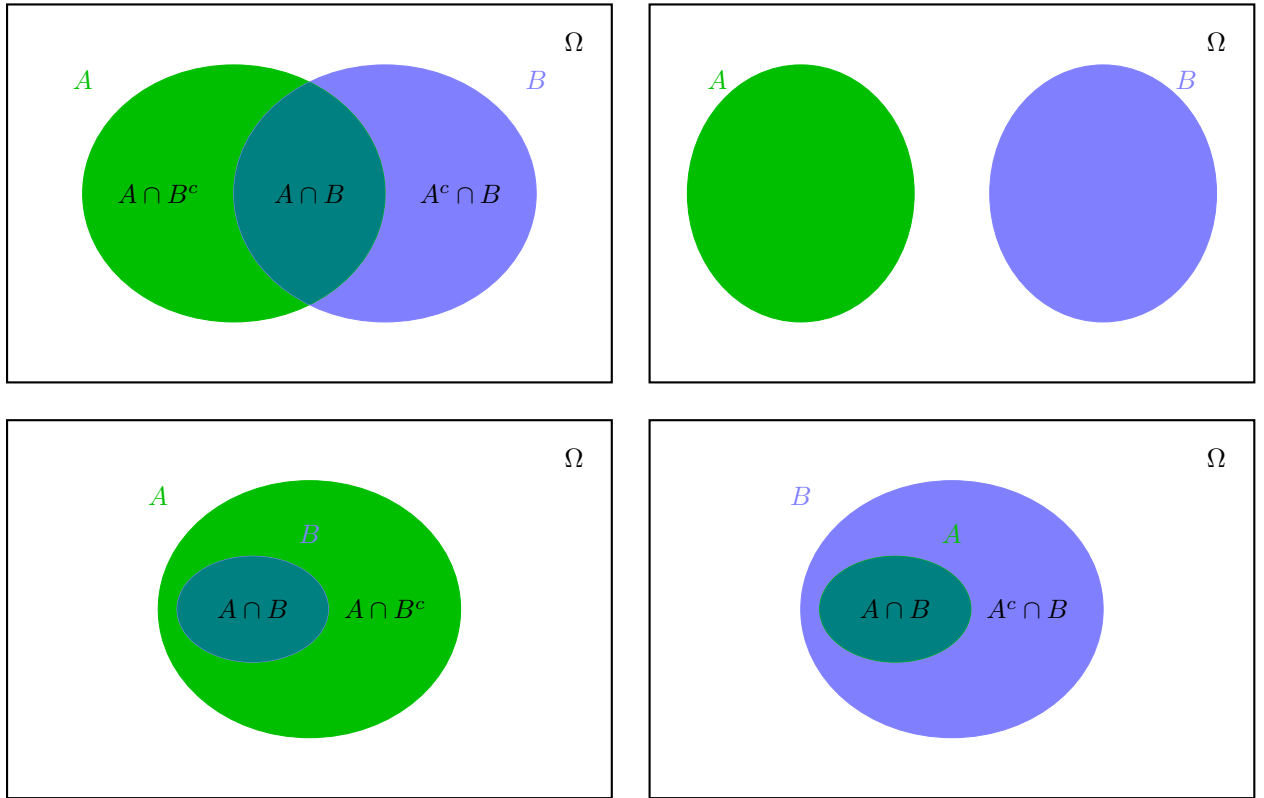


Figure 1: Diferentes relações entre A e B demonstradas por Diagramas de Venn. Note que em todos os casos, $A \cap B^c \in \mathcal{A}$ ou $A \cap B^c = \emptyset \in \mathcal{A}$ ou $A \cap B^c = A \in \mathcal{A}$

As representações por Diagramas de Venn apresentadas na figura 2.2 não é prova formal de que a álgebra \mathcal{A} é fechada para a diferença, mas é um recurso visual que pode auxiliar no entendimento da relação entre os conjuntos.

Definition 2.5. Uma classe \mathcal{A} de conjuntos/subconjuntos de $\Omega \neq \emptyset$, verificando os axiomas Ax_1, Ax_2 e Ax_3 é chamada de σ -álgebra de subconjuntos de Ω .

Note que uma σ -álgebra é sempre uma álgebra. Uma outra forma de construir σ -álgebras é partir de uma álgebra munida dos axiomas de Kolmogorov (Teorema de Carathéodory).

Proposition 2.1. Seja \mathcal{A} uma σ -álgebra de subconjuntos de Ω , se A_1, A_2, \dots , é uma coleção em $\mathcal{A} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

Example 2.3. Seja $\Omega = \{1, 2, 3, 4, 5, 6\}$ (o lançamento de um dado cúbico usual). A σ -álgebra usual é definida da seguinte forma e denotada por $\mathcal{P}(\Omega)$ (chamada de partes de Ω ou *powerset* de Ω):

$$\begin{aligned}\mathcal{A} = \{ & \emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \\ & \{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{1, 6\}, \\ & \{2, 3\}, \{2, 4\}, \dots, \\ & \Omega\end{aligned}$$

Example 2.4. Definamos a σ -álgebra de Borel no intervalo $\Omega = [0, 1]$. Uma possível definição seria:

$$\mathcal{A} = \text{ todos os subconjuntos de } [0, 1] \text{ cujo comprimento esteja bem definido}$$

Podemos, por exemplo, propor uma álgebra para o intervalo $[0, 1]$ dada por:

$$\mathcal{A}_r = \{A \subset [0, 1] : A \text{ é uma união finita de intervalos} \}$$

É possível encontrar um conjunto A tal que $A \notin \mathcal{A}$, por exemplo:

$$A = \left\{ \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, \frac{3}{4}\right) \cup \dots \cup \left(1 - \frac{1}{2^n}, 1 - \frac{1}{2^{n+1}}\right) \cup \dots \right\}$$

Podemos ver que, para qualquer n^* finito, $\lim_{n \rightarrow n^*} \left(1 - \frac{1}{2^{n+1}}\right) \neq 1$, de modo que o conjunto A não cobrirá completamente o intervalo $[0, 1]$. Dessa forma, a σ -álgebra de Borel no intervalo $[0, 1]$ (denotada $\mathcal{B}_{[0,1]}$) é definida como:

$$\mathcal{B}_{[0,1]} = \{A : A \subset [0, 1] \text{ e } A \text{ é boreliano}\}$$

Onde boreliano denota que A é união enumerável (finita ou infinita) de intervalos em $[0, 1]$

2.3 Axiomas de Kolmogorov

Seja $P : \mathcal{A} \rightarrow [0, 1]$, com:

- $Ax_1(K) : P(A) \geq 0, \forall A \in \mathcal{A}$;
- $Ax_2(K) : P(\Omega) = 1$;
- $Ax_3(K) : \text{Se } A_1, A_2, \dots, A_n : A_i \in \mathcal{A} \forall i \text{ e } A_i \cap A_j = \emptyset \forall i, j \in \{1, 2, \dots, n\}, i \neq j \Rightarrow P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k).$

Definition 2.6. Seja Ω um conjunto não-vazio, \mathcal{A} uma σ -álgebra em Ω , com $P : \mathcal{A} \rightarrow [0, 1]$, verificando os axiomas de Kolmogorov, então P é dita finitamente aditiva. Podemos assim, modificar o axioma $Ax_3(K)$ para:

- $Ax'_3(K) : \text{Se } A_1, A_2, \dots \text{ é uma sequência em } \mathcal{A} \text{ tal que } \forall i \neq j, A_i \cap A_j = \emptyset, \text{ tem-se que } P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$ (*propriedade da σ -aditividade*)

Definition 2.7. P definida em uma σ -álgebra \mathcal{A} , satisfazendo os axiomas de Kolmogorov ($Ax_1(K), Ax_2(K), Ax'_3(K)$) é uma medida de probabilidade em \mathcal{A} , constituída pela terna (Ω, \mathcal{A}, P) .

2.4 Propriedades da medida de probabilidade

Proposition 2.2 (Continuidades).

1. Seja $\{A_i\}_{i=1}^{\infty}$ uma sequência (crescente) de eventos tais que $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$, e seja $A = \bigcup_{i=1}^{\infty} A_i$, então $P(A) = \lim_{i \rightarrow \infty} P(A_i)$.
2. Seja $\{B_i\}_{i=1}^{\infty}$ uma sequência (decrescente) de eventos tais que $B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots$, e seja $B = \bigcap_{i=1}^{\infty} B_i$, então $P(B) = \lim_{i \rightarrow \infty} P(B_i)$.

Proof.

1. Note que, sendo $A_0 = \emptyset$, tem-se que $A = (A_1 - A_0) \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \dots$, ou seja, A é união disjunta de eventos $D_i = A_i - A_{i-1}$, de forma que $A_{i-1} \subseteq A_i \Rightarrow P(A_i) = P(A_{i-1}) + P(A_i - A_{i-1}) \Rightarrow P(A_i - A_{i-1}) = P(A_i) - P(A_{i-1})$. Logo, temos que:

$$\begin{aligned}
 A = \bigcup_{i=1}^{\infty} D_i &\xrightarrow{Ax'_3(K)} P(A) = \sum_{i=1}^{\infty} P(D_i) \\
 &= \sum_{i=1}^{\infty} P(A_i - A_{i-1}) \\
 &= \lim_{n \rightarrow \infty} \sum_{i=1}^n [P(A_i) - P(A_{i-1})] \\
 &= \lim_{n \rightarrow \infty} [P(A_1) - P(A_0) + P(A_2) - P(A_1) + P(A_3) - P(A_2) + \dots] \\
 &= \lim_{n \rightarrow \infty} P(A_n)
 \end{aligned}$$

2. Note que, por De Morgan, $B = \bigcap_{i=1}^n B_i = (\bigcup_{i=1}^n B_i^c)^c$. Logo $P(\bigcap_{i=1}^n B_i) = 1 - P(\bigcup_{i=1}^n B_i^c)$. Seja $A = B_i^c$ de modo que:

$$\begin{aligned}
 B_1^c &= \Omega - B_1 = A_1 \\
 B_2^c &= (B_1 - B_2) \cup (\Omega - B_1) = A_2 \\
 &\vdots
 \end{aligned}$$

Assim $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$, e com isso $P(\bigcap_{i=1}^n B_i) = 1 - P(\bigcup_{i=1}^n B_i^c) = 1 - P(\bigcup_{i=1}^n A_i)$. Por outro lado, tem-se que $A = \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i^c \Rightarrow A^c = (\bigcup_{i=1}^{\infty} B_i^c)^c = \bigcap_{i=1}^{\infty} B_i = B$. Logo, temos que:

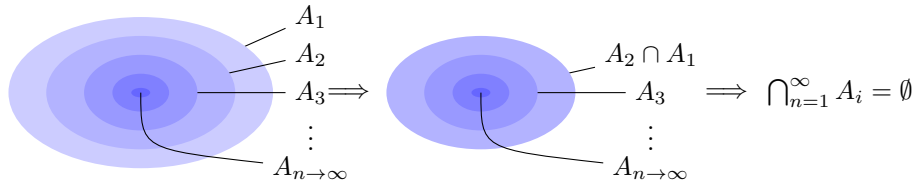
$$P\left(\bigcap_{i=1}^n B_i\right) \xrightarrow{n \rightarrow \infty} (1 - P(A)) = P(A^c) = P(B)$$

□

Definition 2.8 (Continuidade no vazio).

- $Ax_4(K)$: Se $\{A_n\}_{n \geq 1} \subseteq \mathcal{A}$ e $A_n \supseteq A_{n+1} \forall n$ e $\bigcap_{n=1}^{\infty} A_n \neq \emptyset$ então $P(A_n) \xrightarrow{n \rightarrow \infty} 0$

A prova dessa definição é dada pela segunda parte da prova da proposição 2.2. A representação visual é dada pelo seguinte diagrama:



Proposition 2.3. Dados os axiomas $Ax_1(K), Ax_2(K), Ax_3(K)$, o axioma 4 é equivalente ao axioma $Ax'_3(K)$, ou seja, uma probabilidade finitamente aditiva é uma medida de probabilidade se e somente se é contínua no vazio.

A prova de que a σ -aditividade implica o axioma 4 é consequência da prova da proposição anterior, dado que $\bigcap_{n=1}^{\infty} A_n = \emptyset$. Para demonstrar o contrário (que $Ax_1(K) + Ax_2(K) + Ax_3(K) + Ax_4(K) \rightarrow Ax'_3(K)$), tomemos uma sequência infinita de eventos $\{A_i\}_{i \geq 1}$ em \mathcal{A} : $A_i \cap A_j = \emptyset \forall i \neq j$. Devemos ver que $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$. Seja $A = \bigcup_{n=1}^{\infty} A_n = (\bigcup_{n=1}^k A_n) \cup (\bigcup_{n=k+1}^{\infty} A_n)$. Tem-se que:

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

Seja $B_k = \bigcup_{n=k+1}^{\infty} A_n$. Note que $B_k \downarrow \emptyset$ quando $k \rightarrow \infty$ de modo que $P(B_k) \xrightarrow[k \rightarrow \infty]{} 0$, logo:

$$\lim_{k \rightarrow \infty} \sum_{n=1}^k P(A_n) = \sum_{n=1}^{\infty} P(A_n)$$

Corollary 2.1. *Os seguintes sistemas são equivalentes:*

$$Ax_1(K), Ax_2(K), Ax'_3(K) \equiv Ax_1(K), Ax_2(K), Ax_3(K), Ax_4(K)$$

2.5 Propriedades de probabilidade

Seja P uma probabilidade em uma σ -álgebra \mathcal{A} . Suponhamos que todo A abaixo pertença à \mathcal{A} . Então as seguintes propriedades são consequências dos axiomas:

- **P1:** $P(A^c) = 1 - P(A)$;
- **P2:** $0 \leq P(A) \leq 1$;
- **P3:** $A_1 \subset A_2 \Rightarrow P(A_1) \leq P(A_2)$;
- **P4:** $P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$;
- **P5:** $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$;

Com essas propriedades, podemos então definir um modelo probabilístico. Sejam:

- **a)** Um espaço amostral: $\Omega \neq \emptyset$;
- **b)** Uma σ -álgebra em Ω : \mathcal{A} ;
- **c)** Uma medida de probabilidade em \mathcal{A} : P .

