

Notas de Aulas

MI401 Probabilidade

Caio Gomes Alves

Instituto de Matemática, Estatística e Computação Científica

Universidade Estadual de Campinas - UNICAMP

Maio 2025

Conteúdos

1	Prefácio	3
2	Definições Básicas	4
2.1	Modelo Probabilístico	4
2.2	Álgebras de Conjuntos	5
2.3	Axiomas de Kolmogorov	7
2.4	Propriedades da medida de probabilidade	7
2.5	Propriedades de probabilidade	9
2.6	Probabilidade Condicional e Independência	9
2.7	Fórmula de Poincaré e Teorema de Bayes	11
2.8	Exercícios	13
3	Variáveis Aleatórias	30
3.1	Variáveis aleatórias e funções de distribuição	30
3.2	Natureza das variáveis aleatórias	32
3.3	Variáveis aleatórias e σ -álgebra de Borel	34
3.4	Variáveis contínuas	35
3.5	Variáveis aleatórias multidimensionais	36
3.6	Independência	38
3.7	Distribuições de funções de vetores	40
3.8	Método do Jacobiano	44
3.9	Exercícios	48
4	Esperança	49
4.1	Definição	49
4.2	Propriedades da esperança	51
4.3	Esperança de funções de variáveis aleatórias	54
4.4	Momentos de uma variável aleatória	55
4.5	Esperanças e funções de vetores	57
4.6	Convergência	60

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:

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

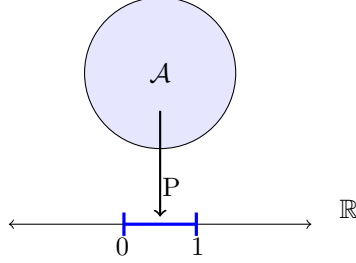
Note que o conjunto A pode ser descrito como a união dos eventos elementares, tais que $A = \{2\} \cup \{4\} \cup \{6\}$. Nesse caso, podemos ver que a probabilidade de A não muda, pois:

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

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: $A \in \mathcal{A}, B \in \mathcal{A}, A - B = A \cap B^c \in \mathcal{A}$.

Prova. 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 $A \cap B = \emptyset \Rightarrow A - B = A \in \mathcal{A}$;
- $A \supset B \Rightarrow A - B = A \cap B^c \in \mathcal{A}$;
- $A \subset B \Rightarrow A - B = \emptyset \in \mathcal{A}$;
- $A = B \Rightarrow A - B = \emptyset \in \mathcal{A}$.

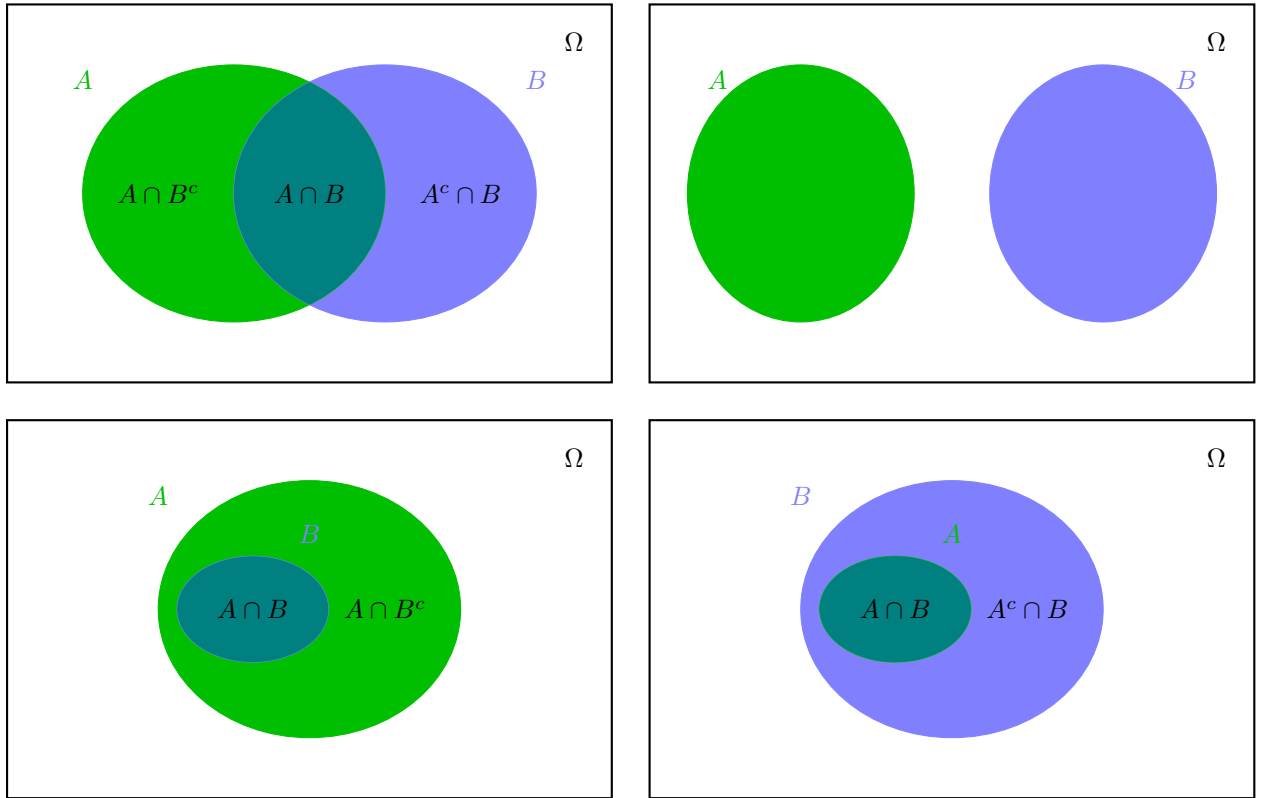


Figure 1: Diferentes relações entre A e B demonstradas por Diagramas de Venn. Note que em todos os casos, $A \cap B^c \in \mathcal{A}$ ou $A \cap B^c = \emptyset \in \mathcal{A}$ ou $A \cap B^c = A \in \mathcal{A}$

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} = todos os subconjuntos de $[0, 1]$ 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 $P(B) = \lim_{i \rightarrow \infty} P(B_i)$.

Prova.

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 $A = \bigcup_{i=1}^{\infty} A_i$ 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:

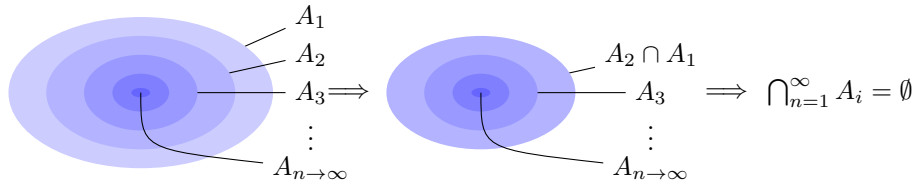
$$P\left(\bigcap_{i=1}^n B_i\right) \xrightarrow{n \rightarrow \infty} (1 - P(A)) = P(A^c) = P(B)$$

□

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:

- $A = \text{"em cada lançamento o valor observado é } \leq 2\text{"}$;
- $B = \text{"a soma dos resultados é igual a 4"}$.

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

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

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:

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

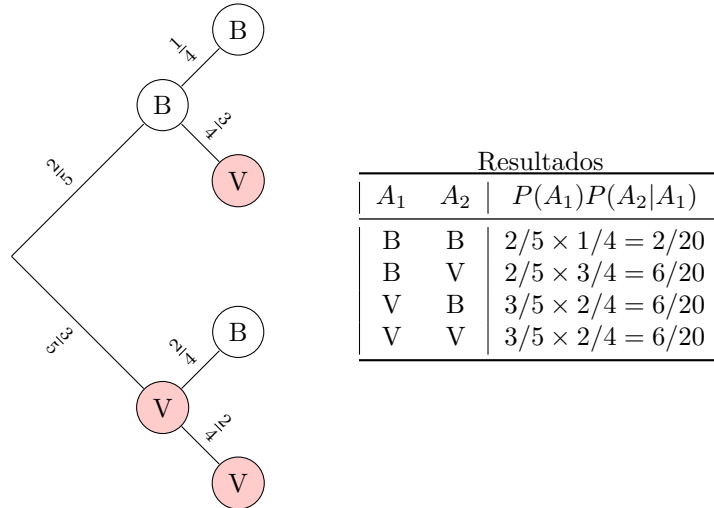
Example 2.5. Considere uma urna com 5 bolas, sendo 3 vermelhas e 2 brancas. O experimento consiste de 2 retiradas sucessivas de uma bola da urna (sem reposição). Considere os eventos $A_1 = \text{Cor da primeira bola}$ e $A_2 = \text{Cor da segunda bola}$:

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

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

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

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:*

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

Prova. 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:

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

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

Resposta. 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 $P(\bigcap_{k=1}^n A_k) \geq 1 - n\epsilon$

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

Resposta. 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 $P(\bigcup_{n=1}^\infty A_n) = 0$.

Resposta. 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 $P(\bigcap_{n=1}^{\infty} A_n) = 1$.

Resposta. 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.

Resposta. 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.

Resposta. 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} .

Resposta. 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 ?

Resposta. 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 $\binom{2n-a-b}{n-a} / \binom{2n}{n}$.

Resposta. 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.

Resposta. 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

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

Resposta. 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.

Resposta. 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.

Resposta. 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:

$$P(V = 1) = \frac{5}{15}$$

$$P(P = 1) = \frac{9}{15}$$

$$P(B = 1) = \frac{1}{15}$$

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.

Resposta. 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.

Resposta. 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.

Resposta. 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.

Resposta. 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

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

Resposta. 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:

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

$$P(B) = P(A \cap B) + P(A^c \cap B) \Rightarrow P(A^c \cap B) = P(B) - P(A \cap B) \quad (13)$$

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.

Resposta. 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*:

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

Resposta. 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:

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

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.

Resposta. 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?

Resposta. 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).

Resposta. 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:

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

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 ”.

Resposta. 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 $n \rightarrow \infty$.

Resposta. 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

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

Resposta. 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:

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

Resposta. 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.

Resposta. 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?

Resposta. 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?

Resposta. 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.

Resposta. 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?

