Notas de Aula - Capítulo 2

Probabilidade

Caio Gomes Alves

24/03/2025

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, $X : \mathcal{F} \to \mathbb{R}$. 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 $(\Omega, \mathcal{F}, P)$ um espaço de probabilidades. Uma função $X : \mathcal{F} \to \mathbb{R}$ 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 $(\Omega, \mathcal{F}, P)$ e $X : \mathcal{F} \to \mathbb{R}$ 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 $r \in [1, 2)$:

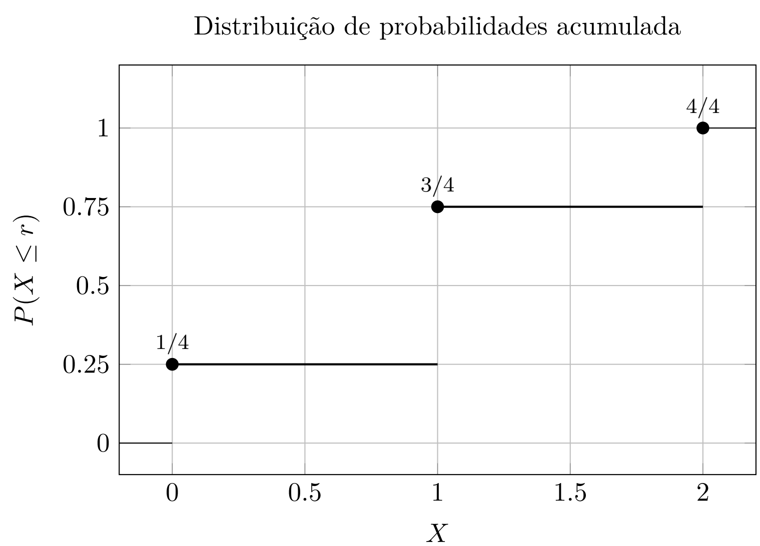
$$F(r) = P([X \leq r]) = P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{3}{4}$$

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 $(\Omega, \mathcal{F}, P)$, 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) $\lim_{t \to -\infty} F(t) = 0$ *e* $\lim_{t \to \infty} F(t) = 1$.

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 $n \to \infty$, 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 $n \to \infty$, logo, pela continuidade da função de probabilidade $P([X \leq x_n]) \downarrow P([X \leq x]), n \to \infty$.

c) Considere agora que $x_n \downarrow -\infty \Rightarrow [X \leq x_n] \downarrow \emptyset,\ n \to \infty \Rightarrow F(x_n) = P([X \leq x_n]) \downarrow P(\emptyset) = 0,\ n \to \infty$. Se $x_n \uparrow \infty \Rightarrow [X \leq x_n] \uparrow \Omega,\ n \to \infty \Rightarrow F(x_n) = P([X \leq x_n]) \uparrow P(\Omega) = 1,\ n \to \infty$.

□

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

a) Existem e são finitos os limites laterais $\lim_{t \to r^-} F(t), \lim_{t \to r^+} F(t), \forall r \in \mathbb{R}$ e $\lim_{t \to r^-} F(t) \leq \lim_{t \to r^+} F(t)$;

b) $\lim_{t \to r^+} F(t) = F(r), \forall r \in \mathbb{R}$;

c) F é descontínua em $r, r \in \mathbb{R}$ se e somente se $\lim_{t \to r^-} F(t) < F(r)$, com um salto de tamanho $F(r) - \lim_{t \to r^-} F(t)$;

d) $\forall r \in \mathbb{R}, P(X = r) = F(r) - \lim_{t \to r^-} F(t)$;

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 $\lim_{t \to r^-} F(t) \leq \lim_{t \to r^+} F(t)$.

c) Como F é monótona não-decrescente, uma descontinuidade só ocorre se e somente se $\lim_{t \to r^-} F(t) < \lim_{t \to r^+} F(t) = F(r)$.

d) Seja $r \in \mathbb{R}$. $[X \leq r] = \bigcap_{n=1}^{\infty}(r - \frac{1}{n} < x \leq r)$, logo:

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

e) Seja $\mathcal{D}$ o conjunto de pontos de descontinuidades de F, e seja $\lim_{t \to x^-} F(t) = F(x^-)$. 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 $x \in \bigcup_{n=1}^{\infty} \mathcal{D}_n \Rightarrow \exists n_1 : x \in \mathcal{D}_n \Rightarrow x \in \mathcal{D}$. $\mathcal{D}$ portanto é a união enumerável de conjuntos finitos, logo é enumerável.