Definition 2.9. Um espaço de probabilidade é uma terna (Ω, \mathcal{A}, P) seguindo **a, b** e **c**.

2.6 Probabilidade Condicional e Independência

Considere o seguinte experimento: um dado é lançado duas vezes e anota-se a dupla de resultados. Temos que:

$$\Omega = \{(i, j) : 1 \leq i \leq 6; 1 \leq j \leq 6; i, j, \in \mathbb{Z}\}$$

Sejam os seguintes eventos:

- $A =$ "em cada lançamento o valor observado é ≤ 2 ";
- $B =$ "a soma dos resultados é igual a 4".

$$A = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$$

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

Já que $\#\Omega = |\Omega| = 36$, e pela equiprobabilidade dos eventos (considerando que os dados são honestos), temos que:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{4}{36}$$

$$P(B) = \frac{|B|}{|\Omega|} = \frac{3}{36}$$

Além disso, $(A \cap B) = \{(2, 2)\}; P(A \cap B) = 1/36$. Suponha que A ocorre com $P(A) > 0$, e que B é o evento de interesse. Assumindo a potencial ocorrência de A , qual é a probabilidade de B ocorrer. Nesse caso $P(B|A) = 1/4$.

Definition 2.10 (Probabilidade condicional). Sejam A e B eventos em \mathcal{A} , com $P(A) > 0$. A probabilidade condicional $P(B|A)$ é definida como:

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

ou equivalentemente:

$$P(A \cap B) = P(B|A)P(A) \quad (4)$$

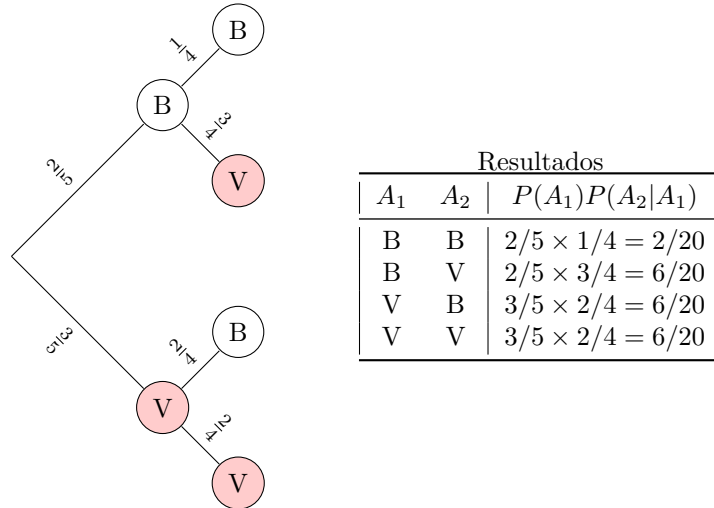
Example 2.5. Considere uma urna com 5 bolas, sendo 3 vermelhas e 2 brancas. O experimento consiste de 2 retiradas sucessivas de uma bola da urna (sem reposição). Considere os eventos $A_1 = \text{Cor da primeira bola}$ e $A_2 = \text{Cor da segunda bola}$:

$$P(A_1 = B) = \frac{2}{5}, \quad P(A_1 = V) = \frac{3}{5}$$

$$P(A_2 = B|A_1 = B) = \frac{1}{4}, \quad P(A_2 = V|A_1 = B) = \frac{3}{4}$$

$$P(A_2 = B|A_1 = V) = \frac{2}{4}, \quad P(A_2 = V|A_1 = V) = \frac{2}{4}$$

Podemos visualizar esse experimento com os seguintes diagrama e tabela de probabilidades:



Definition 2.11 (Eventos independentes).

- **a)** Os eventos A e B são independentes (denotados como $A \perp B$) se $P(A \cap B) = P(A)P(B)$;
- **b)** $\{A_i, i \in \mathbb{I}\}$ são independentes se $P(\bigcap_{i \in \mathcal{J}} A_i) = \prod_{i \in \mathcal{J}} P(A_i)$, \forall subfamílias \mathcal{J} de índices em \mathbb{I} .

Disso segue que, sendo A e B dois eventos, as seguintes propriedades são válidas:

1. Se $P(A) = 0 \Rightarrow P(A \cap B) = 0 \forall B$, ou seja, $A \perp B$;
2. Se $P(B) = 1 \Rightarrow P(A \cap B) = P(A) \forall A$, ou seja, $A \perp B$;
3. A é independente dele mesmo se e somente se $P(A) = 0$ ou $P(A) = 1$;
4. $A \perp B \Rightarrow A \perp B^c, A^c \perp B, A^c \perp B^c$;
5. As seguintes proposições são equivalentes:
 - a) $(A \perp B) \Rightarrow P(B|A) = P(B)$ e $P(B|A^c) = P(B)$;
 - b) $P(B|A) = P(B) \Rightarrow A \perp B$;
 - c) $P(B|A^c) = P(B) \Rightarrow A \perp B$.

Theorem 2.1 (Teorema das Probabilidades Totais).

1. Dados A e B eventos em \mathcal{F} :

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

2. No geral, se B_1, B_2, \dots, B_n é uma partição de Ω , então:

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i) \quad (5)$$

Demonstração: Note que $A = (A \cap B) \cup (A \cap B^c)$ e $(B \cap B^c) = \emptyset$ e $(B \cup B^c) = \Omega$. Além disso, $(A \cap B) \cap (A \cap B^c) = \emptyset$, logo $P(A) = P(A \cap B) + P(A \cap B^c)$. Como, por definição, $P(A|B) = P(A \cap B)/P(B)$ e $P(A|B^c) = P(A \cap B^c)/P(B^c)$, temos que:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

Para o caso geral, temos que $\{B_i\}_{i=1}^n$, $(B_i \cap B_j) = \emptyset \forall i, j$ e $\bigcup_{i=1}^n B_i = \Omega$. Logo:

$$\begin{aligned} A &= (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_n) \\ &\Downarrow \text{Pela } \sigma\text{-aditividade} \\ P(A) &= \sum_{i=1}^n P(A \cap B_i) \end{aligned}$$

E como $P(A|B_i) = P(A \cap B_i) / P(B_i)$:

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$$

2.7 Fórmula de Poincaré e Teorema de Bayes

Theorem 2.2 (Fórmula de Poincaré). *Seja $\{A_i\}_{i \geq 1} \subseteq \mathcal{F}$. Então:*

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \dots \\ &\quad + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n) \end{aligned} \quad (6)$$

A demonstração da fórmula (6) é dada no exercício 2.10.

Theorem 2.3 (Teorema de Bayes). *Seja $\{B_i\}_{i=1}^n$ uma partição de Ω e A um evento em \mathcal{F} , temos que:*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)} \quad (7)$$

O denominador de (7) é derivado do teorema das probabilidades totais, visto que $\{B_i\}_{i=1}^n$ é uma partição de Ω .

Lemma 2.1. *Sejam A_1, A_2, \dots, A_n eventos em \mathcal{F} , logo:*

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1})$$

Proof. Suponha a validade do lema anterior. Logo, seja $D = (\bigcap_{i=1}^n A_i)$:

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

□

2.8 Exercícios

Exercise 2.1 (BJ1). Sejam A, B e C eventos aleatórios. Identifique as seguintes equações e frases, casando cada equação expressa na notação de conjuntos com a correspondente frase na linguagem de eventos:

$A \cap B \cap C = A \cup B \cup C$	A e "B ou C" são incompatíveis.
$A \cap B \cap C = A$	Os eventos A,B e C são idênticos.
$A \cup B \cup C = A$	A ocorrência de A implica a de "B e C".
$(A \cup B \cup C) - (B \cup C) = A$	A ocorrência de A decorre de "B ou C".

Exercise 2.2 (BJ2). A partir dos axiomas, prove a propriedade P5:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n)$$

Proof. Consideremos uma prova por indução para $n \rightarrow \infty$:

Para $n = 2$:

$$P(A_1 \cup A_2) = P(A_1) + P(A_1^c \cap A_2)$$

Considerando que $(A_1^c \cap A_2) \subset A_2$ e o fato de que $(A_1) \cap (A_1^c \cap A_2) = \emptyset$, temos pela propriedade P3 que $P(A_1^c \cap A_2) \leq P(A_2)$, de modo que:

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_1^c \cap A_2) \leq P(A_2) \\ &\leq P(A_1) + P(A_2) \\ &\leq \sum_{i=1}^2 P(A_i) \end{aligned}$$

De modo semelhante, podemos fazer para n :

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= P(A_1) + P(A_1^c \cap A_2) + \dots \\ &\leq P(A_1) + P(A_2) + \dots \\ &\leq \sum_{i=1}^n P(A_i) \end{aligned}$$

Consideremos então uma sequência de eventos $A_i^*, \forall i \in \{n+1, n+2, \dots\}$, disjuntos de A_i . Denotemos ainda $A = (\bigcup_{i=1}^n A_i) \cup (\bigcup_{i=n+1}^{\infty} A_i)$. Pela aditividade infinita (ou ainda pela σ -aditividade), temos que:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{i=1}^n P(A_i) + P\left(\bigcup_{i=n+1}^{\infty} A_i\right)$$

Que por serem disjuntos, pelo axioma Ax_4 tem que $(\bigcup_{i=n+1}^{\infty} A_i) \downarrow \emptyset$, de modo que $P(\bigcup_{i=n+1}^{\infty} A_i) \rightarrow 0$. Logo, tem-se que:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n)$$

□

Exercise 2.3 (BJ3). Sejam A_1, A_2, \dots eventos aleatórios. Mostre que:

a) $P(\bigcap_{k=1}^n A_k) \geq 1 - \sum_{k=1}^n P(A_k^c)$

Proof. Por De Morgan temos que $\bigcap_{k=1}^n A_k = (\bigcup_{k=1}^n A_k^c)^c$, de modo que:

$$\begin{aligned} P\left(\bigcap_{k=1}^n A_k\right) &= P\left(\bigcup_{k=1}^n A_k^c\right)^c \\ &= 1 - P\left(\bigcup_{k=1}^n A_k^c\right) \xrightarrow{\text{Por P4}} P\left(\bigcup_{k=1}^n A_k^c\right) \leq \sum_{k=1}^n P(A_k^c) \\ &\geq 1 - \sum_{k=1}^n P(A_k^c) \end{aligned}$$

□

b) Se $P(A_k) \geq 1 - \epsilon$ para $k = 1, 2, \dots, n$, então $P(\bigcap_{k=1}^n A_k) \geq 1 - n\epsilon$

Proof. É fácil ver que:

$$P(A_k) \geq 1 - \epsilon \Rightarrow P(A_k^c) \leq 1 - (1 - \epsilon) = \epsilon$$

E de modo semelhante ao que foi feito na questão anterior (utilizando De Morgan), temos que:

$$\begin{aligned} P\left(\bigcap_{k=1}^n A_k\right) &= P\left(\bigcup_{k=1}^n A_k^c\right)^c = 1 - P\left(\bigcup_{k=1}^n A_k^c\right) \\ &\geq 1 - \sum_{k=1}^n P(A_k^c) \\ &\geq 1 - \sum_{k=1}^n \epsilon \\ &\geq 1 - n\epsilon \end{aligned}$$

□

c) $P(\bigcap_{k=1}^\infty A_k) \geq 1 - \sum_{k=1}^\infty P(A_k^c)$

Proof. De maneira semelhante ao que foi visto na prova da letra **a**, temos que:

$$\begin{aligned} P\left(\bigcap_{k=1}^\infty A_k\right) &= P\left(\bigcup_{k=1}^\infty A_k^c\right)^c \\ &= 1 - P\left(\bigcup_{k=1}^\infty A_k^c\right) \xrightarrow{\text{Por P5}} P\left(\bigcup_{k=1}^n A_k^c\right) \leq \sum_{k=1}^\infty P(A_k^c) \\ &\geq 1 - \sum_{k=1}^\infty P(A_k^c) \end{aligned}$$

Para ver a demonstração da propriedade P5, vide exercício 2.2.

□

Exercise 2.4 (BJ4). Demonstre as seguintes propriedades:

a) Se $P(A_n) = 0$ para $n = 1, 2, \dots$, então $P(\bigcup_{n=1}^\infty A_n) = 0$.

Proof. Utilizando a propriedade *P5*, temos que:

$$\begin{aligned}
P\left(\bigcup_{n=1}^{\infty} A_n\right) &\leq \sum_{n=1}^{\infty} P(A_n) \\
&\leq \sum_{n=1}^{\infty} 0 \\
&\leq 0 \\
&\downarrow \text{ Por } P2 \\
P\left(\bigcup_{n=1}^{\infty} A_n\right) &= 0
\end{aligned}$$

□

b) Se $P(A_n) = 1$ para $n = 1, 2, \dots$, então $P(\bigcap_{n=1}^{\infty} A_n) = 1$.

Proof. Levando em consideração que se $P(A_n) = 1 \Rightarrow P(A_n^c) = 0$ (pela propriedade *P1*), utilizando De Morgan e a prova da letra **c** do exercício 2.3, temos que:

$$\begin{aligned}
P\left(\bigcap_{n=1}^{\infty} A_n\right) &\geq 1 - \sum_{n=1}^{\infty} P(A_n^c) \\
&\geq 1 - \sum_{n=1}^{\infty} 0 \\
&\geq 1 - 0 \\
&\geq 1 \\
&\downarrow \text{ Por } P2 \\
P\left(\bigcap_{n=1}^{\infty} A_n\right) &= 1
\end{aligned}$$

□

Exercise 2.5 (BJ6). Seja Ω um conjunto não-vazio.

a) Prove: se \mathcal{A} e \mathcal{B} são σ -álgebras de subconjuntos de Ω , então $(\mathcal{A} \cap \mathcal{B})$ também é uma σ -álgebra.

