

Notas de Aula - Capítulo 2

Probabilidade

Caio Gomes Alves

30/03/2025

1 Variáveis Aleatórias

1.1 Variáveis aleatórias e funções de distribuição

Example 1.1. 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 1.1 (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 1.2 (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 1.2. 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 $r \in [1, 2)$:

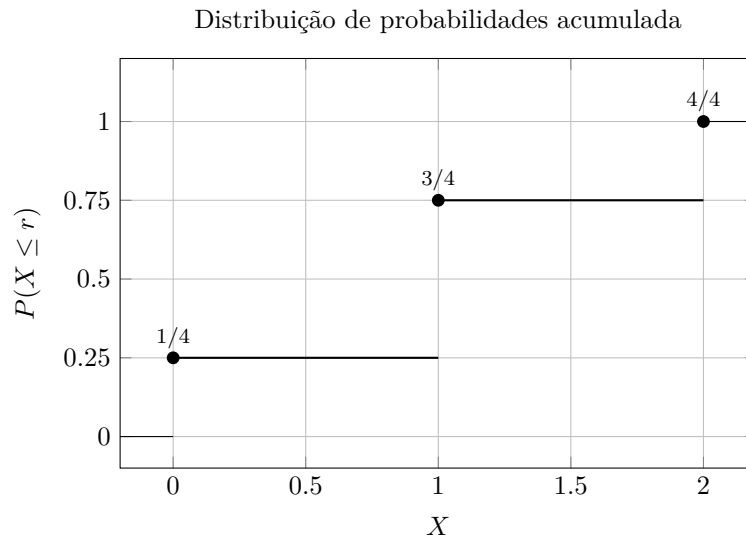
$$F(r) = P([X \leq r]) = P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{3}{4}$$

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:

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



Theorem 1.1 (Propriedades da distribuição acumulada). *Seja X uma variável aleatória definida em (Ω, \mathcal{F}, P) , então a f.d.a. de X (F_X ou F) verifica:*

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

Proof.

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

□

Theorem 1.2. Se F é a f.d.a. da variável aleatória X , então:

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

Proof.

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

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

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

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

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

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

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.

□

1.2 Natureza das variáveis aleatórias

- 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 $P(x_i) = P\{\omega : \omega \in \Omega \text{ e } X(\omega) = x_i\} \forall i \geq 1$.
- 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:

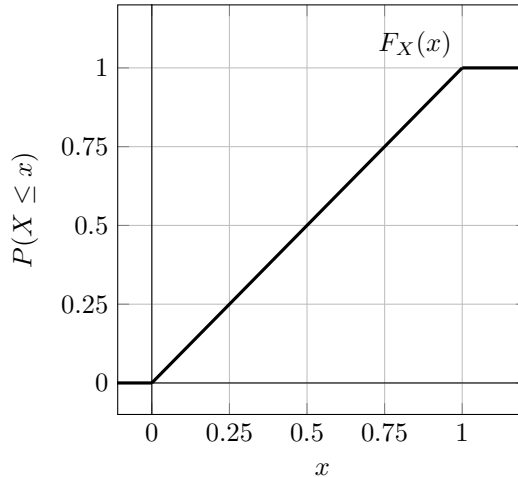
- $x \leq y \Rightarrow F(x) \leq F(y)$;
- Se $x_n \downarrow x \Rightarrow F(x_n) \downarrow F(x)$;
- Se $x_n \downarrow -\infty \Rightarrow F(x_n) \downarrow 0$ e se $x_n \uparrow \infty \Rightarrow F(x_n) \uparrow 1$.

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

- F_X é contínua;
- 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 1.3.

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

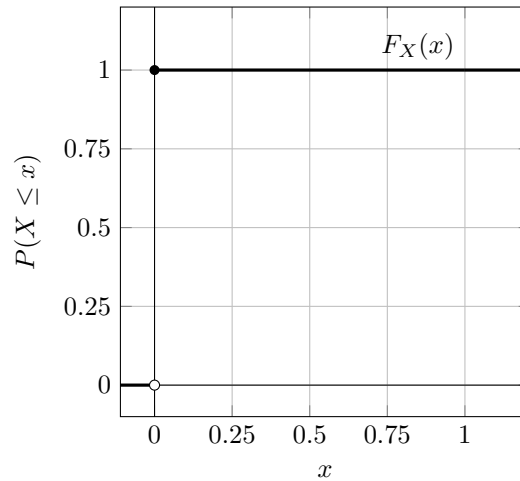


Notas:

- 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:

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

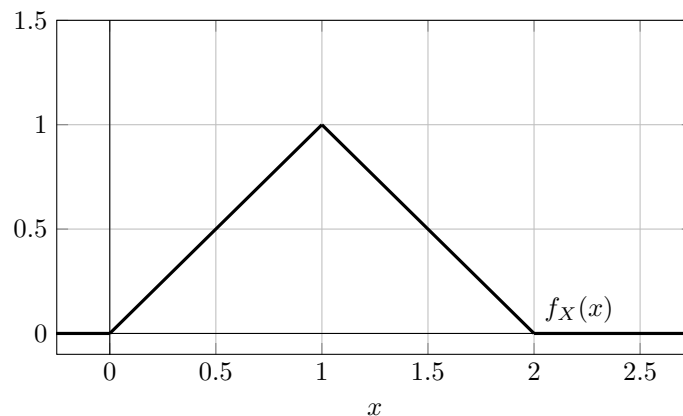


Notas:

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

Example 1.4. Considere a densidade triangular:

$$f_X(x) = \begin{cases} x, & \text{se } 0 \leq x < 1 \\ 2 - x, & \text{se } 1 \leq x < 2 \\ 0 & c.c. \end{cases}$$



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:

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

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

Conjecture 1.1. *Cada função de distribuição se corresponde com apenas uma distribuição? Não.*

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

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

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

□

1.3 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 1.1. *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, $[x \in B] \in \mathcal{A} \forall B \in \mathcal{B}$).*

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 $B_i \cap B_j = \emptyset \forall i \neq j \Rightarrow P_X(\bigcup B_n) = P(X \in \bigcup_n B_n) = P(\bigcup_n [X \in B_n]) = \sum_n P(X \in B_n)$.

Definition 1.3. A probabilidade P_X definida na σ -álgebra de Borel por $P_X(B) = P(X \in B)$ é a distribuição de X .

Proposition 1.2.

- 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 $f \Rightarrow P_X(B) = \int_B f_X dx$.

1.4 Variáveis contínuas

Proposition 1.3. 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.

1.4.1 Exemplos

Example 1.5 (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 1.6 (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 1.7 (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:

$$\begin{aligned} \int_0^\infty g(x) dx &= \int_0^\infty x^{\alpha-1} e^{-x} dx = \Gamma(\alpha) \\ f_X(x) &= \begin{cases} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} & x > 0 \\ 0 & c.c. \end{cases} \end{aligned}$$

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 $\beta = \frac{1}{2}$: $Y \sim \chi^2(n)$

1.5 Variáveis aleatórias multidimensionais

Definition 1.4. 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 1.8. 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 1.5. A função de distribuição acumulada do par de variáveis aleatórias (X, Y) é dada por:

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

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

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

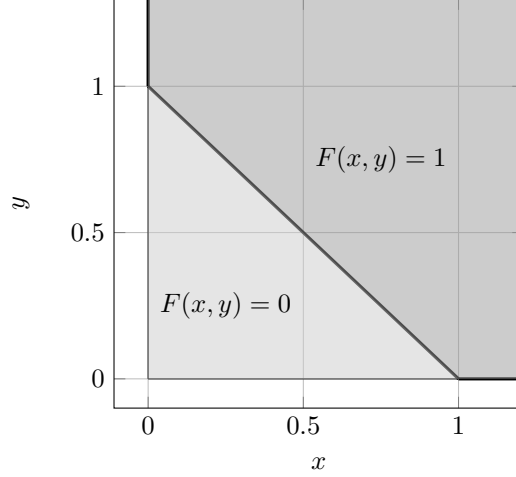
As provas de F_1 e F_2 são de simples construção. Para F_3 temos:

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

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

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

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

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

Definition 1.6. 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 $f(t_1, \dots, t_n) \geq 0 \forall (t_1, \dots, t_n) \in \mathbb{R}^n$ então (X_1, \dots, X_n) é um vetor absolutamente contínuo com densidade f (densidade conjunta).

Definition 1.7. 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 1.4.

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

1.6 Independência

Definition 1.8. 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 $\forall k$.

1.6.1 Critérios ou consequências

Proposition 1.5.

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

Proof.

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

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

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

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

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

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

□

1.6.2 Caso contínuo

Proposition 1.6.

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

Proof.

- **a)** Como consequência da proposição 1.5, 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 1.5: $F_i(x) = F_{X_i}(x_i)$, logo $f_{X_i} = f_i$.

□

1.6.3 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 1.9.

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

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

1.7 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 1.10. Sejam $X \sim U(0, 1)$ e $Y = -\ln(x)$. Temos que $\forall x$ valor de $X : x \in (-\infty, 0] \cup [1, \infty)$ o valor de $f_X(x) = 0$. Seja $x \in (0, 1) \Leftrightarrow -\ln(x) \in (0, \infty)$, logo $\forall y$ valor de $Y : y \in (0, \infty)$. Calculemos $F_Y(y) = P(Y \leq y)$:

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

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

Example 1.11. Sejam $X \perp Y$; $X \sim U(0, 1)$; $Y \sim U(0, 1)$; $Z = \frac{X}{Y}$. Determinar a distribuição de Z :

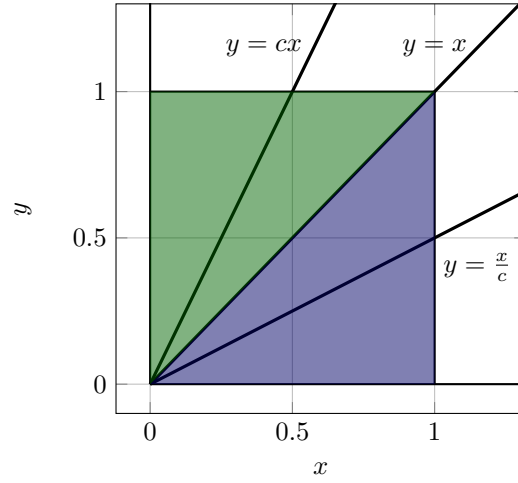
Os valores que geram indefinição de Z são: $X = Y = 0$ e $Y = 0, X > 0$, assim a boa definição de Z é no espaço $[0 < X \leq 1, 0 < Y \leq 1]$. Vejamos se esse intervalo contém toda a massa de probabilidade:

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

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 :

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