Resposta. 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 $r \in [1, 2)$:

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

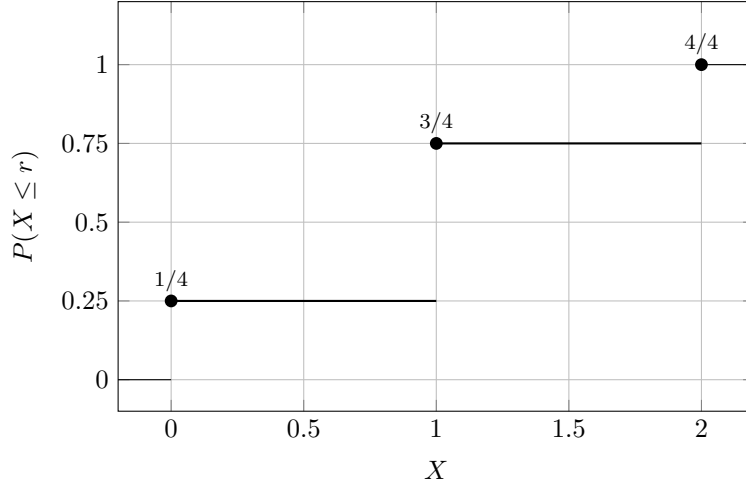
Para $r \geq 2$:

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

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 $\lim_{t \rightarrow \infty} F(t) = 1$.

Prova.

- 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 .*

Prova.

- 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:

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

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:

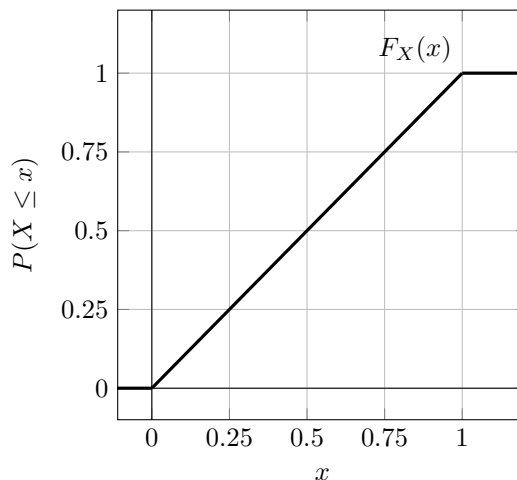
1. $x \leq y \Rightarrow F(x) \leq F(y)$;
2. Se $x_n \downarrow x \Rightarrow F(x_n) \downarrow F(x)$;
3. Se $x_n \downarrow -\infty \Rightarrow F(x_n) \downarrow 0$ e se $x_n \uparrow \infty \Rightarrow F(x_n) \uparrow 1$.

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.

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

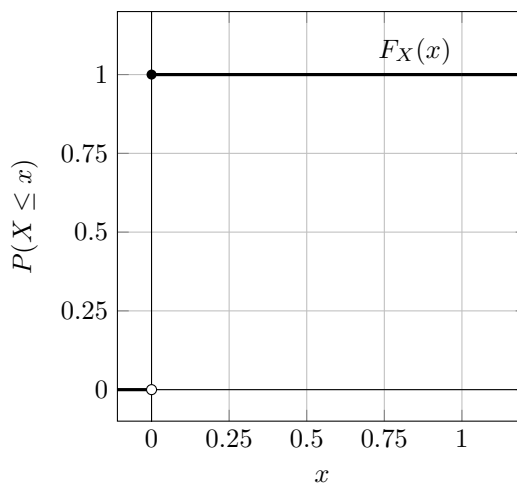


Notas:

- 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:

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

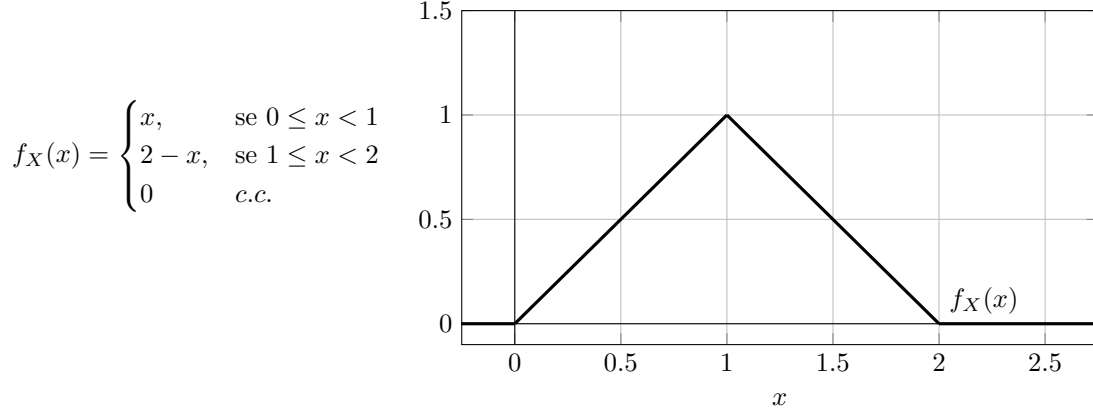


Notas:

- 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.

Prova. 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:

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

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

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 $\lim_{x_i \rightarrow \infty \forall i} F(x_1, \dots, x_n) = 1$.

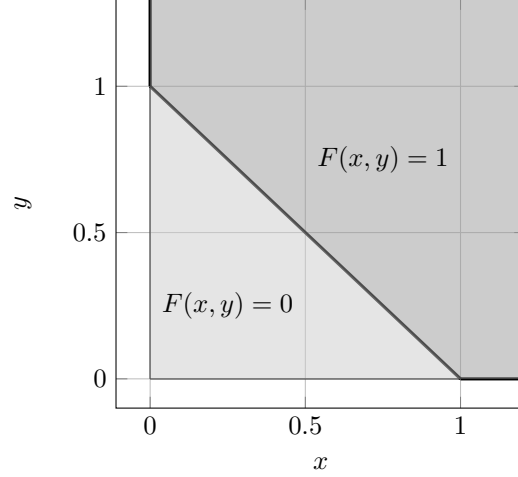
As provas de F_1 e F_2 são de simples construção. Para F_3 temos:

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

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

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 $F_i = F_{X_i}, \forall i$.

Prova.

- **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 $X_i \forall i$.

Prova.

- **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 $Y \sim \text{Exp}(1)$.

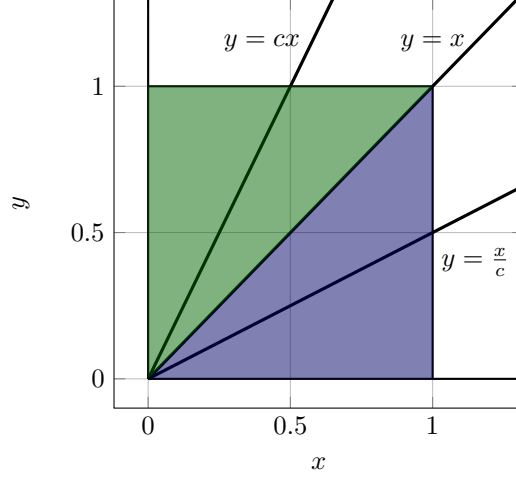
Example 3.11. Sejam $X \perp Y; X \sim U(0, 1); Y \sim U(0, 1); Z = \frac{X}{Y}$. Determinar a distribuição de Z :
Os valores que geram indefinição de Z são: $X = Y = 0$ e $Y = 0, X > 0$, assim a boa definição de Z é no espaço $[0 < X \leq 1, 0 < Y \leq 1]$. Vejamos se esse intervalo contém toda a massa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

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

3.7.1 Distribuição da Soma

Proposition 3.7.

- **a)** Se X e Y tem densidade conjunta $f(x, y) \Rightarrow f_{X+Y}(z) = \int_{-\infty}^{\infty} f(z-t, t) dt = \int_{-\infty}^{\infty} f(t, z-t) dt$;
- **b)** Se $X \perp Y$ e f_X e f_Y são suas marginais, então $f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-t) f_Y(t) dt = \int_{-\infty}^{\infty} f_X(t) f_Y(z-t) dt$.

Prova. Seja $Z = X + Y \Rightarrow [Z \leq z] = [X + Y \leq z] = [(x, y) \in B_z]$. Considerando $B_z = \{(x, y) : x + y \leq z\} = \{(x, y) : x \leq z - y\}$, temos que:

$$F_Z(z) = \int \int_{B_z} f(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f(x, y) dx dy$$

Seja y um valor fixo e defina $s = x + y, ds = dx$. Quando $x = z - y \Rightarrow s = z$, temos:

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^z f(s - y, y) ds dy = \int_{-\infty}^z \int_{-\infty}^{\infty} f(s - y, y) dy ds = \int_{-\infty}^z g(s) ds$$

E g é a densidade de $X + Y$, ou seja, $g(s) = f_{X+Y}(s)$. □

3.7.2 Convolução

Se f_1 e f_2 são densidades de variáveis aleatórias, sua convolução $f_1 * f_2$ é:

$$f_1 * f_2(x) = \int_{-\infty}^{\infty} f_1(x-t)f_2(t)dt$$

Assim, no caso da soma da proposição 3.7, podemos ver que:

$$f_{X+Y}(z) = f_X * f_Y(z)$$

3.7.3 Independência

Proposition 3.8. Se X_1, \dots, X_n são variáveis aleatórias independentes, então funções de famílias disjuntas de $\{X_i\}_{i \geq 1}$ também são independentes.