Proof. Para que $\mathcal{A} \cap \mathcal{B}$ seja uma σ -álgebra, é necessário que cumpram-se os axiomas Ax_1 , Ax_2 e Ax_3 :

- Ax_1 : Sabemos que $\Omega \in \mathcal{A}$ e $\Omega \in \mathcal{B}$, logo sabemos que $\Omega \in (\mathcal{A} \cap \mathcal{B})$;
- Ax_2 : Seja um evento $E \in (\mathcal{A} \cap \mathcal{B})$, sabemos então que $E \in \mathcal{A}$ e $E \in \mathcal{B}$, logo $E^c \in \mathcal{A}$ e $E^c \in \mathcal{B}$, portanto $E^c \in (\mathcal{A} \cap \mathcal{B})$;
- Ax_3 : Sejam dois eventos, $E_1 \in (\mathcal{A} \cap \mathcal{B})$ e $E_2 \in (\mathcal{A} \cap \mathcal{B})$. Com isso, temos que $E_1, E_2 \in \mathcal{A}$ e $E_1, E_2 \in \mathcal{B}$, portanto $(E_1 \cup E_2) \in \mathcal{A}$ e $E_1 \cup E_2 \in \mathcal{B}$, logo $(E_1 \cup E_2) \in (\mathcal{A} \cap \mathcal{B})$.

Como os três axiomas foram cumpridos, temos que $(\mathcal{A} \cap \mathcal{B})$ é uma σ -álgebra. □

b) Generalize o item (a): se $\mathcal{A}_i, i \in \mathcal{I}$, são σ -álgebras de partes de Ω , onde \mathcal{I} é um conjunto não-vazio de índices, então $\bigcap_{i \in \mathcal{I}} \mathcal{A}_i$ também é uma σ -álgebra.

Proof. Como anteriormente, temos que mostrar que $\bigcap_{i \in \mathcal{I}} \mathcal{A}_i$ cumpre os axiomas Ax_1 , Ax_2 e Ax_3 :

- Ax_1 : Sabemos que $\Omega \in \mathcal{A}_i$, $\forall i \in \mathcal{I}$, logo sabemos que $\Omega \in \bigcap_{i \in \mathcal{I}} \mathcal{A}_i$;
- Ax_2 : Seja um evento $E \in \bigcap_{i \in \mathcal{I}} \mathcal{A}_i$, sabemos então que $E \in \mathcal{A}_i$, $\forall i \in \mathcal{I}$, logo $E^c \in \mathcal{A}_i$, $\forall i \in \mathcal{A}$, portanto $E^c \in \bigcap_{i \in \mathcal{I}} \mathcal{A}_i$;
- Ax_3 : Sejam dois eventos, $E_1 \in \bigcap_{i \in \mathcal{I}} \mathcal{A}_i$ e $E_2 \in \bigcap_{i \in \mathcal{I}} \mathcal{A}_i$. Com isso, temos que $E_1, E_2 \in \mathcal{A}_i$, $\forall i \in \mathcal{I}$, portanto $(E_1 \cup E_2) \in \mathcal{A}_i$, $\forall i \in \mathcal{I}$, logo $(E_1 \cup E_2) \in \bigcap_{i \in \mathcal{I}} \mathcal{A}_i$.

Vemos portanto que, por cumprir os axiomas Ax_1 , Ax_2 e Ax_3 , $\bigcap_{i \in \mathcal{I}} \mathcal{A}_i$ é também uma σ -álgebra. \square

c) Seja \mathbb{C} uma classe de subconjuntos de Ω . Mostre que existe *pelo menos uma* σ -álgebra que contém \mathbb{C} .

Proof. É fácil ver que a maior classe de subconjuntos de Ω é o conjunto das partes de Ω , denotado como $\mathcal{P}(\Omega)$ (definido no exemplo 2.3). Assim, temos que $\mathbb{C} \subseteq \mathcal{P}(\Omega)$, de modo que, pelo menos a σ -álgebra formada por $\mathcal{P}(\Omega)$ contém \mathbb{C} . \square

d) Visando a plena utilização dos itens (b) e (c), como você definiria “a menor σ -álgebra contendo \mathbb{C} ”, onde \mathbb{C} é uma classe de subconjuntos de Ω ?

Proof. Considere que temos σ -álgebras de partes de Ω , \mathcal{A}_i com $i \in \mathbb{I}$ (sendo \mathbb{I} um conjunto não-vazio de índices), tais que $\mathbb{C} \in \mathcal{A}_i : \forall i \in \mathbb{I}$. Assim, sabemos que algum dos \mathcal{A}_i é a menor σ -álgebra que contém \mathbb{C} , de modo que $\bigcap_{i \in \mathbb{I}} \mathcal{A}_i$ será a menor σ -álgebra que contém \mathbb{C} . \square

Exercise 2.6 (BJ9). Uma caixa contém $2n$ sorvetes, n do sabor A e n do sabor B . De um grupo de $2n$ pessoas, $a < n$ preferem o sabor A , $b < n$ o sabor B e $2n - (a + b)$ não tem preferência. Demonstre: se os sorvetes são distribuídos ao acaso, a probabilidade de que a preferência de todas as pessoas seja respeitada é de $\binom{2n-a-b}{n-a} / \binom{2n}{n}$.

Proof. Sabendo que a ordem de entrega dos n sorvetes de cada sabor, para as $2n$ pessoas não importa, temos que a quantidade possível de entregas diferentes é:

$$|\Omega| = \binom{2n}{n}$$

Considere que o evento R indica o caso em que todos tiveram sua preferência respeitada. Podemos ver que:

$$P(R) = \frac{|R|}{|\Omega|} = \frac{|R|}{\binom{2n}{n}}$$

Para que R ocorra, é necessário que as a pessoas que preferem A recebam esse sabor, bem como as b pessoas que preferem B . Dessa forma, temos que distribuir os $2n - (a + b)$ sorvetes restantes para as pessoas que não tem preferência. Assim, primeiramente temos os $n - a$ sorvetes do sabor A que não foram alocados, de forma que:

$$\binom{2n-a-b}{n-a} = \frac{(2n-a-b)!}{(2n-a-b-n+a)!(n-a)!} = \frac{(2n-a-b)!}{(n-b)!(n-a)!} \quad (8)$$

E podemos mostrar que, caso fossemos alocar os $n - b$ sorvetes do sabor B para as $2n - (a + b)$ pessoas sem preferência, teríamos:

$$\binom{2n-a-b}{n-b} = \frac{(2n-a-b)!}{(2n-a-b-n+b)!(n-b)!} = \frac{(2n-a-b)!}{(n-a)!(n-b)!} \quad (9)$$

Como (8) e (9) são iguais, podemos ver que a alocação dos sorvetes restantes não depende de qual sabor já foi alocado. Assim, temos que $|R| = \binom{2n-a-b}{n-a} = \binom{2n-a-b}{n-b}$, portanto:

$$P(R) = \frac{|R|}{|\Omega|} = \frac{\binom{2n-a-b}{n-a}}{\binom{2n}{n}}$$

\square

Exercise 2.7 (BJ10). Suponhamos que dez cartas estejam numeradas de 1 até 10. Das dez cartas, retira-se uma de cada vez, ao acaso e sem reposição, até retirar-se o primeiro número par. Conta-se o número de retiradas necessárias. Exiba um bom modelo probabilístico para esse experimento.

Proof. Dada essa formulação, temos que 5 cartas são pares e 5 são ímpares. Assim, considere o evento $\{Y_k : 1 \leq k \leq 6; k \in \mathbb{Z}\}$ em que k indica que a k -ésima retirada contém a primeira carta par. Assim, por exemplo, Y_1 indica o evento em que a primeira carta retirada é par, Y_2 o evento em que a segunda carta retirada é par, e assim por diante.

O nosso espaço amostral é (visto que o número da carta não importa, apenas se é $P = \text{"par"}$ ou $I = \text{"ímpar"}$):

$$\Omega = \{(P), (I, P), (I, I, P), (I, I, I, P), (I, I, I, I, P), (I, I, I, I, I, P)\}$$

É fácil ver que não é possível ter $\{Y_k : k \geq 7\}$, já que as cartas são retiradas sem reposição. Podemos facilmente calcular as probabilidades de cada evento em Ω , como segue:

$$\begin{aligned} P(Y_1) &= \frac{5}{10} = \frac{1}{2} \\ P(Y_2) &= \frac{5}{10} \cdot \frac{5}{9} = \frac{5}{18} \\ P(Y_3) &= \frac{5}{10} \cdot \frac{4}{9} \cdot \frac{5}{8} = \frac{5}{36} \\ P(Y_4) &= \frac{5}{10} \cdot \frac{4}{9} \cdot \frac{3}{8} \cdot \frac{5}{7} = \frac{5}{84} \\ P(Y_5) &= \frac{5}{10} \cdot \frac{4}{9} \cdot \frac{3}{8} \cdot \frac{2}{7} \cdot \frac{5}{6} = \frac{5}{252} \\ P(Y_6) &= \frac{5}{10} \cdot \frac{4}{9} \cdot \frac{3}{8} \cdot \frac{2}{7} \cdot \frac{1}{6} \cdot \frac{5}{5} = \frac{1}{252} \end{aligned}$$

Podemos ver que $\sum_{k=1}^6 P(Y_k) = 1$, e além disso, podemos denotar as probabilidades a partir da seguinte função:

$$P(Y_k) = \frac{5}{11-k} \cdot \prod_{n=1}^{k-1} \frac{6-n}{11-n} \quad (10)$$

A segunda parcela da equação (10) é válida para $k \geq 2$, pois ela representa as $k-1$ cartas ímpares retiradas antes da primeira carta par, caso que só ocorre caso $k \geq 2$. \square

Exercise 2.8 (BJ11). Para cada um dos seguintes experimentos, descreva um espaço de probabilidade que sirva de modelo:

- a) Seleciona-se um ponto, ao acaso, do quadrado unitário

$$\{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$$

Proof. Temos que:

$$\Omega = \{(x, y) \in [0, 1] \times [0, 1] \subset \mathbb{R}^2\}$$

Pela continuidade no vazio, é necessário que a probabilidade de ocorrência de um determinado ponto ser igual a zero, de modo que uma medida de probabilidade possível é por meio de intervalos. Considerando que $x \sim U(0, 1)$ e $y \sim U(0, 1)$ (ou seja, x e y são uniformemente distribuídos), podemos encontrar a probabilidade de $(x, y) \in \mathbb{I}$, com \mathbb{I} sendo um intervalo no cartesiano $[0, 1] \times [0, 1] \in \mathbb{R}^2$, por meio da distribuição de probabilidade conjunta de x e y . \square

- b) Retiram-se cartas sucessivamente de um baralho de 52 cartas, ao acaso e *com* reposição, até retirar-se o primeiro rei. Registra-se o número total de retiradas.

Proof. Considere que $\{Y : Y \in \{1, 2, \dots\}\}$ indica a quantidade de retiradas necessárias até o primeiro rei. O espaço amostral é dado diretamente: $\Omega = \{1, 2, 3, \dots\}$. Temos que, para cada retirada, a probabilidade da carta ser um rei é $4/52 = 1/13$ (considerando que temos 4 reis no baralho), e a probabilidade de não ser é de $48/52 = 12/13$. Assim, a probabilidade de que a primeira retirada seja um rei é de:

$$P(Y = 1) = \frac{1}{13}$$

Caso isso não ocorra, a probabilidade de que o primeiro rei ocorra na segunda retirada é de:

$$P(Y = 2) = \frac{12}{13} \cdot \frac{1}{13}$$

É possível verificar que, para todo $n \in \mathcal{N}$ a probabilidade de que o primeiro rei ocorra na retirada n é de:

$$P(Y = n) = \left(\frac{12}{13}\right)^{n-1} \cdot \left(\frac{1}{13}\right)$$

Esse modelo de probabilidade é denotado modelo geométrico. □

- c) Quinze bolas são retiradas, ao acaso e *com* reposição, de uma urna contendo 5 bolas vermelhas, 9 bolas pretas e uma bola branca. Observa-se o número que ocorre cada cor.

Proof. Sejam os eventos V, P e B o número de vezes que as retiradas foram de bolas vermelhas, pretas e brancas, respectivamente. É necessário (pela definição do modelo) que $V + P + B = 15$, mas consideremos o caso em que o número de retiradas seja n . Assim, para $n = 1$, o espaço amostral Ω é:

$$\Omega = \{(V), (P), (B)\}$$

E as probabilidades de cada evento são:

$$\begin{aligned} P(V = 1) &= \frac{5}{15} \\ P(P = 1) &= \frac{9}{15} \\ P(B = 1) &= \frac{1}{15} \end{aligned}$$

Para $n = 2$ bolas retiradas, temos que o espaço amostral é:

$$\begin{aligned} \Omega = \{ &(V, V), (V, P), (V, B), \\ &(P, V), (P, P), (P, B), \\ &(B, V), (B, P), (B, B)\} \end{aligned}$$

E as probabilidades de cada evento são:

$$\begin{aligned} P(V, V) &= \frac{5}{15} \cdot \frac{5}{15}; P(V, P) = \frac{5}{15} \cdot \frac{9}{15}; P(V, B) = \frac{5}{15} \cdot \frac{1}{15}; \\ P(P, V) &= \frac{9}{15} \cdot \frac{5}{15}; P(P, P) = \frac{9}{15} \cdot \frac{9}{15}; P(P, B) = \frac{9}{15} \cdot \frac{1}{15}; \\ P(B, V) &= \frac{1}{15} \cdot \frac{5}{15}; P(B, P) = \frac{1}{15} \cdot \frac{9}{15}; P(B, B) = \frac{1}{15} \cdot \frac{1}{15} \end{aligned}$$

Aqui é possível ver o padrão que surge para esse problema. Temos que os eventos V, P, B formam uma permutação (com repetição) da quantidade de bolas retiradas. A fórmula para a permutação com repetição de n elementos, em que cada um aparece k_1, k_2, \dots, k_j vezes é dada por:

$$P_n^{k_1, k_2, \dots, k_j} = \frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_j!}$$

Assim, podemos considerar que cada evento irá aparecer uma quantidade $V = v, P = p, B = b$ de vezes, com a seguinte probabilidade:

$$P(V = v, P = p, B = b) = \frac{15!}{v!p!b!} \cdot \left(\frac{5}{15}\right)^v \cdot \left(\frac{9}{15}\right)^p \cdot \left(\frac{1}{15}\right)^b ; \text{ com } v + p + b = 15$$

Caso seja necessário, podemos ainda generalizar para uma quantidade $n : 1 \leq n \leq 15$ de retiradas:

$$P(V = v, P = p, B = b) = \frac{n!}{v!p!b!} \cdot \left(\frac{5}{15}\right)^v \cdot \left(\frac{9}{15}\right)^p \cdot \left(\frac{1}{15}\right)^b ; \text{ com } v + p + b = n$$

Em que verifica-se facilmente que é válido para os casos em que $n = 1$ e $n = 2$ demonstrados anteriormente. \square

d) O experimento (c) é realizado *sem* reposição.

