Introdução à Probabilidade

Thiago Ramos

1 de outubro de 2024

Sumário

Legenda das Caixas

Caixas desta cor representam resultados importantes, como teoremas, exemplos, etc.

Caixas desta cor representam observações importantes, como intuição dos problemas, conexões com outros resultados, etc.

Capítulo 1

Introdução à Probabilidade

As Principais referências usadas foram:

1. [Dur10]
2. [AS16]
3. [Kni09]
4. [Sto14]

1.1 Teoria da Medida

Definição 1.1.1 (π-System)**.** Uma família de conjuntos $\mathcal{F} \subset \Omega$ é um π-system se para todo $A, B \in \mathcal{F}$, temos que $A \cap B \in \mathcal{F}$.

Definição 1.1.2 (λ-System/ Dynkin System)**.** Uma família de conjuntos $\mathcal{F} \subset \Omega$ é um λ-System (ou Dynkin-System) se :

1. $\Omega \in \mathcal{F}$.
2. Se $A, B \in \mathcal{F}$ e $A \subset B$, então $B - A \in \mathcal{F}$.
3. Se $(A_n) \subset \mathcal{F}$ e $A_n \uparrow A$ então $A \in \mathcal{F}$.

Além disso, vamos dizer que $\mathcal{F} \subset \Omega$ é um π-λ-system se este for um π-system e um λ-system.

Observação 1.1.3. É claro que toda σ-álgebra é um π-system e um λ-system.

Lema 1.1.4. *Temos que $\mathcal{F}$ é uma σ-álgebra em Ω se, e somente se, $\mathcal{F} \subset \Omega$ é um π-λ-system.*

Demonstração. A ida é trivial. Suponha que $\mathcal{F}$ é um π-λ-system, vamos mostrar que $\mathcal{F}$ é uma σ-álgebra.

Como $\mathcal{F}$ é um π-system, então $\Omega \in \mathcal{F}$ e como $\mathcal{F}$ é um λ-system, então se $A \in \mathcal{F}$, temos que $A^c = \Omega - A \in \mathcal{F}$. Além disso, dados $A, B \in \mathcal{F}$, temos que

$A \cup B = (A^c \cap B^c)^c \in \mathcal{F}$, logo $\mathcal{F}$ também é fechado por uniões finitas, mas como é um λ-system, tomando $B_n = \cup_i^n A_i \in \mathcal{F}$, temos o resultado, já que $B_n \uparrow \cup A_n$.

■

Definição 1.1.5. Considere a família de subconjuntos $\mathcal{F} \subset \Omega$. Então $\lambda(\mathcal{F})$ é o menor λ-system contendo $\mathcal{F}$.

Dizemos que $\lambda(\mathcal{F})$ é o λ-system gerado por $\mathcal{F}$.

Lema 1.1.6. *Se $\mathcal{F}$ é um π-system, então $\lambda(\mathcal{F}) = \sigma(\mathcal{F})$.*

Demonstração. É claro que $\sigma(\mathcal{F})$ é um λ-system para $\mathcal{F}$ e então,

$$\lambda(\mathcal{F}) \subset \sigma(\mathcal{F}).$$

Portanto, se mostrarmos que $\lambda(\mathcal{F})$ é um $\pi - \lambda$-system, pelo lema anterior vamos ter que $\lambda(\mathcal{F})$ é uma σ-álgebra, logo

$$\sigma(\mathcal{F}) \subset \lambda(\mathcal{F}) \subset \sigma(\mathcal{F}),$$

concluindo o resultado.

Vamos mostrar então que $\lambda(\mathcal{F})$ é π-system, já que claramente já é λ-system.

:=:=

■

1.2 Resultados Básicos

Exemplo 1.2.1 (Uma relação importante)**.** Suponha que $\mathbb{E}(|X|) < \infty$. Sabemos que se $X \geq 0$ tem densidade f_X então

$$\mathbb{E}(X) = \int_0^\infty x f_X(x) dx$$

mas entao, se $F_X(t) = \mathbb{P}(X \leq t)$ é a cumulativa de X, como $f_X(t) = F_X'(t)$ temos que

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^a x f_X(x) dx \\
&= -\int_0^a F_X(t)dt + aF_X(a) \\
&= a - \int_0^a F_X(t)dt - a + aF_X(a) \\
&= \int_0^a 1 - F_X(t)dt + a(F_X(a) - 1) \\
&= \int_0^a P(X > t)dt + a(F_X(a) - 1).
\end{aligned}$$

Tomando o limite $a \to +\infty$ temos que $a(F_X(a) - 1) \to 0$, já que

$$a(F_X(a) - 1) = a\mathbb{P}(X > a) \leq \mathbb{E}(X 1_{X>a}) \to 0,$$

pelo Teorema da Convergência Dominada (lembre-se que pedimos que $\mathbb{E}(|X|) < \infty$). Ficamos então com

$$\mathbb{E}(X) = \int_0^\infty P(X > t)dt.$$

Essa relação fica bem clara no caso discreto, já que temos

$$\begin{array}{lccccc}
\int_0^\infty P(X > t)dt & = \mathbb{P}(X = 1) & +\mathbb{P}(X = 2) & +\mathbb{P}(X = 3) & + \cdots \\
 & 0 & +\mathbb{P}(X = 2) & +\mathbb{P}(X = 3) & + \cdots \\
 & \vdots & & & \\
 & = \mathbb{P}(X = 1) & +2\mathbb{P}(X = 2) & +3\mathbb{P}(X = 3) & + \cdots \\
 & = & \int_0^\infty x P(X = x) = \mathbb{E}(X). & &
\end{array}$$

Exemplo 1.2.2 (Desigualdade Triangular para Probabilidades)**.** Vamos mostrar que

$$\mathbb{P}(|X-Y|>\varepsilon)\leq\mathbb{P}(|X|>\varepsilon/2)+\mathbb{P}(|Y|>\varepsilon/2),$$

isto é, que

$$\{|X-Y|>\varepsilon\}\subset\{|X|>\varepsilon/2\}\cup\{|Y|>\varepsilon/2\}.$$

Para isso, suponha que vale $\{|X-Y|>\varepsilon\}$ mas não $\{|X|>\varepsilon/2\}$, isto é, $\{|X|\leq\varepsilon/2\}$.
Como

$$|X|+|Y|\geq|X-Y|>\varepsilon,$$

temos que

$$|Y|>\varepsilon-\varepsilon/2=\varepsilon/2,$$

concluindo o resultado.

1.3 Distribuições

Teorema 1.3.1. *Seja X uma v.a. e F sua distribuição. Então*

1. *F é não decrescente.*
2. *Temos que*
$$\lim_{x \to +\infty} F(x) = 1$$
e que
$$\lim_{x \to -\infty} F(x) = 0.$$
3. *F é contínua à direita, isto é,*
$$\lim_{y \downarrow x} F(y) = F(X).$$
4. *Se $F(x-) = \lim_{y \uparrow x} F(y)$, então*
$$F(x-) = \mathbb{P}(X < x).$$
5. *Temos que*
$$\mathbb{P}(X = x) = F(x) - F(x-).$$
6. *F tem no máximo uma quantidade enumerável de descontinuidades.*

Demonstração. Vamos provar cada item separadamente.

1. Se $x < y$, então
$$F(y) = \mathbb{P}(X \leq y) = \mathbb{P}(X \leq x) + \mathbb{P}(x < X \leq y) \geq \mathbb{P}(X \leq x) = F(x).$$
2. Note que a família de eventos $\{X \leq n\}$ é crescente e que
$$\bigcup_{i=1}^{n} \{X \leq i\} = \Omega.$$
3. Note que, se $y_n \downarrow x$, então a família de eventos $\{X \leq y_n\}$ é decrescente e que
$$\bigcap_{i=1}^{n} \{X \leq y_i\} = \{X \leq x\}.$$
4. Note que,se $y_n \uparrow x$, então a família de eventos $\{X \leq y_n\}$ é crescente e que
$$\bigcup_{i=1}^{n} \{X \leq y_i\} = \{X < x\}.$$
5. Basta notar que
$$\mathbb{P}(X = x) = \mathbb{P}(X \leq x) - \mathbb{P}(X < x),$$
e usar os itens anteriores.

6. Para cada descontinuidade x de F, temos que
$$F(x-) - F(x+) > 0,$$
tome então $q_x \in \mathbb{Q}$ tal que $F(x-) < q_x < F(x+)$. Como F é não decrescente, é fácil ver que se x e y são pontos de descontinuidade com $x \neq y$, então $q_x \neq q_y$. Mas então existe uma injeção dos pontos de descontinuidade de F em $\mathbb{Q}$, logo só podem existir enumeráveis pontos de descontinuidade.

■

Proposição 1.3.2. *Seja X uma v.a. e suponha que sua distribuição F é **contínua**. Então $Y = F(X)$ tem distribuição uniforme $U(0,1)$, isto é,*
$$P(Y \leq y) = y.$$

Demonstração. Suponha inicialmente que F é estritamente crescente e portanto é inversível. Então o resultado é óbvio, já que
$$\begin{aligned}\mathbb{P}(Y \leq y) &= \mathbb{P}(F(X) \leq y)\\ &= \mathbb{P}(X \leq F^{-1}(y))\\ &= F \circ F^{-1}(y) = y.\end{aligned}$$
No caso em que F é não decrescente, basta considerar
$$F^{-1}(y) = \inf_t \{t \in \mathbb{R} : \ F(t) \geq y\}.$$
Como F é não decrescente e **contínua**, é fácil ver que
$$F \circ F^{-1}(y) = y.$$

■

Note que o resultado anterior não é verdade se pedirmos apenas continuidade à direita de F, já que se x é um ponto de descontinuidade de F e $F(x-) < \alpha < F(x+)$, então $F \circ F^{-1}(\alpha) > \alpha$ (faça um desenho).

Fazendo um desenho, é fácil ver que a inversa generalizada de F é não decrescente e continua à esquerda, sendo que, os pontos de descontinuidade de F vão gerar pedaços constantes em F^{-1} e os pedaços constantes de F vão gerar trechos de descontinuidade em F^{-1}.

Proposição 1.3.3. *Suponha que F seja uma função que satisfaz os itens 1,2 e 3 do teorema anterior, então existe uma v.a. Y com distribuição F.*

Demonstração. Seja $\Omega = (0.1)$ com a medida de Lebesgue, e defina para $t \in \Omega$
$$X(t) = F^{-1}(t) = \inf_y \{y \in \mathbb{R} : \ F(y) \geq t\}.$$
É fácil ver (faça um desenho) que
$$m(t : X(t) \leq w) = F(w).$$

■

1.4 Densidades

Exemplo 1.4.1. Seja $Z = X^2$, onde $X \sim N(0,1)$, vamos calcular a função de densidade de Z. Note que

$$\begin{aligned} F_Z(t) &= \mathbb{P}(Z < t) = \mathbb{P}(\sqrt{t} < X < \sqrt{t}) \\ &= \int_{-\sqrt{t}}^{\sqrt{t}} f_X(x)dx, \end{aligned}$$

onde f_X é a densidade de uma normal comum. Temos então que

$$\begin{aligned} f_Z(t) = \frac{dF_Z(t)}{dt} &= \frac{f_X(\sqrt{t})}{2\sqrt{t}} + \frac{f_X(-\sqrt{t})}{2\sqrt{t}} \\ &= \frac{1}{2\sqrt{t2\pi}}\left(e^{-t/2} + e^{-t/2}\right) \\ &= \frac{1}{\sqrt{2\pi t}}e^{-t/2} = \frac{1}{\sqrt{2\pi}}t^{-1/2}e^{-t/2} \\ &= \frac{1}{\sqrt{2\pi}}t^{1/2-1}e^{-t/2}. \end{aligned}$$

1.5 Independência

Lema 1.5.1. *Suponha que X, X' são i.i.d. e que $X = X'$ q.t.p.. Então X é constante q.t.p..*

Demonstração. Note que, para qualquer evento A

$$\mathbb{P}(X \in A) = \mathbb{P}(X \in A, X' \in A) = \mathbb{P}(X \in A)^2.$$

Logo,

$$\mathbb{P}(X \in A) \in \{0, 1\}.$$

■

Exemplo 1.5.2. Vamos mostrar que conjuntos independentes 2 a 2 não necessariamente são independentes. Para isso, sejam $X_i,\ i = 1, 2, 3$ v.a.'s independentes tal que

$$\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = 1/2.$$

Se $A_i = \{X_j = X_k\},\ i \neq j \neq k$, é fácil ver que

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(X_1 = X_2 = X_3) = 1/4 = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

Mas

$$\mathbb{P}(A_i \cap A_j \cap A_k) = \mathbb{P}(X_1 = X_2 = X_3) = 1/4 \neq 1/8.$$

Exemplo 1.5.3. Suponha que $(X_1, \ldots, X_n)$ tem densidade $f(x_1, \ldots, x_n)$. Vamos mostrar que se $f(x)$ pode ser escrita como $\prod g_i(x_i)$, onde $g_i \geq 0$ são mensuráveis, então X_i são independentes. Note que não estamos assumindo que g_i são densidades.

Vamos supor que temos duas v.a., o caso geral sai por indução. Temos então que

$$\mathbb{P}((X, Y) \in (A, B)) = \int_A g(x)dx \int_B h(y)dy.$$

Supondo $B = \Omega$, temos que

$$\mathbb{P}(X \in A) = \mathbb{P}((X, Y) \in (A, \Omega)) = \int_A g(x)dx \int_\Omega h(y)dy.$$

Isto é, para todo $A \subset \Omega$ mensurável,

$$\int_A f_X(x)dx = \int_A g(x)dx \int_\Omega h(y)dy = C_h \int_A g(x)dx,$$

onde $C_h > 0$. Como isso vale para qualquer A, temos que

$$f_X(x) = g(x)C_h,$$

e da mesma forma

$$f_Y(y) = h(y)C_g.$$

Além disso, tomando $A = B = \Omega$, é fácil ver que $C_g C_h = 1$. Mas então

$$\begin{aligned}\mathbb{P}((X,Y) \in (A,B)) = &\int_A g(x)dx \int_B h(y)dy \\ &\frac{1}{C_h C_g} \int_A f_X(x)dx \int_B f_Y(y)dy \\ &\mathbb{P}(X \in A)\mathbb{P}(Y \in B).\end{aligned}$$

Para o caso geral em que $(X,Y) \in D$, isto é, D não é um retângulo, basta utilizar aqueles argumentos de medida, já que os retângulos geral a σ-álgebra de Borel.

1.6 Algumas Desigualdades Clássicas

Proposição 1.6.1 (Desigualdade de Jensen)**.** *Seja* $\varphi : \mathbb{R} \to \mathbb{R}$ *convexa e* $X : \Omega \to \mathbb{R}$ *uma v.a. qualquer. Então*

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)).$$

Demonstração. Considere uma reta suporte de φ em $y = \mathbb{E}X$, temos então que

$$\varphi(x) \geq \varphi(\mathbb{E}X) + (\mathbb{E}X - x)\alpha, \ \forall x \in \mathbb{R},$$

para algum $\alpha \in \mathbb{R}$ fixado. Integrando a expressão acima, temos que

$$\begin{aligned} \mathbb{E}\varphi(X) &\geq \mathbb{E}[\varphi(\mathbb{E}X) + (\mathbb{E}X - X)\alpha] \\ &\geq \mathbb{E}(\varphi(\mathbb{E}X)) + \mathbb{E}(\mathbb{E}X - X)\alpha \\ &= \varphi(\mathbb{E}X) \end{aligned}$$

■

Exemplo 1.6.2 (Intuição de Jensen)**.** Sabemos que se $\sum_{i=1}^n a_i = 1, \ a_i \geq 0$ então, se φ é convexa:

$$\varphi\left(\sum_{i=1}^n a_i x_i\right) \leq \sum_{i=1}^n a_i \varphi(x_i).$$

Como estamos num espaço de probabilidade, temos que

$$\sum_\omega \mathbb{P}(d\omega) = 1,$$

logo,

$$\begin{aligned} \varphi(\mathbb{E}(X)) &= \varphi\left(\sum_\omega X(\omega)\mathbb{P}(d\omega)\right) \\ &\leq \sum_\omega \varphi(X(\omega))\mathbb{P}(d\omega) = \mathbb{E}(\varphi(X)). \end{aligned}$$

Proposição 1.6.3 (Desigualdade de Young)**.** *Suponha que* $p, q \in (0, \infty)$ *e que* $1/p + 1/q = 1$*. Então para* $a, b > 0$

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Demonstração. Como $x \mapsto e^x$ é convexa, temos que

$$\exp(t\alpha + (1-t)\beta) \leq t\exp(\alpha) + (1-t)\exp(\beta),$$

tomando $t = 1/p$ e $\alpha = \log a^p, \ \beta = \log b^q$, ficamos com

$$\exp((\log a^p)/p + (\log b^q)/q) \leq \frac{1}{p}\exp(\log a^p) + \frac{1}{q}\exp(\log b^q),$$

isto é,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

■

Observação 1.6.4 (Intuição para Young). Suponha que $f \geq 0$ é estritamente crescente e tome $a, b \geq 0$. É fácil ver que (ver figura ??)

$$ab \leq \int_0^a f(x)dx + \int_0^b f^{-1}(x)dx.$$

Para provarmos Young, basta utilizarmos essa observação com $f(x) = x^{p-1}$.

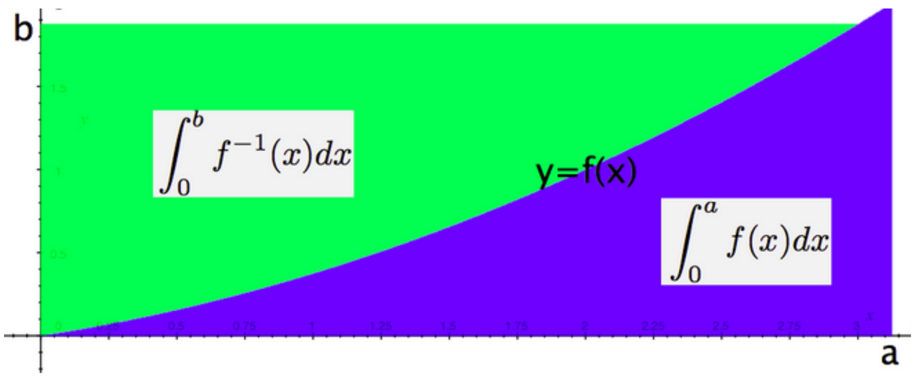


Figura 1.1: Retirado de https://math.stackexchange.com/questions/149901/geometric-interpretation-of-youngs-inequality.

Proposição 1.6.5 (Desigualdade de Holder)**.** *Suponha que* $X \in L^p(\Omega)$ *e* $Y \in L^q(\Omega)$ *tal que* $p, q \in [1, +\infty]$ *e* $\frac{1}{p} + \frac{1}{q} = 1$. *Então,*

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q.$$

Demonstração. O caso em que $p = 1$ e $q = \infty$ é trivial. Para $p, q \in (0, +\infty)$, note que, por Young,

$$\begin{aligned}\mathbb{E}\left(\frac{|X|}{\|X\|_p}\frac{|Y|}{\|Y\|_q}\right) &\leq \frac{1}{p}\mathbb{E}\left(\frac{|X|^p}{\|X\|_p^p}\right) + \frac{1}{q}\mathbb{E}\left(\frac{|Y|^q}{\|Y\|_q^q}\right) \\ &\leq \frac{1}{p} + \frac{1}{q} = 1.\end{aligned}$$

■

Observação 1.6.6. Note que Cauchy–Schwarz é um caso específico para $q = 2$.

Corolário 1.6.7. *Se* $1 \leq p \leq q \leq +\infty$ *então* $L^q(\Omega) \subset L^p(\Omega)$.

Demonstração. Suponha que $X \in L^q(\Omega)$. Se $\alpha = q/p \geq 1$ e $\beta \geq 1$ tal que $1/\alpha + 1/\beta = 1$, então por Holder,

$$\mathbb{E}(|X^p|) = \mathbb{E}(|1 \cdot X^p|) \leq (\mathbb{E}(|X^{\alpha \cdot p}|))^{1/\alpha} \cdot (\mathbb{E}(|1|^{\beta}))^{1/\beta} = (\mathbb{E}(|X^q|))^{p/q},$$

ou seja,

$$\|X\|_p \leq \|X\|_q < \infty.$$

$\blacksquare$

1.7 Função Geradora de Momento

Definição 1.7.1 (Função Geradora de Momento)**.** Seja X uma variável aleatória, definimos a Função Geradora de Momento M_X de X como

$$M_X(t) = \mathbb{E}(e^{Xt}).$$

Proposição 1.7.2. *Suponha que ???? , então*

$$\frac{d}{dt}\mathbb{E}(e^{Xt}) = \mathbb{E}\left(\frac{d}{dt}e^{X}t\right) = \mathbb{E}(Xe^{Xt}).$$

***terminar a prova

Exemplo 1.7.3 (Bernoulli)**.** Seja X Bernoulli com $P(X = 1) = p$, então é fácil ver que

$$M_X(t) = pe^t + 1 - p.$$

Exemplo 1.7.4 (Geométrica)**.** Seja X geométrica com probabilidade de acerto p, então

$$\begin{aligned} M_X(t) &= \sum_{k=1}^{\infty} e^{tk}(1-p)^{k-1}p \\ &= \frac{e^t p}{1-(1-p)e^t} \end{aligned}$$

se $(1-p)e^t < 1$.

Exemplo 1.7.5 (Binomial)**.** Seja X binomial $B(n, p)$, então

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{n} \binom{n}{k} e^{tk}(1-p)^{n-k}p^k \\ &= \sum_{k=0}^{n} \binom{n}{k} (1-p)^{n-k}(e^t p)^k = (1-p+pe^t)^n. \end{aligned}$$

Exemplo 1.7.6 (Uniforme)**.** Seja X uniforme $U[a, b]$, então

$$M_X(t) = \frac{1}{b-a}\int_b^a e^{tx}dx = \frac{e^{tb}-e^{ta}}{t(b-a)}.$$

Exemplo 1.7.7 (Normal)**.** Seja X normal $N[\mu, \sigma^2]$ e suponha que $\mu = 0$ e $\sigma^2 = 1$, então

$$\begin{aligned} M_X(t) &= \frac{1}{\sqrt{2\pi}}\int e^{tx}e^{-x^2/2}dx \\ &= \frac{e^{t^2/2}}{\sqrt{2\pi}}\int e^{(x-t)^2/2}dx \\ &= e^{t^2/2}. \end{aligned}$$

E no caso caso geral, por um simples argumento de mudança de coordenadas, temos que

$$M_X(t) = e^{t\mu + t^2/2\sigma^2}$$

Através da caracterização da F.G.d.M. de uma normal $N(0, \sigma)$, podemos criar a seguinte definição:

Definição 1.7.8 (Sub-Gaussiana)**.** Seja X uma v.a. centrada. Então X é Sub-Gaussiana com fator de variância v se

$$M_X(t) \leq e^{t^2 v/2}.$$

Dizemos que $X \in \mathcal{G}(v)$.

Isto é, uma v.a. X centrada é sub-gaussiana se sua função geradora de momento é dominada pela f.g.d.m. de uma v.a. Y normal centrada com variância v, isto é,

$$M_X(t) \leq M_Y(t).$$

1.8 Exercícios Resolvidos

Exercício 1.8.1 ([GS01])**.** Fixada uma sequência s de caras e coroas de tamanho k , prove que com probabilidade 1, eventualmente essa sequência ocorrerá num lançamento de moedas honestas.

Demonstração. Vamos particionar os inteiros positivos em intervalos de tamanho k, isto é,

$$\mathbb{N} = \bigsqcup_{i=0}^{\infty} A_i, \;\; A_i = [ik, (i+1)k).$$

Portanto, para qualquer $i \in \mathbb{N}$,

$$\mathbb{P}(A_i = s) = (1/2)^k.$$

Note que, se s nunca ocorre, em particular $A_i \neq s$, para todo $i \in \mathbb{N}$, e portanto

$$\begin{aligned}\mathbb{P}(s \text{ não ocorrer em } kn \text{ lançamentos}) &\leq \mathbb{P}(A_i \neq s, \;\; i \leq n) \\ &= \prod_{i}^{n} \mathbb{P}(A_i \neq s) \\ &= (1 - 1/2^k)^n \xrightarrow{n\to\infty} 0.\end{aligned}$$

Ou seja,

$$\mathbb{P}(s \text{ ocorrer eventualmente }) = 1.$$

■

Exercício 1.8.2 ([GS01])**.** Sejam $A_1, \dots, A_n$ eventos com $n \geq 2$, prove que

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

Demonstração. A prova é por indução. Se $n = 2$ já sabemos que isso é verdade. Suponha que a afirmação vale para n, vamos mostrar que então vale para $n+1$.

Temos que

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) &= \mathbb{P}\left(A_n \cup \bigcup_{i=1}^{n} A_i\right) \\ &= \mathbb{P}(A_n) + \mathbb{P}(\bigcup_{i=1}^{n} A_i) - \mathbb{P}(A_n \cap \bigcup_{i=1}^{n} A_i) \\ &= \mathbb{P}(A_n) + \mathbb{P}(\bigcup_{i=1}^{n} A_i) - \mathbb{P}(\bigcup_{i=1}^{n} (A_i \cap A_n)),\end{aligned}$$

e com isso temos o resultado, via hipótese de indução. ■

Exercício 1.8.3 ([GS01])**.** Dados eventos $A_1, \dots, A_n$, suponha que é certo que pelo menos um desses eventos, mas não mais que dois, ocorra. Se $\mathbb{P}(A_i) = p$ e $\mathbb{P}(A_i \cap A_j) = q, \;\; i \neq j$, mostre que $p \geq 1/n$ e $q \leq 2/n$.

Demonstração. Pelo exercício anterior,

$$1 = \mathbb{P}(\bigcup_{i}^{n} A_i) = \sum_{i}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + 0 + \cdots + 0$$
$$= np - qn(n-1)/2.$$

Então, é fácil ver que

$$1 = np - qn(n-1)/2 \leq np.$$

Além disso,

$$qn(n-1)/2 = np - 1 \leq n - 1,$$

já que $p \leq 1$. ■

Exercício 1.8.4 ([GS01])**.** Suponha que A, B são independentes e mostre que A^c, B também são.