Prova: Caso especial. Considere $Y_i = g_i(X_i)$. É necessário provar que $F_{Y_1 \dots Y_n}(y_1, \dots, y_n) = \prod_{i=1}^n F_{Y_i}(y_i)$:

$$\begin{aligned} F_{Y_1 \dots Y_n}(y_1, \dots, y_n) &= P(g_1(x_1) \leq y_1, \dots, g_n(x_n) \leq y_n) \\ &= P(X_1 \in g_1^{-1}((-\infty, y_1]), \dots, X_n \in g_n^{-1}((-\infty, y_n])) \\ &= \prod_{i=1}^n P(X_i \in g_i^{-1}((-\infty, y_i])) \\ &= \prod_{i=1}^n P(g_i(X_i) \in (-\infty, y_i]) = \prod_{i=1}^n F_{Y_i}(y_i) \end{aligned}$$

□

Example 3.12. Considere $X \perp Y$, $X \sim \text{Exp}(1)$ e $Y \sim \text{Exp}(1)$. Determine:

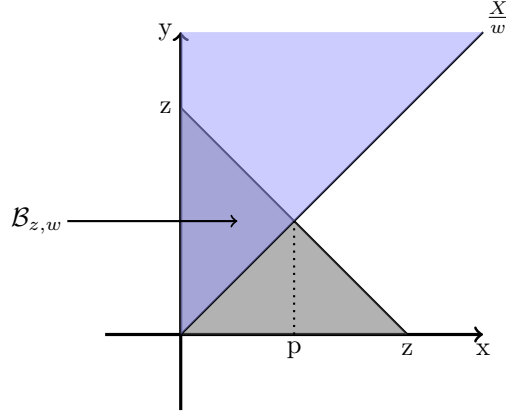
- a) A distribuição de $Z = X + Y$ e $W = \frac{X}{Y}$;
- b) Mostrar que $Z \perp W$.

a)

Como os valores de X e Y são sempre positivos, os valores de Z e W também o serão. Verifiquemos que $F_{ZW}(z, w) = F_Z(z)F_W(w)$:

$$\begin{aligned} P[Z \leq z, W \leq w] &= F_{ZW}(z, w) \\ &= \left[X + Y \leq z, \frac{X}{Y} \leq w \right] \\ &= \left[Y \leq z - X, \frac{X}{w} \leq Y \right] \end{aligned}$$

Vejamos que temos que considerar que $Y \leq z - X$ e que $\frac{X}{w} \leq Y$, ou seja, temos que avaliar as variáveis no seguinte boreliano:



Onde a região em azul claro são os valores onde $Y \geq \frac{X}{w}$, e a região cinza são os valores em que $Y \leq z - X$, o ponto p é dado por:

$$\begin{aligned}\frac{X}{w} &= z - X \Rightarrow z = X \left(\frac{1}{w} + 1 \right) \\ z &= X \left(\frac{w+1}{w} \right) \\ X &= \frac{zw}{w+1}\end{aligned}$$

Assim, estamos interessados em encontrar $P((X, Y) \in \mathcal{B}_{z,w})$, que será:

$$\begin{aligned}P((X, Y) \in \mathcal{B}_{z,w}) &= \int_0^p \int_{\frac{x}{w}}^{z-x} e^{-x} e^{-y} dy dx \\ &= \int_0^{\frac{zw}{w+1}} e^{-x} \left[-e^{-y} \right]_{\frac{x}{w}}^{z-x} dx \\ &= \int_0^{\frac{zw}{w+1}} e^{-x} \left[e^{-\frac{x}{w}} - e^{-z+x} \right] dx \\ &= \int_0^{\frac{zw}{w+1}} e^{-x} \left(\frac{1+w}{w} \right) - e^{-z} dx \\ &= -\frac{w}{(1+w)} e^{-x} \left(\frac{1+w}{w} \right) \Big|_0^{\frac{zw}{w+1}} - e^{-z} x \Big|_0^{\frac{zw}{w+1}} \\ &= \frac{w}{1+w} (1 - e^{-z} - ze^{-z})\end{aligned}$$

Assim, temos que a distribuição de Z e W será dada por:

$$F_{ZW}(z, w) = \begin{cases} 0 & , z \leq 0, w \leq 0 \\ \frac{w}{1+w} (1 - e^{-z} - ze^{-z}) & , z > 0, w > 0 \end{cases}$$

Que é uma distribuição de probabilidade, pois é absolutamente contínua (e por consequência, contínua à direita) e os seguintes limites são bem definidos:

$$\begin{aligned}\lim_{w \rightarrow 0} F_{ZW}(z, w) &= 0 \\ \lim_{z \rightarrow 0} F_{ZW}(z, w) &= 0 \\ \lim_{z \rightarrow \infty, w \rightarrow \infty} F_{ZW}(z, w) &= 1\end{aligned}$$

b)

Temos que as distribuições marginais de Z e W serão:

$$F_Z(z) = \lim_{w \rightarrow \infty} F_{ZW}(z, w) = 1 - e^{-z} - ze^{-z}$$

$$F_W(w) = \lim_{z \rightarrow \infty} F_{ZW}(z, w) = \frac{w}{1+w}$$

E como a distribuição conjunta é o produto das marginais, temos que $Z \perp W$. As densidades serão dadas pelas derivadas da distribuição acumulada conjunta, ou seja:

$$f_{ZW}(z, w) = \frac{\partial}{\partial z} \frac{\partial}{\partial w} \left(\frac{w}{1+w} (1 - e^{-z} - ze^{-z}) \right)$$

$$= \frac{1}{(1+w)^2} ze^{-z} I_{(0,\infty)}(z) I_{(0,\infty)}(w)$$

3.8 Método do Jacobiano

Seja $g : G_0 \rightarrow G$, com $G, G_0 \subseteq \mathbb{R}^n$ e ambos abertos. Então $g(x_1, \dots, x_n) = (g_1(x_1, \dots, x_n), \dots, g_n(x_1, \dots, x_n)) = (y_1, \dots, y_n)$, com g sendo bijetiva, ou seja, para todo y valor de Y , existe \underline{x} valor de X tal que $g(\underline{x}) = y$.

Logo g admite inversa usual $g^{-1} = h$, com $h = (h_1, \dots, h_n)$:

$$x_1 = h_1(y_1, \dots, y_n)$$

$$\vdots$$

$$x_n = h_n(y_1, \dots, y_n)$$

Vamos supor que existem as derivadas parciais $\frac{\partial x_i}{\partial y_j}, \forall i, \forall j$, e que elas são contínuas em G . Desejamos computar: $\int \dots \int_C f_Y(y) dy$, em termos de $\int \dots \int_D f_X(x) dx$.

Example 3.13. Sejam $Y = (Y_1, Y_2) = \left(X_1 + X_2, \frac{X_1}{X_2}\right)$. Teremos então que: $y_1 = g_1(x_1, x_2) = x_1 + x_2$ e $y_2 = g_2(x_1, x_2) = \frac{x_1}{x_2}$. Temos assim os valores dos y 's em termos dos x 's, e desejamos encontrar o contrário:

$$y_1 = x_1 + x_2 \Rightarrow x_1 = y_1 - x_2$$

$$y_2 = \frac{y_1 - x_2}{x_2} \Rightarrow x_2 = \frac{y_1}{y_2 + 1} \Rightarrow x_1 = \frac{y_1 y_2}{y_2 + 1}$$

Agora que temos os valores de X_1 e X_2 em função de Y_1 e Y_2 . Agora, podemos calcular as derivadas parciais de x com relação a y :

$$\frac{\partial x_1}{\partial y_1} = y_2(y_2 + 1)^{-1}$$

$$\frac{\partial x_1}{\partial y_2} = y_1(y_2 + 1)^{-2}$$

$$\frac{\partial x_2}{\partial y_1} = (y_2 + 1)^{-1}$$

$$\frac{\partial x_2}{\partial y_2} = -y_1(y_2 + 1)^{-2}$$

Definimos agora o Jacobiano:

$$J(\underline{x}, \underline{y}) = \det \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix} \quad (25)$$

Dessa forma, o Jacobiano da transformação será:

$$\begin{aligned} J(\underline{x}, \underline{y}) &= \det \begin{bmatrix} y_2(y_2 + 1)^{-1} & y_1(y_2 + 1)^{-2} \\ (y_2 + 1)^{-1} & -y_1(y_2 + 1)^{-2} \end{bmatrix} \\ &= [y_2(y_2 + 1)^{-1}] \cdot [-y_1(y_2 + 1)^{-2}] - [y_1(y_2 + 1)^{-2}] \cdot [(y_2 + 1)^{-1}] \\ &= -y_1(y_2 + 1)^{-2} \end{aligned}$$

Pelo teorema do Jacobiano, temos que:

$$\int \cdots \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n = \int \cdots \int_{g(A)} f(h_1(y_1, \dots, y_n), \dots, h_n(y_1, \dots, y_n)) |J(\underline{x}, \underline{y})| dy_1 \cdots dy_n$$

Se f é integrável em A , com $A \subseteq G_0$ e $h = g^{-1}$. Assim, usando os valores do exemplo 3.12, temos que $X_1 \sim \exp(1)$, $X_2 \sim \exp(1)$, $X_1 \perp X_2$, com densidade conjunta dada por $f_{X_1, X_2}(x_1, x_2) = e^{-x_1 - x_2}$, de modo que:

$$\begin{aligned} f(h_1(y_1, y_2), h_2(y_1, y_2)) |J(\underline{x}, \underline{y})| &= f\left(\frac{y_1 y_2}{y_2 + 1}, \frac{y_1}{y_2 + 1}\right) | -y_1(y_2 + 1)^{-2} | \\ &= \exp\left(-\left[\frac{y_1 y_2}{y_2 + 1} + \frac{y_1}{y_2 + 1}\right]\right) y_1(y_2 + 1)^{-2} \\ &= e^{-y_1} y_1(y_2 + 1)^{-2} \end{aligned}$$

Que é a mesma densidade conjunta encontrada para Z e W no exemplo 3.12.