Proof. Como temos 15 bolas que serão retiradas *sem* reposição, o único evento possível após as 15 serem retiradas é:

$$\Omega = \{(V = 5, P = 9, B = 1)\}$$

E a probabilidade de isso ocorrer é 1 (visto que é o único evento no espaço amostral). Caso consideremos uma quantidade de retiradas $n < 15$, temos que o modelo de probabilidade é diferente. Consideremos que $V + P + B = n$ e que a quantidade de vezes que cada cor aparece é v, p e b , respectivamente. Então, como a ordem com que as cores são retiradas não importa, a probabilidade de aparecer uma quantidade de bolas de cada cor é dada por:

$$P(V = v, P = p, B = b) = \frac{\binom{5}{v} \binom{9}{p} \binom{1}{b}}{\binom{15}{n}}, \quad v + p + b = n$$

Esse modelo de probabilidade é chamado de multinomial hipergeométrico, e é uma generalização do modelo hipergeométrico para mais de duas classes (como é o caso). \square

Exercise 2.9 (BJ12). Retiram-se 4 cartas, ao acaso, de um baralho de 52 cartas. Registra-se o número de reis na amostra. Exiba um bom modelo probabilístico para este experimento se:

a) As retiradas são feitas *sem* reposição.

Proof. Considerando que em um baralho usual tem 52 cartas, e que a ordem com que cada uma das 4 cartas retiradas da amostra não importa (apenas importa a quantidade de reis na amostra), a quantidade total de amostras possíveis é $\binom{52}{4}$.

Como temos 4 reis no baralho, isso implica que há 48 cartas que são “não-reis”. Dessa forma, se na amostra forem coletados k reis, serão coletados também $4 - k$ “não-reis”, com os k reis podendo aparecer de $\binom{4}{k}$ maneiras diferentes (não importa qual o rei foi registrado) e os $4 - k$ “não-reis” podem aparecer de $\binom{48}{4-k}$ maneiras diferentes.

Assim, seja K o evento registrar k reis na amostra, a probabilidade $P(K = k)$ é dada por:

$$P(K = k) = \frac{\binom{4}{k} \binom{48}{4-k}}{\binom{52}{4}} \quad (11)$$

Esse modelo é chamado de hipergeométrico, que vale quando sabemos a quantidades de sucessos totais na população, e queremos contar a quantidade de sucessos coletados em uma amostra finita da população (que também deve ser finita). \square

b) As retiradas são feitas *com* reposição.

Proof. Se as retiradas são feitas com reposição, a probabilidade de registrar um rei em cada retirada é de $4/52$ e a probabilidade de registrar um “não-rei” é de $48/52$. Como a ordem das retiradas não importa, podemos ver que em uma amostra de tamanho 4, os k reis podem aparecer de $\binom{4}{k}$ maneiras diferentes. Além disso, podemos ver que, como irão aparecer k reis na amostra, consequentemente irão aparecer $4 - k$ “não-reis”.

Assim, seja K o evento registrar k reis na amostra, a probabilidade $P(K = k)$ é dada por:

$$P(K = k) = \binom{4}{k} \left(\frac{4}{52}\right)^k \left(\frac{48}{52}\right)^{4-k} \quad (12)$$

Esse modelo é chamado de binomial, e vale quando queremos encontrar a probabilidade de ocorrer k sucessos em uma amostra de tamanho n , dado que a probabilidade de cada sucesso é fixa. \square

c) Determine em que caso, (a) ou (b), é mais provável obter 4 reis.

Proof. Substituindo os valores de k em (11) e (12) para 4, podemos calcular as probabilidades em cada caso. Assim:

$$P(K = k) = \frac{\binom{4}{4} \binom{48}{0}}{\binom{52}{4}} \approx 3.7 \times 10^{-6}$$

$$P(K = k) = \binom{4}{4} \left(\frac{4}{52}\right)^4 \left(\frac{48}{52}\right)^0 \approx 3.5 \times 10^{-5}$$

De modo que é possível ver que no caso com reposição a probabilidade de encontrar 4 reis é maior. \square

Exercise 2.10 (BJ13).

a) Sejam A, B e C eventos aleatórios em um espaço de probabilidade (Ω, \mathcal{A}, P) . Mostre que

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

e

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

Proof. Podemos escrever os eventos A e B como as seguintes uniões de eventos disjuntos:

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

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

Utilizando a propriedade da aditividade finita (P3), temos que:

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \Rightarrow P(A \cap B^c) = P(A) - P(A \cap B) \\ P(B) &= P(A \cap B) + P(A^c \cap B) \Rightarrow P(A^c \cap B) = P(B) - P(A \cap B) \end{aligned} \quad (13)$$

Além disso, podemos escrever o evento $(A \cup B)$ como a seguinte união disjunta de eventos:

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

Por fim, utilizando os resultados de (13) e a aditividade finita, temos que:

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

Para a segunda expressão, podemos levar em consideração que os conjuntos A, B e C podem ser escritos como uniões de eventos disjuntos da seguinte forma:

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

Nos utilizando novamente da aditividade finita, temos que:

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

De maneira similar ao que fizemos na demonstração anterior, podemos isolar as probabilidades à direita, como por exemplo:

$$P(A \cap B \cap C^c) = P(A) - P(A \cap B^c \cap C^c) - P(A \cap B^c \cap C) - P(A \cap B \cap C) \quad (14)$$

Mas vale notar que, por serem eventos disjuntos:

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

De modo que a equação (14) pode ser reescrita como:

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

Assim, podemos denotar as seguintes probabilidades:

$$\begin{aligned}
P(A \cap B \cap C^c) &= P(A \cap B) - P(A \cap B \cap C) \\
P(A \cap B^c \cap C) &= P(A \cap C) - P(A \cap B \cap C) \\
P(A^c \cap B \cap C) &= P(B \cap C) - P(A \cap B \cap C)
\end{aligned} \quad (15)$$

Utilizando os resultados de (15), podemos isolar as outras probabilidades, tais como:

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

De modo que podemos denotar as seguintes probabilidades:

$$\begin{aligned}
P(A \cap B^c \cap C^c) &= P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C) \\
P(A^c \cap B \cap C^c) &= P(B) - P(A \cap B) - P(B \cap C) + P(A \cap B \cap C) \\
P(A^c \cap B^c \cap C) &= P(C) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)
\end{aligned} \quad (16)$$

O evento $(A \cup B \cup C)$ pode ser escrito como a seguinte união de eventos disjuntos (de fácil verificação que são disjuntos dois a dois):

$$\begin{aligned} (A \cup B \cup C) = & (A \cap B \cap C^c) \cup (A \cap B^c \cap C) \cup (A^c \cap B \cap C) \cup \\ & (A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C) \cup \\ & (A \cap B \cap C) \end{aligned} \quad (17)$$

Por fim, valendo-se da aditividade finita e substituindo em (17) os resultados obtidos em (15) e (16), temos que:

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

□

b) Enuncie a generalização do item **(a)** para o caso da união de n eventos aleatórios.

Proof. Podemos ver que as demonstrações anteriores podem ser escritas como:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) = & \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \dots \\ & + (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \end{aligned} \quad (18)$$

Esse é chamado de princípio de inclusão-exclusão.

□

c) Prove as seguintes *desigualdades de Bonferroni*:

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

Proof. Podemos demonstrar a primeira desigualdade utilizando a equação (18):

$$\begin{aligned} \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) & \leq P\left(\bigcup_{i=1}^n A_i\right) \\ 0 & \leq P\left(\bigcup_{i=1}^n A_i\right) - \left(\sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j)\right) \\ 0 & \leq \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \sum_{1 \leq i < j < k < l \leq n} P(A_i \cap A_j \cap A_k \cap A_l) + \dots \\ & + (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \end{aligned} \quad (19)$$

E como $(A_{i_1} \cap \dots \cap A_{i_n}) \subseteq (A_{i_1} \cap \dots \cap A_{i_{n-1}}) \Rightarrow P((A_{i_1} \cap \dots \cap A_{i_n})) \leq P((A_{i_1} \cap \dots \cap A_{i_{n-1}}))$, temos que a expressão (19) é maior que 0. Para a segunda desigualdade, vamos nos valer do mesmo princípio:

$$\begin{aligned}
P\left(\bigcup_{i=1}^n A_n\right) &\leq \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) \\
&\leq \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \left(P\left(\bigcup_{i=1}^n A_n\right)\right) \quad (20) \\
&\leq \sum_{1 \leq i < j < k < l \leq n} P(A_i \cap A_j \cap A_k \cap A_l) - \dots - (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k})
\end{aligned}$$

E da mesma forma que antes, é possível ver que a última expressão em (20) é maior que 0. \square

(ii) Se k é ímpar, $k \leq n$, então:

$$\begin{aligned}
P\left(\bigcup_{i=1}^n A_i\right) &\leq \sum_{i=1}^n P(A_n) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \dots \\
&\quad + (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k});
\end{aligned}$$

se k é par, $k \leq n$, vale \geq nesta última desigualdade.

Proof. Como $k \leq n$, podemos separar a desigualdade em dois casos:

1. $k = n$;
2. $k < n$;

No primeiro caso é fácil ver que a expressão se iguala à generalização para a união dada em (18). Para o segundo caso, temos que:

$$\begin{aligned}
P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \dots \\
&\quad + (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_{k-1} < i_k} P(A_{i_1} \cap \dots \cap A_{i_k}) + (-1)^k \sum_{1 \leq i_1 < \dots < i_k < i_{k+1}} P(A_{i_1} \cap \dots \cap A_{i_{k+1}}) + \dots
\end{aligned}$$

\uparrow
Termo k

\uparrow
Termo k+1

Como k é ímpar, o termo k será positivo e o termo $k+1$ será negativo. Assim, se subtrairmos os $k+j$, $j \in \{1, \dots, n-k\}$ termos de ambos os lados, teremos:

$$\begin{aligned}
P\left(\bigcup_{i=1}^n A_i\right) - \left((-1)^k \sum_{1 \leq i_1 < \dots < i_k < i_{k+1}} P(A_{i_1} \cap \dots \cap A_{i_{k+1}}) + \dots\right) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \dots \\
&\quad + (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_{k-1} < i_k} P(A_{i_1} \cap \dots \cap A_{i_k})
\end{aligned}$$

E podemos ver que:

$$P\left(\bigcup_{i=1}^n A_i\right) - \left((-1)^k \sum_{1 \leq i_1 < \dots < i_k < i_{k+1}} P(A_{i_1} \cap \dots \cap A_{i_{k+1}}) + \dots\right) \geq P\left(\bigcup_{i=1}^n A_i\right)$$

De modo que:

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \dots \\ + (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_{k-1} < i_k} P(A_{i_1} \cap \dots \cap A_{i_k})$$

Se k for par, o termo k será negativo e o termo $k+1$ será positivo, de modo que a desigualdade anterior se inverte, ao fazer a subtração dos $k+j$, $j \in \{1, \dots, n-k\}$ termos na igualdade. \square

Exercise 2.11 (BJ15). Suponha que n cartas numeradas de 1 até n sejam embaralhadas e retiradas uma por uma, sem reposição, até todas as cartas serem retiradas. Qual a probabilidade de que para pelo menos uma carta, o número da carta coincida com o número da retirada?

Proof. Seja A_i o evento em que o número da carta i coincidiu com o número da retirada. Podemos ver que, o caso em que para pelo menos uma delas coincida é equivalente a $\bigcup_{i=1}^n A_i$. Dessa maneira, podemos ver que a probabilidade de isso ocorrer é:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \dots \\ + (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k});$$

O primeiro termo pode ser demonstrado como sendo:

$$\sum_{i=1}^n P(A_i) = P(A_1) + P(A_2) + \dots + P(A_n) = \sum_{i=1}^n \frac{1}{n} = 1$$

Para o termo de intercessão dois a dois, temos que a probabilidade de que o número na primeira carta ser igual a o número da retirada é de $1/n$, e o da segunda carta o ser é de $1/(n-1)$, e como temos $\binom{n}{2}$ combinações diferentes de retiradas, temos que a probabilidade do segundo termo é:

$$\sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) = P(A_1 \cap A_2) + P(A_1 \cap A_3) + \dots + P(A_{n-1} \cap A_n) \\ = \frac{\binom{n}{2}}{n \cdot (n-1)} = \frac{n!}{(n-2)!2!} \cdot \frac{1}{n \cdot (n-1)} = \frac{n!}{n!2!} = \frac{1}{2!}$$

Assim, podemos ver que para qualquer termo teremos:

$$\sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) = \frac{1}{k!}$$

De modo que a probabilidade da união dos eventos se resume à série:

$$P\left(\bigcup_{i=1}^n A_i\right) = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{k-1} \frac{1}{k!}$$

\square

Exercise 2.12 (BJ16). Seja (Ω, \mathcal{A}, P) um espaço de probabilidade e suponha que todos os conjuntos abaixo pertençam a \mathcal{A} . Prove:

- a) Se os A_n são disjuntos e $P(B|A_n) \geq c$ para todo n , então $P(B|\bigcup_{n=1}^k A_n) \geq c$ (pode supor $P(A_n) > 0$ para todo n).