Demonstração. Note que

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = (1 - \mathbb{P}(A^c))\mathbb{P}(B),$$

e isso implica que

$$\mathbb{P}(A^c)\mathbb{P}(B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(A^c \cap B).$$

■

Exercício 1.8.5 (Teorema de Waring)**.** ([GS01]) Sejam $A_1, \ldots, A_n$ eventos quaisquer e considere N uma v.a. que conta quantos dos A_i ocorreram. Prove que

$$\mathbb{P}(N = k) = \sum_{i=0}^{n-k} (-1)^i \binom{k+i}{k} S_{k+i},$$

onde

$$S_j = \sum_{i_i < i_2 < \cdots < i_j} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_j}).$$

Demonstração. ************************ ■

Exercício 1.8.6 (Desigualdade de Kounias)**.** ([GS01]) Mostre que

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \min_k \left\{ \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i \neq k} \mathbb{P}(A_i \cap A_k) \right\}.$$

Demonstração. Já vimos que

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

e portanto

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

Agora, considere a matriz M com coordenadas $\mathbb{P}(A_i \cap A_j)$. Note que

$$\sum_{i<j} \mathbb{P}(A_i \cap A_j)$$

é equivalente a soma das entradas estritamente acima da diagonal de M e que, fixado k,

$$\sum_{i \neq k} \mathbb{P}(A_i \cap A_k) = M_{1,k} + \cdots + M_{k-1,k} + M_{k,j+1} \cdots + M_{k,n}.$$

Portanto, como cada elemento da expressão acima está estritamente acima da diagonal de M,

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &\leq \sum_{i=1} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) \\ &\leq \sum_{i=1} \mathbb{P}(A_i) - \sum_{i \neq k} \mathbb{P}(A_i \cap A_k). \end{aligned}$$

■

Exercício 1.8.7 ([AS16])**.** Sejam $v_1 \ldots, v_n$ vetores em $\mathbb{R}^n$ tal que $|v_i| = 1$ para todo i. Então existem $\sigma_i \in \{+1, -1\}$ tal que

$$|\sigma_1 v_1 + \ldots \sigma_n v_n| \leq \sqrt{n}$$

e da mesma forma, podemos encontrar σ's tal que

$$|\sigma_1 v_1 + \ldots \sigma_n v_n| \geq \sqrt{n}.$$

Demonstração. Considere σ_i v.a. uniformes em $\{-1, 1\}$ e defina

$$X = |\sigma_1 v_1 + \ldots \sigma_n v_n|^2.$$

Note que

$$\mathbb{E}(X) = \sum_i \sigma_i^2 v_i^2 = \sum_i \sigma_i^2 = \sum_i 1 = n.$$

Mas então, deve existir uma configuração de σ's de tal forma que

$$|\sigma_1 v_1 + \ldots \sigma_n v_n|^2 \leq n$$

e uma tal que

$$|\sigma_1 v_1 + \ldots \sigma_n v_n|^2 \geq n.$$

■

Para o próximo exercício, vamos precisar da seguinte definição:

Definição 1.8.8 (Uniformemente Integrável)**.** Seja H uma família de v.a.'s tal que

$$\sup_{X\in H}\mathbb{E}(|X|)<\infty.$$

Dizemos que H é uma família uniformemente integrável se

$$\sup_{X\in H}\int_{X>a}|X|d\mathbb{P}\to 0,\ a\to\infty.$$

Exercício 1.8.9 ([CY12])**.** Dado uma v.a. X, considere a medida limitada

$$\nu_X(A)=\int_A Xd\mathbb{P}.$$

Mostre que dado uma família H de v.a.'s não negativas tal que

$$\sup_{X\in H}\mathbb{E}(|X|)<\infty,$$

então:

1. H é uniformemente integrável.
2. $\nu_X,\ X\in H$ é equi-absolutamente contínua com respeito à $\mathbb{P}$, isto é,
$$\forall\varepsilon>0,\ \exists\delta>0,\ \forall A\text{ mensurável},\ \mathbb{P}(A)\leq\delta\Rightarrow\sup_{X\in H}\nu_X(A)\leq\varepsilon.$$
3. Para qualquer sequência de mensuráveis (A_n) tal que $A_n\downarrow\emptyset$, temos que
$$\lim_{n\to\infty}\left(\sup_{X\in H}\nu_X(A_n)\right)=0.$$
4. Para qualquer sequência disjunta de mensuráveis (B_n), temos que
$$\lim_{n\to\infty}\left(\sup_{X\in H}\nu_X(B_n)\right)=0.$$

Demonstração.

1→2 Suponha que H é uniformemente integrável e tome $\varepsilon>0$. Por hipótese, temos que existe a' grande, tal que

$$\nu_X(X>a')<\varepsilon/2,\ \forall X\in H.$$

Agora note que

$$\begin{aligned}\nu_X(A)&=\nu_X(A\cap(X>a'))+\nu_X(A\cap(X\leq a'))\\&\leq\nu_X(X>a')+\int_{A\cap(X\leq a')}Xd\mathbb{P}\\&\leq\varepsilon/2+a'\mathbb{P}(A\cap(X\leq a'))\\&\leq\varepsilon/2+a'\mathbb{P}(A).\end{aligned}$$

Portanto, escolhendo $\delta=\varepsilon/(2a')$, temos o resultado.

2→3 Como $A_n \downarrow \emptyset$, temos que
$$\lim_n \mathbb{P}(A_n) = 0.$$
Portanto, dado $\varepsilon > 0$, para todo n suficientemente grande, temos que
$$\mathbb{P}(A_n) \leq \delta,$$
onde $\delta > 0$ é como na hipótese do item 2. Logo,
$$\lim_n (\sup_{X \in H} A_n) \leq \varepsilon,$$
e como $\varepsilon > 0$ é qualquer, temos o resultado.

2→4 Como
$$\mathbb{P}(\cup B_n) = \sum \mathbb{P}(B_n) \leq 1,$$
temos que
$$\lim_n \mathbb{P}(B_n) = 0,$$
e isso é suficiente para usarmos o mesmo argumento do item anterior.

4→3 Suponha que exista uma sequência (A_n) tal que $A_n \downarrow \emptyset$ mas para algum $\varepsilon > 0$,
$$\lim_{n \to \infty} \left(\sup_{X \in H} \nu_X(A_n) \right) > \varepsilon.$$
Então existe uma subsequência (X_{n_k}, A_{n_k}) tal que
$$\nu_{X_{n_k}}(A_{n_k}) > \varepsilon/2.$$
Basta considerar então $B_k = A_{n_k} \backslash A_{n_{k+1}}$, essa sequência é disjunta, mas
$$\lim_{k \to \infty} \left(\sup_{X \in H} \nu_X(B_k) \right) > \varepsilon/2,$$
contrariando o item 4.

3→1 Suponha que H não é uniformemente integrável. Então, existe $\varepsilon > 0$ tal que conseguimos uma sequência (X_{n_k}, a_k) com $a_k \to \infty$ tal que
$$\int_{X_{n_k} > a_k} X_{n_k} d\mathbb{P} > \varepsilon/2.$$
Tomando
$$A_k = \bigcup_{i=k}^{\infty} \{X_{n_k} > a_k\},$$
conseguimos uma sequência decrescente, tal que
$$\lim_{k \to \infty} \left(\sup_{X \in H} \nu_X(A_k) \right) > \varepsilon/2.$$

Note que isso ainda não é suficiente para chegarmos numa contradição, já que não sabemos se $A_k \downarrow \emptyset$.

Note que podemos supor que a_k é da forma 2^k, já que se $2^k < a_k$, então

$$\int_{X_{n_k} > 2^k} X_{n_k} d\mathbb{P} \geq \int_{X_{n_k} > a_k} X_{n_k} d\mathbb{P} > \varepsilon/2.$$

Mas,

$$\int_{X_{n_k} > 2^k} X_{n_k} d\mathbb{P} > \int_{X_{n_k} > 2^k} 2^k d\mathbb{P} = 2^k \mathbb{P}(X_{n_k} > 2^k),$$

isto é,

$$\mathbb{P}(X_{n_k} > 2^k) < \frac{\sup_{X \in H} \mathbb{E}(X)}{2^k} = \frac{C}{2^k},$$

já que, por hipótese,

$$\sup_{X \in H} \mathbb{E}(X) < \infty.$$

Temos então que

$$\mathbb{P}(A_n) = \mathbb{P}(\bigcup_{i=n}^{\infty} (X_{n_n} > a_n)) \leq \sum_{i=n}^{\infty} \frac{C}{2^n} \xrightarrow{n \to \infty} 0.$$

Isso quer dizer que $A_n \downarrow A$ onde A é um conjunto com medida nula, e portanto mensurável. Logo, considerando uma nova sequência $A'_n = A_n \backslash A$, temos o resultado.

■

Exercício 1.8.10 (Distribuição LogNormal não é definida por seus momentos)**.** ([CY12]) Seja N uma v.a. com distribuição normal $N(0, \sigma^2)$. Defina

$$X = \exp N.$$

1. Encontre a densidade de X.

2. Prove que para todo $n, p \in \mathbb{Z}$,

$$\mathbb{E}\left[X^n \sin\left(\frac{p\pi}{\sigma^2} N\right)\right] = 0.$$

3. Mostre que existem infinitas probabilidades em $\mathbb{R}_+$ tal que:

   i. Para todo $n \in \mathbb{Z}$,

   $$\int x^n d\mu(x) = \exp\left(\frac{n^2\sigma^2}{2}\right),$$

   ii. μ tem densidade limitada com respeito à lei de $\exp N$.

Demonstração.

1. Para $t > 0$, temos que

$$\mathbb{P}(X \leq t) = \mathbb{P}(\exp N \leq t) = \mathbb{P}(N \leq \log t),$$

portanto, a densidade de X é dada por

$$\frac{d}{dt}\mathbb{P}(N \leq \log t) = \frac{1}{t\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\log t)^2}{2\sigma^2}\right).$$

2. Como

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i},$$

é fácil ver que

$$X^n \sin\left(\frac{p\pi}{\sigma^2} N\right) = \operatorname{Im}\left(\mathbb{E}(e^{zN})\right),$$

onde $z = n + \frac{\pi p}{\sigma^2} i$. Um simples cálculo nos mostra que

$$\mathbb{E}(e^{zN}) = \exp\left(\frac{z^2\sigma^2}{2}\right),$$

já que

$$\exp(zx)\exp\left(-\frac{x^2}{2\sigma^2}\right) = \exp\left(-\frac{(x-\sigma^2 z)^2}{2\sigma^2}\right)\exp\left(\frac{z^2\sigma^2}{2}\right).$$

Disso, é fácil ver que, existe uma constante $C > 0$ tal que

$$\operatorname{Im}\left(\exp\left(\frac{z^2\sigma^2}{2}\right)\right) = C \cdot \sin(np\pi) = 0,$$

já que $n, p \in \mathbb{Z}$.

3. Seja $f_l(x)$ a função de densidade de $X = \exp N$ que encontramos no item 1. Pelo item 2, temos que para todo $n \in \mathbb{Z}$,

$$0 = \mathbb{E}(X^n \sin((p\pi/\sigma^2)\log X)) = \int x^n \sin((p\pi/\sigma^2)\log x) f_l(x)dx.$$

Em particular,

$$\int (1 + \sin((p\pi/\sigma^2)\log x)) f_l(x)dx = 1,$$

ou seja,

$$x \mapsto (1 + \sin((p\pi/\sigma^2)\log x)) f_l(x)$$

é uma densidade. De fato, isso é verdade para

$$x \mapsto (1 + \sum_p c_p \sin((p\pi/\sigma^2)\log x)) f_l(x),$$

onde (c_p) é uma sequência de números reais tal que $\sum_p |c_p| \leq 1$, já que

$$1 + \sum_p c_p \sin(x) \geq 0.$$

Vamos denotar a lei dada pela densidade acima por Q_{c_p}. Portanto, se Y é distribuída de acordo com Q_{c_p}, temos que

$$\begin{aligned}
\mathbb{E}_{Q_{c_p}}(Y^n) &= \int y^n (1 + \sum_p c_p \sin((p\pi/\sigma^2)\log x)) f_l(x) dx \\
&= \int y^n f_l(x) dx \\
&= \mathbb{E}(X^n) \\
&= \mathbb{E}(e^{nN}) \\
&= \exp\left(\frac{n^2\sigma^2}{2}\right).
\end{aligned}$$

Onde a última igualdade vem do item 2.

Mostramos então que para cada sequência (c_p), conseguimos uma lei Q_{c_p} tal que se Y tem lei Q_{c_p}, então **todos os momentos** de Y coincidem com os momentos de $X = e^N$, porém X e Y **possuem leis diferentes!**. Ou seja, saber os momentos de uma v.a. Z não é suficiente para saber a lei de Z.

■

Capítulo 2

Convergência

As Principais referências usadas foram:

1. [vdV00]
2. [Dur10]

2.1 Modos de Convergência

Definição 2.1.1. Seja (X_n) uma sequência de vetores aleatórios e X um vetor aleatório.

1. Dizemos que X_n converge q.t.p. para X se

$$\mathbb{P}\left(\lim_{n\to+\infty} \|X_n - X\| = 0\right) = 1,$$

e simbolizamos tal convergência como $X_n \to_{qtp} X$.

2. Dizemos que X_n converge para X em probabilidade se dado $\varepsilon > 0$,

$$\lim_{n\to+\infty} \mathbb{P}(\|X_n - X\| > \varepsilon) = 0,$$

e simbolizamos tal convergência como $X_n \to_p X$.

3. Dizemos que X_n converge em L^p para X se, fixado $p > 0$,

$$\lim_{n\to+\infty} \mathbb{E}\|X_n - X\|^p = 0.$$

4. Dizemos que X_n converge para X em distribuição (ou em Lei, ou fracamente) se para cada ponto de continuidade de $x \mapsto \mathbb{P}(X \leq x)$, temos que

$$\lim_{n\to+\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x),$$

e simbolizamos tal convergência como $X_n \to_d X$.

Observação 2.1.2. Veremos mais adiante que podemos interpretar a convergência em distribuição como uma **convergência fraca*** no dual das funções contínuas e limitadas (ver Capítulo **??** e/ou Teorema 2.2.1 itens 1 e 2, por exemplo). Apesar de se tratar de uma convergência fraca* e não apenas fraca, por razões histórias abandonamos a * no nome.

Definição 2.1.3. Seja $(A_n)_{n\in\mathbb{N}}$ uma sequência de eventos mensuráveis. Então

$$\begin{aligned}\{A_n,\ i.o.\} &= \bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n \\ &= \limsup A_n\end{aligned}$$

Observe que

$$\mathbb{P}(\{A_n,\ i.o.\}) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{n=m}^{\infty} A_n\right),$$

já que os conjuntos $B_m = \bigcup_{n=m}^{\infty} A_n$ são encaixados.

Lema 2.1.4. *Seja*

$$B_n^{\varepsilon} = \{\omega \in \Omega :\ \|X_n(\omega) - X(\omega)\| > \varepsilon\},$$

então, X_n converge para X q.t.p. se, e só se, para todo $\varepsilon > 0$,

$$\mathbb{P}(B_n^{\varepsilon}\ i.o.) = 0.$$

Demonstração. Dado $\varepsilon > 0$, se

$$\mathbb{P}(B_n^{\varepsilon}\ i.o.) = 0,$$

então para quase todo $\omega \in \Omega$, existe $N(\omega,\varepsilon) \in \mathbb{N}$ tal que, para $n \geq N(\omega,\varepsilon)$,

$$\|X_n(\omega) - X(\omega)\| \leq \varepsilon.$$

Mas então, para quase todo $\omega \in \Omega$, temos que

$$\lim_{n\to+\infty} \|X_n(\omega) - X(\omega)\| \leq \varepsilon,$$

e como $\varepsilon > 0$ é um valor positivo arbitrário, temos que, para quase todo ponto

$$\lim_{n\to+\infty} \|X_n(\omega) - X(\omega)\| = 0,$$

ou equivalentemente,

$$\mathbb{P}\left(\lim_{n\to+\infty} \|X_n - X\| = 0\right) = 1.$$

O outro lado é trivial. ∎

Observação 2.1.5. Esta forma de escrever a convergência q.t.p. será útil quando estudarmos os lemas de Borel-Cantelli.

Proposição 2.1.6. *Se $X_n \to_{q.t.p.} X$, então $X_n \to_p X$.*

Demonstração. Se $X_n \to_{\text{q.t.p.}} X$, então dado $\varepsilon > 0$ já vimos que

$$\mathbb{P}(\|X_n - X\| > \varepsilon,\ i.o.) = 0.$$

Mas então

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(\|X_n - X\| > \varepsilon) &\leq \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=n}^{\infty} \|X_n - X\| > \varepsilon\right)\\ &= \mathbb{P}(\|X_n - X\| > \varepsilon,\ i.o.) = 0.\end{aligned}$$

■

Proposição 2.1.7. *Se $X_n \to_p X$, então $X_n \to_d X$.*

Demonstração. Dado $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}(X_n \leq x) &= \mathbb{P}(X_n \leq x,\ X \leq x + \varepsilon) + \mathbb{P}(X_n \leq x,\ X > x + \varepsilon)\\ &\leq \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(\|X_n - X\| > \varepsilon).\end{aligned}$$

E da mesma forma

$$\begin{aligned}\mathbb{P}(X \leq x - \varepsilon) &= \mathbb{P}(X_n \leq x,\ X \leq x - \varepsilon) + \mathbb{P}(X_n > x,\ X \leq x + \varepsilon)\\ &\leq \mathbb{P}(X_n \leq x) + \mathbb{P}(\|X_n - X\| > \varepsilon).\end{aligned}$$

E portanto, temos que

$$\mathbb{P}(X \leq x - \varepsilon) - \lim_n \mathbb{P}(\|X_n - X\| > \varepsilon) \leq \liminf_n \mathbb{P}(X_n \leq x),$$

e também

$$\limsup_n \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x + \varepsilon) + \lim_n \mathbb{P}(\|X_n - X\| > \varepsilon).$$

Ou seja,

$$\mathbb{P}(X \leq x - \varepsilon) \leq \liminf_n \mathbb{P}(X_n \leq x) \leq \limsup_n \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x + \varepsilon),$$

já que a sequência converge em probabilidade. Como estamos supondo que $F_X(x) = \mathbb{P}(X \leq x)$ é contínua em x e que $\varepsilon >$ é qualquer, temos o resultado. ■

Observação 2.1.8. Temos então que

$$\text{(q.t.p.)} \Rightarrow \text{(probabilidade)} \Rightarrow \text{(distribuição)}$$

2.2 Convergência Fraca em $\mathbb{R}^k$

2.2.1 Teorema Portmanteau

Teorema 2.2.1 (Portmanteau). [a] *Para qualquer sequência (X_n) de vetores aleatórios e X um vetor aleatório, são equivalentes:*

1. $X_n \to_d X$
2. $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ *para toda f contínua e limitada.*
3. $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ *para toda f Lipschitz e limitada.*
4. $\liminf_n \mathbb{E}(f(X_n)) \geq \mathbb{E}(f(X))$ *para toda f contínua e não negativa.*
5. $\liminf_n \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$ *para todo G aberto.*
6. $\limsup_n \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$ *para todo F fechado.*
7. $\mathbb{P}(X_n \in B) \xrightarrow{n\to\infty} \mathbb{P}(X \in B)$ *para todo Boreliano B tal que $\mathbb{P}(X \in \delta B) = 0$.*

[a] Portmanteau significa maleta em francês.

Demonstração do Teorema Portmanteau.

Vamos seguir a demonstração de [vdV00].

$1 \to 2$: Vamos supor que $F_X(x) = \mathbb{P}(X \leq x)$ é contínua para todo x. Utilizando a hipótese, temos que, para qualquer retângulo fechado $I \subset \mathbb{R}^d$,

$$\mathbb{P}(X_n \in I) \xrightarrow{n\to\infty} \mathbb{P}(X \in I).$$

Podemos escolher um retângulo I grande o suficiente de tal forma que

$$\mathbb{P}(X \notin I) < \varepsilon,$$

já que $\mathbb{P}(X \in \mathbb{R}^m) = 1$.

Como f é contínua e portanto absolutamente contínua em I (já que este é compacto), podemos particionar I em uma quantidade finita retângulos menores I_j, com $j = 1, 2, \ldots, k$ de tal forma que para todo $0 \leq j \leq k$,

$$|f(x) - f(y)| < \varepsilon, \ x, y \in I_j.$$

Fixado um ponto $x_j \in I_j$ para cada I_j, defina

$$f_\varepsilon = \sum_j f(x_j) 1_{Ij}.$$

Sem perda de generalidade, podemos supor que $f \in [-1, 1]$, já que f é limitada, e portanto

$$\begin{aligned}|\mathbb{E}(f(X_n)) - \mathbb{E}(f_\varepsilon(X_n))| &= |\mathbb{E}([f(X_n) - f_\varepsilon(X_n)]1_{X_n \in I}) + |\mathbb{E}([f(X_n) - f_\varepsilon(X_n)]1_{X_n \notin I}) \\ &\leq \varepsilon + 2\mathbb{P}(X_n \notin I),\end{aligned}$$

e então, para n grande o suficiente, temos que

$$|\mathbb{E}(f(X_n)) - \mathbb{E}(f_\varepsilon(X_n))| < 3\varepsilon.$$

E o mesmo vale para X, isto é,

$$|\mathbb{E}(f(X)) - \mathbb{E}(f_\varepsilon(X))| \leq \varepsilon + 2\mathbb{P}(X \notin I) \leq 3\varepsilon.$$

Por fim, temos que

$$\begin{aligned}|\mathbb{E}(f(X)) - \mathbb{E}(f(X_n))| &= |\mathbb{E}(f(X)) - \mathbb{E}(f(X_n)) \pm \mathbb{E}(f_\varepsilon(X)) \pm \mathbb{E}(f_\varepsilon(X_n))| \\ &\leq |\mathbb{E}(f(X)) - \mathbb{E}(f_\varepsilon(X))| \\ &+ |\mathbb{E}(f(X_n)) - \mathbb{E}(f_\varepsilon(X_n))| \\ &+ |\mathbb{E}(f_\varepsilon(X)) - \mathbb{E}(f_\varepsilon(X_n))|,\end{aligned}$$

portanto, só precisamos controlar o último termo. Mas note que

$$|\mathbb{E}(f_\varepsilon(X)) - \mathbb{E}(f_\varepsilon(X_n))| \leq \sum_j |\mathbb{P}(X_n \in I_j) - \mathbb{P}(X \in I_j)||f(x_j)| \xrightarrow{n\to\infty} 0.$$

Para a demonstração no caso geral, recomendamos a leitura da Seção 1.3 para melhor entendimento da função de distribuição e depois do item (a) do Teorema 2.2 em [vdV00].

2→ 3: Óbvio.

3→ 5: Note que

$$f_m(x) = (m \cdot \mathrm{dist}(x, G^c)) \wedge 1$$

é Lipschitz, limitada e

$$f_m(x) \uparrow 1_G(x),$$

portanto, como

$$f_m(x) \leq 1_G(x),$$

temos que, *por hipótese*,

$$\liminf_n \mathbb{P}(X_n \in G) \geq \liminf_n \mathbb{E}(f_m(X_n)) = \mathbb{E}(f_m(X)),$$

e pelo teorema da convergência monótona,

$$\lim_m \mathbb{E}(f_m(X)) = \mathbb{E}(1_G(X)) = \mathbb{P}(X \in G).$$

5↔ 6: Basta tomar o complementar.

5+6→ 7: Lembre-se que $\delta B = \overline{B} - B^\circ$. Então por (5) e (6), temos que

$$\mathbb{P}(X \in \overline{B}) \geq \limsup \mathbb{P}(X_n \in \overline{B}) \geq \liminf \mathbb{P}(X_n \in B^\circ) \geq \mathbb{P}(X \in B^\circ).$$

Como $\mathbb{P}(\delta B) = 0$, temos que $\mathbb{P}(X \in \overline{B}) = \mathbb{P}(X \in B^\circ)$ e portanto, provamos o que queríamos, já que $\mathbb{P}(X \in B) \leq \mathbb{P}(X \in \overline{B})$ e $\lim_n \mathbb{P}(X_n \in B)$ está entre os $\limsup$ e $\liminf$ acima.

7→1: Se F_x é contínua em x, então

$$\mathbb{P}(X = x) = \mathbb{P}\left(X \in \delta\left(\prod^{d}(-\infty, x]\right)\right) = 0.$$

2→4: Tome f contínua, não negativa e defina

$$f_m = f \cdot 1_{f \leq m}.$$

Note que $0 \leq f_m \uparrow f$ e f_m é contínua e limitada. Logo, por hipótese,

$$\liminf_n \mathbb{E}(f(X_n)) \geq \liminf_n \mathbb{E}(f_m(X_n)) = \mathbb{E}(f_m(X)).$$

Por fim, utilizando o teorema da convergência monótona, temos que

$$\liminf_n \mathbb{E}(f(X_n)) = \lim_m \liminf_n \mathbb{E}(f(X_n)) \geq \lim_m \mathbb{E}(f_m(X)) = \mathbb{E}(f(X)).$$

4→2: Suponha inicialmente que f é continua, não negativa e que $f \leq M$. Por hipótese, temos que

$$\liminf \mathbb{E}(f(X_n)) \geq \mathbb{E}(f(X)).$$

Como $0 \leq M - f$, também por hipótese, temos que

$$\liminf \mathbb{E}(M - f(X_n)) \quad \geq \mathbb{E}(M - f(X)),$$

isto é,

$$\mathbb{E}(f(X)) \geq \limsup \mathbb{E}(f(X_n)),$$

e portanto

$$\liminf \mathbb{E}(f(X_n)) \geq \mathbb{E}(f(X)) \geq \limsup \mathbb{E}(f(X_n)).$$

concluindo o resultado.

