Notas de Aula - Capítulo 2

Probabilidade

Caio Gomes Alves

23/04/2025

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 $r \in [1, 2)$:

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

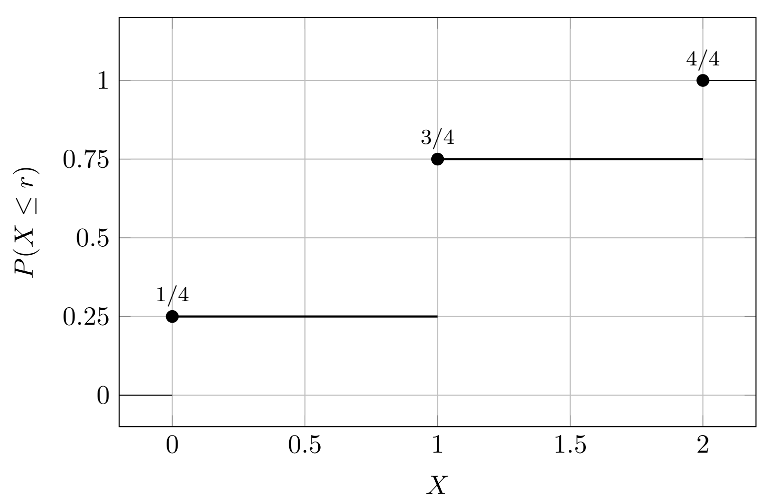
Para $r \geq 2$:

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

Logo, F é dada por:

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

Distribuição de probabilidades acumulada



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

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 $n \to \infty$, 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 $x_n \uparrow \infty \Rightarrow [X \leq x_n] \uparrow \Omega,\ n \to \infty \Rightarrow F(x_n) = P([X \leq x_n]) \uparrow P(\Omega) = 1,\ n \to \infty$.

$\square$

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.

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

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.

$\square$

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 $X : \Omega \to \{x_1, x_2, \dots\}$ (ou seja, $X(\omega) \in \{x_1, x_2, \dots\}, \forall \omega \in \Omega$) e $P : \{x_1, x_2, \dots\} \to [0, 1]$ 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:

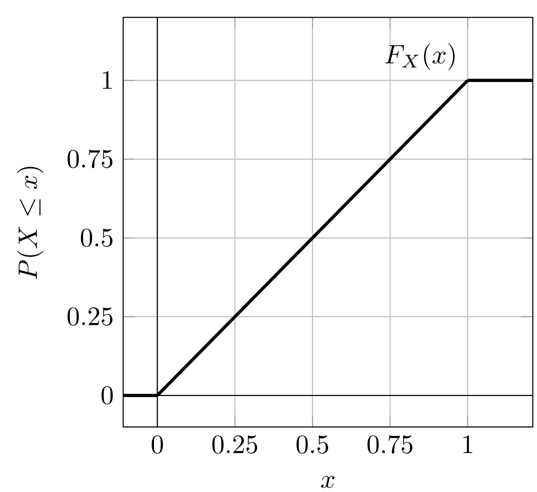
1. $x \leq y \Rightarrow F(x) \leq F(y)$;
2. Se $x_n \downarrow x \Rightarrow F(x_n) \downarrow F(x)$;
3. Se $x_n \downarrow -\infty \Rightarrow F(x_n) \downarrow 0$ e se $x_n \uparrow \infty \Rightarrow F(x_n) \uparrow 1$.

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.

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

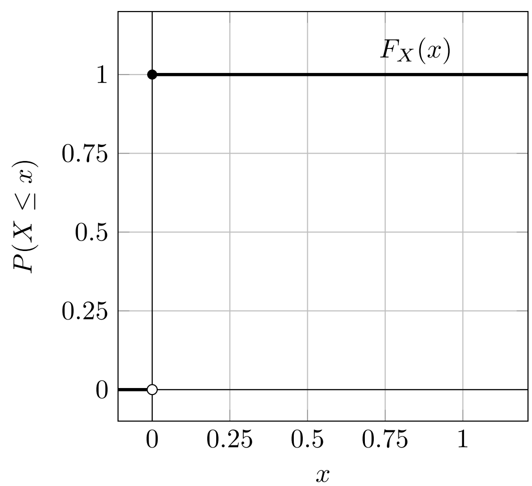


Notas:

- 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:

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

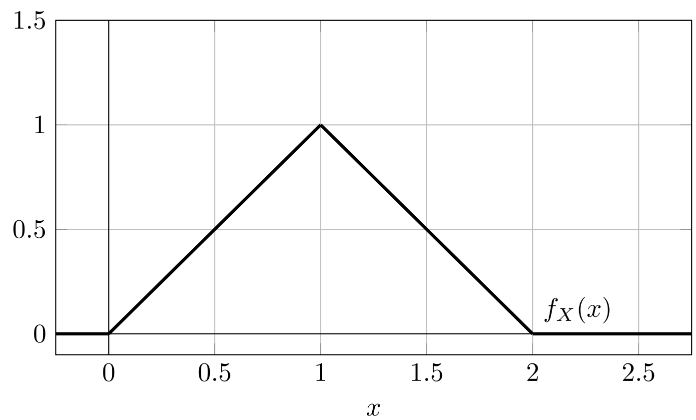


Notas:

- 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.*

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 $P(-X \leq \omega) = \Phi(\omega)$:

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

□

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 $B_1, \ldots, B_n \in \mathcal{B}$, 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 $\{x_1, x_2, \ldots\} \Rightarrow P_X(B) = \sum_{i:x_i \in B} P(x_i)$;
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)**.**

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

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:

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

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

1.5 Variáveis aleatórias multidimensionais

Definition 1.4. A distribuição de probabilidades do vetor aleatório dado por $(x_1, \ldots, x_n)$ é uma tabela que associa a cada valor $(x_1, \ldots, x_n)$ sua probabilidade $P(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$, 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 variáve aleatórias (X, Y) é dada por:

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

Seja $\underline{\text{X}} = (X_1, \ldots, X_n)$ tal que X_i é variável aleatória definida em $(\Omega, \mathcal{A}, P)$ $\forall i$. Então F, a acumulada de $\underline{\text{X}}$ 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: $\lim_{x_i \to -\infty} F(x_1, \ldots, x_n) = 0$ e $\lim_{x_i \to \infty \forall i} F(x_1, \ldots, x_n) = 1$.

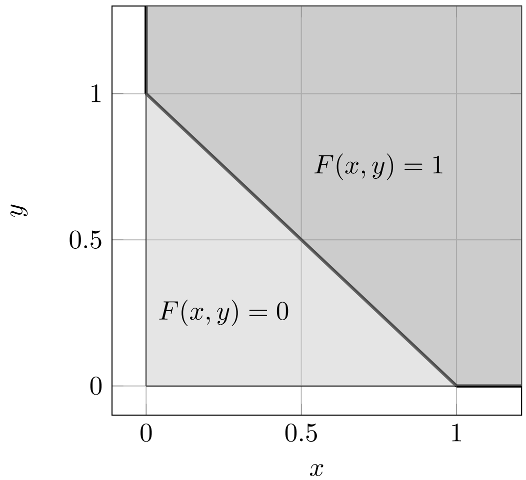
As provas de F_1 e F_2 são de simples construção. Para F_3 temos:

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

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

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 $\Delta_{k,I}(g(x_1, \ldots, x_k)) = g(x_1, \ldots, x_{k-1}, b) - g(x_1, \ldots, x_{k-1}, a)$ onde $g : \mathbb{R}^k \to \mathbb{R}; I = (a, b], a \leq b$. Logo, se $I_1 = (a_1, b_1]$ e $I_2 = (a_2, b_2], F : \mathbb{R}^2 \to \mathbb{R}$. 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: $\Delta_{1,I_1}\Delta_{2,I_2} \ldots \Delta_{n,I_n}(F(x_1, \ldots, x_n)) \geq 0 \ \forall I_k = (a_k, b_k]; a_k \leq b_k, k = 1, \ldots, n$.