Proof. Sabemos que $A_i \cap A_j = \emptyset$, $\forall i, j$. Dito isso, podemos ver que a seguinte relação é válida:

$$P(B|A_n) = \frac{P(A_n \cap B)}{P(A_n)} \geq c$$

$$P(A_n \cap B) \geq cP(A_n) \quad (21)$$

Além disso, podemos desenvolver $P(B|\bigcup_{n=1}^k A_n)$ da seguinte forma:

$$\begin{aligned} P\left(B \mid \bigcup_{n=1}^k A_n\right) &= \frac{P(B \cap (A_1 \cup A_2 \cdots \cup A_k))}{P(\bigcup_{n=1}^k A_n)} \\ &= \frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_k \cap B))}{\sum_{n=1}^k P(A_n)} \\ P\left(B \mid \bigcup_{n=1}^k A_n\right) &= \frac{\sum_{n=1}^k P(A_n \cap B)}{\sum_{n=1}^k P(A_n)} \end{aligned} \quad (22)$$

O denominador de (22) é simplesmente o somatório das probabilidades dos A_n 's pelo fato de que eles são disjuntos (definidos no enunciado da questão). Agora, considerando que a relação (21) é válida para todos os A_n 's, vamos somar todas as probabilidades para os $n \in \{1, 2, \dots, k\}$:

$$\begin{aligned} P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_k \cap B) &\geq cP(A_1) + cP(A_2) + \cdots + cP(A_k) \\ \sum_{n=1}^k P(A_n \cap B) &\geq \sum_{n=1}^k cP(A_n) \\ \sum_{n=1}^k P(A_n \cap B) &\geq c \sum_{n=1}^k P(A_n) \\ \frac{\sum_{n=1}^k P(A_n \cap B)}{\sum_{n=1}^k P(A_n)} &\geq c \\ P\left(B \mid \bigcup_{n=1}^k A_n\right) &\geq c \end{aligned}$$

□

- b) O item (a) com “=” no lugar de “ \geq ”.

Proof. Substituindo o sinal de \geq em (22) por uma igualdade, a prova é igual ao já realizado no item anterior. □

- c) Se $A_n \supset A_{n+1}$ e $P(A_{n+1}|A_n) \leq \frac{1}{2}$ para todo n , então $P(A_n) \rightarrow 0$ quando $n \rightarrow \infty$.

Proof. Consideremos o caso inicial, com A_1 e A_2 . Disso tem-se que:

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} \leq \frac{1}{2}$$

Como $A_1 \supset A_2$, $P(A_1 \cap A_2) = P(A_2)$. Logo:

$$\frac{P(A_2)}{P(A_1)} \leq \frac{1}{2} \Rightarrow P(A_2) \leq \frac{1}{2}P(A_1)$$

Para o caso seguinte, com A_2 e A_3 , temos que:

$$\begin{aligned} P(A_3|A_2) &= \frac{P(A_2 \cap A_3)}{P(A_2)} \leq \frac{1}{2} \\ \frac{P(A_3)}{P(A_2)} &\leq \frac{1}{2} \Rightarrow P(A_3) \leq \frac{1}{2}P(A_2) \end{aligned}$$

E como $P(A_2) \leq \frac{1}{2}P(A_1)$, temos que $P(A_3) \leq \frac{1}{4}P(A_1)$. Assim, já é possível identificar que, para qualquer n temos que:

$$\begin{aligned} P(A_n) &\leq \frac{1}{2^{n-1}}P(A_1) \\ \lim_{n \rightarrow \infty} P(A_n) &\leq \lim_{n \rightarrow \infty} \frac{1}{2^{n-1}}P(A_1) = 0 \end{aligned}$$

Assim, independentemente do valor de $P(A_1)$, o valor $P(A_n) \rightarrow 0$ conforme $n \rightarrow \infty$. □

d) Se os A_n são disjuntos e $P(B|A_n) = P(C|A_n)$ para todo n , então

$$P(B \cup A_n) = P(C \cup A_n)$$

Proof. Como os A_n s são disjuntos, temos que:

$$\begin{aligned} P(B|A_n) &= \frac{P(B \cap (\cup A_n))}{P(\cup A_n)} \\ &= \frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_n \cap B))}{\sum P(A_n)} \\ &= \frac{\sum P(A_n \cap B)}{\sum P(A_n)} \end{aligned}$$

Para C temos a mesma relação:

$$P(C|A_n) = \frac{\sum P(A_n \cap C)}{\sum P(A_n)}$$

E disso temos que:

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

Como, por hipótese, temos que $P(B|A_n) = P(C|A_n) \Rightarrow P(A_n \cap B) = P(A_n \cap C)$, de modo que, como os A_n s são disjuntos, $\sum P(A_n \cap B) = \sum P(A_n \cap C)$, logo:

$$\frac{\sum P(A_n \cap B)}{\sum P(A_n)} = \frac{\sum P(A_n \cap C)}{\sum P(A_n)}$$

□

e) Se A_1, A_2, \dots são disjuntos e $\cup A_n = \Omega$, então:

$$P(B|C) = \sum_n P(A_n|C)P(B|A_n \cap C)$$

Proof. Pelo Teorema da Multiplicação, temos que $P(A_n \cap B \cap C)$ pode ser escrito como:

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

É importante notar que essa representação não é única, mas apenas conveniente para o problema em questão. Podemos então somar para todos os A_n s:

$$\sum P(A_n \cap B \cap C) = \sum P(B|A_n \cap C)P(A_n \cap C)P(C) = P(C) \sum P(B|A_n \cap C)P(A_n \cap C)$$

Como os A_n s formam uma partição de Ω , $\sum P(A_n \cap B \cap C) = P(B \cap C)$. Logo:

$$\begin{aligned} P(B|C) &= \frac{P(B \cap C)}{P(C)} \\ &= \frac{P(C) \sum P(B|A_n \cap C)P(A_n \cap C)}{P(C)} \\ &= \sum P(B|A_n \cap C)P(A_n \cap C) \end{aligned}$$

□

Exercise 2.13 (BJ17). Suponha que a ocorrência ou não de chuva dependa das condições do tempo no dia imediatamente anterior. Admita-se que se chove hoje, choverá amanhã com probabilidade de 0,7 e que se não chove hoje choverá amanhã com probabilidade 0,4. Sabendo-se que choveu hoje, calcule a probabilidade de que choverá depois de amanhã.

Proof. Sejam os eventos C_n = “Choveu no dia de hoje”, NC_n = “Não choveu no dia de hoje”. De maneira similar, C_{n+1} indica que choverá amanhã, C_{n+2} que choverá depois de amanhã e assim por diante. Sabemos pelo enunciado as seguintes probabilidades:

$$\begin{aligned} P(C_{n+1}|C_n) &= 0,7, \quad P(NC_{n+1}|C_n) = 0,3 \\ P(C_{n+1}|NC_n) &= 0,4, \quad P(NC_{n+1}|NC_n) = 0,6 \end{aligned}$$

Além disso, como os eventos Chover e Não-Chover formam uma partição (são eventos complementares), pelo Teorema da Probabilidade Total temos que a probabilidade de chover depois de amanhã é dada por:

$$P(C_{n+2}) = P(C_{n+2}|C_{n+1})P(C_{n+1}) + P(C_{n+2}|NC_{n+1})P(NC_{n+1}) \quad (23)$$

É fácil perceber que $P(C_{n+2}|C_{n+1}) = P(C_{n+1}|C_n)$ e de maneira similar que $P(C_{n+2}|NC_{n+1}) = P(C_{n+1}|NC_n)$. Ainda assim, é necessário encontrar as probabilidades $P(C_{n+1})$ e $P(NC_{n+1})$. Como sabemos que choveu hoje, $P(C_n) = 1$ e $P(NC_n) = 0$, de modo que:

$$\begin{aligned} P(C_{n+1}) &= P(C_{n+1}|C_n)P(C_n) + P(C_{n+1}|NC_n)P(NC_n) \\ &= 0,7 \times 1 + 0,4 \times 0 = 0,7 \\ P(NC_{n+1}) &= P(NC_{n+1}|C_n)P(C_n) + P(NC_{n+1}|NC_n)P(NC_n) \\ &= 0,3 \times 1 + 0,6 \times 0 = 0,3 \end{aligned}$$

Substituindo esses valores em (23), temos:

$$\begin{aligned}
P(C_{n+2}) &= P(C_{n+1}|C_n) \times 0,7 + P(C_{n+1}|NC_n) \times 0,3 \\
&= 0,7 \times 0,7 + 0,4 \times 0,3 = 0,49 + 0,12 = 0,61
\end{aligned}$$

□

Exercise 2.14 (BJ18). Certo experimento consiste em lançar um dado equilibrado duas vezes, independentemente. Dado que os dois números sejam diferentes, qual a probabilidade condicional de:

a) Pelo menos um dos números ser 6?

Proof. Sejam A_1 e A_2 os lançamentos do primeiro e do segundo dado, respectivamente. Sabemos que $P(A_1 = A_2) = 0$. Disso temos que:

$$\begin{aligned}
P((A_1 = 6) \cup (A_2 = 6)) &= P(A_1 = 6) + P(A_2 = 6) - P((A_1 = 6) \cap (A_2 = 6)) \\
&= \frac{1}{6} + \frac{1}{6} - 0 \\
&= \frac{1}{3}
\end{aligned}$$

□

b) A soma dos números ser 8?

Proof. Considere o evento $S = x, x \in \{2, 3, \dots, 12\}$ o resultado da soma dos lançamentos A_1 e A_2 . Utilizando o Teorema da Probabilidade Total, podemos decompor a probabilidade da soma ser igual a 8 da seguinte forma:

$$\begin{aligned}
P(S = 8) &= P(S = 8|A_1 = 1)P(A_1 = 1) + P(S = 8|A_1 = 2)P(A_1 = 2) + \dots + P(S = 8|A_1 = 6)P(A_1 = 6) \\
&= 0 \times \frac{1}{6} + \frac{1}{5} \times \frac{1}{6} + \frac{1}{5} \times \frac{1}{6} + 0 \times \frac{1}{6} + \frac{1}{5} \times \frac{1}{6} + \frac{1}{5} \times \frac{1}{6} \\
&= \frac{1}{30} + \frac{1}{30} + \frac{1}{30} + \frac{1}{30} \\
&= \frac{4}{30}
\end{aligned}$$

□

Exercise 2.15 (BJ19). Em teste de múltipla escolha, a probabilidade do aluno saber a resposta é p . Havendo m escolhas, se ele sabe a resposta ele responde corretamente com probabilidade 1; se não sabe, ele responde corretamente com probabilidade $\frac{1}{m}$. Qual a probabilidade de que ele soubesse a resposta dado que a pergunta foi respondida corretamente? Calcule o limite dessa probabilidade quando (i) $m \rightarrow \infty$ com p fixo e (ii) $p \rightarrow 0$ com m fixo.

Proof. Sejam: $P(S) = p$ a probabilidade de saber a resposta, $P(A|S) = 1$ a probabilidade de acertar, dado que sabia a resposta, $P(A|NS) = \frac{1}{m}$ a probabilidade de acertar, dado que não sabia a resposta e $P(NA|NS) = \frac{m-1}{m}$ a probabilidade de não acertar, dado que não sabe a resposta. Sabemos que os eventos S e NS são complementares, assim como A e NA . Queremos encontrar $P(S|A)$, que é dada por:

$$\begin{aligned}
P(S|A) &= \frac{P(S \cap A)}{P(A)} \\
&= \frac{P(A|S)P(S)}{P(A|S)P(S) + P(A|NS)P(NS)} \\
&= \frac{1 \times p}{1 \times p + \frac{1}{m} \times (1 - p)} \\
&= \frac{p}{\frac{mp+1-p}{m}}
\end{aligned}$$

De modo que, simplificando a última expressão:

$$P(S|A) = \frac{mp}{p(m-1)+1} \quad (24)$$

Agora, calculando os limites temos:

• (i)

$$\lim_{p \rightarrow 0} \frac{mp}{p(m-1)+1} = \frac{0}{1} = 0$$

• (ii)

$$\lim_{m \rightarrow \infty} \frac{mp}{p(m-1)+1} \xrightarrow{\text{L'Hôpital}} \frac{\frac{\partial}{\partial m} mp}{\frac{\partial}{\partial m} p(m-1)+1} = \frac{p}{p} = 1$$

□

Exercise 2.16 (BJ20). Durante o mês de novembro a probabilidade de chuva é de 0,3. O Fluminense ganha um jogo em um dia de chuva com a probabilidade de 0,4; em um dia sem chuva com a probabilidade 0,6. Se ganhou um jogo em novembro, qual é a probabilidade de que choveu neste dia?

Proof. Sejam os seguintes eventos: $P(C) = 0,3$ é a probabilidade de chover em novembro, $P(NC) = 0,7$ é a probabilidade de não chover em novembro, $P(V|C) = 0,4$ é a probabilidade de vitória, dado que choveu no dia, $P(D|C) = 0,6$ é a probabilidade de derrota, dado que choveu no dia, $P(V|NC) = 0,6$ é a probabilidade de vitória, dado que não choveu no dia e $P(D|NC) = 0,4$ é a probabilidade de derrota, dado que não choveu no dia. Pelo Teorema da Probabilidade Total, temos que:

$$\begin{aligned} P(V) &= P(V|C)P(C) + P(V|NC)P(NC) \\ &= 0,4 \times 0,3 + 0,6 \times 0,7 \\ &= 0,54 \end{aligned}$$

Além disso, temos que o evento $P(C \cap V) = P(V|C)P(C)$, logo:

$$P(C \cap V) = P(V|C)P(C) = 0,4 \times 0,3 = 0,12$$

Assim, temos que a probabilidade de ter chovido, dado que o Fluminense ganhou o jogo em novembro é de:

$$P(C|V) = \frac{P(C \cap V)}{P(V)} = \frac{0,12}{0,54} = \frac{2}{9}$$

□

3 Variáveis Aleatórias

3.1 Variáveis aleatórias e funções de distribuição

Example 3.1. Considere um experimento em que uma moeda é lançada duas vezes. Seja X = total de caras nos dois lançamentos. Denotemos o evento cara como H e coroa como T . Logo:

Espaço Amostral (Ω)	X
HT	1
TH	1
HH	2
TT	0

Logo, $X : \mathcal{F} \rightarrow \mathbb{R}$. Vale também que, $\forall x$ valor na imagem de X , $X^{-1}(x) \in \mathcal{F}$. Por exemplo:

$$x = 1 \Rightarrow X^{-1}(1) = \{HT, TH\}$$

$$x = 2 \Rightarrow X^{-1}(2) = \{HH\}$$

$$x = 0 \Rightarrow X^{-1}(0) = \{TT\}$$

Definition 3.1 (Variável aleatória). Seja (Ω, \mathcal{F}, P) um espaço de probabilidades. Uma função $X : \mathcal{F} \rightarrow \mathbb{R}$ é variável aleatória se $[x \in I] \in \mathcal{F}$, $I \in \mathbb{R}$ (ou, equivalentemente, se $\{\omega : X(\omega) \in I\} \in \mathcal{F}$; $X^{-1}(I) \in \mathcal{F}$).