No caso geral, basta escrever f como

$$f = f^+ - f^-,$$

com $f^+ \geq 0$, e $f^- \geq 0$ e concluímos o teorema. ■

Observação 2.2.2 (Como lembrar de Portmanteau). Uma forma de lembrar da direção das inclusões nos itens (5) e (6) é considerando o caso em que $\mathbb{P}(X_n = x_n) = 1$. Para isso, suponha que G é aberto, que $(x_n) \subset G$ e que $x_n \xrightarrow{n\to\infty} x \in \delta G = \overline{G} - G^\circ$, então $X_n \to_d X$, onde $\mathbb{P}(X = x) = 1$, e daí

$$1 = \liminf_n \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G) = 0.$$

Da mesma forma, tomando o fechado $F = G^c$, temos que

$$0 = \limsup_n \mathbb{P}(X_n \in G^c) \leq \mathbb{P}(X \in G^c) = 1.$$

Observação 2.2.3. Note que se $B[0; M]$ é a bola fechada de raio M centrada na origem, então

$$\{\|X\| > M\} = \{X \in B[0; M]^c\}.$$

Primeiramente note que, como $B[0; M]$ é fechado, então $B[0; M]^c$ é aberto, então se $X_n \to_d X$, pelo Teorema 2.2.1, temos que

$$\liminf \mathbb{P}(X_n \in B[0; M]^c) \geq \mathbb{P}(X \in B[0; M]^c),$$

ou seja,

$$\liminf \mathbb{P}(\|X_n\| > M) \geq \mathbb{P}(\|X\| > M).$$

Da mesma forma, se considerarmos,

$$\{\|X\| \geq M\} = \{X \in B(0; M)^c\},$$

onde $B(0; M)$ é a bola aberta de raio M centrada na origem, temos que $B(0; M)^c$ é fechado e portanto se $X_n \to_d X$,

$$\limsup \mathbb{P}(\|X_n\| \geq M) \leq \mathbb{P}(\|X\| \geq M).$$

A proposição a seguir nos mostrar o poder do Teorema Portmanteau 2.2.1.

Proposição 2.2.4 (Algumas consequências de Portmanteau)**.** *Sejam (X_n) e (Y_n) dois vetores aleatórios e c uma constante, então*

1. *$X_n \to_p c$ se, e só se, $X_n \to_d c$.*
2. *Se $X_n \to_d X$ e $d(X_n, Y_n) \to_p 0$, então $Y_n \to_d X$.*
3. *Se $X_n \to_d X$ e $Y_n \to_p c$, então $(X_n, Y_n) \to_d (X, c)$.*
4. *Se $X_n \to_p X$ e $Y_n \to_p Y$, então $(X_n, Y_n) \to_p (X, Y)$.*

Demonstração.

1. A ida é óbvia já que (probabilidade)⇒(distribuição). Agora suponha que $X_n \to_d c$. Por 2.2.1, temos que

$$\limsup_n \mathbb{P}(\|X_n - c\| \geq \varepsilon) = \limsup_n \mathbb{P}(X_n \in B(c, \varepsilon)^c) \quad \leq \mathbb{P}(c \in B(c, \varepsilon)^c) = 0$$

2. Considere uma função f Lipschitz e limitada. Podemos supor sem perda de generalidade que a constante de Lipschitz de f é igual a 1 e que $f \in [0, 1]$.

Temos então que

$$\begin{aligned}
|\mathbb{E}(f(X_n)) - \mathbb{E}(f(Y_n))| &\leq \mathbb{E}|f(X_n) - f(Y_n)| \\
&= \mathbb{E}|f(X_n) - f(Y_n)|1_{d(X_n,Y_n)\leq\varepsilon} \\
&+ \mathbb{E}|f(X_n) - f(Y_n)|1_{d(X_n,Y_n)>\varepsilon} \\
&\leq \varepsilon\mathbb{P}(d(X_n,Y_n) \leq \varepsilon) + 2\mathbb{P}(d(X_n,Y_n) > \varepsilon) \\
&\leq \varepsilon + 2\mathbb{P}(d(X_n,Y_n) > \varepsilon).
\end{aligned}$$

Note que, $\mathbb{P}(d(X_n,Y_n) > \varepsilon) \xrightarrow{n\to\infty} 0$ por hipótese, e como $\varepsilon > 0$ é qualquer,

$$|\mathbb{E}(f(X_n)) - \mathbb{E}(f(Y_n))| \xrightarrow{n\to\infty} 0,$$

isto é, as sequências $\mathbb{E}(f(X_n))$ e $\mathbb{E}(f(Y_n))$ tem o mesmo limite, porém, sabemos que

$$\mathbb{E}(f(X_n)) \xrightarrow{n\to\infty} \mathbb{E}(f(X)),$$

pelo item 3 de 2.2.1. Portanto, mostramos que que para qualquer f Lipschitz e limitada,

$$\mathbb{E}(f(Y_n)) \xrightarrow{n\to\infty} \mathbb{E}(f(X)),$$

o que é equivalente por 2.2.1 a $Y_n \to_d X$.

3. Note que

$$d((X_n,Y_n),(X_n,c)) = d(Y_n,c) \xrightarrow{n\to\infty} 0,$$

logo, pelo item anterior, basta mostrarmos que

$$(X_n,c) \to_d (X,c).$$

Note que dada qualquer

$$f : (x,y) \mapsto f((x,y))$$

contínua e limitada, temos que

$$f_c : x \mapsto f(x,c),$$

é contínua e limitada. Como $X_n \to_d X$, temos que

$$\mathbb{E}(f(X_n,c)) = \mathbb{E}(f_c(X_n)) \xrightarrow{n\to\infty} \mathbb{E}(f_c(X)) = \mathbb{E}(f(X,c)).$$

Como f era qualquer função contínua e limitada, temos o resultado por 2.2.1.

4. Basta notar que

$$\begin{aligned}
d((X_n,Y_n),(X,Y)) &= \sqrt{((X_n - X)^2 + (Y_n - Y)^2)} \\
&\leq d(X_n,X) + d(Y_n,Y).
\end{aligned}$$

■

Teorema 2.2.5 (Teorema da Aplicação Contínua). *Seja $f : \mathbb{R}^k \to \mathbb{R}^m$ uma função contínua e mensurável em quase todo ponto. Então*

1. *Se $X_n \to_{q.t.p.} X$ então $f(X_n) \to_{q.t.p.} f(X)$.*

2. *Se $X_n \to_p X$ então $f(X_n) \to_p f(X)$.*

3. *Se $X_n \to_d X$ então $f(X_n) \to_d f(X)$.*

Demonstração.

1. É óbvio da definição de convergência q.t.p. e do fato de g ser contínua.

2. Fixe $\varepsilon > 0$. Dado $\delta > 0$, defina B_δ como o conjunto formado por todos os pontos x tal que **existe** y satisfazendo

$$(\|x - y\| < \delta) \ \wedge \ (\|f(x) - f(y)\| > \varepsilon).$$

Dado $\omega \in \Omega$, suponha que $\|f(X(\omega)) - f(X_n(\omega))\| > \varepsilon$. Então, ou

$$\|X(\omega) - X_n(\omega)\| < \delta$$

e portanto

$$X(\omega) \in B_\delta,$$

ou então

$$\|X(\omega) - X_n(\omega)\| \geq \delta.$$

Temos então que

$$\mathbb{P}(\omega : \|f(X(\omega)) - f(X_n(\omega))\| > \varepsilon) \leq \underbrace{\mathbb{P}(\omega : X(\omega) \in B_\delta)}_{\text{I}} + \underbrace{\mathbb{P}(\omega : \|X_n(\omega) - X(\omega)\| \geq \delta)}_{\text{II}}.$$

Note que o termo II na expressão acima tende a zero, para qualquer $\delta > 0$, quando n cresce, já que por hipótese temos convergência em probabilidade. Além disso, note que os conjuntos B_δ são encaixados quando diminuímos δ, e que, a menos de um conjunto de medida nula,

$$\bigcap_{\delta \to 0} B_\delta = \emptyset.$$

já que g é continua. Portanto,

$$\mathbb{P}(\omega : X(\omega) \in \bigcap_{\delta} B_\delta) = \lim_{\delta \to 0} \mathbb{P}(\omega : X(\omega) \in B_\delta) = 0.$$

Como $\delta > 0$ foi escolhido qualquer, temos o resultado.

3. Vamos usar o item 6 do Teorema 2.2.1, para provar que
$$f(X_n) \to_d f(X).$$
É claro que
$$\{f(X_n) \in F\} = \{X_n \in f^{-1}(F)\}.$$
Além disso, se F é *fechado qualquer*, e C é o conjunto de medida total tal que f é contínua para todo ponto em C, temos que
$$f^{-1}(F) \subset \overline{f^{-1}(F)} \subset f^{-1}(F) \cup C^c.$$
A segunda inclusão é facilmente provada, usando o fato de que f é contínua. Por 2.2.1, usando o fato de que $X_n \to_d X$,
$$\begin{aligned}
\limsup_n \mathbb{P}(f(X_n) \in F) &= \limsup_n \mathbb{P}(X_n \in f^{-1}(F)) \\
&\leq \limsup_n \mathbb{P}(X_n \in \overline{f^{-1}(F)}) \\
&\leq \mathbb{P}(X \in \overline{f^{-1}(F)}) \\
&\leq \mathbb{P}(X \in f^{-1}(F) \cup C^c) \\
&\leq \mathbb{P}(X \in f^{-1}(F)) + 0 \\
&\leq \mathbb{P}(f(X) \in F).
\end{aligned}$$
Como F é qualquer fechado, pelas equivalências do Teorema 2.2.1, temos o resultado.

■

Teorema 2.2.6 (Slutsky)**.** *Sejam X_n, Y_n, X vetores aleatórios (ou v.a.'s possivelmente). Se $X_n \to_d X$ e $Y_n \to_d c$, para uma constante c, então*

1. $X_n + Y_n \to_d X + c$
2. $Y_n X_n \to_d c \cdot X$
3. $Y_n^{-1} X_n \to_d c^{-1} \cdot X$, *se* $c \neq 0$.

Demonstração. Pela Proposição 2.2.4, temos que $Y_n \to_p c$, e com isso, também temos que
$$(X_n, Y_n) \to_d (X, c).$$
Como, $(x, y) \mapsto x + y$ e $(x, y) \mapsto x \cdot y$ são aplicações contínuas, por 2.2.5, temos que o resultado. ■

Definição 2.2.7 (Distribuição Defeituosa)**.** Uma Função de Distribuição Defeituosa é uma função que satisfaz todas as propriedades de funções de distribuição, porém seu limite em $+\infty$ pode ser menor que 1 e seu limite em $-\infty$ pode ser maior que 0.

********melhorar isso no futuro

2.2.2 Teorema de Prohorov

Teorema 2.2.8 (Teorema de Seleção de Helly)**.** *Toda sequência (F_n) de distribuições em $\mathbb{R}^k$ possui uma subsequência F_{n_k} tal que*

$$F_{n_k}(x) \to F(x),\ k \to \infty$$

para cada ponto de continuidade de uma função de distribuição F, possivelmente defeituosa.

Demonstração. Seja considere uma enumeração $\{q_1, q_2, \dots\}$ de $\mathbb{Q}^k$. Para um j fixado, note que

$$(F_n(q_i))_n \subset [0,1],$$

e portanto conseguimos uma subsequência de $(F_n(q_i))_n$ convergindo para um limite que vamos chamar de $G(q_1)$. Por um simples argumento de diagonalização, conseguimos uma subsequência (F_{n_k}), tal que, para todo q_j, temos que

$$F_{n_l}(q_j) \to G(q_j),\ l \to \infty.$$

Defina

$$F(x) = \inf\{G(q) : q \in \mathbb{Q},\ q > x\}.$$

Como para cada n, F_n é não decrescente, temos que se $q_j \leq q_l$, então

$$G(q_j) \leq G(q_l),$$

e daí é fácil ver que F é não decrescente.

Afirmação: F é contínua à direita.
Prova da afirmação: Note que

$$\begin{aligned}\lim_{x_n \downarrow x} F(x_n) &= \inf\{G(q) : q \in \mathbb{Q},\ q > x_n,\ \text{para algum } n\} \\ &= \inf\{G(q) :\ q > x,\ q \in \mathbb{Q}\} = F(x).\end{aligned}$$

□

Afirmação: (F_{n_k}) converge pontualmente em pontos de continuidade de F
Prova da afirmação: Seja x um ponto de continuidade de F. Dado $\varepsilon > 0$, tome $a < b < x < c$ pontos racionais próximos o suficiente de x tal que

$$F(x) - \varepsilon < F(a) \leq F(b) \leq F(x) \leq F(c) < F(x) + \varepsilon.$$

Sabemos que

$$F_{n_k}(b) \leq F_{n_k}(x) \leq F_{n_k}(c),$$

já que cada F_n é uma distribuição. Agora note que

$$F_{n_k}(b) \to G(b) \geq F(a),$$

já que $a < b$. Além disso, para todo $q > c$, sabemos que

$$G(c) \leq G(q),$$

isto é,

$$G(c) \leq \inf_{q>c} G(q) = F(c),$$

e portanto

$$f_{n_k}(c) \to G(c) \leq F(c).$$

Mas então, tomando k suficientemente grande, as observações anteriores nos dizem que

$$F(x) - \varepsilon < F(a) \leq F_{n_k}(b) \leq F_{n_k}(x) \leq F_{n_k}(c) \leq F(c) \leq F(x) + \varepsilon,$$

ou seja,

$$|F(x) - F_{n_k}(x)| < \varepsilon$$

e como $\varepsilon >$ foi qualquer, temos o resultado. □

Note que o que fizemos até então é suficiente para provar o Teorema no caso em que $k = 1$, já que o fato de F ser não decrescente é suficiente para garantir que

$$\mu((a, b]) = F(b) - F(a) \geq 0,$$

e então conseguimos induzir uma medida finita no nosso espaço.

Porém, no caso em que $k > 1$ temos um trabalho extra de provar que da fato F atribui valores não negativos nos respectivos cubos k-dimensionais (equivalentes aos intervalos em $\mathbb{R}$). Faremos o resto da prova em uma versão futura destas notas, apesar do argumento ser bem simples.

■

Definição 2.2.9 (Família Tight)**.** Uma família $\{X_\alpha : \alpha \in A\}$ é **tight** se para todo $\varepsilon > 0$, existe $M > 0$ tal que

$$\sup_\alpha \mathbb{P}(\|X_\alpha\| > M) < \varepsilon.$$

Isto é, uma família de vetores aleatórios é tight, se existe um compacto grande o suficiente de tal forma que esta família está suportada neste compacto com probabilidade alta.

Teorema 2.2.10 (Teorema de Prohorov)**.** *Seja (X_n) uma sequência de vetores aleatórios em $\mathbb{R}^k$. Então*

1. *Se $X_n \to_d X$ para algum X, então (X_n) é tight.*
2. *Se (X_n) é tight, então existe uma subsequência (X_{n_k}) tal que*

$$X_{n_k} \to_d X$$

para algum X.

Demonstração.

1. Tome M grande o suficiente para que
$$\mathbb{P}(\|X\| \geq M) < \varepsilon.$$
Por Portanteau 2.2.1 (ou Observação 2.2.3), temos que
$$\limsup_n \mathbb{P}(\|X_n\| \geq M) \leq \mathbb{P}(\|X\| \geq M) < \varepsilon.$$
Portanto, para N grande, se $n \geq N$,
$$\mathbb{P}(\|X_n\| \geq M) \leq \varepsilon.$$
Para lidar com os índices menores que N, basta considerar M um pouco maior, se necessário.

2. Seja F_n a função de distribuição de X_n. Pelo Teorema de Seleção de Helly 2.2.8, conseguimos uma subsequência (F_{n_k}) que converge para uma função de distribuição F, possivelmente defeituosa.

   Como (X_n) é tight, dado $\varepsilon_i = 1/i$ com $i = 1, 2 \ldots$, existe M_i tal que, para todo $n \in \mathbb{N}$
$$\mathbb{P}(\|X_n\| \leq M_i) > 1 - \varepsilon_i.$$
É claro que isso implica que, para todo $n \in \mathbb{N}$
$$F_n(M) = \mathbb{P}(X_n \leq M_i) > 1 - \varepsilon_i,$$
já que
$$\{\|X_n\| \leq M_i\} \subset \{X_n \leq M_i\}.$$
Fazendo M_i um pouco maior, se necessário, podemos supor que M_i é um ponto de continuidade de F, e portanto
$$F(M_i) = \lim_k F_{n_k}(M_i) \geq 1 - 1/i.$$
Logo, se $x \to +\infty$, basta tomar uma sequência (M_{i_j}) com $x \geq M_{i_j}$ e $M_{i_j} \to +\infty$ e então temos que $F(x) \to 1$.

   O argumento para quando $x \to -\infty$ é semelhante. (faça um desenho)

$\blacksquare$

2.3 Exercícios Resolvidos

Capítulo 3

Teorema Central do Limite

As Principais referências usadas foram:

1. [Dur10].
2. [Str10].

3.1 Motivação: Teorema de De Moivre-Laplace

Seja $X_1, \ldots, X_n$ uma sequência de v.a.'s i.i.d. com $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ e considere $S_n = X_1 + \cdots + X_n$. Note que

$$\mathbb{P}(S_{2n} = 2k) = \binom{2n}{n+k} 2^{-2n},$$

já que para $S_{2n} = 2k$, precisamos que $1 \cdot (2n + A) - 1 \cdot (2n + B) = 2k$ e que $(2n + A) + (2n + B) = 2n$. Por Stirling A.1, não é difícil mostrar que

$$\binom{2n}{n+k} 2^{-2n} \sim \left(1 + \frac{k}{n}\right)^{-n-k} \left(1 - \frac{k}{n}\right)^{-n+k} \left(1 + \frac{k}{n}\right)^{-1/2} \left(1 - \frac{k}{n}\right)^{-1/2} (\pi n)^{-1/2}.$$

Além disso,

$$\left(1 + \frac{k}{n}\right)^{-n-k} \left(1 - \frac{k}{n}\right)^{-n+k} = \left(1 - \frac{k^2}{n^2}\right)^{-n} \left(1 + \frac{k}{n}\right)^{-k} \left(1 - \frac{k}{n}\right)^{k}.$$

Lema 3.1.1. *Suponha que $a_i \to +\infty$, que $c_i \to 0$ e $a_i c_i \to \lambda$, então*

$$(1 + c_i)^{a_i} \to e^{\lambda}.$$

Demonstração. Vamos usar a identidade

$$\frac{\log(1 + x)}{x} \to 1, \;\; x \to 0.$$

Temos então que

$$\frac{\log(1+c_j)}{c_j} = \frac{a_j}{a_j}\frac{\log(1+c_j)}{c_j} \to 1,$$

ou seja,

$$a_j \log(1+c_j) = \lim_j a_j c_j = \lambda.$$

Tomando exp dos dois lados, temos o resultado. ■

Lema 3.1.2. *Suponha que*

1. $\max_{1\le j\le n} |c_{j,n}| \xrightarrow{n\to\infty} 0,$
2. $\sum_{j=1}^{n} c_{j,n} \xrightarrow{n\to\infty} \lambda,$
3. $\sup_n \sum_{j=1}^{n} c_{j,n} < \infty.$

Então,

$$\prod_{j=1}^{n}(1+c_{j,n}) \to e^{\lambda}.$$

Demonstração. Por 1,

$$\log(1+c_{j,n}) \sim c_{j,n}.$$

Por 3,

$$\frac{\sum_{j=1}^{n}\log(1+c_{j,n})}{\sum_{j=1}^{n} c_{j,n}} \sim 1.$$

E então, por 2

$$\sum_{j=1}^{n} c_{j,n}\frac{\sum_{j=1}^{n}\log(1+c_{j,n})}{\sum_{j=1}^{n} c_{j,n}} \sim \sum_{j=1}^{n} c_{j,n} \sim \lambda.$$

E daí basta aplicar a exponencial dos dois lados.

■

Utilizando o lema 3.1.1, se considerarmos $x = k/\sqrt{n/2}$, é fácil ver que

$$\begin{aligned}
\left(1-\frac{k^2}{n^2}\right)^{-n} &\longrightarrow \exp(x^2/2)\\
\left(1+\frac{k}{n}\right)^{-k} &\longrightarrow \exp(-x^2/2)\\
\left(1-\frac{k}{n}\right)^{k} &\longrightarrow \exp(-x^2/2)\\
\left(1+\frac{k}{n}\right)^{-1/2}\left(1-\frac{k}{n}\right)^{-1/2} &\longrightarrow 1.
\end{aligned}$$

E portanto, se $\dfrac{k}{\sqrt{n/2}} \to x$, então

$$\mathbb{P}(S_{2n} = 2k) \sim \frac{1}{\sqrt{n\pi}} e^{-x^2/2}. \tag{3.1.1}$$

Note que então

$$\mathbb{P}(a\sqrt{2n} \leq S_{2n} \leq b\sqrt{2n}) = \sum_{m \in [a\sqrt{2n}, b\sqrt{2n}] \cap 2\mathbb{Z}} \mathbb{P}(S_{2s} = m),$$

isto é,

$$\mathbb{P}(a \leq S_{2n}/\sqrt{2n} \leq b) = \sum_{x \in [a,b] \cap 2\mathbb{Z}/\sqrt{2n}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \sqrt{2/n},$$

e portanto, se n é arbitrariamente grande, temos que $\sqrt{2/n} \approx dx$ e ficamos com

$$\mathbb{P}(a \leq S_n/\sqrt{n} \leq b) \xrightarrow{n \to \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Ou seja, temos que $S_n/\sqrt{n}$ converge em distribuição para uma normal com mesma média e variância que X_1.

3.2 Funções Características

Definição 3.2.1. Seja X uma v.a., então definimos sua função característica $\varphi_X(t)$ como

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \mathbb{E}(\cos tX) + i\mathbb{E}(\sin tx).$$

Note que se X admite uma densidade f, então φ_X nada mais é que a transformada de Fourier de f.

Proposição 3.2.2. *Seja φ_X a func. carct. de X, então*

1. $\varphi_X(0) = 1$.
2. $\varphi_X(-t) = \overline{\varphi_X(t)}$.
3. $|\varphi_X(t)| = |\mathbb{E}e^{itX}| \leq \mathbb{E}|e^{itX}| = 1$.
4. $|\varphi_X(t+h) - \varphi_X(t)| \leq \mathbb{E}|e^{ihX} - 1|$.
5. $\varphi_X(t)$ *é uniformemente contínua.*
6. $\mathbb{E}e^{it(aX+b)} = e^{itb}\varphi_X(at)$.

> *Observação* 3.2.3. Note que o item 3 da proposição anterior nos garante que a função característica de X *sempre existe*, ao contrário, por exemplo, da função geradora de momento de X.

Demonstração.

1. Óbvio.
2. Óbvio.
3. Por Jensen
$$|\mathbb{E}e^{itX}|^2 = (\mathbb{E}\cos tX)^2 + (\mathbb{E}\sin tX)^2 \leq \mathbb{E}(\cos^2 tX) + \mathbb{E}(\sin^2 tX) = 1.$$
4. Trivial.
5. Precisamos mostrar que para todo $\varepsilon > 0$, existe $\delta > 0$ tal que
$$|t - s| < \delta \Rightarrow |\varphi_X(t) - \varphi_X(s)| < \varepsilon.$$
Tome $\varepsilon > 0$ qualquer e defina $h = t - s$. Pelo item anterior, temos que
$$|\varphi_X(t) - \varphi_X(s)| \leq \mathbb{E}|e^{ihX} - 1|.$$
Como $e^{ihX} - 1 \to 0$, se $h \to 0$, então $e^{ihX} - 1$ é limitado e portanto, pelo teorema da convergência dominada, temos que $\mathbb{E}|e^{ihX} - 1| \to 0$, se $h \to 0$.
Portanto, para todo s, t tal que $s - t = h$ satisfaz $\mathbb{E}|e^{ihX} - 1| < \varepsilon$, temos o resultado.

6. Trivial.

■

Teorema 3.2.4. *Sejam X, Y são independentes com funções características.* φ_X *e* φ_Y*, respectivamente. Então,*

$$\varphi_{X+Y} = \varphi_X \varphi_Y.$$

Demonstração. Basta lembrar que se X, Y são independentes e f é mensurável, então

$$\mathbb{E}(f(X)f(Y)) = \mathbb{E}(f(X))\mathbb{E}(f(Y)),$$

já que $f(X), f(Y)$ são independentes. ■

Exemplo 3.2.5. Seja $X \sim N(\mu, \sigma)$, então

$$\varphi_X(t) = e^{+it\mu - \sigma^2 t^2/2}.$$

Teorema 3.2.6 (Fórmula de Inversão)**.** *Seja* $\varphi(t) = \int e^{itx} d\mu(x)$*, onde* μ *é uma probabilidade. Então, se* $a < b$*,*

$$\lim_{T \to +\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = \mu((a,b)) + \frac{\mu(\{a,b\})}{2}.$$

Observação 3.2.7. Note que a fórmula de inversão nos diz que a função característica de uma v.a. determina unicamente sua distribuição.

Demonstração. Temos que

$$\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \int e^{itx} d\mu(x) dt.$$

Note que,

$$\left| \frac{e^{-ita} - e^{-itb}}{it} \right| = \left| \int_a^b e^{-ity} dy \right| \leq |b - a|,$$

e então,

$$\int \left| e^{itx} \frac{e^{-ita} - e^{-itb}}{it} \right| d\mu \leq |b - a| < +\infty.$$

Portanto, por Fubini, temos que

$$\begin{aligned}
\int_{-T}^{T} \int \frac{e^{-ita} - e^{-itb}}{it} e^{itx} d\mu dt &= \int \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} e^{itx} dt d\mu \\
&= \int \int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt d\mu \\
&= \int \int_{-T}^{T} \left(\frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t} \right) dt d\mu.
\end{aligned}$$

Do cálculo complexo, sabemos que

$$\int_0^\infty \frac{\sin y}{y} dy = \frac{\pi}{2}.$$

Fazendo uma mudança de coordenadas,

$$\lim_{T\to+\infty} \int_{-T}^{T} \frac{\sin(t(x-a))}{t} dt = 2\cdot \operatorname{sgn}(x-a) \int_0^\infty \frac{\sin y}{y} dy = \operatorname{sgn}(x-a)\pi.$$

Temos então que

$$\lim_{T\to+\infty} \frac{1}{2\pi} \int_{-T}^{T} \left(\frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t} \right) dt = \begin{cases} 1 & \text{se } a<x<b \\ 0 & \text{se } x<a \text{ ou } x>b \\ 1/2 & \text{se } x=a \text{ ou } x=b. \end{cases}$$

Além disso, como $(\sin y)/y$ é contínua em $[0,\infty)$, e $\int_0^\infty (\sin y)/y dy < \infty$, a função

$$F(T) = \int_0^T \frac{\sin y}{y} dy$$

é contínua para $T \geq 0$ e limitada. Portanto, pelo teorema da convergência dominada,

$$\lim_{T\to+\infty} \int \int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt d\mu = \int \int_{-\infty}^{+\infty} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt d\mu.$$

Com isso temos o resultado. ■

Proposição 3.2.8. *Seja* $\varphi(t) = \int e^{itx} d\mu(x)$. *Se* $\int |\varphi(t)| dt < \infty$, *então* μ *tem como função de densidade,*

$$f(x) = \frac{1}{2\pi} \int e^{-itx} \varphi(t) dt,$$

com f *contínua e limitada.*

Demonstração. Pela demonstração do teorema anterior,

$$\left| \frac{e^{-ita} - e^{-itb}}{it} \right| = \left| \int_a^b e^{-ity} dy \right| \leq |b-a|,$$

logo, pela conclusão do teorema anterior

$$\mu((a,b)) + \frac{\mu(\{a,b\})}{2} \leq \frac{b-a}{2\pi} \int |\varphi(t)| dt,$$

portanto, $\mu(\{a\}) = \mu(\{b\}) = 0$.