Definition 1.6. Seja $F : \mathbb{R}^n \to \mathbb{R}$ 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 $(X_1, \ldots, X_n)$ toma valores em um conjunto discreto, o vetor é discreto;
- **b)** Se para o vetor aleatório $(X_1, \ldots, X_n)$, F é dada pela forma $F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \ldots \int_{-\infty}^{x_1} f(t_1, \ldots, t_n) dt_n \ldots dt_1$, $\forall (x_1, \ldots, x_n)$ onde $f(t_1, \ldots, t_n) \geq 0 \ \forall (t_1, \ldots, t_n) \in \mathbb{R}^n$ então $(X_1, \ldots, X_n)$ 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 $\underline{X} = (X_1, \ldots, X_n)$, 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, $P_{\underline{X}}(B) = P(\underline{X} \in B) = \int \ldots \int_B f(x_1, \ldots, x_n) dx_n \ldots dx_1$.*

1.6 Independência

Definition 1.8. As variáveis aleatórias são (coletivamente) independentes se:

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

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

1.6.1 Critérios ou consequências

Proposition 1.5.

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

Prova.

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

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

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

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

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

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

□

1.6.2 Caso contínuo

Proposition 1.6.

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

Prova.

- **a)** Como consequência da proposição 1.5, temos que: $F_{X_1 \ldots X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \forall (x_1, \ldots, x_n)$. Logo, por definição temos:

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

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

- **b)** Considere:

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

Defina $F_i(x) = \int_{-\infty}^{x_i} f_i(t)dt$. 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 $F_i(x) \to 1, x \to \infty$, 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 $F_X(x) = \lim_{y\to\infty} F(x,y) = F(x,\infty)$ 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, $(X,Y) \sim N_2\left(\binom{\mu_1}{\mu_2}, \begin{bmatrix}\sigma_1 & \rho \\ \rho & \sigma_2\end{bmatrix}\right)$, 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 $(\Omega,\mathcal{A},P)$. 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:

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

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:

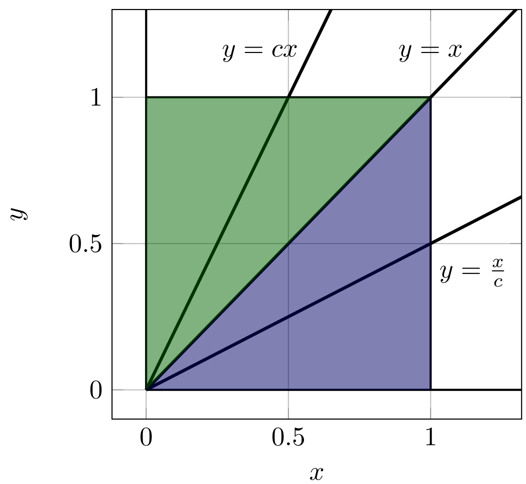
Os valores que geram indefinição de Z são: $X = Y = 0$ e $Y = 0, X > 0$, assim a boa definição de Z é no espaço $[0 < X \leq 1, 0 < Y \leq 1]$. Vejamos se esse intervalo contém toda a massa de probabilidade:

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

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:

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

1.7.1 Distribuição da Soma

Proposition 1.7.

- ***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:

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

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

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

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

1.7.2 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 1.7, podemos ver que:

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

1.7.3 Independência

Proposition 1.8. *Se $X_1, \ldots, X_n$ 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 $F_{Y_1 \ldots Y_n}(y_1, \ldots, y_n) = \prod_{i=1}^{n} F_{Y_i}(y_i)$:

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

□

Example 1.12. Considere $X \perp Y$, $X \sim Exp(1)$ e $Y \sim Exp(1)$. Determine:

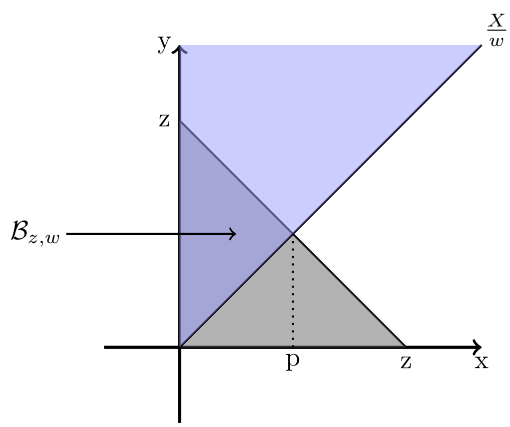
- a) A distribuição de $Z = X + Y$ e $W = \frac{X}{Y}$;
- b) Mostrar que $Z \perp W$.