□

1.2 Natureza das variáveis aleatórias

a) X é uma variável aleatória discreta se os valores que ela toma pertencem a um conjunto enumerável, logo $X : \Omega \to \{x_1, x_2, \ldots\}$ (ou seja, $X(\omega) \in \{x_1, x_2, \ldots\}, \forall \omega \in \Omega$) e $P : \{x_1, x_2, \ldots\} \to [0, 1]$ 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:

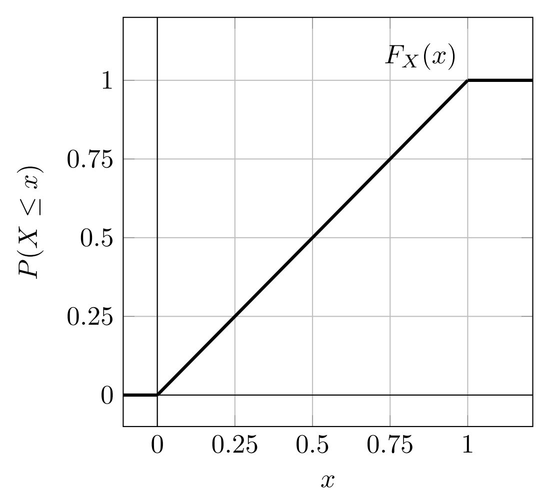
1. $x \leq y \Rightarrow F(x) \leq F(y)$;
2. Se $x_n \downarrow x \Rightarrow F(x_n) \downarrow F(x)$;
3. Se $x_n \downarrow -\infty \Rightarrow F(x_n) \downarrow 0$ e se $x_n \uparrow \infty \Rightarrow F(x_n) \uparrow 1$.

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 1.3.

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

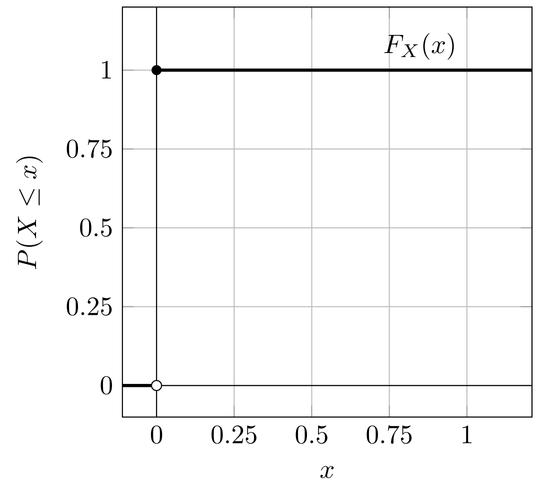


Notas:

- 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:

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

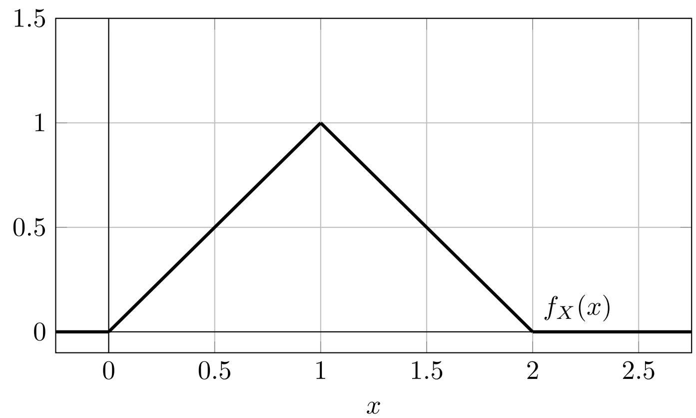


Notas:

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

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:

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

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 $(\Omega, \mathcal{A}, P)$, 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 $(\Omega, \mathcal{A}, P)$ 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 $(\Omega, \mathcal{A}, P)$, então o evento $[x \in \mathcal{B}] = \{\omega : \omega \in \Omega$ 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 $B = (a, b) \Rightarrow B = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right]$ Sabemos que os eventos $\left(a < X \leq b - \frac{1}{n}\right] \in \mathcal{A}$ e as suas uniões também pertencem à $\mathcal{A}$. Quanto à probabilidade, temos $P(X \in B) = P\left(\bigcup_{n=1}^{\infty}\left(a < X \leq b - \frac{1}{n}\right]\right) = \lim_{n\to\infty} P\left(\left(a < X \leq b - \frac{1}{n}\right]\right) = \lim_{n\to\infty} F_X\left(b - \frac{1}{n}\right) - F_X(a) = F_X(b^-) - F_X(a)$;
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)**.**

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

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

Example 1.6 (Distribuição Cauchy)**.**

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

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:

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

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

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