Além disso, novamente pelo teorema anterior,

$$\begin{aligned} \mu((x,x+h)) &= \frac{1}{2\pi} \int \frac{e^{-itx} - e^{-it(x+h)}}{it} \varphi(t) dt \\ &= \frac{1}{2\pi} \int \left(\int_x^{x+h} e^{-ity} dy \right) \varphi(t) dt, \end{aligned}$$

como a a expressão dentro dos parênteses tem módulo integrável, podemos usar Fubini e portanto

$$\begin{aligned}\mu((x,x+h)) &= \frac{1}{2\pi}\int\left(\int_x^{x+h} e^{-ity}dy\right)\varphi(t)dt \\ &= \frac{1}{2\pi}\int_x^{x+h}\int e^{-ity}\varphi(t)dtdy \\ &= \int_x^{x+h} f(y)dy.\end{aligned}$$

■

3.2.1 Momentos e Derivadas

Lema 3.2.9. *Temos que*

$$\left|e^{ix}-\sum_{m=0}^{n}\frac{(ix)^m}{m!}\right| \leq \min\left(\frac{|x|^{n+1}}{(n+1)!},\frac{2|x|^n}{n!}\right).$$

Demonstração. É fácil ver que

$$e^{ix}-\sum_{m=0}^{n}\frac{(ix)^m}{m!} = \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is}ds,$$

por exemplo, basta notar que o termo à direita nada mais é que o resto da expansão de Taylor de e^{ix}, e a ideia para chegar nessa expressão é a mesma que se usa para encontrar o resto em Taylor.

Portanto, reduzimos nosso trabalhar a entender

$$\frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is}ds.$$

Uma primeira estimativa surge ao notar que $|e^{is}| \leq 1$, já que $s \in \mathbb{R}$, logo

$$\begin{aligned}\left|\frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is}ds\right| &\leq \frac{1}{n!}\int_0^x |(x-s)^n|ds \\ &\leq \frac{1}{n!}\frac{1}{n+1}|x|^{n+1} \\ &= \frac{1}{(n+1)!}|x|^{n+1}.\end{aligned}$$

Note que essa estimativa é boa quando $|x|$ é pequeno.

Vamos tentar encontrar uma outra estimativa que funcione melhor quando $|x|$ é grande. A ideia é reduzir a ordem da expressão acima para $|x|^n$ ao invés de $|x|^{n+1}$.

Para isso, note que

$$\frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}e^{is}ds = \frac{(ix)^n}{n!}+\frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is}ds,$$

isto é,

$$\int_0^x (x-s)^{n-1} e^{is} ds = \frac{x^n}{n} + \frac{i}{n} \int_0^x (x-s)^n e^{is} ds,$$

onde essa igualdade sai por integração por partes, ou apenas observando a diferença entre a expansão de Taylor de e^{ix} de ordem $n-1$ e n.

Então, como

$$\frac{x^n}{n} = \int_0^x (x-s)^{n-1} ds,$$

temos que

$$\frac{i}{n} \int_0^x (x-s)^n e^{is} ds = \int_0^x (x-s)^{n-1} (e^{is} - 1) ds.$$

E por fim, como $|e^{is} - 1| \leq 2$, temos que

$$\left| \frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is} ds \right| \leq \frac{2}{n!} |x|^n.$$

■

Observação 3.2.10. A prova desse lema pode parecer abstrata a primeira vista, porém ela é super intuitiva pensando na expansão de Taylor de e^{ix}. Todos os passos que fazemos na demonstração são réplicas de resultados envolvendo estimativa de erros em Taylor.

Corolário 3.2.11. *Temos que*

$$\left| \mathbb{E} e^{itX} - \sum_{m=0}^{n} \mathbb{E} \frac{(itX)^m}{m!} \right| \leq \mathbb{E} \min \left(\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right).$$

Demonstração. Basta aplicar a esperança dos dois lados, isto é

$$\begin{aligned} \left| \mathbb{E} e^{itX} - \sum_{m=0}^{n} \mathbb{E} \frac{(itX)^m}{m!} \right| &\leq \mathbb{E} \left| e^{itX} - \sum_{m=0}^{n} \frac{(itX)^m}{m!} \right| \\ &\leq \mathbb{E} \min \left(\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right). \end{aligned}$$

Em particular, podemos encontrar um limitante mais simples apenas jogando o denominador fora, isto é,

$$\left| \mathbb{E} e^{itX} - \sum_{m=0}^{n} \mathbb{E} \frac{(itX)^m}{m!} \right| \leq \mathbb{E} \min \left(|tX|^{n+1}, 2|tX|^n \right).$$

■

Corolário 3.2.12. *Se* $\mathbb{E}(X^2) < \infty$*, então*

$$\varphi_X(t) = 1 + it\mathbb{E}X - t^2 \mathbb{E}X^2/2 + o(t^2).$$

Demonstração. Pelo Corolário anterior,

$$\left|\mathbb{E}e^{itX} - 1 - itX + t^2\mathbb{E}(X^2)/2\right| \leq t^2\mathbb{E}\min\left(|t|\cdot|X|^3, 2|X|^2\right).$$

Se

$$R(t) = t^2\min\left(|t|\cdot|X|^3, 2|X|^2\right),$$

é claro que $0 \leq R(t) \leq 2t^2X^2$ e está última é integrável.

Como

$$\lim_{t\to 0} R(t)/t^2 = 0,$$

pelo Teorema da Convergência Dominada, temos que

$$\lim_{t\to 0}\mathbb{E}(R(t))/t^2 = 0,$$

e assim concluímos o corolário. ■

Observação 3.2.13. Mais uma vez, todos os resultados anteriores servem para formalizar o fato intuitivo (via Taylor) de que

$$\mathbb{E}(e^{itX}) \approx 1 + itX - t^2\mathbb{E}X^2/2.$$

Note que apenas aplicar a expansão de Taylor sem nenhuma justificativa prévia **não** é suficiente, pois estamos pedindo apenas que $\mathbb{E}X^2 < \infty$ e não colocamos restrição nenhuma sobre o terceiro momento, que apareceria no termo de erro caso utilizássemos expansão de Taylor!

Proposição 3.2.14. *Se* $\mathbb{E}|X|^n < \infty$, *então*

$$\varphi^{(n)}(t) = \int (ix)^n e^{itx} d\mu(x).$$

Demonstração. Suponha $n = 1$. Por 3.2.9,

$$\frac{|e^{itx}(e^{ihx}-1)|}{|h|} = \frac{|e^{ihx}-1|}{|h|} \leq |x|,$$

e este último termo é integrável por hipótese. Mas então, pelo teorema da convergência dominada

$$\begin{aligned}\varphi'(t) = \lim_{h\to 0}\frac{\varphi(t+h)-\varphi(t)}{h} &= \lim_{h\to 0}\int\frac{e^{ihx}-1}{h}e^{itx}d\mu(x)\\ &= \int\lim_{h\to 0}\frac{e^{ihx}-1}{h}e^{itx}d\mu(x)\\ &= \int ixe^{itx}d\mu(x).\end{aligned}$$

O caso geral sai facilmente pode indução. ■

Corolário 3.2.15. *Se* $\mathbb{E}(|X|^n) < \infty$, *então*

$$\mathbb{E}(X^n) = \frac{\varphi^{(n)}(0)}{i^n}.$$

3.2.2 Teorema de Continuidade de Lévy

Teorema 3.2.16 (Teorema de Continuidade de Lévy)**.** *Seja (μ_n) uma sequência de medidas de probabilidade com respectivas funções características φ_n. Então*

1. *Se $\mu_n \to_d \mu$, então para todo $t \in \mathbb{R}$, temos que $\varphi_n(t) \to \varphi(t)$, onde φ é a f.c. de μ.*
2. *Se para todo $t \in \mathbb{R}$, vale $\varphi_n(t) \to \varphi(t)$, tal que φ é contínua em 0 então a sequência μ_n é tight e converge em distribuição para uma medida μ com f.c. φ.*

Demonstração.

1. Sabemos que e^{ix} é contínua e limitada, então o item 1 é verdadeiro pela equivalência de convergência em distribuição via funções continuas e limitadas.
2. Suponha que (μ_n) é tight. Então, pelo Teorema de Prohorov 2.2.10, existe uma subsequência (μ_{n_k}) tal que

   $$\mu_{n_k} \to_d \mu,$$

   para alguma medida μ com f.c. $\hat{\varphi}$. Pelo item anterior, temos que

   $$\varphi_{n_k}(t) \to \hat{\varphi}(t),$$

   mas como por hipótese, $\varphi_n(t) \to \varphi$, então $\hat{\varphi} = \varphi$.

   Tomando qualquer subsequência de (μ_k), conseguimos repetir esse argumento e mostrar que existe uma subsubsequência (μ_{k_l}) convergindo em distribuição para μ. Isso quer dizer que dada uma função f contínua e limitada qualquer, e a uma subsequência da sequência números reais $(\int f d\mu_n)$, conseguimos uma nova subsequência satisfazendo

   $$\int f d\mu_{n_{k_l}} \to \int f d\mu,$$

   mas daí basta usar o Lema 4.2.7 e a equivalência 2 de Portemanteau 2.2.1.

   Portanto, só nos falta mostrar que de fato (μ_n) é tight.

   Afirmação: (μ_n) é tight.

   Prova da afirmação: Fixe $\varepsilon > 0$ qualquer. Note que,

$$
\begin{aligned}
\frac{1}{2T}\int_{-T}^{T}\varphi_n(t)dt &= \frac{1}{2T}\int_{-T}^{T}\int e^{itx}d\mu(x)dt \\
&\underset{\text{Fubini}}{=} \frac{1}{2T}\int\int_{-T}^{T} e^{itx}dtd\mu_n(x) \\
&= \frac{1}{2T}\int\int_{-T}^{T}(\cos(tx)+i\sin(tx))dtd\mu_n(x) \\
&= \frac{1}{2T}\int \frac{2\sin(Tx)}{x}d\mu_n(x) \\
&= \int \frac{\sin(Tx)}{Tx}d\mu_n(x) \\
&= \int_{|x|\leq M}\frac{\sin(Tx)}{Tx}d\mu_n(x) + \int_{|x|>m}\frac{\sin(Tx)}{Tx}d\mu_n(x) \\
&= \int_{|x|\leq M}\left|\frac{\sin(Tx)}{Tx}\right|d\mu_n(x) + \int_{|x|>M}\left|\frac{\sin(Tx)}{Tx}\right|d\mu_n(x) \\
&\leq \mu_n(x:|x|\leq M) + \frac{\mu_n(x:|x|>M)}{TM} \\
&\leq (1-\mu_n(x:|x|>M)) + \frac{\mu_n(x:|x|>M)}{TM} \\
&= 1-\left(1-\frac{1}{TM}\right)\mu_n(x:|x|>M).
\end{aligned}
$$

Portanto, tomando $M = 2/T$,

$$\mu_n(x:|x|>2/T) \leq \frac{1}{T}\int_{-T}^{T}(1-\varphi_n(t))dt.$$

Como $\varphi(t)$ é contínua em $t=0$, escolhendo T suficientemente pequeno, temos que, para $|t|<T$,

$$|1-\varphi(t))| < \varepsilon/2.$$

Além disso, para todo t temos que

$$\varphi(t)-\varphi_n(t) \xrightarrow{n\to\infty} 0,$$

e como $\varphi(t)-\varphi_n(t)$ é limitada, e estamos interessados em integrar tal função num intervalo limitado, podemos utilizar o teorema da convergência dominada. Isto é, para algum N suficientemente grande, se $n>N$ então

$$\int_{-T}^{T}(\varphi(t)-\varphi_n(t))dt < \varepsilon/2$$

Temos então que, para $n > N$,

$$\begin{aligned}
\mu_n(x : |x| > 2/T) &\leq \frac{1}{T}\int_{-T}^{T}(1-\varphi_n(t))dt \\
&= \frac{1}{T}\int_{-T}^{T}(1-\varphi(t))dt + \frac{1}{T}\int_{-T}^{T}(\varphi(t)-\varphi_n(t))dt \\
&\leq \varepsilon/2 + \varepsilon/2 = \varepsilon.
\end{aligned}$$

Como existe apenas uma quantidade finite de elementos com $n \leq N$, conseguimos lidar com eles individualmente escolhendo um T possivelmente menor (ou seja, $2/T$ possivelmente maior) e finalmente conseguimos provar a afirmação.

$\square$

$\blacksquare$

3.3 Teorema(s) Central(is) do Limite

3.3.1 TCL - Caso i.i.d.

Teorema 3.3.1 (Caso i.i.d.). *Seja (X_n) uma sequência i.i.d. com média μ e variância $\sigma^2 \in (0,\infty)$. Se*

$$S_n = X_1 + \cdots + X_n,$$

então

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \longrightarrow_d Z,$$

onde $Z \sim N(0,1)$.

Demonstração. Sem perda de generalidade, vamos supor que $\mu = 0$. Definindo

$$Y_n := \frac{S_n}{\sqrt{n\sigma^2}},$$

a ideia da prova é mostrar que para todo $t \in \mathbb{R}$,

$$\varphi_{Y_n}(t) \xrightarrow{n\to\infty} \varphi_Z(t),$$

e então concluir utilizando o Teorema 3.2.16.

Pelo Lema 3.2.12 e o fato de que as v.a's são i.i.d.,

$$\varphi_{Y_n}(t) = \left(1 - \frac{t^2}{2n} + o(n^{-1})\right)^n.$$

É tentador pensar em usar 3.1.1, porém, como estamos trabalhando com números complexos, precisamos expandir aquele resultado de forma mais geral.[1]

Lema 3.3.2. *Sejam $z_1,\ldots,z_n$ e $w_1,\ldots,w_n$ números complexos com módulo menor que θ. Então,*

$$\left|\prod_{m=1}^n z_m - \prod_{m=1}^n w_m\right| \le \theta^{n-1}\sum_{m=1}^n |z_m - w_m|.$$

Demonstração. Para $n = 1$ isso é verdade. Suponha que vale para $n-1$ e note que

$$\begin{aligned}
\left|\prod_{m=1}^n z_m - \prod_{m=1}^n w_m\right| &= \left|\prod_{m=1}^n z_m - \prod_{m=1}^n w_m + z_1\prod_{m=2}^n w_m - z_1\prod_{m=2}^n w_m\right| \\
&\le \left|\prod_{m=1}^n z_m - z_1\prod_{m=2}^n w_m\right| + \left|-\prod_{m=1}^n w_m + z_1\prod_{m=2}^n w_m\right| \\
&\le |z_1|\left|\prod_{m=2}^n z_m - \prod_{m=2}^n w_m\right| + \left|\prod_{m=2}^n w_m\right||z_1 - w_1| \\
&\le \theta\theta^{n-2}\sum_{m=2}^n |z_m - w_m| + \theta^{n-1}|z_1 - w_1|.
\end{aligned}$$

[1]Não estou completamente convencido de que esta etapa é realmente necessária.

■

Lema 3.3.3. *Seja b um número complexo com $|b| \leq 1$. Então $|e^b - (1+b)| \leq |b^2|$.*

Demonstração. Basta notar que

$$\begin{aligned} |e^b - (1-b)| &= \left| \sum_{k=2}^{\infty} \frac{(ib)^k}{k!} \right| \\ &\leq \frac{|b|^2}{2} \left(\frac{1}{1} + \frac{2}{3!} + \frac{2}{4!} + \frac{2}{5!} + \ldots \right) \end{aligned}$$

Como para $n \geq 4$, vale que $2^n \leq n!$, então

$$\begin{aligned} |e^b - (1-b)| &\leq \frac{|b|^2}{2} \left(1 + \frac{1}{3} + \frac{1}{2^3} + \frac{1}{2^4} \ldots \right) \\ &\leq \frac{|b|^2}{2} \left(1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} \ldots \right) \\ &= |b|^2. \end{aligned}$$

■

Proposição 3.3.4. *Seja $c_n \xrightarrow{n\to\infty} c \in \mathbb{C}$, então*

$$\left(1 + \frac{c_n}{n}\right)^n \xrightarrow{n\to\infty} e^c.$$

Demonstração. Como $c_n \xrightarrow{n\to\infty} c$, existe $\lambda > 0$ tal que $|c_n| \leq \lambda$ para n grande. Utilizando o fato de que $1 + x \leq e^x$, temos que $1 + \lambda/n \leq e^{\lambda/n}$. Pelos lemas anteriores,

$$\begin{aligned} |(1 + c_n/n)^n - (e^{c_n/n})^n| &\leq (e^{\lambda/n})^{n-1} \sum_{m=1}^{n} |1 + c_n/n - e^{c_n/n}| \\ &\leq (e^{\lambda/n})^n \frac{1}{e^{\lambda/n}} \sum_{m=1}^{n} |1 + c_n/n - e^{c_n/n}| \\ &\leq e^{\lambda} n \left| \frac{c_n}{n} \right|^2 \xrightarrow{n\to\infty} 0. \end{aligned}$$

Portanto,

$$\lim_{n\to\infty} (1 + c_n/n)^n = \lim_{n\to\infty} e^{c_n} = e^c.$$

■

Fim da Demonstração de 3.3.1: Tínhamos que

$$\varphi_{Y_n}(t) = \left(1 - \frac{t^2}{2n} + o(n^{-1})\right)^n.$$

Tomando

$$c_n = -t^2/2 + n \cdot o(1/n) = c_n = -t^2/2 + \frac{o(n)}{1/n} \xrightarrow{n\to\infty} -t^2/2,$$

temos o resultado pela proposição anterior e o Teorema 3.2.16.

■

3.3.2 TCL - Lindeberg-Feller

Teorema 3.3.5 (Lindeberg-Feller)**.** *Para cada n, sejam $X_{n,m}$, $0 \leq m \leq n$ v.a.'s* ***independentes*** *com $\mathbb{E}_{n,m} = 0$. Suponha que*

1. $\sum_{m=1}^{n} \mathbb{E}X_{n,m}^2 \to \sigma^2 > 0$.

2. *Para todo $\varepsilon > 0$*

$$\sum_{m=1}^{n} \mathbb{E}(|X_{n,m}^2| 1_{|X_{n,m}|>\varepsilon}) \xrightarrow{n\to\infty} 0.$$

Então,

$$S_n = X_{n,1} + \cdots + X_{n,n} \longrightarrow_d \sigma Z, \; n \to \infty.$$

Demonstração. A ideia é a mesma que utilizamos na versão i.i.d.3.3.1, isto é, vamos mostrar que

$$\varphi_{S_n}(t) = \prod_{m=1}^{n} \varphi_{n,m}(t) \xrightarrow{n\to\infty} \exp\left(\frac{-t^2\sigma^2}{2}\right)$$

e concluir utilizando o Teorema de Continuidade de Levy 3.2.16.

Para isso, vamos estudar como se comporta o erro de segunda ordem para $\varphi_{n,m}$. Por 3.2.12, temos que

$$|\varphi_{n,m}(t) - 1 + \sigma_{n,m}^2 t^2/2| \leq \mathbb{E}\min(|tX_{n,m}|^3, 2|tX_{n,m}|^2),$$

onde

$$\sigma_{n,m}^2 = \mathbb{E}(X_{n,m}^2).$$

E portanto

$$\begin{aligned}
|\varphi_{n,m}(t) - 1 + \sigma_{n,m}^2 t^2/2| &\leq \mathbb{E}\min(|tX_{n,m}|^3, 2|tX_{n,m}|^2) \\
&= \mathbb{E}1_{|X_{n,m}|>\varepsilon}\min(|tX_{n,m}|^3, 2|tX_{n,m}|^2) \\
&+ \mathbb{E}1_{|X_{n,m}|\leq\varepsilon}\min(|tX_{n,m}|^3, 2|tX_{n,m}|^2) \\
&\leq \mathbb{E}(2|tX_{n,m}|^2, |X_{n,m}| > \varepsilon) + \mathbb{E}(|tX_{n,m}|^3, |X_{n,m}| \leq \varepsilon) \\
&= 2t^2\mathbb{E}(|X_{n,m}|^2, |X_{n,m}| > \varepsilon) + \varepsilon t^3\mathbb{E}(|X_{n,m}|^2, |X_{n,m}| \leq \varepsilon) \\
&= 2t^2\mathbb{E}(|X_{n,m}|^2, |X_{n,m}| > \varepsilon) + \varepsilon t^3\mathbb{E}(X_{n,m}^2).
\end{aligned}$$

Somando em m, temos que

$$\sum_{m=1}^{n} |\varphi_{n,m}(t) - 1 + \sigma_{n,m}^2 t^2/2| \leq \varepsilon t^3 \sum_{m=1}^{n} \mathbb{E}(X_{n,m}^2) + 2t^2 \sum_{m=1}^{n} \mathbb{E}(|X_{n,m}|^2, |X_{n,m}| > \varepsilon).$$

Logo, tomando o limite $\xrightarrow{n\to\infty}$,

$$\sum_{m=1}^{\infty} |\varphi_{n,m}(t) - 1 + \sigma_{n,m}^2 t^2/2| \leq \varepsilon t^3 \sigma^2$$

e como $\varepsilon > 0$ é qualquer, temos que

$$\sum_{m=1}^{\infty} |\varphi_{n,m}(t) - 1 + \sigma_{n,m}^2 t^2/2| = 0.$$

Lembrando agora de 3.3.2, se conseguirmos limitar o módulo de $\varphi_{n,m}(t)$ e $(1 - t^2\sigma_{n,m}^2/2)$ por 1, pelo resultado anterior teremos que

$$\left| \prod_{m=1}^{n} \varphi_{n,m}(t) - \prod_{m=2}^{n} (1 - t^2\sigma_{n,m}^2/2) \right| \xrightarrow{n\to\infty} 0$$

e como por hipótese $\sum \sigma_{n,m}^2 = \sigma^2 > 0$,

1. $\max\limits_{1\leq j\leq n} |t^2\sigma_{n,m}^2| \xrightarrow{n\to\infty} 0$,

2. $\displaystyle\sum_{j=1}^{n} t^2\sigma_{n,m}^2 \xrightarrow{n\to\infty} \lambda$,

3. $\displaystyle\sup_n \sum_{j=1}^{n} t^2\sigma_{n,m}^2 < \infty$,

por 3.1.2, teremos o resultado.

Afirmação: Para cara para n suficientemente grande e $1 \leq m \leq n$, temos que $|\varphi_{n,m}(t)| \leq 1$ e $|1 - t^2\sigma_{n,m}^2/2| \leq 1$.
Prova da afirmação: É claro que $|\varphi_{n,m(t)}| \leq 1$. Além disso, note que

$$\sigma_{n,m}^2 \leq \varepsilon/2 + \mathbb{E}(|X_{n,m}|^2, |X_{n,m}| > \varepsilon/2),$$

e portanto, pela hipótese 2, temos que para n suficientemente grande,

$$\sigma_{n,m}^2 \leq \varepsilon,$$

e como ε é qualquer, basta escolher n tal que

$$-2 \leq \frac{-t^2\sigma_{n,m}^2}{2} \leq 0.$$

□

Com isso, temos que

$$\prod_{m=1}^{n} \varphi_{n,m}(t) \sim \prod_{m=2}^{n} (1 - t^2\sigma_{n,m}^2/2) \sim e^{-t^2\sigma^2/2}$$

e daí basta utilizar o teorema de continuidade de Lévy 3.2.16.

■

Para finalizar essa seção, vamos mostrar que é possível mostrar o caso com v.a.'s i.i.d. estudado em 3.3.1 utilizando a versão de Lindeberg-Feller.

Considere então (Y_n) uma sequência de v.a.'s i.i.d. com $\mathbb{E}Y_1 = 0$ e $\mathbb{V}\mathrm{ar}(Y_1) = \sigma^2 < \infty$ e defina:

$$X_{n,m} := \frac{Y_m}{\sqrt{n}}.$$

Note que então

1. $\displaystyle\sum_{m=1}^{n} \mathbb{E}X_{n,m}^2 \to \sigma^2 > 0.$

2. Para todo $\varepsilon > 0$ fixado

$$\begin{aligned}\sum_{m=1}^{n} \mathbb{E}(|X_{n,m}^2|, |X_{n,m}| > \varepsilon) &= n\mathbb{E}(|Y_1/\sqrt{n}|^2, |Y_1/\sqrt{n}| > \varepsilon) \\ &= \mathbb{E}(|Y_1|^2, |Y_1| > \varepsilon\sqrt{n}) \xrightarrow{n\to\infty} 0,\end{aligned}$$

onde limite acima é verdade pelo teorema da convergência dominada e o fato de que Y_1 tem segundo momento finito.

3.4 Teorema Central do Limite - Rota Alternativa

Nesta seção, vamos exibir um outro caminho em direção a prova do Teorema Central do Limite. O método que utilizaremos para provar o TCL é conhecido como método de Lindeberg e não exige o uso de funções características. Os únicos resultados que utilizaremos é a equivalência de convergência em distribuição via funções continuas e limitadas e alguns fatos sobre a distribuição normal.

Teorema 3.4.1 (Lindeberg-Feller - [Str10]). *Para cada n, sejam $X_{n,m}$, $0 \leq m \leq n$ v.a.'s **independentes** com $\mathbb{E}_{n,m} = 0$. Suponha que*

1. $\displaystyle\sum_{m=1}^{n} \mathbb{E}X_{n,m}^2 = \sigma^2 > 0.$

2. *Para todo $\varepsilon > 0$*

$$\sum_{m=1}^{n} \mathbb{E}(|X_{n,m}^2| 1_{|X_{n,m}|>\varepsilon}) \xrightarrow{n\to\infty} 0.$$

Então,

$$S_n = X_{n,1} + \cdots + X_{n,n} \longrightarrow_d \sigma Z, \;\; n \to \infty.$$

Onde $Z \sim N(0,1)$.

Demonstração. Suponha $\sigma = 1$. Considere $W_{n,1}, \ldots, W_{n,n}$ v.a.'s normais centradas com

$$\mathbb{V}\text{ar}(W_{n,m}) = \mathbb{E}X_{n,m}^2 = \sigma_{n,m}^2,$$

que sejam independentes entre si e também independentes de $(X_{n,m})$.