3.8.1 Notas

1. Sendo f a densidade de X_1, \dots, X_n e $P((X_1, \dots, X_n) \in G_0) = 1$, se $Y_i = g_i(x_1, \dots, x_n); i = 1, \dots, n$, e $\mathcal{B} \subseteq G$, com \mathcal{B} boreliano. Então:

$$\begin{aligned} P((Y_1, \dots, Y_n) \in \mathcal{B}) &= P((X_1, \dots, X_n) \in h(\mathcal{B})) \\ &= \int \cdots \int_{h(\mathcal{B})} f(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= \int \cdots \int_{\mathcal{B}} f(h_1(x_1, \dots, x_n), \dots, h_n(x_1, \dots, x_n)) |J(\underline{x}, \underline{y})| dy_1 \cdots dy_n \end{aligned}$$

2. $P((Y_1, \dots, Y_n) \in G) = P((X_1, \dots, X_n) \in h(G)) = P((X_1, \dots, X_n) \in G_0) = 1$. De modo análogo:

$$\begin{aligned} P((Y_1, \dots, Y_n) \in \mathcal{B}) &= P((Y_1, \dots, Y_n) \in \mathcal{B} \cap G) \\ &= \int \cdots \int_{\mathcal{B} \cap G} f(h(y)) |J(\underline{x}, \underline{y})| dy_1 \cdots dy_n \end{aligned}$$

Theorem 3.3. Sob as condições impostas no início da seção, a densidade conjunta de (Y_1, \dots, Y_n) é dada por:

$$f_{Y_1, \dots, Y_n} = \begin{cases} f_X(h_1(y_1, \dots, y_n), \dots, h_n(y_1, \dots, y_n)) |J(\underline{x}, \underline{y})| & , y \in G \\ 0 & , c.c. \end{cases}$$

3.8.2 Propriedades do Jacobiano

Podemos inverter a ordem das variáveis no Jacobiano, seguindo a seguinte propriedade:

$$J(\underline{x}, \underline{y}) = (J(\underline{y}, \underline{x}))^{-1} \Big|_{\underline{x}=h(\underline{y})} \quad (26)$$

Example 3.14. Retornando ao problema apresentado no exemplo 3.12:

$$\begin{aligned} y_1 &= x_1 + x_2 & y_2 &= x_1 x_2^{-1} \\ \frac{\partial y_1}{\partial x_1} &= 1 & \frac{\partial y_1}{\partial x_2} &= 1 \\ \frac{\partial y_2}{\partial x_1} &= x_2^{-1} & \frac{\partial y_2}{\partial x_2} &= -x_1(x_2)^{-2} \end{aligned}$$

De modo que podemos agora encontrar o Jacobiano com relação aos valores das derivadas parciais dos y 's, e invertê-lo para encontrar o Jacobiano dos x 's:

$$\begin{aligned} J(\underline{y}, \underline{x}) &= \det \begin{bmatrix} 1 & 1 \\ x_2^{-1} & -x_1(x_2)^{-2} \end{bmatrix} = (x_2)^{-2}(x_2 + x_1)(-1) \\ &= \left(\frac{y_2 + 1}{y_1} \right)^2 \left(\frac{y_1}{y_2 + 1} + \frac{y_1 y_2}{y_2 + 1} \right) (-1) \\ &= \frac{(y_2 + 1)^2}{(y_1)^2} \frac{y_1(y_2 + 1)}{y_2 + 1} (-1) \\ &= -\frac{(y_2 + 1)^2}{y_1} = -y_1^{-1}(y_2 + 1)^2 = \frac{1}{J(\underline{x}, \underline{y})} \end{aligned}$$

Temos que, se $g : G_0 \rightarrow G$, com $G_0, G \subseteq \mathbb{R}^n$ abertos, se $g(x_1, \dots, x_n) = (y_1, \dots, y_n)$, então g é bijetiva e $h = g^{-1}$.

Example 3.15. Seja $X \sim U(0, 1)$ e $Y = -\ln(X)$. Temos que $G_0 = (0, 1)$, e $g(x) = -\ln(x)$, de modo que $G = (0, \infty)$. Então:

$$\begin{aligned} g^{-1}(y) &= h(y) = \exp(-y) = e^{-y} \\ \frac{\partial}{\partial y}(g^{-1}(y)) &= -e^{-y} = J(x, y) \end{aligned}$$

Assim, para encontrar $P(Y \leq y)$, teremos:

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

Pelo Jacobiano, teremos:

$$f_Y(y) = f_X(h(y)) \cdot |J| = 1 \cdot e^{-y}$$

Theorem 3.4. Sejam G_1, G_2, \dots, G_k disjuntos tais que $P\left(\underline{X} \in \bigcup_{i=1}^k G_i\right) = 1$, tal que $g|_{G_l}$ é 1:1 para todo $l = 1, \dots, k$. Denotamos por $h^{(l)}$ a inversa de g em G_l , e definimos assim o Jacobiano local $J_l(\underline{x}, \underline{y})$ como:

$$f_Y(y) = \begin{cases} \sum_{l=1}^k f(h^{(l)}(y)) |J_l(x, y)| & ; y \in G_l \\ 0 & c.c. \end{cases}$$

Example 3.16. Sejam $X \sim N(0, 1)$ e $Y = X^2$. Sabemos que $y = x^2$ não é bijetiva, mas podemos considerar a seguinte partição em que essa função seja localmente bijetiva: $G_1 = (-\infty, 0)$ e $G_2 = (0, \infty)$. Então, em $G_1, h^{(1)}(y) = -\sqrt{y}$, e em $G_2, h^{(2)}(y) = \sqrt{y}$, de modo que os jacobianos locais serão:

$$J_1(x, y) = \frac{\partial}{\partial y} h^{(1)}(y) = -\frac{1}{2\sqrt{y}}$$

$$J_2(x, y) = \frac{\partial}{\partial y} h^{(2)}(y) = \frac{1}{2\sqrt{y}}$$

Assim, a densidade de Y será dada por:

$$\begin{aligned} f_Y(y) &= f_X(h^{(1)}(y)) |J_1(x, y)| + f_X(h^{(2)}(y)) |J_2(x, y)| \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y\right) \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y\right) \frac{1}{2\sqrt{y}} \\ &= \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} & , y > 0 \\ 0 & , c.c. \end{cases} \end{aligned}$$

Ou seja, $Y \sim Gama\left(\frac{1}{2}, \frac{1}{2}\right)$, ou $Y \sim \chi^2(1)$.

Notas:

- Se X_1, \dots, X_n são iid, com $X_i \sim N(0, 1) \Rightarrow X_1^2 + \dots + X_n^2 \sim \chi^2(n)$;
- Se $X \sim N(0, 1), Y \sim \chi^2(n)$, com $X \perp Y \Rightarrow \frac{x}{\sqrt{y/n}} \sim t(n)$;
- Sejam X_1, \dots, X_n , iid, com $X_i \sim N(0, 1)$, com $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$:
 1. $\frac{\bar{x}\sqrt{n}}{\sigma} \sim N(0, 1)$;
 2. $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$;
 3. $\frac{\bar{x}\sqrt{n}}{s} \sim t(n-1)$;
 4. $\bar{x} \perp s^2$.
- Se $X \sim \chi^2(k), Y \sim \chi^2(n), X \perp Y \Rightarrow \frac{X/k}{Y/n} \sim F(k, n)$;
- Se $T \sim t(n) \Rightarrow T^2 \sim F(1, n)$.

3.9 Exercícios

TODO

4 Esperança

4.1 Definição

Definition 4.1. Se X é uma variável aleatória com distribuição F , a esperança de X é definida por $E(X) = \int_{-\infty}^{\infty} x dF(x)$, sempre que a integral estiver bem definida.

Convenção: Se $E(X) < \infty$, então X é integrável.

Nota: $\int_{-\infty}^{\infty} x dF(x)$ é bem definida se $\int_0^{\infty} x dF(x)$ ou $\int_{-\infty}^0 x dF(x)$ for finita, já que $\int_{-\infty}^{\infty} x dF(x) = \underbrace{\int_{-\infty}^0 x dF(x)}_{\text{I} \leq 0} + \underbrace{\int_0^{\infty} x dF(x)}_{\text{II} \geq 0}$. Assim, podemos separar em quatro casos:

1. Se **I** e **II** são finitos, então X é integrável;
2. Se **I** é finito e **II** = $+\infty$, então $E(X) = +\infty$;
3. Se **II** é finito e **I** = $-\infty$, então $E(X) = -\infty$;
4. Se **I** = $-\infty$ e **II** = $+\infty$, então $E(X)$ é indefinida.

Propriedade: $E(|X|) = \int |x| dF(x)$. Logo, X é integrável se e somente se $E(|X|) < \infty$.

Example 4.1. $X \sim U(0, 1)$, $Y = \min(X, \frac{1}{2})$:

$$\begin{aligned} P\left(Y = \frac{1}{2}\right) &= P\left(X > \frac{1}{2}\right) = 1 - F_X\left(\frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2} = P_Y\left(Y = \frac{1}{2}\right) \\ E(Y) &= \int_{-\infty}^{\infty} y dF(y) = \int_0^{1/2} y \cdot 1 dy + \frac{1}{2} P_Y\left(Y = \frac{1}{2}\right) \\ &= \left. \frac{y^2}{2} \right|_0^{1/2} + \frac{1}{4} \\ &= \frac{1}{8} + \frac{1}{4} = \frac{3}{8} \end{aligned}$$

Proposition 4.1. $E(X) = \int_0^{\infty} (1 - F(x)) dx - \int_{-\infty}^0 F(x) dx$. Disso, temos que:

- **a)** $\int_0^{\infty} x dF(x) = \int_0^{\infty} (1 - F(x)) dx$;
- **b)** $\int_{-\infty}^0 x dF(x) = - \int_{-\infty}^0 F(x) dx$;

Prova. Vejamos **(a)**: considere que $d(xF(x)) = F(x)dx + x dF(x) \Rightarrow x dF(x) = d(xF(x)) - F(x)dx$. Seja um $b > 0$:

$$\begin{aligned} \int_0^b x dF(x) &= \int_0^b d(xF(x)) - \int_0^b F(x) dx \\ &= xF(x) \Big|_0^b - \int_0^b F(x) dx \\ &= bF(b) - \int_0^b F(x) dx \\ &= \int_0^b [F(b) - F(x)] dx \end{aligned}$$

Note que $\int_0^b x dF(x) \leq \int_0^{\infty} [1 - F(x)] dx$, $\forall b > 0$. Basta notar que $F(b) - F(x) \leq 1 - F(x)$ e que $\int_0^b [1 - F(x)] dx \leq \int_0^{\infty} [1 - F(x)] dx$. Logo:

$$\int_0^{\infty} x dF(x) = \lim_{b \rightarrow \infty} \int_0^b x dF(x) \leq \int_0^{\infty} [1 - F(x)] dx \Rightarrow \int_0^{\infty} x dF(x) \leq \int_0^{\infty} [1 - F(x)] dx$$

Considere $\lambda > 0$ e $b > 0$, tais que:

$$\begin{aligned}\int_0^b [F(b) - F(x)]dx &\geq \int_0^\lambda [1 - F(x)]dx = \int_0^\lambda [F(b) - 1]dx + \int_0^\lambda [1 - F(x)]dx \\ &= \lambda[F(b) - 1] + \int_0^\lambda [1 - F(x)]dx \\ \int_0^b [F(b) - F(x)]dx &\geq \lambda[F(b) - 1] + \int_0^\lambda [1 - F(x)]dx\end{aligned}$$

Logo, como $\int_0^\infty x dF(x) = \lim_{b \rightarrow \infty} \int_0^b [F(b) - F(x)]dx \geq \lim_{b \rightarrow \infty} \{\lambda[F(b) - 1] + \int_0^\lambda [1 - F(x)]dx\} = \int_0^\lambda [1 - F(x)]dx$. Assim:

$$\int_0^\infty x dF(x) \geq \lim_{\lambda \rightarrow \infty} \int_0^\lambda [1 - F(x)]dx = \int_0^\infty [1 - F(x)]dx$$

E como $\int_0^\infty x dF(x) \leq \int_0^\infty [1 - F(x)]dx$, temos que $\int_0^\infty x dF(x) = \int_0^\infty [1 - F(x)]dx$ □

Corollary 4.1. Se X é tal que $X(\omega) \geq 0 \forall \omega \in \mathbb{R} \Rightarrow E(X) = \int_0^\infty [1 - F(x)]dx = \int_0^\infty P(X \geq x)dx$.