Definition 3.2 (Distribuição Acumulada). Considere um espaço de probabilidades (Ω, \mathcal{F}, P) e $X : \mathcal{F} \rightarrow \mathbb{R}$ uma variável aleatória, defina $F(r) = P(X \leq r) = P(\{\omega : X(\omega) \leq r\})$.

Example 3.2. Seja X = número de caras em dois lançamentos de moeda (honesta). Temos que as probabilidades de X são dadas por:

$$P(X = 0) = P(\{TT\}) = \frac{1}{4}$$

$$P(X = 1) = P(\{TH, HT\}) = \frac{2}{4}$$

$$P(X = 2) = P(\{HH\}) = \frac{1}{4}$$

Para encontrarmos a função de distribuição acumulada, podemos particinar o espaço e “acumular” as probabilidades. Para $r < 0$:

$$F(r) = P([X \leq r]) = P(\emptyset) = 0$$

Para $r \in [0, 1)$:

$$F(r) = P([X \leq r]) = P(X \leq 0) = \frac{1}{4}$$

Para $r \in [1, 2)$:

$$F(r) = P([X \leq r]) = P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{3}{4}$$

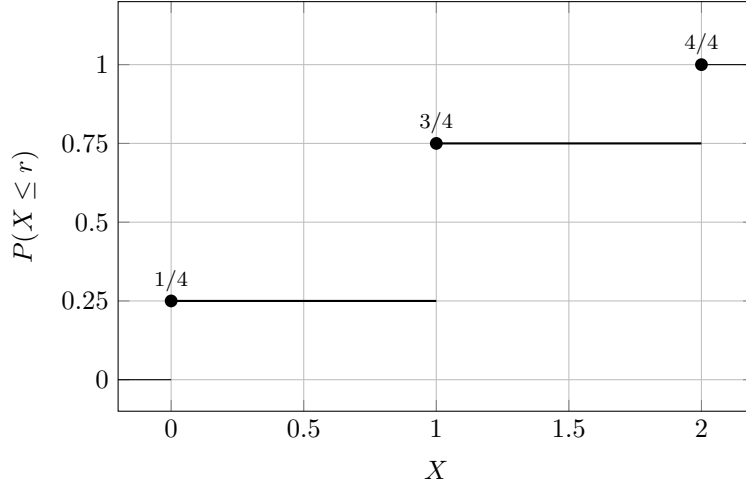
Para $r \geq 2$:

$$F(r) = P([X \leq r]) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 1$$

Logo, F é dada por:

$$F(r) = \begin{cases} 0, & r < 0 \\ \frac{1}{4}, & r \in [0, 1) \\ \frac{3}{4}, & r \in [1, 2) \\ 1, & r \geq 2 \end{cases}$$

Distribuição de probabilidades acumulada



Theorem 3.1 (Propriedades da distribuição acumulada). *Seja X uma variável aleatória definida em (Ω, \mathcal{F}, P) , então a f.d.a. de X (F_X ou F) verifica:*

- a) F é monótona não decrescente;
- b) F é contínua à direita;
- c) $\lim_{t \rightarrow -\infty} F(t) = 0$ e $\lim_{t \rightarrow \infty} F(t) = 1$.

Proof.

- a) Dados $a, b \in \mathbb{R} : a \leq b$; $[X \leq a] \subseteq [X \leq b] \Rightarrow P([X \leq a]) \leq P([X \leq b]) \Rightarrow F(a) \leq F(b)$.
- b) Se $X_n \downarrow x$, quando $n \rightarrow \infty$, temos que $\{[X \leq x_n]\}_{n \geq 1}$ é tal que $\bigcap_{n \geq 1} [X \leq x_n] = [X \leq x]$. Isso significa que $[X \leq x]$ acontece se e somente se $[X \leq x_n] \forall n$. Além disso, $[X \leq x_n] \downarrow [X \leq x]$ quando $n \rightarrow \infty$, logo, pela continuidade da função de probabilidade $P([X \leq x_n]) \downarrow P([X \leq x]), n \rightarrow \infty$.
- c) Considere agora que $x_n \downarrow -\infty \Rightarrow [X \leq x_n] \downarrow \emptyset, n \rightarrow \infty \Rightarrow F(x_n) = P([X \leq x_n]) \downarrow P(\emptyset) = 0, n \rightarrow \infty$. Se $x_n \uparrow \infty \Rightarrow [X \leq x_n] \uparrow \Omega, n \rightarrow \infty \Rightarrow F(x_n) = P([X \leq x_n]) \uparrow P(\Omega) = 1, n \rightarrow \infty$.

□

Theorem 3.2. *Se F é a f.d.a. da variável aleatória X , então:*

- a) *Existem e são finitos os limites laterais $\lim_{t \rightarrow r-} F(t), \lim_{t \rightarrow r+} F(t), \forall r \in \mathbb{R}$ e $\lim_{t \rightarrow r-} F(t) \leq \lim_{t \rightarrow r+} F(t)$;*
- b) $\lim_{t \rightarrow r+} F(t) = F(r), \forall r \in \mathbb{R}$;
- c) F é descontínua em $r, r \in \mathbb{R}$ se e somente se $\lim_{t \rightarrow r-} F(t) < F(r)$, com um salto de tamanho $F(r) - \lim_{t \rightarrow r-} F(t)$;
- d) $\forall r \in \mathbb{R}, P(X = r) = F(r) - \lim_{t \rightarrow r-} F(t)$;
- e) *Existem no máximo um total enumerável de descontinuidades em F .*

Proof.

- a) F é monótona e limitada ($0 \leq F \leq 1$). Logo, os limites laterais existem e são limitados.
- b) Como F é monótona não-decrescente, $\forall x, y : x \leq y \Rightarrow F(x) \leq F(y)$. Logo $\lim_{t \rightarrow r^-} F(t) \leq \lim_{t \rightarrow r^+} F(t)$.
- c) Como F é monótona não-decrescente, uma descontinuidade só ocorre se e somente se $\lim_{t \rightarrow r^-} F(t) < \lim_{t \rightarrow r^+} F(t) = F(r)$.
- d) Seja $r \in \mathbb{R}$. $[X \leq r] = \bigcap_{n=1}^{\infty} (r - \frac{1}{n} < x \leq r)$, logo:

$$\begin{aligned}
P([X = r]) &= P\left(\bigcap_{n=1}^{\infty} \left(r - \frac{1}{n} < x \leq r\right)\right) \\
&\Downarrow (\text{Teorema da continuidade}) \\
&= \lim_{n \rightarrow \infty} P\left(\left(r - \frac{1}{n} < x \leq r\right)\right) \\
&= \lim_{n \rightarrow \infty} \left(F(r) - F\left(r - \frac{1}{n}\right)\right) \\
&= F(r) - \lim_{n \rightarrow \infty} F\left(r - \frac{1}{n}\right) \\
P([X = r]) &= F(r) - \lim_{t \rightarrow r^-} F(t)
\end{aligned}$$

- e) Seja \mathcal{D} o conjunto de pontos de descontinuidades de F , e seja $\lim_{t \rightarrow x^-} F(t) = F(x^-)$. Logo:

$$\mathcal{D} = \{x \in \mathbb{R} : F(x) - F(x^-) > 0\}$$

Seja \mathcal{D}_n o conjunto de pontos para os quais a amplitude do salto é maior ou igual a $\frac{1}{n}$. Logo:

$$\mathcal{D}_n = \left\{x \in \mathbb{R} : F(x) - F(x^-) \geq \frac{1}{n}\right\} \Rightarrow \#\mathcal{D} = |\mathcal{D}| \leq n$$

Se $x \in \mathcal{D} \Rightarrow \exists n_0 > 1 : F(x) - F(x^-) \geq \frac{1}{n_0} \Rightarrow x \in \bigcup_{n=1}^{\infty} \mathcal{D}_n$. Se $x \in \bigcup_{n=1}^{\infty} \mathcal{D}_n \Rightarrow \exists n_1 : x \in \mathcal{D}_{n_1} \Rightarrow x \in \mathcal{D}$. \mathcal{D} portanto é a união enumerável de conjuntos finitos, logo é enumerável. \square

3.2 Natureza das variáveis aleatórias

- a) X é uma variável aleatória discreta se os valores que ela toma pertencem a um conjunto enumerável, logo $X : \Omega \rightarrow \{x_1, x_2, \dots\}$ (ou seja, $X(\omega) \in \{x_1, x_2, \dots\}, \forall \omega \in \Omega$) e $P : \{x_1, x_2, \dots\} \rightarrow [0, 1]$ é dado por $P(x_i) = P\{\omega : \omega \in \Omega \text{ e } X(\omega) = x_i\} \forall i \geq 1$.
- b) X é uma variável aleatória absolutamente contínua se $\exists f$ (uma função) tal que $f(x) \geq 0, \forall x \in \mathbb{R}$ e $F_X(x) = \int_{-\infty}^x f(t)dt$ (onde f é chamada de densidade de X).

Sob **(a)** temos que $[X \leq x] = \bigcup_{i: x_i \leq x} [X = x_i]$. Logo $F_X(x) = \sum_{i: x_i \leq x} P(x_i)$.

Sob **(b)** estamos afirmando que F_X é a integral de f (ou seja, f é a sua derivada) para todo x exceto em um conjunto de medida de Lebesgue nula, ou seja, se seu comprimento for zero ($\int_a^a f(t)dt = 0$). Ainda sob

(b), se f é uma função de densidade podemos definir $F(x) = \int_{-\infty}^x f(t)dt$ e F verifica:

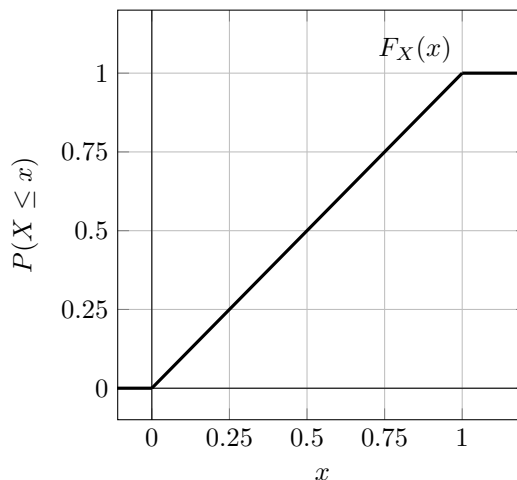
1. $x \leq y \Rightarrow F(x) \leq F(y)$;
2. Se $x_n \downarrow x \Rightarrow F(x_n) \downarrow F(x)$;
3. Se $x_n \downarrow -\infty \Rightarrow F(x_n) \downarrow 0$ e se $x_n \uparrow \infty \Rightarrow F(x_n) \uparrow 1$.

Dada uma variável aleatória com distribuição F_X , X tem densidade se:

- (i) F_X é contínua;
- (ii) F_X é derivável por partes (ou derivável no interior de um número finito ou enumerável de intervalos fechados cuja união é igual a \mathbb{R}), ou derivável para todo x exceto um número finito (enumerável) de pontos.

Example 3.3.

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x, & x \in [0, 1] \\ 1, & x > 1 \end{cases}$$

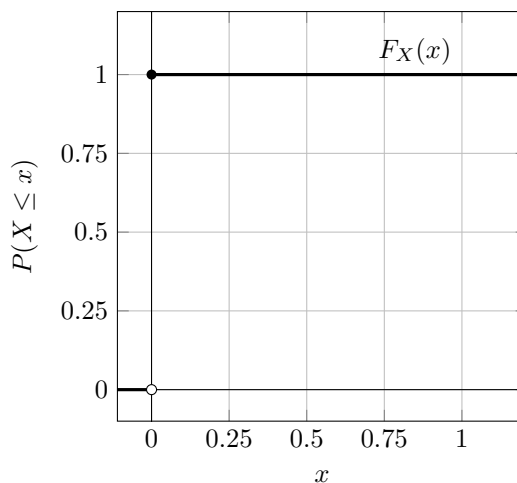


Notas:

- F_X é contínua;
- $\{0, 1\}$ são pontos sem derivada;
- Podemos definir os seguintes intervalos em que F_X é derivável: $(-\infty, 0)$, $(0, 1)$, $(1, \infty)$;
- $F'_X(x) = \begin{cases} 1, & x \in (0, 1) = f_X(x); \\ 0, & c.c. \end{cases}$;
- $f(0)$ e $f(1)$ podem ser definidos como zero ou um, já que tais definições não alteram $F_X(x) = \int_{-\infty}^x f(t)dt$.

Em contrapartida, considere:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

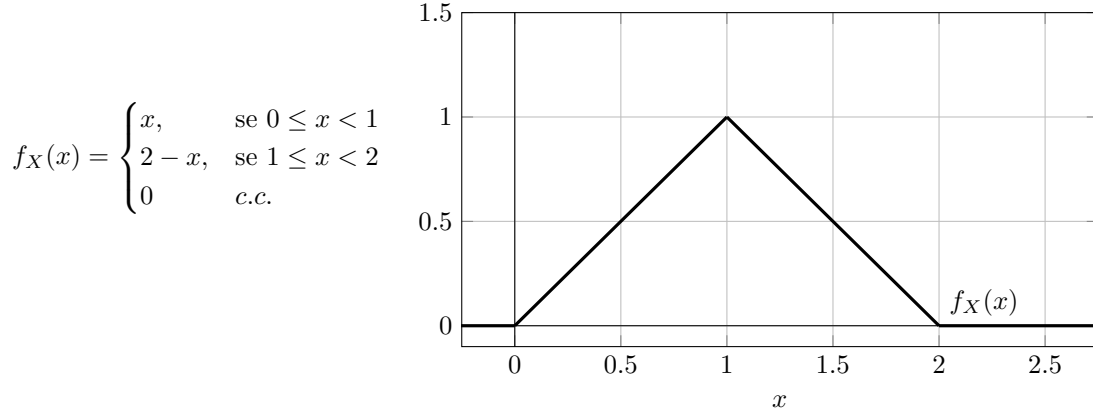


Notas:

- F_X não é contínua;

- $P(X = 0) = \lim_{x \rightarrow 0^+} F_X(x) - \lim_{x \rightarrow 0^-} F_X(x) = 1$.