A ideia é provar que para para qualquer f contínua e limitada, temos que

$$\mathbb{E}f(X_{n,1} + \ldots, X_{n,1}) - \mathbb{E}f(W_{n,1} + \ldots, W_{n,1}) \xrightarrow{n\to\infty} 0$$

e assim concluir o teorema. (Precisa mostrar que a soma infinita de normais nesse caso converge)***

Para isso, vamos supor sem perda de generalidade que $f \in C_c^\infty(\mathbb{R})$ e o resultado geral sai por um argumento de aproximação.

Defina

$$\mathbb{E}^{(j)}(\,\cdot\,) := \mathbb{E}(\,\cdot\,| X_{n,1}, X_{n,2}, \ldots, X_{n,j-1}, W_{n,j+1}, \ldots, W_{n,n}).$$

Utilizando a mesma técnica do Lema 3.2.9, se

$$Z_{n,j} = X_{n,1} + \cdots + X_{n,j-1} + 0 + W_{n,j+1} + \cdots + W_{n,n}$$

conseguimos mostrar que para qualquer v.a. Y, existem constantes $A, B > 0$ tal que,

$$\left| \mathbb{E}\mathbb{E}^{(j)} \left(f(Z_{n,j} + Y) - f(Z_{n,j}) - f'(Z_{n,j})Y - \frac{1}{2} f''(Z_{n,j})Y^2 \right) \right|$$
$$\leq \mathbb{E} \min\{AY^2, BY^3\}.$$

Portanto, a menos de constantes,

$$
\begin{aligned}
&\mathbb{E}\mathbb{E}^{(j)}\left(f(Z_{n,j}+X_{n,j})-f(Z_{n,j}+W_{n,j})\right)\\
&\leq \mathbb{E}\min\{X_{n,j}^2, X_{n,j}^3\}+\mathbb{E}\min\{W_{n,j}^2, W_{n,j}^3\}\\
&+\mathbb{E}\left(f'(Z_{n,j})(X_{n,j})-\frac{1}{2}f''(Z_{n,j})X_{n,j}^2-f'(Z_{n,j})(W_{n,j})+\frac{1}{2}f''(Z_{n,j})W_{n,j}^2\right)\\
&=\mathbb{E}\min\{X_{n,j}^2, X_{n,j}^3\}+\mathbb{E}\min\{W_{n,j}^2, W_{n,j}^3\}\\
&\leq \varepsilon\mathbb{E}X_{n,j}^2+\mathbb{E}(X_{n,j}^2, |X_{n,j}|>\varepsilon)+\mathbb{E}W_{n,j}^3.
\end{aligned}
$$

já que $X_{n,j}$ e $W_{n,j}$ tem mesma média e variância e são independentes de $Z_{n,j}$.

Mas note que

$$
\begin{aligned}
\mathbb{E}f(X_{n,1}+\ldots,X_{n,1})-\mathbb{E}f(W_{n,1}+\ldots,W_{n,1}) &= \mathbb{E}f(Z_{n,n}+X_{n,n})-\mathbb{E}f(Z_{n,1}+W_{n,1})\\
&=\sum_{j=1}^{n}\mathbb{E}f(Z_{n,j}+X_{n,j})-\mathbb{E}f(Z_{n,j}+W_{n,j})
\end{aligned}
$$

Portanto, utilizando a desigualdade e o fato de que $(X_{n,m})$ satisfaz a condição de Lindeberg, temos o resultado.

■

3.5 Exercícios Resolvidos

Exercício 3.5.1 ([CY12]). O objetivo deste exercício é mostrar que se vale

$$\begin{cases} (X-Y) \text{ e } X \text{ são independentes} \\ (X-Y) \text{ e } Y \text{ são independentes,} \end{cases}$$

então $X-Y$ é constante q.t.p.. Vamos dividir esse problema em itens.

1. Prove que o resultado é verdadeiro se X e Y tem segundo momento finito.

* Para os próximos itens, não vamos supor integrabilidade.

2. Se φ_X é a f.c. de X, e H é a f.c. de $(X-Y)$, mostre que

$$\varphi(x)(1-|H(x)|^2) = 0, \ \forall x \in \mathbb{R}.$$

3. Mostre que se vale o item 2, então para $\varepsilon >$ suficientemente pequeno

$$|H(x)| = 1, \ \forall x \in (-\varepsilon, \varepsilon).$$

4. Mostre que se vale o item anterior, então $X-Y$ é constante q.t.p.

Demonstração. 1. Suponha sem perda de generalidade que $\mathbb{E}(X) = \mathbb{E}(Y) = 0$. Então

$$\begin{aligned} \mathbb{E}((X-Y)^2) &= \mathbb{E}((X-Y)X) - \mathbb{E}((X-Y)Y) \\ &= \mathbb{E}(X-Y)\mathbb{E}(X) - \mathbb{E}(X-Y)\mathbb{E}(Y) = 0. \end{aligned}$$

Isto é, $\mathbb{V}\text{ar}(X-Y) = 0$ e portanto $X-Y$ é constante q.t.p..

2. Basta notar que, por independência,

$$\begin{cases} \varphi_Y(t) = \varphi_{X+-(X-Y)}(t) & = \varphi_X(t)\varphi_{-(X-Y)}(t) \\ \varphi_X(t) = \varphi_{Y+(X-Y)}(t) & = \varphi_Y(t)\varphi_{X-Y}(t), \end{cases}$$

portanto,

$$\begin{aligned} \varphi_X(t) &= \varphi_Y(t)H(t) \\ &= \varphi_X(t)\varphi_{-(X-Y)}(t)\varphi_{X-Y}(t) \\ &= \varphi_X(t)H(t)\overline{H(t)} \\ &= \varphi_X(t)|H(t)|^2. \end{aligned}$$

3. Como $\varphi_X(0) = 1$ e φ_X é contínua em 0, temos o resultado.

Lema 3.5.2. *Suponha que para uma v.a. Z*

$$|\varphi_Z(t)|^2 = 1, \ |t| < \varepsilon.$$

Então Z é constante q.t.p..

Demonstração. Considere então Z' uma v.a. i.i.d. com relação à Z. Temos então que

$$\begin{aligned}\mathbb{E}(e^{it(Z-Z')}) &= \varphi_{Z-Z'}(t) \\ &= \varphi_Z(t)\overline{\varphi_Z(t)} = 1.\end{aligned}$$

Portanto,

$$\mathbb{E}(\cos(t(Z-Z')) = 1,$$

isto é,

$$t(Z-Z') \in 2\pi\mathbb{Z},\ q.t.p..$$

Disso, é fácil concluir que $Z = Z'$, já que podemos dividir tudo por t e então tomar $t \to 0$ para chegar numa contradição.

Mas então, como Z, Z' são independentes, note que, para qualquer evento A,

$$\mathbb{P}(Z \in A) = \mathbb{P}(Z \in A, Z' \in A) = \mathbb{P}(Z \in A)^2.$$

Logo, $\mathbb{P}(Z \in A) \in \{0,1\}$. ■

4. Basta usar o Lema 3.5.2 e o item 3.

■

Exercício 3.5.3 ([Dur10])**.** Se $\varphi_X(t)$ é real, então X e $-X$ tem a mesma lei.

Demonstração. Se $\varphi_X(t)$ é real, então

$$\varphi_X(t) = \overline{\varphi_X(t)},$$

mas então, por 3.2.2, temos que

$$\varphi_X(t) = \overline{\varphi_X(t)} = \varphi_X(-t) = \varphi_{-X}(t),$$

e portanto, pela fórmula de inversão 3.2.6, temos o resultado. ■

Exercício 3.5.4 ([Dur10])**.** Sejam $X_i,\ i \in \{1,2\}$ independentes com distribuição normal com média 0 e variância $\sigma_i^2,\ i = 1,2$, respectivamente. Então $X_1 + X_2$ tem distribuição normal com média 0 e variância $\sigma_1^2 + \sigma_2^2$.

Demonstração. Temos que

$$\begin{aligned}\varphi_{X_1+X_2}(t) &= \varphi_{X_1}(t)\varphi_{X_2}(t) \\ &= e^{-\sigma_1^2t^2/2}e^{-\sigma_2^2t^2/2} \\ &= e^{-(\sigma_1^2+\sigma_2^2)t^2/2} \\ &= \varphi_Y(t),\end{aligned}$$

onde Y tem distribuição normal com média 0 e variância $\sigma_1^2+\sigma_2^2$. O resultado segue pela fórmula de inversão 3.2.6. ■

Capítulo 4

Lei(s) dos Grandes Números

As Principais referências usadas foram:

1. [Dur10]

4.1 Lei Fraca dos Grandes Números

A nossa principal referência para esse tópico é [Dur10].

Nesta seção, estamos tentando o entender o seguinte problema:

Sejam $(X_i)_{i\leq n}$ v.a. e $b_n > 0$ uma sequência positiva. Vamos estudar as condições necessárias para que

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(S_n)}{b_n}\right| > \varepsilon\right) \xrightarrow{n\to\infty} 0. \tag{4.1.1}$$

4.1.1 Brainstorm

Suponha que $(X_i)_{i\leq n}$ sejam v.a. 2-2 independentes, e com variância finita. Então, por Chebyshev:

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(S_n)}{b_n}\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\frac{\sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i)}{b_n^2}.$$

Portanto, se

$$\frac{\sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i)}{b_n^2} \xrightarrow{n\to\infty} 0,$$

conseguimos garantir que vale 4.1.1.

Isso automaticamente nos leva ao seguinte resultado:

Teorema 4.1.1 (Lei Fraca dos Grandes Números: i.i.d.+Variância Finita). *Sejam $(X_i)_{i\leq n}$ sejam v.a. i.i.d. com variância finita, então*

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(S_n)}{n}\right| > \varepsilon\right) \xrightarrow{n\to\infty} 0.$$

Demonstração. Por Chebyshev:

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(S_n)}{n}\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\frac{n\mathbb{V}\mathrm{ar}(X_1)}{n^2} \xrightarrow{n\to\infty} 0.$$

■

Observação 4.1.2. Note que acima escolhemos $b_n = n$, mas o resultado ainda vale se $b_n = n^{1/2+\alpha}$, $\alpha > 0$.

Da análise acima, temos que no caso em que X_i tem segundo momento finito ou, mais geralmente, que conseguimos mostrar que,

$$\frac{\mathbb{V}\mathrm{ar}(S_n)}{b_n^2} \xrightarrow{n\to\infty} 0,$$

temos total entendimento de 4.1.1, então o nosso próximo passo é estudar o que acontece quando não pedimos segundo momento finito.

Uma primeira ideia seria truncarmos as v.a.'s X_i considerando uma nova sequência dada por

$$\bar{X}_i = X_i 1_{|X_i|\leq c}.$$

Dessa forma, se

$$\bar{S}_n = \bar{X}_1 + \cdots + \bar{X}_n,$$

temos o seguinte:

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(S_n)}{b_n}\right| > \varepsilon\right) &= \mathbb{P}\left(\left\{\left|\frac{S_n - \mathbb{E}(S_n)}{b_n}\right| > \varepsilon\right\} \cap \{S_n = \bar{S}_n\}\right) \\
&\quad + \mathbb{P}\left(\left\{\left|\frac{S_n - \mathbb{E}(S_n)}{b_n}\right| > \varepsilon\right\} \cap \{S_n \neq \bar{S}_n\}\right) \\
&\leq \underbrace{\mathbb{P}\left(\left|\frac{\bar{S}_n - \mathbb{E}(\bar{S}_n)}{b_n}\right| > \varepsilon\right)}_{\mathbf{I}} + \underbrace{\mathbb{P}(S_n \neq \bar{S}_n)}_{\mathbf{II}}.
\end{aligned}$$

Note que pela discussão anterior, sabemos exatamente o comportamento do termo **I**, já que as v.a.'s são limitadas por c e portanto tem segundo momento finito. Vamos tentar entender o que acontece com o termo **II**.

Um possível bound para o termo **II** seria

$$\begin{aligned}
\mathbb{P}(S_n \neq \bar{S}_n) &\leq \mathbb{P}(\cup\{X_i \neq \bar{X}_i\}) \\
&\leq \sum_{i=1}^{\infty} \mathbb{P}(|X_i| > c).
\end{aligned}$$

Note que isso nos causa um problema sério, já que como $c > 0$ é qualquer e não temos que X_i são necessariamente limitadas, dificilmente vamos conseguir provar que

$$\sum_{i=1}^{n} \mathbb{P}(|X_i| > c) \xrightarrow{n\to\infty} 0.$$

O que podemos fazer para tentar corrigir tal problema é trocarmos $c > 0$ por uma sequência positiva $c_n \to +\infty$ de tal forma que

$$\sum_{i=1}^{n} \mathbb{P}(|X_i| > c_n) \xrightarrow{n\to\infty} 0,$$

isso parece ser bem promissor, já que, por exemplo, se X_i são i.i.d. tal que $\mathbb{E}(|X_1|) < \infty$, então tomando $c_n = n$,

$$\sum_{i=1}^{n} \mathbb{P}(|X_i| > c_n) = n\mathbb{P}(|X_1| > n) \xrightarrow{n\to\infty} 0,$$

como já vimos anteriormente (basta utilizar convergência dominada).

Note que as observações acima não são suficientes para concluirmos o problema 4.1.1 no caso em que não temos informação de segundo momento, já que trocarmos o limitante $c > 0$ por uma sequência positiva e então precisamos tomar certo cuidado com o termo **I**. Mas como veremos a seguir, isso não será um grande problema.

Observação 4.1.3. Note que faz sentido utilizarmos

$$\mathbb{P}(\cup_k^n \{X_k \neq \bar{X}_{n,k}\})$$

como um bound para **II**, já que tomando c_n suficientemente grande, temos que a probabilidade do evento $\{X_k \neq \bar{X}_{n,k}\}$ deve diminuir.

4.1.2 Formalização

Teorema 4.1.4 (Lei Fraca para arrays Triangulares)**.** *Para cada n, seja $X_{n,k}$, $1 \leq k \leq n$,* ***independentes****. Seja $c_n > 0$ com $c_n \to \infty$ e $\bar{X}_{n,k} = X_{n,k}1_{(|X_{n,k}| \leq c_n)}$, então se*

1. $\sum_{k=1}^{n} \mathbb{P}(|X_{n,k}| > c_n) \xrightarrow{n\to\infty} 0$ *e*
2. $\frac{1}{b_n^2} \sum_{k=1}^{n} \mathbb{E}(\bar{X}_{n,k}^2) \xrightarrow{n\to\infty} 0,$

então

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(\bar{S}_n)}{b_n}\right| > \varepsilon\right) \xrightarrow{n\to\infty} 0,$$

onde $S_n = X_{n,1} + \cdots + X_{n,n}$ e $\bar{S}_n = \bar{X}_{n,1} + \cdots + \bar{X}_{n,n}$.

Demonstração. Como já vimos anteriormente, temos que

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(\bar{S}_n)}{b_n}\right| > \varepsilon\right) \leq \underbrace{\mathbb{P}\left(\left|\frac{\bar{S}_n - \mathbb{E}(\bar{S}_n)}{b_n}\right| > \varepsilon\right)}_{\mathbf{I}} + \underbrace{\mathbb{P}(S_n \neq \bar{S}_n)}_{\mathbf{II}}$$

(I): Por Chebyshev, temos que

$$\begin{aligned}\mathbb{P}\left(\left|\frac{\bar{S}_n - \mathbb{E}(\bar{S}_n)}{b_n}\right| > \varepsilon\right) &\leq \frac{1}{\varepsilon^2}\frac{\mathbb{V}\mathrm{ar}(\bar{S}_n)}{b_n^2} \\ &\leq \frac{1}{\varepsilon^2}\frac{\sum_{k=1}^n \bar{X}_{n,k}^2}{b_n^2} \xrightarrow{n\to\infty} 0,\end{aligned}$$

por hipótese.

(II): Temos que

$$\begin{aligned}\mathbb{P}(S_n \neq \bar{S}_n) &\leq \mathbb{P}(\cup_k^n\{X_k \neq \bar{X}_{n,k}\}) \\ &\leq \sum_{k=1}^n \mathbb{P}(|X_{n,k}| > c_n) \xrightarrow{n\to\infty},\end{aligned}$$

por hipótese.

Portanto,

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(\bar{S}_n)}{b_n}\right| > \varepsilon\right) \xrightarrow{n\to\infty} 0.$$

■

Teorema 4.1.5 (Lei Fraca dos Grandes Números). *Sejam X_i v.a.'s i.i.d. com*

$$x\mathbb{P}(|X_1| > x) \xrightarrow{x\to\infty} 0,$$

e $\bar{X}_k = X_k 1_{(|X_k| \leq n)}$. Então

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(\bar{S}_n)}{n}\right| > \varepsilon\right) \xrightarrow{n\to\infty} 0.$$

> *Observação* 4.1.6. Note que o que aparece na nossa expressão é $\mathbb{E}(\bar{S}_n)$ e não $\mathbb{E}(S_n)$!.

Demonstração. Vamos utilizar o Teorema 4.1.4 com $X_{n,k} = X_k$, e $b_n = c_n = n$. Então só precisamos mostrar que

1. $\sum_{k=1}^n \mathbb{P}(|X_k| > n) \xrightarrow{n\to\infty} 0$ e

2. $\dfrac{1}{n^2}\sum_{k=1}^n \mathbb{E}(\bar{X}_k^2) \xrightarrow{n\to\infty} 0.$

Note que o item (1) é satisfeito, já que

$$\sum_{k=1}^n \mathbb{P}(|X_k| > n) = n\mathbb{P}(|X_1| > n) \xrightarrow{n\to\infty} 0,$$

por hipótese.

Para o item (2), note que só precisamos mostrar que

$$\frac{1}{n}\mathbb{E}(\bar{X}_1^2) \xrightarrow{n\to\infty} 0,$$

já que as v.a.'s são i.i.d.. Mas

$$\begin{aligned}
\frac{1}{n}\mathbb{E}(\bar{X}_1^2) &= \frac{1}{n}\int_0^\infty \mathbb{P}(|\bar{X}_1^2| > t)dt \\
&= \frac{1}{n}\int_0^\infty 2t\mathbb{P}(|\bar{X}_1| > t)dt \\
&= \frac{2}{n}\int_0^n t\mathbb{P}(|\bar{X}_1| > t)dt + \frac{2}{n}\int_n^\infty t\mathbb{P}(|\bar{X}_1| > t)dt \\
&= \frac{2}{n}\int_0^n t\mathbb{P}(|\bar{X}_1| > t)dt \\
&= \frac{2}{n}\left(\int_0^n t\mathbb{P}(|X_1| > t)dt - n\mathbb{P}(|X_1| > n)\right) \\
&\leq \frac{2}{n}\int_0^n t\mathbb{P}(|X_1| > t)dt.
\end{aligned}$$

Já que

$$t\mathbb{P}(|X_1| > t) \to 0,$$

deve existir $t_0 > 0$ tal que

$$t\mathbb{P}(|X_1| > t) \leq 1, \ \forall t > t_0,$$

portanto, como

$$t\mathbb{P}(|X_1| > t) \leq t,$$

temos que,

$$t\mathbb{P}(|X_1| > t) \leq t_0, \ \text{para } t \leq t_0,$$

isto é,

$$t\mathbb{P}(|X_1| > t) \leq \max\{t_0, 1\}, \ t > 0.$$

Logo, existe $M < \infty$ tal que

$$M = \sup_{t>0} t\mathbb{P}(|X_1| > t).$$

Defina

$$m_K = \sup_{t>K} t\mathbb{P}(|X_1| > t),$$

então temos que

$$\begin{aligned}
\frac{1}{n}\mathbb{E}(\bar{X}_1^2) &\leq \frac{2}{n}\int_0^n t\mathbb{P}(|X_1| > t)dt \\
&= \frac{2}{n}\int_0^K t\mathbb{P}(|X_1| > t)dt + \frac{2}{n}\int_K^n t\mathbb{P}(|X_1| > t)dt \\
&\leq \frac{2}{n}K \cdot M + \frac{2}{n}(n-K)m_K \\
&= \frac{2}{n}K \cdot M + \frac{2}{n}nm_K - \frac{2}{n}Km_K \\
&\xrightarrow{n\to\infty} 2m_K.
\end{aligned}$$

Mas K foi arbitrário e

$$m_K = \sup_{t>K} t\mathbb{P}(|X_1| > t),$$

então, como

$$n\mathbb{P}(|X_1| > n) \xrightarrow{n\to\infty} 0,$$

é claro que tomando $K \to \infty$, temos que

$$m_K \to 0,$$

concluindo o resultado.

■

Teorema 4.1.7 (Lei Fraca - i.i.d.+$\mathbb{E}(|X_1|) < \infty$)**.** *Sejam X_i v.a.'s i.i.d. com $\mathbb{E}(|X_1|) < \infty$. Então*

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}(S_n)}{n}\right| > \varepsilon\right) \xrightarrow{n\to\infty} 0.$$

Observação 4.1.8. Note que, agora sim, quem aparece na expressão é $\mathbb{E}(S_n)$ e não sua versão truncada.

Observação 4.1.9. Como as v.a.'s são i.i.d., o resultado acima pode ser escrito como

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| > \varepsilon\right) \xrightarrow{n\to\infty} 0.$$

Demonstração. A estratégia básica é que, como temos integrabilidade de X, conseguimos utilizar o Teo. da Conv. Dominada (TCD) para passarmos os limites para dentro das integrais.

Note que por TCD,

$$x\mathbb{P}(|X_1| > x) \leq \mathbb{E}(|X_1|1_{|X_1|>x}) \xrightarrow{x\to\infty} 0,$$

e portanto pela Lei Fraca dos Grandes Números (Teorema 4.1.5), temos que a média da soma converge para a média truncada de x_1 em probabilidade, isto é,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(\bar{S_n})\right| > \varepsilon\right) \xrightarrow{n\to\infty} 0,$$

mas utilizando a integrabilidade de X_1 e o TCD novamente, temos que

$$\mathbb{E}(\bar{X}_1) = \mathbb{E}(X_1 1_{|X_1|\leq n}) \xrightarrow{x\to\infty} \mathbb{E}(X_1),$$

então, pela desigualdade triangula para probabilidades

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| > \varepsilon\right) &= \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1) + \mathbb{E}(\bar{X}_1) - \mathbb{E}(\bar{X}_1)\right| > \varepsilon\right) \\
&\leq \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(\bar{X}_1)\right| > \varepsilon/2\right) + \mathbb{P}\left(\left|\mathbb{E}(\bar{X}_1) - \mathbb{E}(X_1)\right| > \varepsilon/2\right) \\
&\xrightarrow{n\to\infty} 0.
\end{aligned}$$

■

4.2 Lemas de Borel-Cantelli

Teorema 4.2.1 (Primeiro Lema de Borel-Cantelli)**.** *Seja* $(A_n)_{n\in\mathbb{N}}$ *uma sequência de eventos mensuráveis. Se*

$$\sum_{i=1}^{\infty} \mathbb{P}(A_n) < \infty,$$

então

$$\mathbb{P}(A_n,\ i.o.) = 0.$$

Demonstração. Pelo teorema da convergência monótona,

$$\mathbb{E}\left(\sum_{i=1}^{\infty} 1_{A_i}\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty.$$

Isso implica que

$$\sum_{i=1}^{\infty} 1_{A_i} < \infty, \text{ q.t.p.},$$

mas

$$\sum_{i=1}^{\infty} 1_{A_i}(\omega) < \infty,$$

se, e só se, $\omega \in A_i$ para apenas finitos $i's$. ■

Teorema 4.2.2 (Segundo Lema de Borel-Cantelli)**.** *Seja* $(A_n)_{n\in\mathbb{N}}$ *uma sequência de eventos mensuráveis* ***independentes****. Se*

$$\sum_{i=1}^{\infty} \mathbb{P}(A_n) = \infty,$$

então

$$\mathbb{P}(A_n,\ i.o.) = 1.$$

Demonstração. Já vimos que

$$\mathbb{P}(\{A_n,\ i.o.\}) = \mathbb{P}(\lim_{m\to\infty} B_m) = \lim_{m\to\infty} \mathbb{P}(B_m),$$

onde

$$B_m = \bigcup_{n=m}^{\infty} A_n.$$

Utilizando a identidade $1 - x \leq e^{-x}$ e o fato de que $\mathbb{P}(D^c) = 1 - \mathbb{P}(D)$ para D mensurável qualquer, temos que

$$\begin{aligned}
\mathbb{P}(B_m) &= \mathbb{P}(\bigcup_{n=m}^{\infty} A_n) \\
&= 1 - \mathbb{P}(\bigcap_{n=m}^{\infty} A_n^c) \\
&= 1 - \prod_{n=m}^{\infty} (\mathbb{P}(A_n^c)) \\
&= 1 - \prod_{n=m}^{\infty} (1 - \mathbb{P}(A_n)) \\
&\leq 1 - \exp(-\sum_{n=m}^{\infty} \mathbb{P}(A_n)) = 1.
\end{aligned}$$

Logo,

$$\mathbb{P}(A_n,\ i.o.) = 1.$$

■

Corolário 4.2.3 (Lei 0-1). *Seja $(A_n)_{n\in\mathbb{N}}$ uma sequência de eventos mensuráveis* ***independentes****. Então*

$$\mathbb{P}(A_n,\ i.o.) \in \{0, 1\}.$$

Demonstração. Basta notar que a série

$$\sum_{i=1}^{\infty} \mathbb{P}(A_n)$$

converge ou diverge para $+\infty$. ■

Exemplo 4.2.4. Seja (X_n) uma sequência i.i.d. de v.a.'s Bernoulli tal que $\mathbb{P}(X_n = 1) = 1/n^{\alpha}$, $\alpha > 0$. Vamos estudar a probabilidade do evento

$$A = \{X_n = 1, \text{ i.o.}\},$$

isto é, o evento de $X_n = 1$ infinitas vezes. Pelos lemas de Borel-Cantelli, temos que

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n = 1) = \sum_{n=1}^{\infty} = \frac{1}{n^{\alpha}},$$

e portanto se $\alpha \leq 1$ temos que $\mathbb{P}(A) = 0$ e caso contrário $\mathbb{P}(A) = 1$.

Exemplo 4.2.5. Seja $\Omega = (0, 1)$ com a medida de Lebesgue e considere $A_n = (0, 1/n)$. É fácil ver que

$$\sum_{i=1}^{\infty} \mathbb{P}(A_n) = \infty,$$

mas

$$\mathbb{P}(A_n, \text{i.o.}) = 0.$$

Note que isso não contraria o segundo lema de Borel-Cantelli, já que os eventos A_n não são independentes!