Example 4.2. Seja $X \sim \text{Exp}(\lambda)$, qual a $E(X)$? Como o suporte de X é $(0, \infty)$, aplica-se o corolário anterior, de modo que:

$$\begin{aligned}F_X(x) &= 1 - e^{-\lambda x} \Leftrightarrow P(X > x) = e^{-\lambda x} \\ E(X) &= \int_0^\infty e^{-\lambda x} dx = -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^\infty = \frac{1}{\lambda}\end{aligned}$$

Nota: Suponha X discreta e $X(\omega) \geq 0 \forall \omega$. Então:

$$\begin{aligned}E(X) &= \sum_{n=0}^\infty P(X > n) = \sum_{n=0}^\infty P(X \geq n+1) \\ &= \sum_{n=1}^\infty P(X \geq n)\end{aligned}$$

Example 4.3. Considere o lançamento de uma moeda até a 1ª cara. Suponha p = probabilidade de cara e $(1-p)$ = probabilidade de coroa, e X = número de lançamentos até a primeira cara. Tome o evento $[X \geq n]$, logo:

$$E(X) = \sum_{n=1}^\infty (1-p)^{n-1} = \sum_{n=0}^\infty (1-p)^n = \frac{1}{p}$$

Nota: Sendo X uma variável aleatória, temos pelo corolário 4.1 que:

$$\begin{aligned}E(|X|) &= \int_0^\infty P(|X| > x)dx \\ &= \int_0^\infty [P(X > x) + P(X < -x)]dx \\ &= \int_0^\infty P(X > x)dx + \int_0^\infty P(X < -x)dx \\ &= \int_0^\infty (1 - F(x))dx + \int_0^\infty F((-x)^-)dx\end{aligned}$$

Onde $F((-x)^-) = \lim_{u \uparrow -x} F(u)$, que caso F seja contínua, coincide com $F(-x)$. Logo:

$$E(|X|) = \int_0^\infty (1 - F(x))dx + \int_0^\infty F(-x)dx$$

Já que F pode ser descontínua em uma coleção enumerável de pontos. Agora, tomando a transformação de variável $y = -x \Leftrightarrow dy = -dx$:

$$\begin{aligned} E(|X|) &= \int_0^\infty (1 - F(x))dx + \int_{-\infty}^0 F(y)dy \\ &= \int_0^\infty (1 - F(x))dx + \int_{-\infty}^0 F(x)dx \end{aligned}$$

Utilizando os resultados **a** e **b** da proposição 4.1, temos que:

$$\begin{aligned} E(|X|) &= \int_0^\infty x dF(x) - \int_{-\infty}^0 x dF(x) \\ &= \int_0^\infty |x| dF(x) + \int_{-\infty}^0 |x| dF(x) \\ &= \int_{-\infty}^\infty |x| dF(x) \end{aligned}$$

Onde F é a acumulada de X , ao invés de $|X|$. Assim, a integrabilidade de X depende da finitude de $\int_0^\infty x dF(x)$ e $\int_{-\infty}^0 x dF(x)$, logo X é integrável se $E(|X|) < \infty$.

4.2 Propriedades da esperança

- **E₁**: Se $X = c$, com c uma constante, $E(X) = c$;
- **E₂(monotonia)**: Se X e Y são variáveis aleatórias, com $X \leq Y \Rightarrow E(X) \leq E(Y)$, caso ambas as esperanças estejam bem definidas;

Prova. Seja z um valor fixo. Se $Y \leq z \Rightarrow X \leq z$, logo $[Y \leq z] \subseteq [X \leq z]$, assim:

$$\begin{aligned} P(Y \leq z) &\leq P(X \leq z) \\ F_Y(z) &\leq F_X(z) \iff 1 - F_Y(z) \geq 1 - F_X(z) \end{aligned}$$

E pela proposição 4.1, temos que:

$$\begin{aligned} E(Y) &= \int_0^\infty [1 - F_Y(z)]dz - \int_{-\infty}^0 F_Y(z)dz \geq \int_0^\infty [1 - F_X(z)]dz - \int_{-\infty}^0 F_X(z)dz = E(X) \\ E(Y) &\geq E(X) \end{aligned}$$

□

- **E₃(linearidade)**:

- (i) Se $E(X)$ é bem definida, $a, b \in \mathbb{R}$, então $E(aX + b) = aE(X) + b$;
- (ii) $E(aX + bY) = aE(X) + bE(Y)$, caso o termo $aE(X) + bE(Y)$ esteja bem definido;
- Note que se $E(X) = \infty \Rightarrow E(X - X) \neq E(X) - E(X)$.

Prova. Quando $a = 0$; $E(aX + b) = E(b) = b = 0E(X) + b$.

Quando $a > 0, b > 0$; $F_{aX+b}(x) = P(aX + b \leq x) = P(X \leq \frac{x-b}{a}) = F_X(\frac{x-b}{a})$. Logo:

$$\begin{aligned}
E(aX + b) &= \int_0^\infty [1 - F_{aX+b}(x)] dx - \int_{-\infty}^0 F_{aX+b}(x) dx \\
&= \int_0^\infty \left[1 - F_X\left(\frac{x-b}{a}\right) \right] dx - \int_{-\infty}^0 F_X\left(\frac{x-b}{a}\right) dx
\end{aligned}$$

Tome $y = \frac{x-b}{a} \Rightarrow dy = \frac{1}{a} dx$. Então:

$$\begin{aligned}
E(aX + b) &= \int_{-b/a}^\infty a[1 - F_X(y)] dy - \int_{-\infty}^{-b/a} aF_X(y) dy \\
&= a \left\{ \int_{-b/a}^\infty [1 - F_X(y)] dy - \int_{-\infty}^{-b/a} F_X(y) dy \right\} \\
&= a \int_0^\infty [1 - F_X(y)] dy - a \int_{-\infty}^0 F_X(y) dy + a \int_{-b/a}^0 [1 - F_X(y)] dy + a \int_{-b/a}^0 F_X(y) dy \\
&= aE(X) + a \int_{-b/a}^0 dy \\
&= aE(X) + a \frac{b}{a} \\
&= aE(X) + b
\end{aligned}$$

□

- **E₄(Desigualdade de Jansen):** Seja φ uma função convexa, definida na reta, com X integrável, então:

$$E(\varphi(X)) \geq \varphi(E(X)) \quad (27)$$

Nota: Caso φ seja côncava:

$$E(\varphi(X)) \leq \varphi(E(X))$$

Prova para convexa. Tome x_0 e $\varphi(x_0)$. Então existe uma reta L tal que L passe por $\varphi(x_0)$ e φ fica por cima de L . Logo temos a seguinte equação da reta:

$$L(x) = \varphi(x_0) + \lambda(x - x_0)$$

Onde λ é alguma constante apropriada. Então para todo x temos:

$$\begin{aligned}
\varphi(x) &\geq L(x) = \varphi(x_0) + \lambda(x - x_0) \\
&\Downarrow \mathbf{E_2} \\
E(\varphi(x)) &\geq E(L(x)) \stackrel{\mathbf{E_1}, \mathbf{E_3}}{=} \varphi(x_0) + \lambda[E(x) - x_0]
\end{aligned}$$

Que vale para $x_0 = E(x)$, de modo que $E(\varphi(x)) \geq \varphi(E(x)) + \lambda[E(x) - E(x)]$, então:

$$E(\varphi(x)) \geq \varphi(E(x))$$

A prova para funções côncavas segue a mesma metodologia, com a inversão da desigualdade. □

4.2.1 Critério de integrabilidade

Suponha que X é uma variável aleatória dominada por Y (ou seja, $X \leq Y$), sendo Y uma variável aleatória integrável. X é integrável? Temos que:

$$X \leq Y \Rightarrow E(X) \leq E(Y)$$

Se X e Y são tais que $Y \geq 0$ e Y é integrável e $|X| \leq Y \Rightarrow 0 \leq |X| \leq Y$, e como consequência:

$$0 \leq E(X) \leq E(Y) < \infty \Rightarrow X \text{ é integrável}$$

De maneira similar, seja X uma variável aleatória qualquer. Então:

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E(|X|) \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n)$$

Assim, X é integrável se e somente se $\sum_{n=1}^{\infty} P(|X| \geq n) < \infty$.