Example 3.4. Considere a densidade triangular:



Por definição, $f(x) \geq 0 \forall x$. Para verificarmos que a probabilidade total é igual a um, podemos realizar a seguinte integração por partes:

$$\begin{aligned} \int_{-\infty}^x f_X(x) dx &= \int_0^2 f_X(x) dx \\ &= \int_0^1 x dx + \int_1^2 (2 - x) dx \\ &= \frac{x^2}{2} \Big|_0^1 + 2x \Big|_1^2 - \frac{x^2}{2} \Big|_1^2 \\ &= 1 \end{aligned}$$

O que demonstra que $f_X(x)$ é densidade de probabilidade.

Conjecture 3.1. Cada função de distribuição se corresponde com apenas uma distribuição? Não.

Proof. Considere, por exemplo, que a variável aleatória $X \sim N(0, 1)$. Logo, a sua função distribuição de probabilidade é dada por $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ e $\Phi(x)$ é sua acumulada. Vejamos que $X \sim N(0, 1) \iff -X \sim N(0, 1)$:

Seja ω um possível valor de $-X$, devemos calcular $P(-X \leq \omega)$ e provar que $P(-X \leq \omega) = \Phi(\omega)$:

$$P(-X \leq \omega) = P(X \geq -\omega) = 1 - P(X \leq -\omega) = 1 - \Phi(-\omega) = 1 - (1 - \Phi(\omega)) = \Phi(\omega)$$

□

3.3 Variáveis aleatórias e σ -álgebra de Borel

Se X é uma variável aleatória em (Ω, \mathcal{A}, P) , cada evento $[X \leq x] \in \mathcal{A} \forall x \in \mathbb{R}$. Isto é, $[X \in \mathcal{B}]$, onde $[X \in \mathcal{B}] = [X \leq x]$ é um evento e $P(X \in \mathcal{B})$ é bem definido. No entanto, a operacionalidade do sistema (Ω, \mathcal{A}, P) pode ser estendido a todo boreliano (ou seja, a todos os elementos da σ -álgebra de Borel, que é a menor σ -álgebra contendo os intervalos cujos comprimentos estejam bem definidos).

Proposition 3.1. Se X é uma variável aleatória em (Ω, \mathcal{A}, P) , então o evento $[x \in \mathcal{B}] = \{\omega : \omega \in \Omega \text{ e } X(\omega) \in \mathcal{B}\}$ é um evento aleatório para todo \mathcal{B} boreliano (ou seja, $[x \in \mathcal{B}] \in \mathcal{A} \forall \mathcal{B} \in \mathcal{B}$).

Podemos ver que diferentes tipos de intervalos (leia-se borelianos) podem ser mostrados como pertencentes à σ -álgebra, de modo que variáveis aleatórias que operam sobre esses intervalos estarão bem definidas:

1. Se $B = (-\infty, b] \Rightarrow [X \in B] \in \mathcal{A}$ de acordo com a definição de variável aleatória;
2. Se $B = (a, \infty)$, podemos fazer $B = (-\infty, a]^c$. Como o evento $[X \leq a] \in \mathcal{A}$ por definição, sendo \mathcal{A} uma σ -álgebra, deve ocorrer que $[X \leq a]^c = B \in \mathcal{A}$, ou seja, $B \in \mathcal{A}$;
3. Se $B = (a, b] \Rightarrow [X \in B] = [X \in (a, b]] = [X \leq b] - [X \leq a]$. Como $[X \leq b] \in \mathcal{A}$ e $[X \leq a] \in \mathcal{A}$, então $P(X \in B) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$;
4. Se $B = (a, b) \Rightarrow B = \bigcup_{n=1}^{\infty} (a, b - \frac{1}{n}]$. Sabemos que os eventos $(a < X \leq b - \frac{1}{n}) \in \mathcal{A}$ e as suas uniões também pertencem à \mathcal{A} . Quanto à probabilidade, temos $P(X \in B) = P(\bigcup_{n=1}^{\infty} (a < X \leq b - \frac{1}{n})) = \lim_{n \rightarrow \infty} P((a < X \leq b - \frac{1}{n})) = \lim_{n \rightarrow \infty} F_X(b - \frac{1}{n}) - F_X(a) = F_X(b^-) - F_X(a)$;
5. Se $B = \bigcup_{i=1}^n B_i : B_i \in \mathcal{A} \forall i$, e sendo os B_i 's disjuntos, temos que $[X \in B] = \bigcup_{i=1}^n [X \in B_i] \Rightarrow P([X \in B]) = \sum_{i=1}^n P(X \in B_i)$.

Podemos assim reformular os axiomas de Kolmogorov:

- $Ax_1(K)$: $P_X(B) = P(X \in B) \geq 0$;
- $Ax_2(K)$: $P_X(\mathbb{R}) = P(X \in \mathbb{R}) = 1$;
- $Ax_3(K)$: Se $B_1, \dots, B_n \in \mathcal{B}$, com $B_i \cap B_j = \emptyset \forall i \neq j \Rightarrow P_X(\bigcup B_n) = P(X \in \bigcup B_n) = P(\bigcup_n [X \in B_n]) = \sum_n P(X \in B_n)$.

Definition 3.3. A probabilidade P_X definida na σ -álgebra de Borel por $P_X(B) = P(X \in B)$ é a distribuição de X .

Proposition 3.2.

- a) Se X é uma variável aleatória discreta com valores em $\{x_1, x_2, \dots\} \Rightarrow P_X(B) = \sum_{i: x_i \in B} P(x_i)$;
- b) Se X é absolutamente contínua com densidade $f \Rightarrow P_X(B) = \int_B f_X dx$.

3.4 Variáveis contínuas

Proposition 3.3. Se $X \sim f_X$, $y = bx + c$, $b > 0$ e $c \in \mathbb{R} \Rightarrow Y \sim f_Y$ onde $f_Y(y) = \frac{1}{b} f_X(\frac{y-c}{b})$; $y \in \mathbb{R}$, onde c é dito um parâmetro de posição (muitas vezes de posição central) e b um parâmetro de escala.

3.4.1 Exemplos

Example 3.5 (Distribuição Normal).

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Rightarrow f_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Aqui, μ representa a média (posição central) da distribuição e σ^2 a sua variância.

Example 3.6 (Distribuição Cauchy).

$$f(x) = \frac{1}{\pi(1+x^2)} \Rightarrow f_{b,M}(x) = \frac{1}{b} \frac{1}{\pi \left(1 + \left(\frac{x-M}{b}\right)^2\right)} = \frac{b}{\pi(b^2 + (x-M)^2)}$$

Neste caso, M é a mediana da distribuição e b representa a distância entre M e o 1º quartil da distribuição.

Example 3.7 (Distribuições Exponencial e Gamma). Considere $g(x) = e^{-x} I_{0,\infty}(x)$. Sabemos que g é uma distribuição de probabilidade pois:

$$\begin{cases} g(x) \geq 0 \forall x \in (0, \infty) \\ \int_0^{\infty} e^{-x} dx = 1 \end{cases}$$

Vamos agora incluir no formato do tipo exponencial um componente polinomial. Dado $\alpha > 0$, defina $g(x) = x^{\alpha-1} e^{-x}$. Podemos ver que g é integrável, de modo que:

$$\int_0^\infty g(x)dx = \int_0^\infty x^{\alpha-1}e^{-x}dx = \Gamma(\alpha)$$

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x} & x > 0 \\ 0 & \text{c.c.} \end{cases}$$

Defina agora $y = \frac{X}{\beta}$ onde $X \sim \text{Gamma}(\alpha, 1)$ e $\beta > 0$. A densidade de Y pode ser encontrada por meio de:

$$P(Y \leq y) = P\left(\frac{X}{\beta} \leq y\right) = P(X \leq \beta y) \Rightarrow F_Y(y) = F_X(\beta y)$$

$$f_Y(y) = \beta f_X(\beta y) = \beta \frac{(\beta y)^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta y} = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$$

Nesse caso (conhecido como distribuição Gama) $\frac{1}{\beta}$ é um parâmetro de escala e α é um parâmetro de forma. Temos alguns casos especiais, como:

- Se $\alpha = 1$: $Y \sim \text{Exp}(\beta)$;
- Se $\alpha = \frac{n}{2}$, com n inteiro e $\beta = \frac{1}{2}$: $Y \sim \chi^2(n)$

3.5 Variáveis aleatórias multidimensionais

Definition 3.4. A distribuição de probabilidades do vetor aleatório dado por (x_1, \dots, x_n) é uma tabela que associa a cada valor (x_1, \dots, x_n) sua probabilidade $P(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$, onde p é a distribuição conjunta.

Example 3.8. Considere o conjunto de 32 cartas para poker: 7,8,9,10,J,Q,K,A, dos 4 naipes. Duas cartas são retiradas aleatoriamente, sem reposição, e X = número de ases que a pessoa recebe e Y = número de cartas de copas que a pessoa recebe. Qual a probabilidade $P(X = 0, Y = 0)$?

$$P(X = 0, Y = 0) = \frac{\binom{21}{2}}{\binom{32}{2}} = \frac{210}{496}$$

Definition 3.5. A função de distribuição acumulada do par de variáveis aleatórias (X, Y) é dada por:

$$F(X, Y) = P(X \leq x, Y \leq y) = \sum_{\{i: x_i \leq x\}} \sum_{\{j: y_j \leq y\}} P(X = x_i, Y = y_j)$$

Seja $\underline{X} = (X_1, \dots, X_n)$ tal que X_i é variável aleatória definida em $(\Omega, \mathcal{A}, P) \forall i$. Então F , a acumulada de \underline{X} verifica:

- F_1 : F é não decrescente em cada uma das coordenadas;
- F_2 : F é contínua à direita em cada uma das coordenadas;
- F_3 : $\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_n) = 0$ e $\lim_{x_i \rightarrow \infty \forall i} F(x_1, \dots, x_n) = 1$.

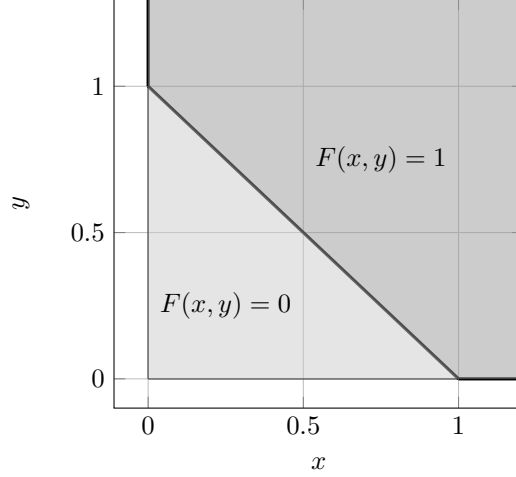
As provas de F_1 e F_2 são de simples construção. Para F_3 temos:

Proof. Considere i fixo e o evento $[X_1 \leq x_1, \dots, X_{i-1} \leq x_{i-1}, X_i \leq -m, X_{i+1} \leq x_{i+1}, \dots, X_n \leq x_n]$. Logo, $F(x_1, \dots, x_{i-1}, -m, x_{i+1}, \dots, x_n) \xrightarrow{m \rightarrow \infty} 0$.

Por outro lado, note que $[X_1 \leq x_1, \dots, X_{i-1} \leq x_{i-1}, X_i \leq m, X_{i+1} \leq x_{i+1}, \dots, X_n \leq x_n] \xrightarrow{m \rightarrow \infty} [X_1 \leq x_1, \dots, X_{i-1} \leq x_{i-1}, X_{i+1} \leq x_{i+1}, \dots, X_n \leq x_n]$ (que é o evento marginal sem o X_i). Já se $x_i \rightarrow \infty \forall i$: $\bigcap_{i=1}^n [X_i \leq x_i] \uparrow \Omega \Rightarrow F(x_1, \dots, x_n) = P(\bigcap_{i=1}^n [X_i \leq x_i]) \uparrow 1, x_i \rightarrow \infty \forall i$. \square

F_1, F_2 e F_3 não são condições suficientes para que F seja uma função de distribuição acumulada. Vejamos um exemplo que segue F_1, F_2 e F_3 e que não é função de distribuição acumulada:

Seja $F_0(x, y) = \begin{cases} 1 & \text{se } x \geq 0, y \geq 0, x + y \geq 1 \\ 0 & \text{c.c.} \end{cases}$. Gráficamente, temos:



É fácil ver que F_0 segue F_1, F_2 e F_3 , mas vejamos que F_0 atribui probabilidade negativa a certos eventos, a ver $[0 \leq X \leq 1, 0 \leq Y \leq 1]$:

$$\begin{aligned}
 F_0(0, 0) &= P(X \leq 0, Y \leq 0) \\
 F_0(1, 1) &= P(X \leq 1, Y \leq 1) \\
 F_0(1, 1) - F_0(1, 0) &= P(X \leq 1, Y \leq 1) - P(X \leq 1, Y \leq 0) = P(X \leq 1, 0 \leq Y \leq 1) \\
 F_0(0, 1) - F_0(0, 0) &= P(X \leq 0, Y \leq 1) - P(X \leq 0, Y \leq 0) = P(X \leq 0, 0 \leq Y \leq 1) \\
 F_0(1, 1) - F_0(1, 0) - F_0(0, 1) - F_0(0, 0) &= P(X \leq 1, 0 \leq Y \leq 1) - P(X \leq 0, 0 \leq Y \leq 1) \\
 &= P(0 \leq X \leq 1, 0 \leq Y \leq 1) = -1
 \end{aligned}$$

Defina $\Delta_{k,I}(g(x_1, \dots, x_k)) = g(x_1, \dots, x_{k-1}, b) - g(x_1, \dots, x_{k-1}, a)$ onde $g : \mathbb{R}^k \rightarrow \mathbb{R}; I = (a, b], a \leq b$. Logo, se $I_1 = (a_1, b_1]$ e $I_2 = (a_2, b_2]$, $F : \mathbb{R}^2 \rightarrow \mathbb{R}$. Então:

$$\begin{aligned}
 \Delta_{1,I_1}(\Delta_{2,I_2}(F(x, y))) &= \Delta_{1,I_1}(F(x, b_2) - F(x, a_2)) \\
 &= F(b_1, b_2) + F(a_1, a_2) - F(a_1, b_2) - F(b_1, a_2) \geq 0 \\
 &= P(a_1 < X \leq b_1, a_2 < Y \leq b_2) \geq 0
 \end{aligned}$$

No geral:

$$\bullet \quad F_4: \Delta_{1,I_1} \Delta_{2,I_2} \dots \Delta_{n,I_n}(F(x_1, \dots, x_n)) \geq 0 \quad \forall I_k = (a_k, b_k]; a_k \leq b_k, k = 1, \dots, n.$$

Definition 3.6. Seja $F : \mathbb{R}^n \rightarrow \mathbb{R}$ seguindo F_1, F_2, F_3 e F_4 , logo F é uma função de distribuição acumulada n-dimensional (ou n-variada).