Exemplo 4.2.6. Seja (X_n) uma sequência i.i.d. de v.a.'s Bernoulli tal que $\mathbb{P}(X_n = 1) = p_n$. É fácil ver que

$$X_n \to_p 0$$

se, e só se, $p_n \xrightarrow{n\to\infty} 0$.

Agora, vamos mostrar que $X_n \to_{\text{q.t.p.}} 0$ se, e só se, $\sum p_n < \infty$. Para isso, note que dado $1 > \varepsilon > 0$,

$$\mathbb{P}(X_n > \varepsilon) = \mathbb{P}(X_n = 1) = p_n$$

e portanto, pelos lemas de Borel-Cantelli, temos que

$$\mathbb{P}(X_n > \varepsilon, \text{ i.o.}) = 0, \text{ se } \sum p_n < \infty$$

e que

$$\mathbb{P}(X_n > \varepsilon, \text{ i.o.}) = 1, \text{ se } \sum p_n = \infty.$$

Concluímos então que existe uma sequência de v.a.'s que converge em probabilidade mas não q.t.p., já que podemos considerar $p_n = 1/n$.

4.2.1 Aplicações dos Lemas de Borel-Cantelli

Lema 4.2.7. *Seja Ω um espaço métrico e y_n uma sequência tal que para qualquer subsequência $y_{n(m)}$ existe outra subsequência $y_{n(m_k)}$ tal que*

$$y_{n(m_k)} \to y.$$

Então $y_n \to y$.

Demonstração. Se y_n não converge para y, existe uma bola B em torno de y e uma subsequência $y_n(m)$ tal que $y_n(m) \notin B$ para todo $n(m)$, o que contradiz a hipótese. ∎

Proposição 4.2.8. *Temos que X_n converge para X em probabilidade se, e só se, para toda subsequência $X_{n(m)}$ existe uma subsequência $X_{n(m_k)}$ desta última tal que $X_{n(m_k)}$ converge q.t.p. para X.*

Demonstração. Suponha que X_n converge para X em probabilidade. Seja a_k uma sequência tal que $a_k \to 0$ quando $k \to \infty$, considere $n(m_k)$ tal que

$$\mathbb{P}(|X_{n(m_k)} - X| > a_k) \leq \frac{1}{2^k}.$$

Temos então pelo primeiro lema de Borel-Cantelli que

$$\mathbb{P}(|X_{n(m_k)} - X| > a_k, \text{ i.o.}) = 0,$$

concluindo a primeira parte. Para a segunda parte, dado $\varepsilon > 0$, considere

$$y_n = \mathbb{P}(|X_n - X| > \varepsilon)$$

e aplique o lema 4.2.7. ∎

Observação 4.2.9. Como já vimos que existem sequências que convergem em probabilidade mas não q.t.p., o lema 4.2.7 nos diz que a convergência q.t.p. **não** é originária de nenhuma métrica (na verdade, nenhuma topologia).

Corolário 4.2.10. *Se X_n converge para X em probabilidade e f é contínua, então $f(X_n)$ converge para $f(X)$ em probabilidade. Além disso, se f é limitada, então $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$.*

Demonstração. Como X_n converge para X em probabilidade e f é contínua, pela volta da proposição 4.2.8, temos a primeira parte do resultado. Além disso, se f é limitada, pelo Teorema da Convergência Limitada, temos que para qualquer subsequência, existe outra subsequência tal que

$$\mathbb{E}(f(X_{n(m_k)})) \to \mathbb{E}(f(X)),$$

e daí, definindo

$$y_n = \mathbb{E}(f(X_n)),$$

basta utilizar o lema 4.2.7. ■

Teorema 4.2.11 (Lei Forte - $\mathbb{E}(X^4) < \infty$)**.** *Seja (X_n) uma sequência v.a.'s i.i.d. tal que $\mathbb{E}(X) = \mu$ e $\mathbb{E}(X^4) < \infty$.*
Então se $S_n = X_1 + \cdots + X_n$, temos que

$$\frac{S_n}{n} \xrightarrow{n\to\infty}_{q.t.p.} \mu.$$

Demonstração. Sem perda de generalidade, podemos supor que $\mu = 0$, então dado $\varepsilon > 0$, o nosso objetivo é mostrar que

$$\mathbb{P}(|S_n/n - \mu| > \varepsilon,\ i.o.) = 0.$$

Note que

$$\mathbb{E}(X^4) = \sum_{i,j,k,l} \mathbb{E}(X_i X_j X_k X_l),$$

e por independência temos que $\mathbb{E}(X_i^2 X_j X_k) = \mathbb{E}(X_i^3 X_k) = \mathbb{E}(X_i X_j X_k X_l) = 0$, se os índices acima são diferentes. Portanto, só precisamos estudar os casos em que $\mathbb{E}(X_i^4)$ e $\mathbb{E}(X_i^2 X_j^2)$, com índices diferentes.

É claro que temos n possibilidades para $\mathbb{E}(X_i^4)$, e para $\mathbb{E}(X_i^2 X_j^2) = \mathbb{E}(X_i^2)^2$ temos $n(n-1)6/2 = 3n(n-1)$, já que devemos escolher um índice, depois um índice diferente onde a ordem não importa, resultando em $n(n-1)/2$ possibilidades, e com os índices fixados, temos 6 formas de fazer aparecer dois quadrados. Ou seja,

$$\mathbb{E}(X^4) = n\mathbb{E}(X_1^4) \leq Cn^2.$$

Mas então, usando a desigualdade de Markov,

$$\mathbb{P}(|S_n/n| > \varepsilon) \leq \frac{Cn^2}{(\varepsilon n)^4} = \frac{C}{\varepsilon^4 n^2},$$

e portanto,

$$\sum_{i=1}^{\infty} \mathbb{P}(|S_n/n| > \varepsilon) < \infty,$$

concluindo assim o teorema já que o primeiro lema de Borel-Cantelli nos diz que

$$\mathbb{P}(|S_n/n - \mu| > \varepsilon,\ i.o.) = 0.$$

■

Exemplo 4.2.12 (Long Run)**.** Seja (X_n) uma sequência i.i.d. Bernoulli tal que $\mathbb{P}(X_1 = 1) = 1/2$. Considere

$$l_n = \max\{m : X_{n-m+1} = \cdots = X_n = 1\},$$

e defina

$$L_n = \max_{1 \leq m \leq n} l_n,$$

isto é, L_n é a maior sequência constante igual à 1 de lançamentos consecutivos (long run) até o tempo n.

É fácil ver que

$$\mathbb{P}(l_n \geq (1+\varepsilon)\log_2 n) \leq \frac{1}{n^{1+\varepsilon}},$$

logo por Borel-Cantelli, para quase todo $\omega \in \Omega$ e $\varepsilon > 0$, existe N_ε tal que para todo $n \geq N_\varepsilon$,

$$l_n(\omega) \leq (1+\varepsilon)\log_2 n$$

e portanto

$$\limsup_{n\to\infty} L_n(\omega)/\log_2 n \leq 1+\varepsilon.$$

Como isso vale para quase todo ω e para todo $\varepsilon > 0$, temos que

$$\limsup_{n\to\infty} L_n/\log_2 n \leq 1,\ \text{q.t.p.}$$

Agora precisamos mostrar que

$$\liminf_{n\to\infty} L_n/\log_2 n \geq 1,\ \text{q.t.p.},$$

para isso, vamos dividir a sequência $[n]$, com n grande, em $(1-\varepsilon)\log_2 n + 1$ pedaços iguais. Logo, a probabilidade de um desses blocos ser constante igual à 1 é

$$\left(\frac{1}{2}\right)^{1+(1-\varepsilon)\log_2 n} = \left(\frac{1}{2n}\right)^{1-\varepsilon}.$$

Então,
$$\mathbb{P}(L_n \leq (1-\varepsilon)\log_2 n) \leq (1 - 1/2n^{1-\varepsilon})^{n/(1-\varepsilon)\log_2 n},$$
já que se $L_n \leq (1-\varepsilon)\log_2 n$, em particular cada um dos blocos que criamos acima **não** é constante igual à 1. Portanto, usando que $1 - x \leq e^{-x}$,

$$\begin{aligned}
\mathbb{P}(L_n \leq (1-\varepsilon)\log_2 n) &\leq (1 - 1/2n^{1-\varepsilon})^{n/(1-\varepsilon)\log_2 n} \\
&\leq e^{-1/2n^{1-\varepsilon} n/(1-\varepsilon)\log_2 n} \\
&= \exp\left(\frac{-n}{(1-\varepsilon)2n^{1-\varepsilon}\log_2 n}\right) \\
&= \exp\left(\frac{-n^{\varepsilon}}{(1-\varepsilon)2\log_2 n}\right)
\end{aligned}$$

que é somável. Temos então, por Borel-Cantelli, que

$$\mathbb{P}(L_n \leq (1-\varepsilon)\log_2 n,\ i.o.) = 0,$$

e então, como $\varepsilon > 0$ é qualquer, temos o resultado.

Exemplo 4.2.13. Suponha que (A_n) é uma sequência de eventos independentes, tal que $\mathbb{P}(A_n) < 1$ e $\mathbb{P}(\cup A_n) = 1$, então

$$\mathbb{P}(A_n,\ i.o.) = 1.$$

Demonstração. É fácil ver, que

$$\prod_{n=1}^{\infty}(1 - \mathbb{P}(A_n)) = 0.$$

Como $1 - \mathbb{P}(A_n) > 0$, devem existir infinitos índices tal que $\mathbb{P}(A_n) > 0$ e portanto, para qualquer m grande o suficiente

$$\prod_{n=m}^{\infty}(1 - \mathbb{P}(A_n)) = 0,$$

ou seja,

$$\mathbb{P}(\bigcup_{n=m}^{\infty} A_n) = 1,$$

mas então, basta lembrar que

$$\mathbb{P}(\{A_n,\ i.o.\}) = \lim_{m\to\infty} \mathbb{P}(B_m),$$

onde

$$B_m = \bigcup_{n=m}^{\infty} A_n.$$

■

O próximo Teorema é importante, pois sua demonstração nos revela um fato muito interessante.

Teorema 4.2.14. *Sejam $A_1, A_2, \ldots$ eventos dois a dois independentes e*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty,$$

então temos que

$$\frac{\sum_{m=1}^{\infty} 1_{A_m}}{\sum_{m=1}^{\infty} \mathbb{P}(A_m)} \xrightarrow{n\to\infty} 1, \ \ q.t.p..$$

Demonstração. Como os eventos são dois a dois disjuntos, se

$$S_n = 1_{A_1} + \cdots + 1_{A_n},$$

então temos que

$$\mathbb{V}\mathrm{ar}(S_n) = \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(X_i),$$

logo, dado $\varepsilon > 0$, por Chebychev,

$$\begin{aligned}
\mathbb{P}(|S_n - \mathbb{E}(S_n)| > \varepsilon\mathbb{E}(S_n)) &\leq \frac{\mathbb{V}\mathrm{ar}(S_n)}{(\varepsilon\mathbb{E}(S_n))^2} \\
&\leq \frac{\sum \mathbb{E}(S_i^2)}{(\varepsilon\mathbb{E}(S_n))^2} \\
&= \frac{\sum \mathbb{E}(S_i)}{(\varepsilon\mathbb{E}(S_n))^2} \\
&= \frac{1}{\varepsilon^2\mathbb{E}(S_n)} \xrightarrow{n\to\infty} 0.
\end{aligned}$$

Ou seja, mostramos que

$$\frac{S_n}{\mathbb{E}(S_n)} \xrightarrow{n\to\infty}_p 1.$$

É claro que ainda não temos o que queríamos, já que precisamos de convergência q.t.p.. Para isso, considere

$$n_k = \inf\{n : \ S_k \geq k^2\},$$

e defina $T_k = S_{n_k}$. Note que como $1_{A_m} \in \{0,1\}$, a coma S_{n_k} cresce a cada termo em no máximo 1 unidade, e portanto tempos que

$$k^2 \leq \mathbb{E}(T_k) \leq k^2 + 1.$$

Utilizando a mesma ideia anterior, por Chebychev, temos que

$$\mathbb{P}(|T_k - \mathbb{E}(T_k)| > \varepsilon\mathbb{E}(T_k)) \leq \frac{1}{\varepsilon^2 k^2},$$

e então, por Borel-Cantelli, temos que

$$\frac{T_k}{\mathbb{E}(T_k)} \xrightarrow{n\to\infty}_{\text{q.t.p.}} 1.$$

Agora, considere então ω satisfazendo a convergência acima, temos que para $n_k \leq n \leq n_{k+1}$

$$\frac{T_k(\omega)}{\mathbb{E}(T_{k+1})} \leq \frac{S_n(\omega)}{\mathbb{E}(S_n)} \leq \frac{T_{k+1}(\omega)}{\mathbb{E}(T_k)}.$$

que podemos reescrever como

$$\frac{\mathbb{E}(T_k)}{\mathbb{E}(T_{k+1})}\frac{T_k(\omega)}{\mathbb{E}(T_k)} \leq \frac{S_n(\omega)}{\mathbb{E}(S_n)} \leq \frac{T_{k+1}(\omega)}{\mathbb{E}(T_{k+1})}\frac{\mathbb{E}(T_{k+1})}{\mathbb{E}(T_k)},$$

e portanto, só precisamos mostrar que

$$\frac{\mathbb{E}(T_{k+1})}{\mathbb{E}(T_k)} \longrightarrow 1.$$

Mas isso é verdade, já que

$$k^2 \leq T_k \leq T_{k+1} \leq (k+1)^2 + 1.$$

■

Observação 4.2.15. Note que o teorema anterior nos diz que se queremos mostrar que

$$\frac{X_n}{c_n} \xrightarrow{n\to\infty}_{\text{q.t.p.}} 1,$$

com $X_n \geq 0$ e ambas sequências X_n e c_n crescentes, então é suficiente mostrarmos que o resultado vale para uma subsequência (n_k) tal que

$$\frac{c_{n_{k+1}}}{c_{n_k}} \longrightarrow 1,$$

e a prova desse fato é realizada seguindo exatamente os passos da demonstração to teorema anterior.

Proposição 4.2.16. *Sejam $X_0 \leq X_1 \leq X_2 \leq \ldots$ v.a.'s com $\mathbb{E}(X_n) \sim an^\alpha$, com $a, \alpha > 0$ e tal que $\mathbb{V}ar(X_n) \leq Bn^\beta$, com $\beta < 2\alpha$. Então*

$$\frac{X_n}{n^\alpha} \xrightarrow{n\to\infty}_{q.t.p.} a.$$

Demonstração. Vamos utilizar a técnica descrita em 4.2.15.

Note que, para $c_n = n^{2/(2\alpha-\beta)}$, temos que

$$\begin{aligned}
\mathbb{P}(|X_{c_n}/c_n^\alpha - a| > \varepsilon) &\leq \frac{\mathbb{V}\mathrm{ar}(X_{c_n})}{c_n^{2\alpha}\varepsilon^2} \\
&= \frac{B}{e^2 c_n^{-\beta+2\alpha}} \\
&= \frac{B}{e^2 c_n^{-\beta+2\alpha}} \\
&= \frac{B}{e^2 n^2}
\end{aligned}$$

.

Note que c_n é crescente em n já que $2\alpha - \beta > 0$ e que

$$\frac{c_{n_{k+1}}}{c_{n_k}} \longrightarrow 1.$$

■

4.3 Convergência de Séries Aleatórias e Lei Forte dos Grandes Números

Definição 4.3.1 (Eventos de Cauda)**.** Seja (X_n) uma sequência de v.a.'s e $\mathcal{F}'_n = \sigma(X_n, X_{n+1}, \dots)$[1]

Definimos a σ-álgebra da cauda da sequência (X_n) como

$$\mathcal{T} = \bigcap_{i=n}^{\infty} \mathcal{F}'_n.$$

Intuitivamente, $A \in \mathcal{T}$ se, e só se, mudar uma quantidade finita de elementos de A não afeta a ocorrência desse evento.

Teorema 4.3.2 (Lei 0-1 de Kolmogorov)**.** *Seja (X_n) uma sequência de v.a's independentes e $A \in \mathcal{T}$, então*

$$\mathbb{P}(A) \in \{0, 1\}.$$

Demonstração. A ideia é mostrar que se $A \in \mathcal{T}$ então o evento A independente de si mesmo, e portanto $\mathbb{P}(A) \in \{0, 1\}$.

Para uma prova formal, precisamos de alguns lemas técnicos envolvendo π-systems, que eu farei no futuro. Entretanto, os passos são bem intuitivos se abrirmos mão de de formalismo. Estes são:

1. Se $A \in \sigma(X_1, \dots, X_k)$ e $B = \sigma(X_{k+1}, X_{k+2}, \dots)$ então A é, intuitivamente, independente de B, já que (X_n) é uma sequência de v.a.'s independentes.

2. Se $A \in \sigma(X_1, X_2, \dots)$ e $B \in \mathcal{T}$, então A, B são independentes. Isso acontece pelo fato de ser suficiente mostrar no caso em que $A \in \sigma(X_1, \dots, X_k)$ para algum k. Nesse caso, como $B \in \mathcal{T} \subset \sigma(X_{k+1}, X_{k+2}, \dots)$, o resultado saí pelo item 1.

   Como $\mathcal{T} \in \sigma(X_1, X_2, \dots)$, temos o resultado.

■

4.3.1 Teorema das Três Séries

Teorema 4.3.3 (Desigualdade Maximal de Kolmogorov)**.** *Suponha que $X_1, \dots, X_n$ são independentes, centradas e com variância finita. Então*

$$\mathbb{P}\left(\max_{1 \le k \le n} |S_k| \ge x\right) \le \frac{\mathbb{V}ar(S_n)}{x^2}.$$

Demonstração. Defina

$$A_k = \{|S_1| < x, |S_2| < x, \dots, |S_k| \ge x\} \subset \Omega,$$

[1] Podemos pensar em $\mathcal{F}'_n$ como o futuro a partir do tempo n.

e note que os conjuntos acima são disjuntos. Além disso, é fácil ver que

$$\left\{\max_{1\leq k\leq n}|S_k|\geq x\right\}\subset\bigsqcup_{1\leq k\leq n}A_k.$$

Logo,

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(S_n)=\mathbb{E}S_n^2 &\geq \sum_{k=1}^n\int_{A_k}S_n^2d\mathbb{P}\\
&=\sum_{k=1}^n\int_{A_k}S_k^2-2S_k(S_k-S_n)+\underbrace{(S_n-S_k)^2}_{\geq 0}d\mathbb{P}\\
&\geq\sum_{k=1}^n\int_{A_k}S_k^2-2S_k(S_k-S_n)d\mathbb{P}\\
&\geq\sum_{k=1}^n\int_{A_k}S_k^2d\mathbb{P}+2\underbrace{\mathbb{E}(1_{A_k}S_k)\mathbb{E}(S_n-S_k)}_{(*)}\\
&\geq\sum_{k=1}^n\int_{A_k}S_k^2d\mathbb{P}\\
&\geq x^2\sum_{k=1}^n\mathbb{P}(A_k)\geq x^2\mathbb{P}\left(\max_{1\leq k\leq n}|S_k|\geq x\right).
\end{aligned}$$

Onde, em $(*)$ utilizamos o fato de que $1_{A_k}S_k\in\sigma(X_1,\ldots,X_k)$ e $S_n-S_k\in\sigma(X_{k+1},\ldots,X_n)$ e portanto são independentes, e também o fato de que $\mathbb{E}(S_n-S_k)=0$.

■

Teorema 4.3.4 (Teorema de uma Série)**.** *Sejam $X_1,\ldots,X_n$ independentes e centradas. Se*

$$\sum_{n=1}^\infty\mathbb{V}ar(X_k)<\infty,$$

então

$$\sum_{n=1}^\infty X_n,\ \textit{converge q.t.p.}.$$

Demonstração. Pelo teorema anterior 4.3.3,

$$\mathbb{P}\left(\max_{M\leq k\leq N}|S_k-S_M|\geq x\right)\leq\frac{\mathbb{V}\mathrm{ar}(S_N)-\mathbb{V}\mathrm{ar}(S_M)}{x^2}=\frac{1}{x^2}\sum_{k=M}^N\mathbb{V}\mathrm{ar}(X_k).$$

Portanto, tomando o limite em N, ficamos com

$$\mathbb{P}\left(\max_{M\leq k}|S_k-S_M|\geq x\right)\leq\frac{1}{x^2}\sum_{k=M}^\infty\mathbb{V}\mathrm{ar}(X_k).$$

e então, é claro que conseguimos deixar o lado direito da expressão acima arbitrariamente pequeno fazendo M arbitrariamente grande.

Note que, para $k, l \geq M$

$$|S_k - S_l| \leq |S_k - S_M| + |S_l - S_M|$$

e portanto

$$\max_{k,l \geq M} |S_k - S_l| \leq 2 \max_{k \geq M} |S_k - S_M|.$$

Temos então que

$$\mathbb{P}\left(\max_{M \leq k,l} |S_k - S_l| \geq 2x\right) \leq \mathbb{P}\left(\max_{M \leq k} |S_k - S_M| \geq x\right) \stackrel{M\to\infty}{\longrightarrow} 0.$$

Temos então, que existe uma subsequência $(M_j)_j$ tal que

$$\max_{M_j \leq k,l} |S_k - S_l| \stackrel{j\to\infty}{\longrightarrow} 0, \text{ q.t.p..}$$

Mas então, dado ω no conjunto de medida total satisfazendo a propriedade acima e $\varepsilon > 0$ qualquer, existe J tal que

$$\max_{M_J \leq k,l} |S_k(\omega) - S_l(\omega)| < \varepsilon,$$

isto é, a sequência $(S_n(\omega))_n$ é de Cauchy para quase todo $\omega \in \Omega$, e portanto temos que a série do enunciado converge para quase todo ponto.

■

Teorema 4.3.5 (Teorema das Três Séries de Kolmogorov)**.** *Sejam* $X_1, \ldots, X_n$ *independentes. Seja* $A > 0$ *e*

$$Y_i = X_i 1_{|X_i| \leq A}.$$

Então a série

$$\sum_{n=1}^{\infty} X_n,$$

converge q.t.p. ***se, e somente se,*** *valem*

1. $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > A) < \infty$*;*
2. $\sum_{n=1}^{\infty} \mathbb{E} Y_n$ *converge;*
3. $\sum_{n=1}^{\infty} \mathbb{V}ar(Y_n) < \infty$*.*

Demonstração. (⇒) Suponha que valem os 1,2 e 3. Pelo teorema anterior e o item 3, temos que

$$\sum_{n=1}^{\infty}(Y_n - \mathbb{E}(Y_n)),$$

converge q.t.p. e portanto, pelo item 2, temos que $\sum_{n=1}^{\infty} Y_n$ também converge q.t.p..

Como

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > A) < \infty,$$

o Primeiro Lema de Borel-Cantelli nos diz que

$$\mathbb{P}(|X_n| > A, \text{ i.o.}) = 0,$$

isto é, $X_n \neq Y_n$ apenas para uma quantidade finita de índices, e como Y_n converge q.t.p., é claro que $\sum_{n=1}^{\infty} X_n$ converge q.t.p..

(⇐) Suponha que a série converge, vamos provar a necessidade de cada item separadamente.

1. É claro que 1 deve valer, já que caso contrário, pelo Segundo Lema de Borel-Cantelli, teríamos que
$$\mathbb{P}(|X_n| > A, \text{ i.o.}) = 1,$$
o que impossibilidade a convergência da série, já que por exemplo o termo geral nunca vai convergir para 0.

2. Suponha que vale o item 1, mas não o item 3. Considere
$$c_n = \sum_{m=1}^{n} \mathbb{V}\text{ar}(Y_m)$$
e também
$$W_{n,m} = \frac{(Y_m - \mathbb{E}Y_m)}{\sqrt{c_n}}.$$
Note que $\mathbb{E}(W_{n,m}) = 0$, e também que
$$\sum_{m=1}^{n} \mathbb{E}(W_{n,m}^2) = 1, \forall n.$$
Além disso, dado $\varepsilon > 0$
$$\lim_{n\to\infty} \sum_{m=1}^{n} \mathbb{E}(|W_{n,m}|^2, |W_{n,m}| > \varepsilon) = 0$$
já que, $|Y_m - \mathbb{E}Y_m| \leq 2A$, e portanto, se n é grande o suficiente,
$$\frac{2A}{\sqrt{c_n}} < \varepsilon$$

e a soma então é nula. Logo, pelo Teorema Central do Limite 3.4.1, temos que

$$\sum_{m=1}^{n} \frac{(Y_m - \mathbb{E}Y_m)}{\sqrt{c_n}} \longrightarrow_d Z,$$

onde Z tem distribuição normal com média zero e variância 1.

Pelo item 1, temos que

$$\mathbb{P}(X_n \neq Y_n \text{ i.o.}) = 0$$

e portanto, como estamos supondo que $\sum X_n$ converge q.t.p. é claro que $\sum Y_n$ converge q.t.p.. Logo

$$\sum_{m=1}^{n} \frac{Y_m}{\sqrt{c_n}} \longrightarrow_d 0.$$

Por Slutsky 2.2.6, temos que

$$-\sum_{m=1}^{n} \frac{\mathbb{E}Y_m}{\sqrt{c_n}} = \sum_{m=1}^{n} \frac{(Y_m - \mathbb{E}Y_m)}{\sqrt{c_n}} - \sum_{m=1}^{n} \frac{Y_m}{\sqrt{c_n}} \longrightarrow_d Z + 0 = Z,$$

o que é um absurdo, já que o lado esquerdo é determinístico e o lado direito não.

3. Sabemos então que os itens 1 e 2 valem. Pelo teorema de uma Série 4.3.4 e o item 2, temos que

$$\sum_{m=1}^{\infty} (Y_m - \mathbb{E}Y_m)$$

converge q.t.p.. Como estamos supondo que $\sum X_n$ converge q.t.p., pelo item 1, temos que $\sum Y_m$ também converge q.t.p. e daí basta subtrairmos a duas séries.