Prova. Seja $x \geq 0$. Tome $[x]$ como a parte inteira de x . Então $[x] = k$ se $k \leq x < k+1$. Então:

$$\begin{aligned} 0 \leq [x] &\leq x \leq [x] + 1 \\ &\Downarrow \mathbf{E_2, E_3} \\ 0 \leq E([x]) &\leq E(x) \leq E([x]) + 1 \end{aligned}$$

Pelo corolário 4.1, como $[x]$ é discreta e não-negativa, temos que:

$$\begin{aligned} E([x]) &= \sum_{n=1}^{\infty} P([x] \geq n) \\ &= \sum_{n=1}^{\infty} P(x \geq n) \leq E(x) \leq \sum_{n=1}^{\infty} P(x \geq n) + 1 \end{aligned}$$

□

4.2.2 Casos de interesse

a) (Consistência absoluta) $\varphi(X) = |X|$:

$$E(|X|) \geq |E(X)|$$

b) (Consistência quadrática) $\varphi(X) = X^2$:

$$E(X^2) \geq [E(X)]^2$$

c) (Consistência absoluta de ordem p) $\varphi(X) = |X|^p, p \geq 1$:

$$E(|X|^p) \geq |E(X)|^p$$

Nota: φ só precisa ser convexa (ou côncava) em uma região de probabilidade 1. Por exemplo, se X é uma variável aleatória, tal que $P(X > 0) = 1$, ou o suporte da distribuição de X é $(0, \infty)$, $\varphi(X) = \frac{1}{X}$ é convexa em $(0, \infty) \Rightarrow E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)}$. De modo análogo, se $P(X > 0) = 1$ e $\varphi(X) = \ln(X)$, φ é côncava em $(0, \infty)$ logo $E(\ln(X)) \leq \ln(E(X))$.

4.3 Esperança de funções de variáveis aleatórias

Seja X uma variável aleatória, φ uma função mensurável e $Y = \varphi(X)$. Assim, Y é uma variável aleatória, cuja esperança é $E(Y) = \int y dF_{\varphi(X)}(y) = \int_0^\infty [1 - F_{\varphi(X)}(y)] dy - \int_{-\infty}^0 F_{\varphi(X)}(y) dy$.

Theorem 4.1. Se X é uma variável aleatória e φ uma função mensurável, com $Y = \varphi(X)$:

$$E(Y) = E(\varphi(X)) = \int \varphi(x) dF_X(x)$$

Prova para caso $\varphi(x) = x^k$. Note que a prova já foi feita para $\varphi(x) = |x|$. Vejamos que a prova é válida para $\varphi(x) = x^k$, com $k = 1, 2, \dots$, em 2 casos: k par e k ímpar:

k par:

$$\begin{aligned} E(X^k) &= \int_0^\infty P(X^k > t) dt \\ &= \int_0^\infty P(X > \sqrt[k]{t}) dt + \int_0^\infty P(X < -\sqrt[k]{t}) dt \\ &= \int_0^\infty [1 - F_X(\sqrt[k]{t})] dt + \int_0^\infty F_X(-\sqrt[k]{t}) dt \end{aligned}$$

Apliquemos as seguintes mudanças de variáveis: $s = t^{\frac{1}{k}}, ds = \frac{1}{k} t^{\frac{1}{k}-1} dt, dt = \frac{(ds)ks^k}{s}, u = -s, du = -ds$:

$$\begin{aligned} E(X^k) &= \int_0^\infty [1 - F_X(s)] ks^{k-1} ds + \int_0^\infty F_X(-s) ks^{k-1} ds \\ &= \int_0^\infty [1 - F_X(s)] ks^{k-1} ds - \int_{-\infty}^0 F_X(u) ku^{k-1} du \\ &= k \left\{ \int_0^\infty [1 - F_X(s)] s^{k-1} ds - \int_{-\infty}^0 F_X(u) u^{k-1} du \right\} \end{aligned}$$

Agora, mostremos que $E(X^k) = \int x^k dF_X(x)$:

$$\begin{aligned} \int_{-\infty}^\infty x^k dF_X(x) &\stackrel{Def}{=} \int_0^\infty [1 - F_X(x)] d(x^k) - \int_{-\infty}^0 F_X(x) d(x^k) \\ &= k \left\{ \int_0^\infty [1 - F_X(x)] x^{k-1} dx - \int_{-\infty}^0 F_X(x) x^{k-1} dx \right\} \\ &= E(X^k) \end{aligned}$$

□

Nota: A propriedade é também válida para polinômios, visto que a esperança opera de maneira linear.

Example 4.4. Seja $X \sim Exp(\lambda)$, vimos que $E(X) = \frac{1}{\lambda}$:

Calcular $E(X^2)$:

$$\begin{aligned} E(X^2) &= 2 \int_0^\infty x e^{-\lambda x} dx = \frac{2}{\lambda} \int_0^\infty x \lambda e^{-\lambda x} dx \\ &= \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2} \end{aligned}$$

Calcular $E(X^3)$:

$$\begin{aligned}
E(X^3) &= 3 \int_0^\infty x^2 e^{-\lambda x} dx = \frac{3}{\lambda} \int_0^\infty x^2 \lambda e^{-\lambda x} dx \\
&= \frac{3}{\lambda} E(X^2) = \frac{3}{\lambda^3}
\end{aligned}$$

De modo que podemos observar o padrão emergente, e definir $E(X^k) = \frac{k!}{\lambda^k}$.

4.4 Momentos de uma variável aleatória

- a) $E([X - b]^k)$: k -ésimo momento de X em torno de b ;
- b) $E(X^k)$: k -ésimo momento em torno de 0;
- c) Se em (a), $b = E(X)$, o momento é central;
- d) $t > 0$, $E(|X|^t)$: t -ésimo momento absoluto de X .

Definition 4.2 (Variância de uma variável aleatória).

$$\text{Var}(X) = E\{(X - E(X))^2\} \iff \text{Var}(X) = E(X^2) - (E(X))^2$$

Proposition 4.2. Se X é uma variável aleatória, $f(t) = [E(|X|^t)]^{\frac{1}{t}}$ é não-decrescente em $t, t > 0$.

Prova. Devemos provar que, se $0 < s < t$, $f(s) \leq f(t)$ (ou $\{E(|X|^s)\}^{\frac{1}{s}} \leq \{E(|X|^t)\}^{\frac{1}{t}}$). Para tanto, consideremos dois casos: **a)** $E(|X|^s) < \infty$, **b)** $E(|X|^s) = \infty$:

a) Defina $\varphi(y) = |y|^{\frac{t}{s}}$ (caso $\frac{t}{s} > 1$, φ será convexa). Pela Desigualdade de Jansen:

$$\begin{aligned}
E(\varphi(Y)) &\geq \varphi(E(Y)) \\
E\left(|Y|^{\frac{t}{s}}\right) &\geq |E(Y)|^{\frac{t}{s}}
\end{aligned}$$

Tome $Y = |X|^s$. Substituindo temos:

$$\begin{aligned}
E\left(|X|^t\right) &\geq |E(|X|^s)|^{\frac{t}{s}} \\
E(|X|^t) &\geq \{E(|X|^s)\}^{\frac{t}{s}} \\
\{E(|X|^t)\}^{\frac{1}{t}} &\geq \{E(|X|^s)\}^{\frac{1}{s}}
\end{aligned}$$

b) Como $t > s > 0$, sabemos que $|X|^s \leq 1 + |X|^t$. Como $E(|X|^s) = \infty$, então:

$$\infty = E(|X|^s) \leq 1 + E(|X|^t) = \infty$$

□

Corollary 4.2. Se $E(|X|^t) < \infty \forall t \in (0, \infty) \Rightarrow E(|X|^s) < \infty \forall s$, com $0 < s < t$.

4.4.1 Propriedades

- **E₅**: Se $X = c$, com c uma constante, $\text{Var}(X) = 0$;
- **E₆**: $\text{Var}(X + b) = \text{Var}(X)$, $\text{Var}(aX + b) = a^2 \text{Var}(X)$, com $a, b \in \mathbb{R}$;

Prova.

$$\begin{aligned}
\text{Var}(aX + b) &= E\left\{[aX + b - E(aX + b)]^2\right\} \\
&= E\left\{[aX + b - aE(X) - b]^2\right\} \\
&= E\left\{a^2[X - E(X)]^2\right\} \\
&= a^2 E\left\{[X - E(X)]^2\right\} = a^2 \text{Var}(X)
\end{aligned}$$

□

- **E₇(Desigualdade de Tchebychev):** Seja X uma variável aleatória, com $X \geq 0$. Para todo $\lambda > 0$:

$$P(X \geq \lambda) \leq \frac{E(X)}{\lambda} \quad (28)$$

Prova. Seja $Y = I_{[X \geq \lambda]} \lambda = \begin{cases} \lambda & , X \geq \lambda \\ 0 & , c.c. \end{cases}$. Por definição, $0 \leq Y \leq X \Rightarrow E(Y) \leq E(X)$ e $E(Y) = \lambda P(X \geq \lambda)$, de modo que:

$$\lambda P(X \geq \lambda) \leq E(X) \Leftrightarrow P(X \geq \lambda) \leq \frac{E(X)}{\lambda}$$

□

4.4.2 Consequências

a) Para todo $\lambda > 0$:

$$P(|X - E(X)| \geq \lambda) \leq \frac{\text{Var}(X)}{\lambda^2}$$

b) **(Desigualdade de Markov)** Seja X uma variável aleatória, para todo t :

$$P(|X| \geq \lambda) \leq \frac{E(|X|^t)}{\lambda^t} \quad (29)$$

c) Se Z é uma variável aleatória, com $Z \geq 0$ e $E(Z) = 0$:

$$P(Z = 0) = 1 \quad (\text{i.e., } Z = 0 \text{ quase certamente})$$

Provas. a) Se $Y = [X - E(X)]^2$, aplicamos **E₇** usando $\lambda^2 : P(Y \geq \lambda^2) \leq \frac{E(Y)}{\lambda^2}$. Note que $E(Y) = E([X - E(X)]^2) = \text{Var}(X)$. Logo:

$$P(|X - E(X)| \geq \lambda) = P(|X - E(X)|^2 \geq \lambda^2) = P(Y \geq \lambda^2) \leq \frac{E(Y)}{\lambda^2} = \frac{\text{Var}(X)}{\lambda^2}$$

b) Seja $Y = |X|^t$, aplicamos **E₇** a Y e $\lambda^t : P(Y \geq \lambda^t) \leq \frac{E(Y)}{\lambda^t}$. Note que $E(Y) = E(|X|^t)$ e que $P(Y \geq \lambda^t) = P(|X|^t \geq \lambda^t) = P(|X| \geq \lambda)$. Logo:

$$P(|X| \geq \lambda) \leq \frac{E(|X|^t)}{\lambda^t}$$

c) $Z = 0$ quase certamente, usamos **E₇** na variável Z e em $\lambda = \frac{1}{n}$, então:

$$P\left(Z \geq \frac{1}{n}\right) \leq E(Z) \cdot n \stackrel{Hip}{=} 0$$

Temos que $[Z > 0] = \bigcup_n [Z \geq \frac{1}{n}]$, de modo que:

$$P(Z > 0) = P\left(\bigcup_n \left[Z \geq \frac{1}{n}\right]\right) = \lim_{n \rightarrow \infty} P\left(Z \geq \frac{1}{n}\right) = 0 \Rightarrow P(Z = 0) = 1 - P(Z > 0) = 1$$

□

Nota: Se X é uma variável tal que $\text{Var}(X) = 0$, temos que $\text{Var}(X) = 0 \Leftrightarrow E([X - E(X)]^2) = 0$, ou seja, se definirmos $Z = [X - E(X)]^2, Z \geq 0$ e $E(Z) = 0$. Logo, por **c)**, $P(Z = 0) = 1$, ou seja, $P([X - E(X)]^2 = 0) = 1 \Leftrightarrow P(X = E(X)) = 1$, ou seja, $X = E(X)$ quase certamente.