- **a)** Se o vetor aleatório (X_1, \dots, X_n) toma valores em um conjunto discreto, o vetor é discreto;
- **b)** Se para o vetor aleatório (X_1, \dots, X_n) , F é dada pela forma $F(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f(t_1, \dots, t_n) dt_n \dots dt_1$, $\forall (x_1, \dots, x_n)$ onde $f(t_1, \dots, t_n) \geq 0 \quad \forall (t_1, \dots, t_n) \in \mathbb{R}^n$ então (X_1, \dots, X_n) é um vetor absolutamente contínuo com densidade f (densidade conjunta).

Definition 3.7. A probabilidade definida em \mathcal{B}^n (borelianos em \mathbb{R}^n) por $P(\underline{X} \in B)$ (com $B \in \mathcal{B}^n$) é chamada de distribuição conjunta de $\underline{X} = (X_1, \dots, X_n)$, com notação: $P_{\underline{X}}(B) = P(\underline{X} \in B)$.

Proposition 3.4.

- **a)** Se o vetor aleatório \underline{X} é discreto, $P_{\underline{X}}(B) = \sum_{\{i: x_i \in B\}} P(X_i = x_i) \quad \forall B \in \mathcal{B}^n$;
- **b)** Se \underline{X} é absolutamente contínuo com densidade f , $P_{\underline{X}}(B) = P(\underline{X} \in B) = \int \dots \int_B f(x_1, \dots, x_n) dx_n \dots dx_1$.

3.6 Independência

Definition 3.8. As variáveis aleatórias são (coletivamente) independentes se:

$$P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n P(X_i \in B_i), \forall B_i \in \mathcal{B}^n, \forall i = 1, \dots, n$$

Se X_1, \dots, X_n são coletivamente independentes, então X_{i1}, \dots, X_{ik} são coletivamente independentes $\forall k$.

3.6.1 Critérios ou consequências

Proposition 3.5.

- **a)** Se X_1, \dots, X_n são independentes, então $F_{X_1 \dots X_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i), \forall (x_1, \dots, x_n) \in \mathbb{R}^n$;
- **b)** Se existem funções F_1, \dots, F_n tais que $\lim_{n \rightarrow \infty} F_i(x) = 1, \forall i$ e $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_i(x_i), \forall (x_1, \dots, x_n) \in \mathbb{R}^n \Rightarrow X_1, \dots, X_n$ são independentes e $F_i = F_{X_i}, \forall i$.

Proof.

- **a)** Se X_1, \dots, X_n são coletivamente independentes e tomamos $[X_i \leq x_i] = (-\infty, x_i] = B_i$. Então:

$$\begin{aligned} F_{X_1 \dots X_n}(x_1, \dots, x_n) &= P(X_1 \leq x_1, \dots, X_n \leq x_n) \\ &= P(X_1 \in B_1, \dots, X_n \in B_n) \\ &\stackrel{Ind}{=} \prod_{i=1}^n P(X_i \in B_i) \\ &= \prod_{i=1}^n P(X_i \leq x_i) = \prod_{i=1}^n F_{X_i}(x_i) \quad \forall (x_1, \dots, x_n) \end{aligned}$$

- **b)** Para cada i , $F_{X_i}(x_i) = P(X_i \leq x_i) = \lim_{m \rightarrow \infty} P(X_1 \leq m, \dots, X_{i-1} \leq m, X_i \leq x_i, X_{i+1} \leq m, \dots, X_n \leq m)$, de modo que:

$$\begin{aligned} F_{X_i}(x_i) &= \lim_{m \rightarrow \infty} F_{X_1 \dots X_n}(m, \dots, m, x_i, m, \dots, m) \\ &\stackrel{Hip}{=} \lim_{m \rightarrow \infty} \left(\prod_{j=1}^{i-1} F_j(m) \times F_i(x_i) \times \prod_{j=i+1}^n F_j(m) \right) \\ &= F_i(x_i) \end{aligned}$$

Logo, a marginal de X_i é precisamente $F_i, \forall i$. Devemos ainda verificar que $P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n P(X_i \in B_i) \forall B_i \in \mathcal{B}^n$. Considere $B_i = (a_i, b_i], a_i \leq b_i, a_i, b_i \in \mathbb{R}$. Temos que:

$$\begin{aligned} P(X_1 \in B_1, \dots, X_n \in B_n) &= P(a_1 < X_1 \leq b_1, \dots, a_n < X_n \leq b_n) \\ &= \Delta_{1, I_1} \dots \Delta_{n, I_n} (F_{X_1 \dots X_n}(x_1, \dots, x_n)) \\ &\stackrel{Ind}{=} \Delta_{1, I_1} \dots \Delta_{n, I_n} (F_{X_1}(x_1) \dots F_{X_n}(x_n)) \\ &= [F_{X_1}(b_1) - F_{X_1}(a_1)] \times \dots \times [F_{X_n}(b_n) - F_{X_n}(a_n)] \\ &= \prod_{i=1}^n P(a_i < X_i \leq b_i) = \prod_{i=1}^n P(X_i \in B_i) \end{aligned}$$

□

3.6.2 Caso contínuo

Proposition 3.6.

- **a)** Se X_1, \dots, X_n são independentes e possuem densidades f_{X_1}, \dots, f_{X_n} , respectivamente, então $f_{X_1 \dots X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i) \forall (x_1, \dots, x_n) \in \mathbb{R}^n$ é a densidade conjunta de X_1, \dots, X_n ;
- **b)** Se X_1, \dots, X_n tem densidade conjunta $f_{X_1 \dots X_n}(x_1, \dots, x_n) : f_{X_1 \dots X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i) \forall (x_1, \dots, x_n) \in \mathbb{R}^n$, onde $f_i(x) \geq 0 \forall x : \int_{-\infty}^{\infty} f_i(x) dx = 1 \forall i$, então X_1, \dots, X_n são independentes e f_i é a densidade marginal de $X_i \forall i$.

Proof.

- **a)** Como consequência da proposição 3.5, temos que: $F_{X_1 \dots X_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i), \forall (x_1, \dots, x_n)$. Logo, por definição temos:

$$\prod_{i=1}^n F_{X_i}(x_i) = \prod_{i=1}^n \int_{-\infty}^{x_i} f_{X_i}(t) dt = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{X_1}(t_1) \dots f_{X_n}(t_n) dt_n \dots dt_1$$

Assim, f_{X_1}, \dots, f_{X_n} é a densidade conjunta.

- **b)** Considere:

$$\begin{aligned} F_{X_1 \dots X_n}(x_1, \dots, x_n) &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{X_1 \dots X_n}(t_1, \dots, t_n) dt_n \dots dt_1 \\ &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_1(t_1) \dots f_n(t_n) dt_n \dots dt_1 \\ &= \prod_{i=1}^n \int_{-\infty}^{x_i} f_i(t_i) dt_i \end{aligned}$$

Defina $F_i(x) = \int_{-\infty}^x f_i(t) dt$. Sendo assim:

$$\prod_{i=1}^n \int_{-\infty}^{x_i} f_i(t_i) dt_i = \prod_{i=1}^n F_i(x_i)$$

Note que, pela hipótese nas f_i 's, as F_i 's são acumuladas em particular, e $F_i(x) \rightarrow 1, x \rightarrow \infty$, e pela proposição 3.5: $F_i(x) = F_{X_i}(x_i)$, logo $f_{X_i} = f_i$.

□

3.6.3 Propriedades

- **a)** Se $F(x, y)$ é a função de distribuição acumulada conjunta de (X, Y) , então $F_X(x) = \lim_{y \rightarrow \infty} F(x, y) = F(x, \infty)$ é a função de distribuição acumulada marginal de X ;
- **b)** Se $f(x, y)$ é a função de densidade conjunta de (X, Y) , então $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$ é a densidade marginal de X .

Example 3.9.

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 - 2\rho \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) \right] \right\}$$

Sendo $\sigma_i > 0, i = 1, 2; -1 < \rho < 1; \mu_i \in \mathbb{R}, i = 1, 2$. Logo, $(X, Y) \sim N_2 \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix} \right)$, onde, caso $\rho = 0$, X e Y são independentes.

3.7 Distribuições de funções de vetores

Seja $\underline{X} = (X_1, \dots, X_n)$ um vetor aleatório em (Ω, \mathcal{A}, P) . Seja $Y = g(X_1, \dots, X_n)$. Qual a distribuição de Y ?

- **Nota 1:** Para que Y seja variável aleatória cada $B \in \mathcal{B}$ é necessário que $g^{-1}(B)$ seja mensurável, ou seja:

$$\begin{aligned} g^{-1}(B) &= \{x : g(x) \in B\} \\ &\Downarrow \\ F_Y(y) &= P(g(x) \leq y) \end{aligned}$$

Generalizando, se $Y = g(X_1, \dots, X_n)$:

$$F_Y(y) = P(g(X_1, \dots, X_n) \leq y) = P((X_1, \dots, X_n) \in B_y) = P_{\underline{X}}(B_y)$$

Onde $B_y = \{(x_1, \dots, x_n) : g(x_1, \dots, x_n) \leq y\}$.

- **Nota 2:** Se \underline{X} for discreto:

$$P_Y(y_j) = \sum_{\{i: g(x_i) = y_j\}} P_{\underline{X}}(x_i)$$

Example 3.10. Sejam $X \sim U(0, 1)$ e $Y = -\ln(x)$. Temos que $\forall x$ valor de $X : x \in (-\infty, 0] \cup [1, \infty)$ o valor de $f_X(x) = 0$. Seja $x \in (0, 1) \Leftrightarrow -\ln(x) \in (0, \infty)$, logo $\forall y$ valor de $Y : y \in (0, \infty)$. Calculemos $F_Y(y) = P(Y \leq y)$:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(-\ln(X) \leq y) \\ &= P(\ln(X) \geq -y) \\ &= P(X \geq e^{-y}) \\ &= 1 - P(X < e^{-y}) = 1 - e^{-y} \end{aligned}$$

Assim, temos que $Y \sim Exp(1)$.

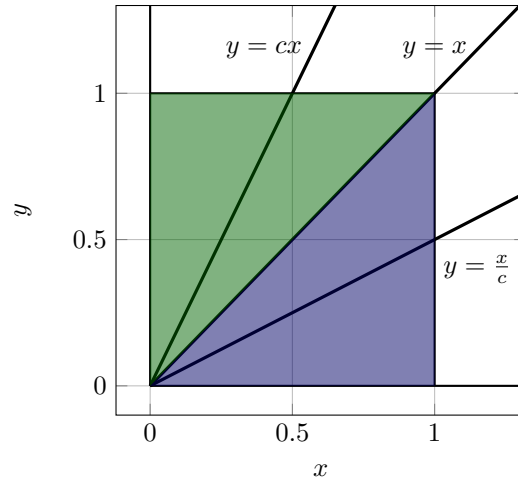
Example 3.11. Sejam $X \perp Y; X \sim U(0, 1); Y \sim U(0, 1); Z = \frac{X}{Y}$. Determinar a distribuição de Z :
Os valores que geram indefinição de Z são: $X = Y = 0$ e $Y = 0, X > 0$, assim a boa definição de Z é no espaço $[0 < X \leq 1, 0 < Y \leq 1]$. Vejamos se esse intervalo contém toda a massa de probabilidade:

$$P([0 < X \leq 1, 0 < Y \leq 1]) = P(0 < X \leq 1) \times P(0 < Y \leq 1) = 1 \times 1 = 1$$

Logo, basta avaliar o conjunto $[0 < X \leq 1, 0 < Y \leq 1] \Rightarrow [Z \in (0, \infty)]$. Assim, calculemos $F_Z(z)$:

$$F_Z(z) = P(Z \leq z) = P\left(\frac{X}{Y} \leq z\right) \Rightarrow \left[\frac{X}{Y} \leq z\right] = [X \leq zY] = \left[\frac{X}{z} \leq Y\right]$$

Sabemos que X e Y pertencem ao intervalo $(0, 1] \times (0, 1]$, de modo que temos duas regiões genéricas para explorar: $z < 1$ e $z > 1$. De maneira gráfica, temos as seguintes regiões (considere $c > 1$):



Podemos ver que a região azul corresponde aos casos onde $z > 1$ e a região verde corresponde aos casos onde $z < 1$. Assim:

- $z < 1$:

$$F_Z(z) = \int_0^z \int_0^{\frac{x}{z}} dy dx = \int_0^z y \Big|_0^{\frac{x}{z}} dx = \int_0^z \frac{x}{z} dx = \frac{1}{z} \times \frac{x^2}{2} \Big|_0^z = \frac{z^2}{2z} = \frac{z}{2}$$

- $z > 1$:

$$F_Z(z) = 1 - \frac{1}{2z}$$

De modo que a distribuição acumulada de Z é dada por:

$$F_Z(z) = \begin{cases} 0 & , z \in (-\infty, 0] \\ \frac{z}{2} & , z \in (0, 1) \\ 1 - \frac{1}{2z} & , z \in [1, \infty) \end{cases}$$

Assim, $F_Z(z) = P\left(\frac{X}{Y} \leq z\right) = P((X, Y) \in B_z)$, onde os conjuntos B_z podem ter formatos diferentes dependendo de z . A densidade será dada pela derivada de $F_Z(z)$ com relação a z :

$$f_Z(z) = \begin{cases} 0 & , z \leq 0 \\ \frac{1}{2} & , z \in (0, 1) \\ \frac{1}{2z^2} & , z \geq 1 \end{cases}$$