■

Teorema 4.3.6 (Lema de Cesaro). *Se* $\sum_{m=1}^{n} a_m \xrightarrow{n\to\infty} \infty$ *e* $x_n \to L$ *então*

$$\frac{\sum_{m=1}^{n} a_m x_m}{\sum_{m=1}^{n} a_m} \xrightarrow{n\to\infty} L.$$

Demonstração. Basta notar que para $K < n$ suficientemente grande e $\varepsilon > 0$,

$$\begin{aligned}
\left| \frac{\sum_{m=1}^{n} a_m x_m}{\sum_{m=1}^{n} a_m} - L \right| &= \left| \frac{\sum_{m=1}^{n} a_m x_m - L\sum_{m=1}^{n} a_m}{\sum_{m=1}^{n} a_m} \right| \\
&= \left| \frac{\sum_{m=1}^{n} a_m (x_m - L)}{\sum_{m=1}^{n} a_m} \right| \\
&\leq \left| \frac{\sum_{m=1}^{K} a_m (x_m - L)}{\sum_{m=1}^{n} a_m} \right| + \left| \frac{\sum_{m=K}^{n} a_m (x_m - L)}{\sum_{m=1}^{n} a_m} \right| \\
&\approx \left| \frac{\sum_{m=1}^{K} a_m (x_m - L)}{\sum_{m=1}^{n} a_m} \right| + \left| \frac{\sum_{m=K}^{n} a_m \varepsilon}{\sum_{m=1}^{n} a_m} \right| \\
&\xrightarrow{n\to\infty} 0 + \varepsilon.
\end{aligned}$$

■

Teorema 4.3.7 (Kronecker). *Se $a_n \uparrow \infty$ e $\sum_{n=1}^{\infty} \frac{x_n}{a_n}$ converge, então*

$$\frac{1}{a_n}\sum_{m=1}^{\infty} x_m \xrightarrow{n\to\infty} 0.$$

Demonstração. Seja $b_m = \sum^m x_k/a_k$ e note que

$$x_m = a_m(b_m - b_{m-1}).$$

Então

$$\begin{aligned}\frac{1}{a_n}\sum_{m=1}^{n} x_m &= \frac{1}{a_n}\left\{a_n b_n + \sum_{m=2}^{n} a_{m-1}b_{m-1} - \sum_{m=1}^{n} a_m b_{m-1}\right\}\\ b_n &- \frac{\sum_{m=1}^{n}(a_m - a_{m-1})b_{m-1}}{\sum_{m=1}^{n}(a_m - a_{m-1})} \xrightarrow{n\to\infty} 0\end{aligned}$$

onde o limite acima é verdade pelo Lema de Cesaro. ■

Teorema 4.3.8 (Lei Forte dos Grandes Números). *Sejam $X_1, \dots X_n$ v.a's i.i.d. com $\mathbb{E}|X_1| < \infty$. Se $\mathbb{E}X_1 = \mu$, então*

$$\frac{S_n}{n} \xrightarrow{n\to\infty} \mu, \ \ q.t.p..$$

Demonstração. Considere $Y_k = X_k 1_{|X_k|\le k}$ e note que

$$\sum_{k}^{\infty} \mathbb{P}(|X_k| > k) \le \int_0^{\infty} \mathbb{P}(|X_1| > t)dt = \mathbb{E}|X_1| < \infty$$

e portanto, por B.C., temos que

$$\mathbb{P}(X_k \neq Y_k, \ i.o.) = 0.$$

Então basta provarmos o teorema para o caso truncado, isto é, podemos trocar S_n por

$$T_n = Y_1 + \cdots + Y_n.$$

Antes de continuarmos com a prova, precisamos de um lema técnica:

Lema 4.3.9.

$$\sum_{k=1}^{\infty} \mathbb{V}ar(Y_k)/k^2 \le 4\mathbb{E}|X_1| < \infty.$$

Prova do Lema. Note que

$$\mathbb{V}\mathrm{ar}(Y_k) \le \mathbb{E}Y_k^2 = \int_0^{\infty} 2t\mathbb{P}(|Y_k| > t)dt \le \int_0^{k} \mathbb{P}(|X_1| > t)dt.$$

Como o termo acima é todo positivo, por Fubini, temos que

$$\sum_{k=1}^{\infty} \mathbb{E}(Y_k^2)/k^2 \leq \sum_{k=1}^{\infty} k^{-2} \int_0^{\infty} 2t 1_{t<k} \mathbb{P}(|X_1| > t) dt$$
$$\int_0^{\infty} \sum_{k=1}^{\infty} k^{-2} 2t 1_{t<k} \mathbb{P}(|X_1| > t) dt$$

Como

$$\sum_{k \geq t} \frac{1}{k^2} \leq \sum_{k \geq t} \frac{2}{k(k+1)} = \sum_{k \geq t} \frac{2}{k} - \frac{2}{k+1} \leq \frac{2}{t},$$

temos o resultado, já que $\mathbb{E}|X_1| = \int_0^{\infty} \mathbb{P}(|X_1| > t) dt$. ■

Defina agora $Z_k = Y_k - \mathbb{E}Y_k$. É claro que

$$\mathbb{V}\text{ar} Z_k = \mathbb{V}\text{ar} Y_k \leq \mathbb{E} Y_k^2$$

e portanto, pelo lema anterior, temos que

$$\sum_{k=1}^{\infty} \mathbb{V}\text{ar}(Z_k)/k^2 < \infty.$$

Pelo Teorema de uma Série, temos que

$$\sum_{k=1}^{\infty} Z_k/k$$

converge q.t.p.. E portanto, pelo teorema de Kronecker, temos que

$$n^{-1} \sum_{k=1}^{n} Z_k = n^{-1} \sum_{k=1}^{n} (Y_k - \mathbb{E}Y_k) \xrightarrow{n \to \infty} 0, \ q.t.p.,$$

isto é,

$$\frac{T_n}{n} - \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}Y_k \xrightarrow{n \to \infty} 0, \ q.t.p..$$

Pelo Teorema da Convergência Dominada, temos que $\mathbb{E}Y_n \xrightarrow{n \to \infty} \mu$. Considerando então $a_n = 1$, temos que $\sum a_n = \infty$ e portanto, pelo teorema de Cesaro

$$\frac{\sum_{k=1}^{n} 1 \cdot \mathbb{E}Y_k}{\sum^n 1} = \frac{\sum_{k=1}^{n} \mathbb{E}Y_k}{n} \xrightarrow{n \to \infty} \mu,$$

e assim, concluímos o teorema. ■

4.4 Exercícios Resolvidos

Capítulo 5

Martingales

As Principais referências usadas foram:

1. [Dur10]

2. [Wil91].

5.1 Esperança Condicional

Definição 5.1.1 (Esperança Condicional)**.** Seja $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ e $\mathcal{G}$ uma sub σ-álgebra de $\mathcal{F}$. Definimos a esperança condicional de X em $\mathcal{G}$, denotada por $\mathbb{E}(X \mid G)$, como sendo a única v.a. Z tal que

1. Z é $\mathcal{G}$ mensurável.

2. Temos que
$$\mathbb{E}(XY) = \mathbb{E}(ZY) \Leftrightarrow \mathbb{E}((X - Z)Y) = 0$$
para toda Y função característica de $\mathcal{G}$.

Observação 5.1.2. Note que se $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ e $Z = \mathbb{E}(X|\mathcal{G})$, então a Definição 5.1.1 diz que $\forall\ Y \in \mathcal{G}$:

$$\begin{array}{lcll} \mathbb{E}(XY) & = & \mathbb{E}(Z, Y) & \Leftrightarrow \\ \mathbb{E}((X - Z)Y) & = & 0 & \Leftrightarrow \\ \langle X - Z, Y\rangle_{L^2} & = & 0, & \end{array}$$

isto é, estamos projetando X ortogonalmente em $\mathcal{G}$. No caso geral, precisamos nos restringir às funções características, já que X não precisa ser quadrado integrável e daí não podemos falar em projeções ortogonais.
Note que poderíamos trocar Y de funções características para funções em $C_c^\infty(\Omega)$ que são quadradas integráveis e densas em L^1 e conseguiríamos estender a definição do caso L^2 para o caso geral por um argumento de aproximação.

Observação 5.1.3. Outra forma de enxergar a esperança condicional de uma v.a. X com relação à uma σ-álgebra $\mathcal{G}$ (e que de certo modo caminha na mesma direção da observação anterior) é que $\mathbb{E}(X|\mathcal{G})$ é a melhor aproximação de X dado informações extras advindas de $\mathcal{G}$.

Antes de continuarmos, precisamos checar alguns pontos da Definição 5.1.1.

Lema 5.1.4 (A definição faz sentido). *Se existe Z como em 5.1.1, então Z é integrável.*

Demonstração. Precisamos mostrar que $\mathbb{E}(|Z|) < \infty$. Seja $A = \{Z \geq 0\}$, então

$$\int_A Z d\mathbb{P} = \int_A X d\mathbb{P} \leq \int_A |X| d\mathbb{P}$$
$$-\int_{A^c} Z d\mathbb{P} = -\int_{A^c} X d\mathbb{P} \leq \int_{A^c} |X| d\mathbb{P},$$

e somando as duas desigualdades temos o resultado. ■

Lema 5.1.5 (Existência). *Existe Z como definido em 5.1.1.*

Demonstração. Vamos supor que $X \geq 0$, sem perda de generalidade. Defina a medida em $\mathcal{G}$

$$\mu(A) = \mathbb{E}(X1_A),$$

e perceba que esta é absolutamente contínua com relação à $\mathbb{P}$ restrita à $\mathcal{G}$. Logo, por Radon-Nikodym, existe $Z \geq 0$ tal que

$$\mu(A) = \mathbb{E}(X1_A) = \mathbb{E}(Z1_A).$$

■

Lema 5.1.6 (Unicidade). *Se $\hat{Z}$ é outra v.a. satisfazendo 5.1.1 então $Z = \hat{Z}$ (q.t.p.).*

Demonstração. Seja $A_\varepsilon = \{Z - \hat{Z} \geq \varepsilon > 0\}$. E note que

$$0 = \int_{A_\varepsilon} X - X d\mathbb{P} = \int_{A_\varepsilon} Z - \hat{Z} d\mathbb{P} \geq \varepsilon \mathbb{P}(A_\varepsilon),$$

concluindo que $\mathbb{P}(A_e) = 0$ para todo $\varepsilon > 0$. Daí é fácil concluir o resultado trocando os sinais em A_ε ■

Agora que provamos que nossa definição faz sentido, estamos prontos para provar alguns resultados práticos sobre esperança condicional.

Proposição 5.1.7 (Propriedades Básicas). *Valem os seguintes resultados:*

1. *Se $X \in \mathcal{G}$ então $\mathbb{E}(X|\mathcal{G}) = X$.*
2. *Se X é independente de $\mathcal{B}$ então $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$.*

3. *Se* $\alpha \in \mathbb{R}$ *então* $\mathbb{E}(X + \alpha Y|\mathcal{G}) = \mathbb{E}(X|\mathcal{G}) + \alpha\mathbb{E}(Y|\mathcal{G})$.

4. *Se* $X \leq Y$ *então* $\mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(Y|\mathcal{G})$.

5. *Se* $X_n \geq 0$ *e* $X_n \uparrow X$*, com* $\mathbb{E}X < \infty$*, então*
$$\mathbb{E}(X_n|\mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G})$$

6. *Vale a desigualdade de Jensen, isto é, se* $\varphi : \mathbb{R} \to \mathbb{R}$ *é convexa então*
$$\mathbb{E}(\varphi(X)|\mathcal{G}) \geq \varphi(\mathbb{E}(X|\mathcal{G})).$$

7. *Se* $\mathcal{H} \subset \mathcal{G}$ *então*
$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{H}).$$

8. *Se* $\mathcal{H} \subset \mathcal{G}$ *então*
$$\mathbb{E}(\mathbb{E}(X|\mathcal{H})|\mathcal{G}) = E(X|\mathcal{H}).$$

9. *Suponha que* $X \in \mathcal{G}$*, então*
$$\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G}),$$
se todas as integrais envolvendo X *e* Y *fizerem sentido.*

Observação 5.1.8. Os itens 1 e 2 coincide com nossa intuição sobre projeções ortogonais/melhor aproximação dada anteriormente. Os itens 3,4,5 e 6 nos dizem que o operador $\mathbb{E}(\cdot|\mathcal{G})$ se comporta de forma parecida com o operador $\mathbb{E}(\cdot)$. Os itens 8 e 9 nos dizem que se criarmos uma cadeia de esperanças condicionais, a que prevalece no final será a menor. Esse fato é bem útil, já que sempre temos a seguinte relação:
$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|\mathcal{F})).$$

Demonstração. Vamos provar os itens separadamente:

1. Pela Definição 5.1.1 esse resultado é óbvio, já que X é $\mathcal{G}$ mensurável e $\mathbb{E}(XY) = \mathbb{E}(XY)$ para qualquer Y. Intuitivamente, estamos projetando ortogonalmente um vetor X em seu próprio espaço vetorial e então $\mathbb{E}(X|\mathcal{G}) = X$.

2. No caso em que $X = 1_B$ temos que
$$\int_A X d\mathbb{P} = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \int_A p(B) d\mathbb{P} = \int_A \mathbb{E}(X) d\mathbb{P}.$$
E no caso X qualquer o argumento saí pelo argumento canônico de aproximação por funções simples.

3. Basta verificar o item *2* de 5.1.1. Note que para $A \in \mathcal{G}$,

$$\int_A \mathbb{E}(X + \alpha Y|\mathcal{G})d\mathbb{P} = \int_A X + \alpha Y d\mathbb{P} \tag{5.1.1}$$
$$= \int_A X d\mathbb{P} + \int_A \alpha Y d\mathbb{P} \tag{5.1.2}$$
$$= \int_A \mathbb{E}(X)|\mathcal{G})d\mathbb{P} + \alpha \int_A \mathbb{E}(Y|\mathcal{G})d\mathbb{P} \tag{5.1.3}$$
$$= \int_A \mathbb{E}(X)|\mathcal{G}) + \alpha\mathbb{E}(Y|\mathcal{G})d\mathbb{P} \tag{5.1.4}$$
$$\tag{5.1.5}$$

   Logo, por unicidade, vale o resultado.

4. Seja $A = \{\mathbb{E}(X|\mathcal{G}) - \mathbb{E}(Y|\mathcal{G}) \geq \varepsilon > 0\}$, então

$$0 \geq \int X - Y d\mathbb{P} = \int_A \mathbb{E}(X|\mathcal{G}) - \mathbb{E}(Y|\mathcal{G})d\mathbb{P} \geq \varepsilon\mathbb{P}(A)$$

   concluindo que $\mathbb{P}(A) = 0$, para todo $\varepsilon > 0$.

5. Como $X_n \leq X_{n+1} \leq X$, temos que

$$\mathbb{E}(X_n|\mathcal{G}) \leq \mathbb{E}(X_{n+1}|\mathcal{G}) \leq \mathbb{E}(X|\mathcal{G}).$$

   Seja Y o limite de da sequência $\mathbb{E}(X_n|\mathcal{G})$. Como $\mathbb{E}(X|\mathcal{G})$ é integrável, pelo Teorema da Convergência Dominada, temos que para $A \in \mathcal{G}$,

$$\begin{aligned}\int_A Y d\mathbb{P} &= \int_A \lim \mathbb{E}(X_n|\mathcal{G})d\mathbb{P}\\ &= \lim \int_A \mathbb{E}(X_n|\mathcal{G})d\mathbb{P}\\ &= \lim \int_A X_n d\mathbb{P}\\ &= \int_A X d\mathbb{P} = \int_A \mathbb{E}(X|\mathcal{G})d\mathbb{P}.\end{aligned}$$

   Logo, como a expressão acima vale para todo $A \in \mathcal{G}$, temos que

$$Y = \lim \mathbb{E}(X_n|\mathcal{G}) = \mathbb{E}(X|\mathcal{G}).$$

6. Um resultado clássico de análise convexa é que, se φ é convexa, então

$$\varphi(x) = \sup\{\ l(x) = ax + b \mid l(x) \leq \varphi(x)\}.$$

   Se $l(x) = ax+b \leq \varphi(x)$, então pelo resultado de linearidade e monotonicidade da condicional, temos que

$$\mathbb{E}(\varphi(X)|\mathcal{G}) \geq \mathbb{E}(aX + b|\mathcal{G}) = l(E(X|\mathcal{G})),$$

   portanto, basta tomar o sup sobre as funções l, e temos o resultado.

7. Seja $A \in \mathcal{H} \subset \mathcal{G}$, então, pela definição de condicional
$$\begin{aligned}\int_A \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H})d\mathbb{P} &= \int_A \mathbb{E}(X|\mathcal{G})d\mathbb{P}\\ &= \int_A X d\mathbb{P},\end{aligned}$$
isto é, tanto $\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H})$ quanto $\mathbb{E}(X|\mathcal{H})$ são $\mathcal{H}$ mensuráveis e suas integrais coincidem em todos os mensuráveis de $\mathcal{H}$, logo por unicidade da condicional, temos que vale o resultado.

8. Como $\mathcal{H} \subset \mathcal{G}$, temos que $E(X|\mathcal{H})$ é $\mathcal{G}$ mensurável. E então, temos o resultado pelo item 1 dessa proposição.

9. Suponha que $X = 1_A$, com $A \in \mathcal{G}$. Então para $B \in \mathcal{G}$,
$$\begin{aligned}\int_B \mathbb{E}(XY|\mathcal{G})d\mathbb{P} &= \int_B XY d\mathbb{P}\\ &= \int_{B\cap A} Y d\mathbb{P}\\ &= \int_{B\cap A} \mathbb{E}(Y|\mathcal{G})d\mathbb{P}\\ &= \int_B X\mathbb{E}(Y|\mathcal{G})d\mathbb{P},\end{aligned}$$
e portanto vale o resultado.

■

Proposição 5.1.9 (Esperança Condicional como Projeção Ortogonal)**.** *Suponha que* $\mathbb{E}X^2 < \infty$*, então* $\mathbb{E}(X|\mathcal{G})$ *é a variável* $Y \in \mathcal{G}$ *que minimiza o erro quadrático* $\mathbb{E}(X-Y)^2$.

Demonstração. Sabemos que $H := L^2(\Omega, \mathcal{F}, \mathbb{P})$ é Hilbert e que nesse caso, $\|X-Y\|_2$ é minimizado quando $Y = P_H(X)$, onde P_H é a projeção ortogonal em H. ALém disso, Y é único já que H é um subespaço fechado.

Portanto, só precisamos mostrar que $Y = \mathbb{E}(X|\mathcal{G})$, então $\langle X-Y, Z\rangle_2 = 0$ para todo $Z \in \mathcal{G}$. Mas isso é claro já que,
$$\begin{aligned}\langle X-Y, Z\rangle_2 &= \int (X-Y)Z d\mathbb{P}\\ &= \int XZ - YZ d\mathbb{P}\\ &= \int XZ - \mathbb{E}(X|\mathcal{G})Z d\mathbb{P}\\ &= \int XZ - \mathbb{E}(ZX|\mathcal{G})d\mathbb{P}\\ &= \int XZ - ZX d\mathbb{P} = 0,\end{aligned}$$
onde usamos o fato de que como $Z \in \mathcal{G}$ então $\mathbb{E}(XZ|\mathcal{G}) = Z\mathbb{E}(X|\mathcal{G})$. ■

Definição 5.1.10 (Independência Condicional)**.** Dizemos que X, Y são independentes condicionalmente à $\mathcal{F}$ se

$$\mathbb{P}(X \in A, Y \in B|\mathcal{F}) = \mathbb{P}(X \in A|\mathcal{F})\mathbb{P}(Y \in B|\mathcal{F})$$

para quaisquer, A, B mensuráveis.

Proposição 5.1.11. *Suponha que X, Y são independentes condicionalmente à $\mathcal{F}$. Então*

$$\mathbb{E}(XY|\mathcal{F}) = \mathbb{E}(X|\mathcal{F})\mathbb{E}(Y|\mathcal{F})$$

Demonstração. Seja $F \in \mathcal{F}$ e suponha que $X = 1_{X \in A}$ e $Y = 1_{Y \in B}$, então

$$\begin{aligned}\mathbb{E}(XY|F) &= \mathbb{E}(1_{X\in A}1_{Y\in B}|\mathcal{F})\\ &= \mathbb{E}(1_{X\in A, Y\in B}|\mathcal{F})\\ &= \mathbb{P}(X \in A, Y \in B|\mathcal{F})\\ &= \mathbb{P}(X \in A|\mathcal{F})\mathbb{P}(Y \in B|\mathcal{F})\\ &= \mathbb{E}(X|F)\mathbb{E}(Y|F).\end{aligned}$$

E o resultado geral sai pelo argumento usual de aproximação por funções simples. ∎

Exemplo 5.1.12. Suponha que X, Y tem densidade conjunta $f_{X,Y}(x, y)$. Então

$$\mathbb{E}(g(X)|Y) = h(Y),$$

onde h satisfaz:

$$h(y)\int f_{X,Y}(x, y)dx = \int g(x)f_{X,Y}(x, y)dx.$$

Demonstração. Perceba que $h(Y) \in \sigma(Y)$, logo só falta mostrar que as integrais coincidem para todos os eventos da forma $\{Y \in A\} \in \sigma(Y)$, com $A \in \mathcal{B}(\mathbb{R})$. Mas note que

$$\begin{aligned}\mathbb{E}(g(X)1_{Y\in A}) &= \int_A\int g(x)f_{X,Y}(x, y)dxdy\\ &= \int_A\left[\int g(x)f_{X,Y}(x, y)dx\right]dy\\ &= \int_A h(y)\int f_{X,Y}(x, y)dxdy\\ &= \int_A\int h(y)f_{X,Y}(x, y)dxdy\\ &= \mathbb{E}(h(Y)1_{Y\in A}).\end{aligned}$$

∎

Se supormos que $f_{X,Y} > 0$, então

$$h(y) = \frac{\int g(x) f_{X,Y}(x,y)dx}{\int f_{X,Y}(x,y)dx}.$$

Para lembrarmos dessa fórmula, basta lembrarmos do caso discreto, isto é:

$$\begin{aligned}
\mathbb{E}(g(X)|Y) &= \sum_x g(x)\mathbb{P}(X = x|Y = y) \\
&= \frac{\sum_x g(x)\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} \\
&= \frac{\sum_x g(x)\mathbb{P}(X = x, Y = y)}{\sum_x \mathbb{P}(X = x, Y = y)} \\
&= \frac{\sum_x g(x) f_{X,Y}(x,y)}{\sum_x f_{X,Y}(x,y)} \\
\text{``} &= \text{''} \frac{\int g(x) f_{X,Y}(x,y)dx}{\int f_{X,Y}(x,y)dx}
\end{aligned}$$

Exemplo 5.1.13. Suponha X_1 e X_2 v.a. independentes, então

$$\mathbb{E}(g(X_1, X_2)|X_1) = h(X_1),$$

onde

$$h(x) = \mathbb{E}(g(x, X_2)).$$

Demonstração. Da mesma forma que no exemplo anterior, vamos mostrar que as esperanças coincidem para eventos da forma $\{X_1 \in A\}$. Note que

$$\begin{aligned}
\mathbb{E}(g(X_1, X_2)1_{X_1 \in A}) &= \int_A \int g(x_1, x_2) f_{X_1,X_2}(x_1, x_2)dx_2 dx_1 \\
&= \int_A \int g(x_1, x_2) f_{X_2}(x_2)dx_2 f_{X_1}(x_1)dx_1 \\
&= \int_A \mathbb{E}(g(x_1, X_2)) f_{X_1}(x_1)dx_1 \\
&= \int_A h(x_1) f_{X_1}(x_1)dx_1 \\
&= \mathbb{E}(h(X_1)1_{X_1 \in A}).
\end{aligned}$$

■

Mostramos que se $X_1, \ldots, X_n$ são independentes, então

$$\mathbb{E}(g(X_1, \ldots, X_n)|X_1, \ldots, X_i)(\omega) = \mathbb{E}(g(X_1(\omega), \ldots, X_i(\omega), X_{i+1}, \ldots, X_n)),$$

isto é, condicionar em $X_1, \ldots, X_i$ nos permite fixar valores na integral nas coordenadas $1, 2, \ldots, i$.

Exemplo 5.1.14. Seja $(X_i)_{i\in\mathbb{N}}$ uma sequência de v.a.s i.i.d. com suporte em $\mathbb{N}$. Defina

$$S_N = X_1 + \cdots + X_N,$$

onde $N = X_1+1$, o nosso objetivo é calcular a FGdM de S_N. Note que não podemos apenas tomar o produto das FGdM, já que S_N depende de X_1. Entretanto, temos que

$$\begin{aligned}
\mathbb{E}(e^{tS_N}) &= \mathbb{E}(\mathbb{E}(e^{tS_n}|X_1)) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(X_1 = k)\mathbb{E}(e^{t(k+X_2+\cdots+X_{k+1})}) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(X_1 = k)e^{tk}\mathbb{E}(e^{t(X_2+\cdots+X_{k+1})}) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(X_1 = k)e^{tk}e^{k\ln(M_{X_1}(t))}) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(X_1 = k)e^{k(t+\ln(M_{X_1}(t)))}) \\
&= M_{X_1}(t + \ln(M_{X_1}(t))).
\end{aligned}$$

5.2 Martingales

5.2.1 Fatos Básicos

Definição 5.2.1 (Martingale)**.** Seja $(\mathcal{F}_n)$ uma filtração. Se (X_n) é um processo adaptado à filtração $(\mathcal{F}_n)$, tal que $\mathbb{E}|X_n| < \infty$ então

- Se $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n, \forall n$ então (X_n) é um **Martingale**.
- Se $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq X_n, \forall n$ então (X_n) é um **Supermartingale**.
- Se $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n, \forall n$ então (X_n) é um **Submartingale**.

Observação 5.2.2. Ao longo do texto, a menos que haja alguma menção contrária, vamos assumir que

$$\mathcal{F}_n = \sigma(X_1, \dots, X_n).$$

Note que essa filtração é uma escolha natural para o nosso problema, já que ela é a menor filtração que deixa (X_n) um processo adaptado. Isso implica, por exemplo, que se (X_n) é um Martingale com respeito à uma filtração $(\mathcal{G}_n)$ tal que $\mathcal{F}_n \subset \mathcal{G}_n$ então (X_n) também é um Martingale adaptado à $(\mathcal{F}_n)$.

Proposição 5.2.3. *Suponha que (X_n) é um Martingale. Se $n > m$ então*

$$\mathbb{E}(X_n|\mathcal{F}_n) = X_m.$$

Além disso, o resultado vale, tomando as devidas alterações, para submartingales e supermartingales.