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:

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

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

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

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

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

b)

Temos que as distribuições marginais de Z e W serão:

$$
\begin{aligned}
F_Z(z) &= \lim_{w\to\infty} F_{ZW}(z,w) = 1-e^{-z}-ze^{-z} \\
F_W(w) &= \lim_{z\to\infty} F_{ZW}(z,w) = \frac{w}{1+w}
\end{aligned}
$$

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:

$$
\begin{aligned}
f_{ZW}(z,w) &= \frac{\partial}{\partial z}\frac{\partial}{\partial w}\left(\frac{w}{1+w}\left(1-e^{-z}-ze^{-z}\right)\right) \\
&= \frac{1}{(1+w)^2}ze^{-z}I_{(0,\infty)}(z)I_{(0,\infty)}(w)
\end{aligned}
$$

1.8 Método do Jacobiano

Seja $g : G_0 \to G$, 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 $h = (h_1, \dots, h_n)$:

$$
\begin{aligned}
x_1 &= h_1(y_1, \dots, y_n) \\
&\vdots \\
x_n &= h_n(y_1, \dots, y_n)
\end{aligned}
$$

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 1.13. 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:

$$
\begin{aligned}
y_1 &= x_1 + x_2 \Rightarrow x_1 = y_1 - x_2 \\
y_2 &= \frac{y_1 - x_2}{x_2} \Rightarrow x_2 = \frac{y_1}{y_2 + 1} \Rightarrow x_1 = \frac{y_1 y_2}{y_2 + 1}
\end{aligned}
$$

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:

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

Definimos agora o Jacobiano:

$$
J(\underline{x}, \underline{y}) = \det \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix} \tag{1}
$$

Dessa forma, o Jacobiano da transformação será:

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

Pelo teorema do Jacobiano, temos que:

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

Se f é integrável em A, com $A \subseteq G_0$ e $h = g^{-1}$. Assim, usando os valores do exemplo 1.12, 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:

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

Que é a mesma densidade conjunta encontrada para Z e W no exemplo 1.12.

1.8.1 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:

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

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:

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

Theorem 1.3. *Sob as condições impostas no início da seção, a densidade conjunta de* $(Y_1, \dots, Y_n)$ *é dada por:*

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

1.8.2 Propriedades do Jacobiano

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

$$J(\underline{x}, \underline{y}) = (J(\underline{y}, \underline{x}))^{-1} \big|_{\underline{x}=h(y)} \tag{2}$$

Example 1.14. Retornando ao problema apresentado no exemplo 1.12:

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

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

Example 1.15. 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:

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

Theorem 1.4. *Sejam $G_1, G_2, \ldots, G_k$ disjuntos tais que $P\left(\underline{X} \in \bigcup_{i=1}^{k} G_i\right) = 1$, tal que $g\big|_{G_l}$ é 1:1 para todo $l = 1, \ldots, k$. Denotamos por $h^{(l)}$ a inversa de g em G_l, e definimos assim o Jacobiano local $J_l(\underline{x}, \underline{y})$ como:*

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

Example 1.16. 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:

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

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

Notas:

- Se $X_1, \ldots, X_n$ são iid, com $X_i \sim N(0,1) \Rightarrow X_1^2 + \ldots + X_n^2 \sim \chi^2(n)$;
- 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 $X_1, \ldots, X_n$, 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 $T \sim t(n) \Rightarrow T^2 \sim F(1,n)$.

1.9 Exercícios

TODO