Example 4.5. Se X e Y são variáveis aleatórias tais que $E(|X|^t) < \infty$ e $E(|Y|^t) < \infty$, então $E(|X+Y|^t) < \infty$:

- (i) A finitude de $E(|X|^t)$ leva à finitude de $E(|aX|^t)$;
- (ii) Se X e Y forem integráveis (com $t = 1$), então $X + Y$ é integrável. Se X e Y tem variâncias finitas ($t = 2$), então $X + Y$ tem variância finita.

Proposition 4.3. *Seja X uma variável aleatória integrável e $\mu = E(X) \Rightarrow \mu$ minimiza $E([X - c]^2)$, com $c \in \mathbb{R}$.*

Prova. Temos que $(X - c)^2 = (X - \mu + \mu - c)^2 = (X - \mu)^2 + 2(\mu - c)(X - \mu) + (\mu - c)^2$. Logo, pelas propriedades lineares do valor esperado:

$$\begin{aligned} E([X - c]^2) &= E([X - c]^2) + 2(\mu - c)E(X - \mu) + (\mu - c)^2 \\ &= \text{Var}(X) + (\mu - c)^2 \end{aligned}$$

□

Proposition 4.4. *Seja X uma variável aleatória e m sua mediana. Assim, m minimiza $E(|X - c|)$, $c \in \mathbb{R}$. Ou seja:*

$$E(|x - m|) = \min_{c \in \mathbb{R}} E(|X - c|)$$

Prova. Considere a definição de mediana: $P(X \leq m) = P(X > m) = \frac{1}{2}$. Suponha que X é integrável, logo $X - c$ também o será para todo c constante real. Vamos ver que, com $m < c$ (o caso em que $m > c$ segue analogamente):

- $X \leq m \Rightarrow |X - c| - |X - m| = \lambda$, onde $\lambda = c - m$;
- $X > m \Rightarrow |X - c| - |X - m| \geq \lambda$.

Seja c tal que $m < c$. Defina $\lambda = c - m > 0$. Então:

- Se $x \leq m \Rightarrow |x - c| = |x - m| + \lambda \Rightarrow |x - c| - |x - m| = \lambda$;
- Se $x > m$ e $x > c$ (os casos intermediários são decorrências), então $x > c \Rightarrow \lambda + |x - c| = |x - m| \Rightarrow |x - c| - |x - m| \geq -\lambda$.

Defina $Y = |X - c| - |X - m| = \begin{cases} \lambda & , \text{se } X \leq m, \\ y & , \text{se } X > m. \end{cases}$ Assim, temos que $y \geq -\lambda$, e que $Y =$

$\begin{cases} \lambda & , P(X \leq m) \geq \frac{1}{2}, \\ y & , P(X > m) \leq \frac{1}{2} \end{cases}$. Logo, $Y \geq \lambda I_{[X \leq m]} - \lambda I_{[X > m]} \Rightarrow E(Y) \geq \lambda E(I_{[X \leq m]}) - \lambda E(I_{[X > m]}) = \lambda P(X \leq m) - \lambda P(X > m) \geq 0$.

Como $E(Y) \geq 0 \Rightarrow E(|X - c|) \geq E(|X - m|) \forall c$, com $m < c$. □

4.5 Esperanças e funções de vetores

Theorem 4.2. *Seja $\underline{X} = (X_1, \dots, X_n)$ um vetor aleatório e $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ mensurável. Então:*

$$\begin{aligned} E(\varphi(\underline{X})) &= \int y dF_{\varphi(\underline{X})}(y) = \int_{\mathbb{R}^n} \varphi(x) dF_{\underline{X}}(x) \\ &= \int \cdots \int \varphi(x_1, \dots, x_n) dF_{\underline{X}}(x_1, \dots, x_n) \end{aligned}$$

Caso discreto: Seja \underline{X} discreto, tomando valores $\underline{X}_i = (x_{i1}, \dots, x_{in})$, com probabilidade $P(\underline{X}_i) = \sum_i P(x_i) = 1$. Então:

$$E(\varphi(\underline{X})) = \sum_i \varphi(\underline{x}_i) P(\underline{x}_i)$$

Caso contínuo: Seja \underline{X} contínuo, com densidade $f(x_1, \dots, x_n)$. Então:

$$E(\varphi(\underline{X})) = \int \cdots \int f(x_1, \dots, x_n) dx_1 \dots dx_n$$

Example 4.6. Lembrando a propriedade $E_3 : E(X + Y) = E(X) + E(Y)$ desde que existam $E(X)$ e $E(Y)$. Seja $\varphi(x, y) = x + y$ e defina $\varphi_1(x, y) = x$ e $\varphi_2(x, y) = y$. Teremos pelo teorema que:

$$\begin{aligned} E(X + Y) &= E(\varphi(x, y)) = \int \int (x + y) dF_{X,Y}(x, y) = \int \int x dF_{X,Y}(x, y) + \int \int y dF_{X,Y}(x, y) \\ &= E(\varphi_1(x, y)) + E(\varphi_2(x, y)) \\ &= E(X) + E(Y) \end{aligned}$$

Se $\{X_i\}_{i=1}^n$ é conjuntamente independente, com densidades f_1, \dots, f_n , sendo a densidade conjunta dada por $f = \prod_{i=1}^n f_i$, então:

$$\begin{aligned} E(\varphi(\underline{X})) &= \int \cdots \int \varphi(x_1, \dots, x_n) f_1(x_1) \dots f_n(x_n) dx_1 \dots dx_n \\ &= \int \cdots \int \varphi(x_1, \dots, x_n) dF_{X_1}(x_1) \dots dF_{X_n}(x_n) \end{aligned}$$

Proposition 4.5. *Sejam $\{X_i\}_{i=1}^n$ conjuntamente independentes e integráveis. Então:*

$$E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i)$$

Prova (para $n = 2$). Seja $\varphi(X, Y) = X.Y$:

$$\begin{aligned} E(X.Y) &= E(\varphi(X, Y)) = \int \int \varphi(x, y) dF_X(x) dF_Y(y) \\ &= \int [y.x dF_X(x)] dF_Y(y) \\ &= \int y E(X) dF_Y(y) \\ &= E(X) \int y dF_Y(y) = E(X) E(Y) \end{aligned}$$

□

Definition 4.3. A covariância entre X e Y será definida por:

$$\text{Cov}(X, Y) = E\left[(X - E(X))(Y - E(Y))\right]$$

Sempre que X e Y sejam integráveis. Assim:

$$\begin{aligned} \text{Cov}(X, Y) &= E\{XY - E(Y)X - E(X)Y + E(X)E(Y)\} \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

Note que X e Y podem ter $\text{Cov}(X, Y) = 0$ e mesmo assim $X \not\sim Y$.

Notas:

- A existência da covariância entre variáveis integráveis depende da existência de $E(XY)$;
- $Cov(X, Y) = 0$ é interpretado como “ X e Y são não-correlacionados”;
- Há casos onde $Cov(X, Y) = 0$ implica independência, como na Normal Bivariada, por exemplo.

Proposition 4.6. *Sejam X_1, \dots, X_n variáveis aleatórias integráveis tais que $Cov(X_i, X_j) = 0 \forall i \neq j$. Então*

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i)$$

Prova.

$$\begin{aligned} \text{Var} \left(\sum_{i=1}^n X_i \right) &= E \{ [X_1 + \dots + X_n] - E[X_1 + \dots + X_n] \}^2 \\ &= E \left\{ \sum_{i=1}^n [X_i - E(X_i)]^2 + 2 \sum_{i < j} (X_i - E(X_i))(X_j - E(X_j)) \right\} \\ &= \sum_{i=1}^n E [(X_i - E(X_i))^2] + 2 \sum_{i < j} E[(X_i - E(X_i))(X_j - E(X_j))] \\ &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Var}(X_i) \end{aligned}$$

□

Corollary 4.3. *Sejam X_1, \dots, X_n variáveis aleatórias independentes e integráveis. Então:*

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i)$$

Definition 4.4. Para X e Y variáveis aleatórias, o coeficiente de correlação de Pearson é definido por:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Com $\sigma_X = \sqrt{\text{Var}(X)}$ e $\sigma_Y = \sqrt{\text{Var}(Y)}$, sempre que $\text{Var}(X)$ e $\text{Var}(Y)$ sejam finitas e maiores que 0.

Proposition 4.7. *Sob os supostos da definição 4.4:*

- **a)** $-1 \leq \rho_{X,Y} \leq 1$;
- **b)** $\rho_{X,Y} = 1 \Leftrightarrow P(Y = aX + b) = 1$, para algum $a > 0$ e $b \in \mathbb{R}$;
- **c)** $\rho_{X,Y} = -1 \Leftrightarrow P(Y = aX + b) = 1$, para algum $a < 0$ e $b \in \mathbb{R}$.

Prova. Note primeiramente que $\text{Cov}(X, Y) = E \{ (X - E(X))(Y - E(Y)) \}$, logo:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = E \left\{ \frac{(X - E(X))}{\sigma_X} \frac{(Y - E(Y))}{\sigma_Y} \right\}$$

Observe que $E \left(\frac{X - E(X)}{\sigma_X} \right) = 0$ e $\text{Var} \left(\frac{X - E(X)}{\sigma_X} \right) = 1$, e analogamente para Y . Assim:

a)

$$\begin{aligned}
0 &\leq \left(\frac{(X - E(X))}{\sigma_X} - \frac{(Y - E(Y))}{\sigma_Y} \right)^2 \\
0 &\leq E \left\{ \left(\frac{(X - E(X))}{\sigma_X} - \frac{(Y - E(Y))}{\sigma_Y} \right)^2 \right\} \\
0 &\leq E \left(\left[\frac{X - E(X)}{\sigma_X} \right]^2 \right) + E \left(\left[\frac{Y - E(Y)}{\sigma_Y} \right]^2 \right) - \frac{2}{\sigma_X \sigma_Y} E((X - E(X))(Y - E(Y))) \\
0 &\leq \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} - \frac{2\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \\
0 &\leq 2 - 2\rho_{X,Y} \\
\rho_{X,Y} &\leq 1
\end{aligned}$$

Tomando a diferença ao invés da soma, chegamos que $\rho_{X,Y} \geq -1$.

b)

Suponha $\rho_{X,Y} = 1 \Leftrightarrow E \left\{ \left[\frac{X - E(X)}{\sigma_X} - \frac{Y - E(Y)}{\sigma_Y} \right]^2 \right\} = 0$. Pela propriedade E_7 , temos que:

$$P \left(\frac{X - E(X)}{\sigma_X} = \frac{Y - E(Y)}{\sigma_Y} \right) = 1$$

Ou seja, $Y \stackrel{q.c.}{=} E(Y) + \frac{\sigma_Y}{\sigma_X}(X - E(X))$, então $a = \frac{\sigma_Y}{\sigma_X} > 0, b = E(Y) - \frac{\sigma_Y}{\sigma_X}E(X)$. □

Nota: Se $P(Y = aX + b) = 1$, sendo $a \neq 0$, pelo desenvolvimento da prova de **(a)**, temos que:

$$\begin{aligned}
\rho_{X,Y} &= E \left\{ \left(\frac{X - E(X)}{\sigma_X} \right) \left(\frac{aX + b - aE(X) - b}{\sqrt{a^2 \sigma_X^2}} \right) \right\} \\
&= \frac{a}{|a|} E \left\{ \left[\frac{X - E(X)}{\sigma_X} \right]^2 \right\} \\
&= \frac{a}{|a|} = \text{sgn}(a) = \pm 1
\end{aligned}$$

4.6 Convergência

Theorem 4.3 (Teorema da convergência monótona). *Sejam X_1, X_2, \dots e X variáveis aleatórias em (Ω, \mathcal{A}, P) . Se $0 \leq X_n \uparrow_{n \rightarrow \infty} X$ (ou seja, $X_n(\omega) \geq 0, \forall \omega \in \Omega$ e $X_n(\omega) \uparrow_{n \rightarrow \infty} X(\omega), \forall \omega \in \Omega$). Então $E(X_n) \uparrow_{n \rightarrow \infty} E(X)$.*

Prova. Como $X_n \uparrow_{n \rightarrow \infty} X, X_n \geq 0$, por E_2 temos que $0 \leq E(X_n) \leq E(X)$. Devemos então provar que

$\forall \epsilon > 0, \lim_{n \rightarrow \infty} E(X_n) \geq E(X) - \epsilon$ (ou seja, $\forall \epsilon > 0 \exists n_0(\epsilon) : E(X_n) \geq E(X) - \epsilon, \forall n : n \geq n_0(\epsilon)$).

Defina $Y = \sum_{n=0}^{\infty} n\epsilon I_{B_n}$, onde $B_n = [n\epsilon < X \leq (n+1)\epsilon], n = 0, 1, \dots$. Assim:

$$\begin{aligned}
n = 0 : B_0 &= [0 < X \leq \epsilon] \\
n = 1 : B_1 &= [\epsilon < X \leq 2\epsilon] \\
n = 2 : B_2 &= [2\epsilon < X \leq 3\epsilon] \\
&\vdots \\
Y &= \begin{cases} n\epsilon & , \text{ se } n\epsilon < X \leq (n+1)\epsilon \\ 0 & , \text{ se } X = 0 \end{cases}
\end{aligned}$$

Temos então que mostrar que $X - \epsilon \leq Y \leq X$. Como $Y = n\epsilon < X$, o lado direito é dado diretamente. Para o lado esquerdo temos que:

$$X \leq (n+1)\epsilon = n\epsilon + \epsilon \Leftrightarrow X - \epsilon \leq n\epsilon = Y \Rightarrow E(X) - \epsilon \leq E(Y) \leq E(X)$$

Note que, se $E(Y) \leq \lim_{n \rightarrow \infty} E(X_n)$, então teremos provado o resultado. Seja $A_k = [X_k \geq Y]$, com k grande o suficiente. Se ω é tal que $X_k(\omega) \geq Y(\omega) \Rightarrow X_{k+1}(\omega) \geq Y(\omega) \Rightarrow A_k \subseteq A_{k+1} \subseteq A_{k+2} \subseteq \dots$

Formalmente, pelo limite $X_k \uparrow X$ se k é suficientemente grande, $X_k(\omega) \geq Y(\omega)$. Assim, $\bigcup A_k = [Y \leq X] = \Omega$, e:

$$YI_k = \begin{cases} Y(\omega) & , \text{se } \omega \in A_k \\ 0 & , \text{se } \omega \notin A_k \end{cases} = \begin{cases} n\epsilon & , \text{se } \omega \in B_n \cap A_k, n = 0, 1, \dots \\ 0 & , \text{se } \omega \notin \bigcup_{n=0}^{\infty} B_n \cap A_k \end{cases}$$

Logo, $0 \leq YI_k \leq X_k \Rightarrow 0 \leq E(YI_k) \leq E(X_k)$. Assim, $E(YI_k)$ será:

$$E(YI_k) = \sum_{n=0}^{\infty} n\epsilon P(B_n \cap A_k) \geq \sum_{n=0}^m n\epsilon P(B_n \cap A_k)$$

Logo, $\lim_{k \rightarrow \infty} E(X_k) \geq \lim_{k \rightarrow \infty} \sum_{n=0}^m n\epsilon P(B_n \cap A_k) = \sum_{n=0}^m n\epsilon P(B_n), \forall m$. Como m é arbitrário, $\lim_{k \rightarrow \infty} E(X_k) \geq \sum_{n=0}^{\infty} n\epsilon P(B_n) = E(Y)$. \square

Theorem 4.4 (Teorema da convergência dominada). *Sejam Y, X_1, X_2, \dots, X variáveis aleatórias em (Ω, \mathcal{A}, P) , tais que Y é integrável, $|X_n| \leq Y \forall n$ e $X_n \rightarrow X$ (ou seja, dado $\omega : X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega)$). Então X e X_n são integráveis e $E(X_n) \xrightarrow[n \rightarrow \infty]{} E(X)$.*

Prova. Por hipótese, temos a integrabilidade de X e X_n , tais que:

$$|X| = \lim_{n \rightarrow \infty} |X_n| \stackrel{\text{hip}}{\leq} Y$$

Assim, X e X_n são dominadas, então por E_2 , X e X_n são integráveis. Defina $Y_n = \inf_{k \geq n} X_k$. Tomar o ínfimo provoca a sequência a se movimentar pela esquerda, de modo que:

$$X(\omega) \stackrel{\text{hip}}{=} \lim_{n \rightarrow \infty} X_n(\omega) = \lim_{n \rightarrow \infty} \inf_{k \geq n} X_k(\omega) = \lim_{n \rightarrow \infty} \left(\inf_{k \geq n} X_k(n) \right) = \lim_{n \rightarrow \infty} Y_n$$

E por definição de $Y_n : Y_n \uparrow X \Rightarrow (Y_n + Y) \uparrow (X + Y)$. Aplicando o teorema da convergência monótona, temos que $|X_n| \leq Y \forall n \Rightarrow -Y \leq X_n$, logo $\inf_{k \geq n} X_k \geq -Y \Rightarrow Y_n \geq -Y \Rightarrow Y_n + Y \geq 0$. Logo:

$$E(Y_n + Y) \uparrow_{n \rightarrow \infty} E(X + Y)$$

Defina $Z_n(\omega) = \sup_{k \geq n} X_k(\omega)$. Note que $Z_n(\omega) \downarrow_{n \rightarrow \infty} X(\omega)$, logo $(Y - Z_n) \uparrow_{n \rightarrow \infty} (Y - X)$. Note que $|X_n| \leq Y, \forall n \Rightarrow X_n \leq Y$, de modo que:

$$\sup_{k \geq n} X_k \leq Y \Rightarrow Z_n \leq Y \Rightarrow 0 \leq Y - Z_n$$

Agora que temos a positividade e a monotonia do crescimento, utilizamos o teorema da convergência monótona, de modo que:

$$E(Y - Z_n) \uparrow_{n \rightarrow \infty} E(Y - X) \Rightarrow E(Z_n) \downarrow_{n \rightarrow \infty} E(X)$$

Juntando as convergências de Y_n e Z_n :

$$Y_n = \inf_{k \geq n} X_k \leq X_n \leq \sup_{k \geq n} X_k = Z_n \Rightarrow E(Y_n) \leq E(X_n) \leq E(Z_n)$$

De modo que $E(X_n) \rightarrow E(X)$. \square

4.6.1 Observações sobre o teorema da convergência dominada

1. Há casos nos quais $X_n \rightarrow X$, no entanto $E(X_n) \not\rightarrow E(X)$;

Example 4.7. $X \sim \text{Cauchy} \rightarrow f(x) = \frac{1}{\pi(1+x^2)}$. Sabemos que $E(X) = \infty$. Seja $X_n = I_{[-n \leq x \leq n]} =$
 $\begin{cases} X & , -n \leq x \leq n \\ 0 & , |x| > n \end{cases}$. Então:

- i. $X_n \xrightarrow{n \rightarrow \infty} X$