Demonstração. Note que,

$$\mathbb{E}(X_n|\mathcal{F}_m) = \mathbb{E}(\mathbb{E}(X_n|\mathcal{F}_{n-1})|\mathcal{F}_m) = \mathbb{E}(X_{n-1}|F_m),$$

e daí o resultado sai por indução e a prova para os outros casos é análoga. ∎

Proposição 5.2.4. *Seja (X_n) um martingale, então para todo n*

$$\mathbb{E}(X_n) = \mathbb{E}(X_1).$$

E o mesmo vale para supermartingales e submartingales, fazendo as devidas alterações.

Demonstração. Note que

$$\mathbb{E}(X_n|\mathcal{F}_1) = X_1,$$

e portanto

$$\mathbb{E}(X_n) = \mathbb{E}\mathbb{E}(X_n|\mathcal{F}_1) = \mathbb{E}(X_1).$$

∎

Isto é, num jogo neutro, o lucro esperado no final de n apostas é o mesmo do lucro que você começou; num jogo desfavorável, o lucro esperado sempre diminui; num jogo favorável, o lucro esperado sempre aumenta.

Proposição 5.2.5. *Se dois processos são* ***adaptados à mesma filtração*** $\mathcal{G}_n$*, então as afirmações abaixo são verdadeiras:*

1. *(Martingale) + (Martingale) = (Martingale).*
2. *(Martingale) + (Subartingale) = (Subartingale).*
3. *(Martingale) + (Supermartingale) = (Superartingale).*

Demonstração. Trivial. ∎

Lema 5.2.6. *Se* (X_n) *é um submartingale, então* $(-X_n)$ *é um supermartingale (e vice-versa).*

Demonstração. Como $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$, temos que

$$\mathbb{E}(-X_{n+1}|\mathcal{F}_n) = -\mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq -X_n.$$

∎

5.2.2 Alguns Exemplos

Exemplo 5.2.7 (Random Walk Simples - Aposta Honesta)**.** Seja (X_n) uma sequência i.i.d. tal que $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ e considere

$$S_n = X_1 + \cdots + X_n.$$

Vamos mostrar que S_n é um Martingale com respeito à σ-álgebra

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n).$$

De fato, note que

$$\begin{aligned}\mathbb{E}(S_{n+1}|\mathcal{F}_n) &= \mathbb{E}(X_1 + X_2 + \cdots + X_n + X_{n+1}|\mathcal{F}_n)\\ &= X_1 + X_2 + \cdots + X_n + \mathbb{E}(X_{n+1}|\mathcal{F}_n)\\ &= X_1 + X_2 + \cdots + X_n + \mathbb{E}(X_{n+1})\\ &= X_1 + X_2 + \cdots + X_n + 0\\ &= S_n.\end{aligned}$$

Observação 5.2.8. Seja (X_n) uma sequência de v.a.'s tal que X_i representa o ganho na i-ésima aposta num jogo de apostas. Se (X_n) é um Martingale, então estamos dizendo que o o melhor chute do ganho esperado na $(n+1)$-ésima jogada é X_n, ou seja, é esperado que seu lucro seja o mesmo em todas as rodadas.
Da mesma forma, se (X_n) é um supermartingale, então o ganho esperado na $(n+1)$-ésima jogada é menor que o valor na jogada anterior, ou seja, o esperado é que percamos dinheiro neste jogo. Como descrito em [Dur10], ao contrário do esperado, um supermartingale não tem nada de 'super'.

Exemplo 5.2.9. Seja (X_n) uma sequência de v.a.'s independentes e não negativas, com

$$\mathbb{E}(X_i) = 1, \ \forall i \in \mathbb{N}.$$

Tomando $(\mathcal{F}_n)$ a filtração natural, e

$$Y_n = X_1 X_2 \dots X_n,$$

temos que Y_n é um martingale, já que, por independência

$$\mathbb{E}(Y_n|\mathcal{F}_{n-1}) = \mathbb{E}(Y_n|\mathcal{F}_{n-1}) = Y_{n-1}\mathbb{E}(X_n|\mathcal{F}_{n-1}) = Y_{n-1}\mathbb{E}(X_n) = Y_{n-1}.$$

Exemplo 5.2.10. Seja $(\mathcal{F}_n)$ uma filtração e $X \in L^1(\Omega, \mathcal{F}_\infty, \mathbb{P})$. Então

$$Y_n = \mathbb{E}(X|\mathcal{F}_n)$$

é um Martingale, já que

$$\mathbb{E}(Y_n|\mathcal{F}_{n-1}) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_n)|\mathcal{F}_{n-1}) = \mathbb{E}(X|\mathcal{F}_{n-1}) = Y_{n-1}.$$

5.2.3 Martingales e Funções Convexas

Teorema 5.2.11. *Seja (X_n) um Martingale e φ uma função convexa com $\mathbb{E}(|\varphi(X)|) < \infty$. Então $(\varphi(X_n))$ é um submartingale. Consequentemente, se $p \geq 1$ e $E|\varphi(X)|^p < \infty$ então $(|X_n|^p)$ é um submartingale.*

Demonstração. Note que, por Jensen condicional,

$$\mathbb{E}(\varphi(X_{n+1})|\mathcal{F}_n) \geq \varphi(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) = \varphi(X_n).$$

Além disso, é claro que $x \mapsto |x|^p$ é convexa se $p \geq 1$. ∎

Teorema 5.2.12. *Seja (X_n) um submartingale e φ uma função convexa crescente com $\mathbb{E}(|\varphi(X)|) < \infty$. Então $(\varphi(X_n))$ é um submartingale. Consequentemente, se (X_n) é um submartingale, então $(X_n - a)^+$ é um submartingale e se (X_n) é um supermartingale, então $X_n \wedge a$ é um supermartingale.*

Demonstração. Note que, por Jensen condicional,

$$\mathbb{E}(\varphi(X_{n+1})|\mathcal{F}_n) \geq \varphi(\mathbb{E}(X_{n+1}|\mathcal{F}_n)).$$

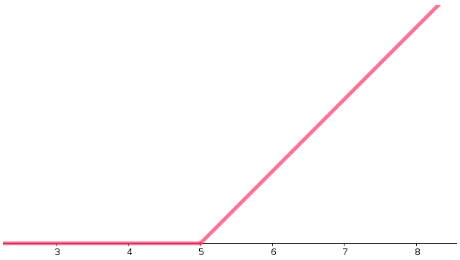
Como (X_n) é submartingale,

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n,$$

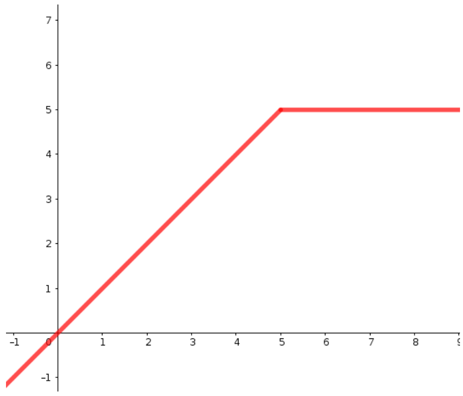
e então, como φ é crescente, temos o resultado.

É fácil ver que $x \mapsto (x-a)^+$ é convexa e crescente e que $x \mapsto -(x \wedge a)$ é convexa e crescente. Abaixo exibimos imagens no caso em que $a = 5$.

∎



(a) $x \mapsto (x-5)^+$



(b) $x \mapsto x \wedge 5$

5.2.4 Sequências Previsíveis

Definição 5.2.13 (Sequência Previsível)**.** Seja $(\mathcal{F}_n)$ uma filtração, dizemos que a sequência (H_n) é previsível com relação à $(\mathcal{F}_n)$ se $H_n \in \mathcal{F}_{n-1}$. Além disso, vamos definir

$$(H,X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1}).$$

Observação 5.2.14. Note que podemos pensar em

$$(H,X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1})$$

como o ganho esperado na n-ésima rodada num jogo de aposta, sendo que temos uma estratégia H_n para a rodada n que depende apenas das rodadas anteriores. O que o teorema acima quer dizer, intuitivamente, é que num jogo desfavorável, independente da estratégia utilizada, o jogo sempre será desfavorável.

Teorema 5.2.15. *Seja X_n um supermartingale, se $H_n \geq 0$ é previsível e cada H_n é limitado, então $(H,X)_n$ é um supermartingale.*

Demonstração. Basta notar que, como $H_n \geq 0$,

$$\begin{aligned}
(H,X)_{n+1} &= \mathbb{E}\left(\sum_{m=1}^{n+1} H_m(X_m - X_{m-1})\,|\mathcal{F}_n\right)\\
&= \left(\sum_{m=1}^{n} H_m(X_m - X_{m-1})\right) + \mathbb{E}(H_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n)\\
&= (H,X)_n + \mathbb{E}(H_{n+1}X_{n+1}|\mathcal{F}_n) - \mathbb{E}(H_{n+1}X_n|\mathcal{F}_n)\\
&= (H,X)_n + H_{n+1}\mathbb{E}(X_{n+1}|\mathcal{F}_n) - H_{n+1}\mathbb{E}(X_n|\mathcal{F}_n)\\
&= (H,X)_n + H_{n+1}\mathbb{E}(X_{n+1}|\mathcal{F}_n) - H_{n+1}\mathbb{E}(X_n|\mathcal{F}_n)\\
&\leq (H,X)_n + H_{n+1}X_n - H_{n+1}X_n = 0.
\end{aligned}$$

E como H é limitado, todos os passos anteriores fazem sentido.

■

Observação 5.2.16. Note que **a mesma demonstração funciona para Martingales e submartingales**, sendo que no primeiro caso não precisamos que $H_n \geq 0$. Isso quer dizer que independentes da estratégia, um jogo neutro continua neutro e um jogo vantajoso continua vantajoso.

5.2.5 Tempo de Parada

Definição 5.2.17 (Tempo de Parada)**.** Seja $\mathcal{F}_n$ a filtração natural dada a sequência (X_n) de v.a.'s. Uma variável aleatório N com valores em $\mathbb{N} \cup \{\infty\}$ é um tempo de parada se para todo $n < \infty$, temos que

$$\{N = n\} \in \mathcal{F}_n.$$

Observação 5.2.18. Intuitivamente, se N é uma v.a. que representa o momento em que um apostador deve parar de apostar, temos que $N \in \mathcal{F}_n$, já que essa decisão só depende do resultado da n-ésima rodada.

Lema 5.2.19. *Seja N um tempo de parada, então*

$$H_n = 1_{n \leq N}$$

é previsível.

Demonstração. Note que

$$\{N \leq n - 1\} \in \mathcal{F}_{n-1},$$

já que N é tempo de parada. Como $\mathcal{F}_{n-1}$ é σ-álgebra, temos que

$$\{N \geq n\} = \{N \leq n - 1\}^c \in \mathcal{F}_{n-1},$$

■

Teorema 5.2.20. *Se X_n é um supermartingale e N é um tempo de parada, então*

$$Y_n = X_{N \wedge n}$$

é um supermartingale . O mesmo vale para martingales e submartingales.

Demonstração. Basta usar o Teorema 5.2.15 com $H_n = 1_{N \geq n}$, já que

$$(H, X)_n = \sum_{m=1}^{n} 1_{N \geq n}(X_m - X_{m-1}) = X_{N \wedge n} - X_0$$

já que vamos apostar só enquanto $n \leq N$, isto é, ainda não for tempo de parada. É claro que sempre $N \geq 0$.

Agora observe que $Y_n = X_0$ (sequência constante) é um supermartingale adaptado à $\mathcal{F}_n$ e que soma de supermartingales adaptados à $\mathcal{F}_n$ é supermartingale adaptado à $\mathcal{F}_n$, logo temos o resultado, já que

$$X_{N \wedge n} = (H, X)_n + X_0.$$

■

Observação 5.2.21. Isso quer dizer que em uma aposta não justa, mesmo impondo um tempo de parada, a aposta ainda é injusta. O mesmo vale para martingales.

5.2.6 Desigualdade dos Cruzamentos

Sejam $a, b \in \mathbb{R}$ e (X_n) um submartingale tal que $a < b$. Defina $N_0 = -1$

$$\begin{aligned} N_{2k-1} &= \inf\{m : \ m > N_{2k-2}, \ X_m \leq a\}, \\ N_{2k} &= \inf\{m : \ m > N_{2k-1}, \ X_m \geq b\} \end{aligned}$$

Intuitivamente, podemos pensar em (N_k) como

$$\begin{aligned} N_1 &= \text{Menor tempo em que o processo é menor que } a, \\ N_2 &= \text{Menor tempo em que o processo é maior que } b \\ &\quad \text{(zeramos o relógio)} \\ N_3 &= \text{Menor tempo em que o processo é menor que } a, \\ N_4 &= \text{Menor tempo em que o processo é maior que } b \\ &\quad \text{(zeramos o relógio)} \\ N_5 &= \text{Menor tempo em que o processo é menor que } a, \\ N_6 &= \text{Menor tempo em que o processo é maior que } b \\ &\vdots \end{aligned}$$

Ou seja, cada par N_{2k-1}, N_{2k} define um cruzamento completo do intervalo (a, b). Abaixo exibimos uma ilustração.

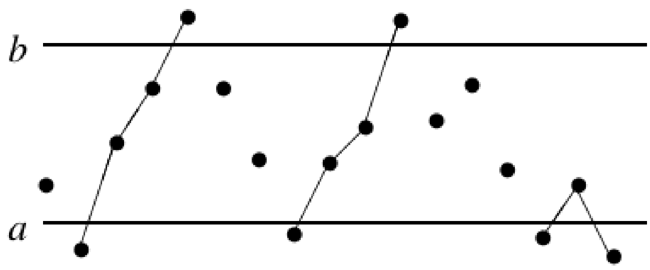


Figura 5.1: Retirado de [Dur10].

Lema 5.2.22. *N_k, $k = 1, 2, \dots$ como acima são tempos de parada.*

Demonstração. Intuitivamente é claro, já que $\{N_j = k\}$ só depende do valor de (X_n) no tempo k. Para uma prova formal, observe que

$$\begin{aligned} \{N_1 = k_1\} &= \{X_1 > a, X_2 > a, \dots, X_{k_1} \leq a\} \in \mathcal{F}_{k_1} \\ \{N_2 = k_2\} &= \bigcup_{k_1 < k_2} (\{N_1 = k_1\} \cap \{X_{k_1+1} < b, \dots, X_{k_2} \geq b\}) \in \mathcal{F}_{k_2}. \\ \{N_3 = k_3\} &= \bigcup_{k_1 < k_2 < k_3} (\{N_1 = k_1\} \cap \{N_2 = k_2\} \cap \{X_{k_2+1} > a, \dots, X_{k_3} \leq a\}) \in \mathcal{F}_{k_1}. \end{aligned}$$

E é fácil generalizar esse argumento indutivamente. ■

Lema 5.2.23. *Defina*

$$H_m = \begin{cases} 1, & se\ N_{2k-1} < m \leq N_{2k},\ para\ algum\ k, \\ 0, & caso\ contrário\ . \end{cases}$$

Então (H_n) é uma sequência previsível.

Demonstração. Note que

$$\begin{aligned} \{N_{2k-1} < m \leq N_{2k}\} &= \{N_{2k-1} < m\} \cap \{m \leq N_{2k}\} \\ &= \{N_{2k-1} \leq m-1\} \cap \{N_{2k} > m-1\}^c. \end{aligned}$$

■

Observação 5.2.24. Note que (H_n) é a estratégia que só aposta caso estejamos num tempo intermediário entre cruzamentos do intervalo (a, b). Isso faz muito sentido pensando num mercado de ações, por exemplo.

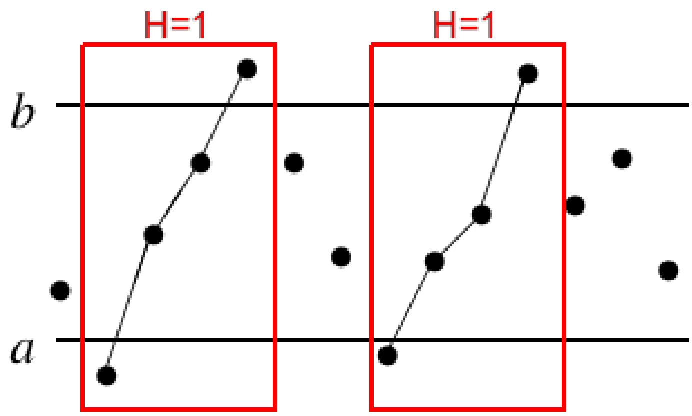


Figura 5.2: Fora dos retângulos vermelhos, $H = 0$.

Teorema 5.2.25 (Desigualdade dos Cruzamentos - [Dur10])**.** *Seja (X_n) um **submartingale** e $a < b$. Então*

$$\begin{aligned} (b-a)\mathbb{E}U_n &\leq \mathbb{E}(a + (X_n - a)^+) - \mathbb{E}(a + (X_0 - a)^+) \\ &= \mathbb{E}(X_n - a)^+ - \mathbb{E}(X_0 - a)^+, \end{aligned}$$

onde

$$U_n = \sup\{k : \ N_{2k} \leq n\},$$

isto é, U_n conta a quantidade de cruzamentos do intervalo (a, b) até o tempo n.

Demonstração. Seja

$$Y_n = a + (X_n - a)^+,$$

isto é,

$$\begin{cases} Y_n & = a \text{ se } X_n \leq a \\ Y_n & = X_n \text{ se } X_n \geq a. \end{cases}$$

Como (X_n) é submartingale, já vimos que $(X_n - a)^+$ é um submartingale. E como o processo constante (a) é um martingale, temos que Y_n **é um submartingale**. Considere $K = 1 - H$, isto é, a estratégia complementar de H (aposta quando H não aposta), onde H é como no Lema 5.2.23. É fácil ver que

$$Y_n - Y_0 = (H, Y)_n + (K, Y)_n.$$

Como (Y_n) define um jogo favorável, sabemos que $(H, Y)_n$ e $(K, Y)_n$ são estratégias favoráveis, já que $H, K \geq 0$ e ambos são limitados. Temos então que $Y_n - Y_0$ é um submartingale.

Agora note que

$$U_n(b - a) \leq (H, Y)_n$$

já que cada cruzamento contribui com pelo menos $(b - a)$. E mais, se n é impar, ou seja, não terminamos um ciclo de aposta, então a contribuição desse último ciclo é não negativa, pois $Y_n \geq a$ e H só aposta se estivermos com ganho ≥ 0. Além disso, como $(K, Y)_n$ é submartingale,

$$\mathbb{E}((K, Y)_n) \geq \mathbb{E}((K, Y)_0) \geq 0.$$

Por fim, juntando todos esses fatos,

$$\begin{aligned} (b - a)\mathbb{E}(U_n) &\leq \mathbb{E}((H, Y)_n) + 0 \\ &\leq \mathbb{E}((H, Y)_n) + \mathbb{E}((K, Y)_n) \\ &= \mathbb{E}((H, Y)_n + (K, Y)_n) \\ &= \mathbb{E}(Y_n - Y_0). \end{aligned}$$

■

5.2.7 Convergência de Martingales

Teorema 5.2.26 (Teorema da Convergência de Martingales-[Dur10])**.** *Seja* (X_n) *um* ***submartingale*** *com*

$$\sup \mathbb{E} X_n^+ < \infty.$$

Então,

$$X_n \xrightarrow{n \to \infty} X, \ q.t.p.$$

para uma v.a. X *com* $\mathbb{E}|X| < \infty$.

Demonstração. Primeiramente, fixe $a < b$ e note que (U_n) é uma sequência monótona e positiva. Logo, pelo teorema da convergência monótona

$$\mathbb{E}(U) = \lim_n \mathbb{E}(U_n),$$

onde $U = \lim_n U_n$, é a v.a. que conta todos os cruzamentos da sequência (X_n).

$$\mathbb{E}(U) = \lim_n \mathbb{E}U_n \leq \lim_n \frac{\mathbb{E}(X_n - a)^+}{b-a} \leq \lim_n \frac{\mathbb{E}X_n^+ + |a|}{b-a} \leq \frac{\sup_n \mathbb{E}X_n^+ + |a|}{b-a} < \infty.$$

Portanto, temos que $U < \infty$ q.t.p.

Agora, supondo que $a, b \in \mathbb{Q}$, observe que

$$A_{a,b} = \{\limsup X_n \geq b > a \geq \liminf X_n\}$$

tem medida nula, pois se não, haveria um conjunto de medida positiva com infinitos cruzamentos do intervalo (a, b), contrariando o fato de que $U < \infty$.

É claro que, então

$$A = \bigcup_{a,b \in \mathbb{Q}, a<b} A_{a,b}$$

tem medida nula. Em particular

$$\{\limsup X_n > \liminf X_n\} = \{\limsup X_n = \liminf X_n\}^c$$

tem medida nula, isto é, (X_n) converge para uma v.a. X q.t.p.

Como

$$\begin{aligned} |X_n| &= X_n^+ + X_n^- \\ X_n &= X_n^+ - X_n^-, \end{aligned}$$

temos que

$$|X_n| = 2X_n^+ - X_n.$$

Como (X_n) é submartingale, temos que

$$-\mathbb{E}(X_n) \leq -\mathbb{E}(X_0),$$

e portanto, utilizando o lema de Fatou,

$$\begin{aligned} \mathbb{E}(|X|) &\leq \liminf \mathbb{E}(|X_n|) \\ &\leq \liminf 2\mathbb{E}(X_n^+) - \mathbb{E}(X_0) \\ &\leq \limsup 2\mathbb{E}(X_n^+) - \mathbb{E}(X_0) < \infty, \end{aligned}$$

provando então que $\mathbb{E}|X| < \infty$. ■

Teorema 5.2.27. *Seja (X_n) um supermartingale com $X_n \geq 0$, então*

$$X_n \xrightarrow{n\to\infty} X, \text{ q.t.p.}$$

e $\mathbb{E}X \leq \mathbb{E}X_0$.

Demonstração. Temos que $(-X_n)$ é um submartingale, com $-X_n \leq 0$ e portanto $(-X_n)^+ = 0$. Pelo teorema 5.2.26, temos que $X_n \xrightarrow{n\to\infty} X$ q.t.p. para uma v.a. X com $\mathbb{E}|X| < \infty$.

Como (X_n) é um supermartingale, temos que

$$\mathbb{E}(X_n) \leq \mathbb{E}(X_0)$$

e o resultado segue pelo lema de Fatou. ■

Exemplo 5.2.28 (O teorema anterior **NÃO** garante convergência L^1)**.** Considere o caso de um passeio aleatório simétrico (S_n) com $S_0 = 1$. Defina

$$N = \inf\{n : S_n = 0\},$$

e também,

$$X_n = S_{N \wedge n}.$$

Como vimos no teorema 5.2.20, X_n é um Martingale. Além disso, X_n é não negativo devido ao fato de que $S_0 = 1$ e a a condição do tempo de parada. Portanto, pelo teorema anterior, $X_n \xrightarrow{n\to\infty} X$, *q.t.p.*. Note que $X = 0$, q.t.p., pois caso contrário, se $X = k > 0$, teríamos que para n grande

$$X_n = k > 0 \text{ e } X_n = k \pm 1.$$

Resumindo, temos que $X_n \xrightarrow{n\to\infty} 0$, q.t.p., porém $\mathbb{E}(X_n) = \mathbb{E}(X_0) = 1$, e então não podemos ter convergência em L^1.

Exemplo 5.2.29. Considere S_n um random walk simples e tome $a \leq 0$ qualquer. Considere também o tempo de parada

$$N = \inf\{k : S_k = a\}.$$

Então, já vimos que o processo $(Y_n) = (S_{n \wedge N})$ é um martingale. É claro que

$$Y_n - a \geq 0$$

e portanto, pelo corolário anterior, temos que

$$Y_n \xrightarrow{n\to\infty} Y, \text{ q.t.p.}$$

com Y v.a. finita q.t.p. já que $\mathbb{E}|Y| < \infty$. Em particular,

$$|Y_{n+1} - Y_n| \xrightarrow{n\to\infty} 0$$

mas como os valores de (Y_n) são discretos, isso significa que para n suficientemente grande $Y_{n+1} = Y_n$ o que se pode acontecer se $N < \infty$ q.t.p..

5.3 Convergência de Martingales: Rota Alternativa

Abaixo exibimos uma formulação alternativa para a desigualdade dos cruzamentos e para o teorema de convergência. A principal diferença é que vamos pedir que $\sup \mathbb{E}|X_n| < \infty$ ao contrário da versão anterior em que pedimos $\sup \mathbb{E}(X_n)^+ < \infty$

Teorema 5.3.1 (Desigualdade dos Cruzamentos - [Wil91]). *Seja (X_n) um* ***supermartingale*** *e $a < b$. Então*

$$(b-a)\mathbb{E}U_n \leq \mathbb{E}(X_n - a)^-$$

onde

$$U_n = \sup\{k : \; N_{2k} \leq n\},$$

isto é, U_n conta a quantidade de cruzamentos no intervalo $[a, b]$ até o tempo n.

Demonstração. Considere H como no Lema 5.2.23. Note que

$$(H, X)_n \geq (b-a)U_n - (X_n - a)^-,$$

já que o lucro total $((H, X)_n)$ é pelo menos $(b-a)$ vezes a quantidade de cruzamentos U_n, porém precisamos descontar um possível prejuízo da rodada n já que esta possivelmente foi interrompida antes de haver um cruzamento.

Como (X_n) é supermartingale, temos que $(H, X)_n$ é supermartingale, logo

$$\mathbb{E}(H, X)_n \leq \mathbb{E}(H, X)_0 := 0$$

e portanto

$$(b-a)\mathbb{E}(U_n) \leq \mathbb{E}(X_n - a)^-.$$

■

Teorema 5.3.2 (Teorema da Convergência de Martingales-[Wil91]). *Seja (X_n) um* ***supermartingale*** *com*

$$\sup \mathbb{E}|X_n| < \infty.$$

Então,

$$X_n \xrightarrow{n\to\infty} X, \;\; q.t.p.$$

para uma v.a. X com $\mathbb{E}|X| < \infty$.

Demonstração. Primeiramente, fixe $a < b$ e note que (U_n) é uma sequência monótona e positiva. Logo, pelo teorema da convergência monótona

$$\mathbb{E}(U) = \lim_n \mathbb{E}(U_n),$$

onde $U = \lim_n U_n$, é a v.a. que conta todos os cruzamentos da sequência (X_n).

$$\mathbb{E}(U) = \lim_n \mathbb{E}U_n \leq \lim_n \frac{\mathbb{E}(X_n - a)^-}{b-a} \leq \frac{\sup_n \mathbb{E}|X_n| + |a|}{b-a} < \infty.$$

Portanto, temos que $U < \infty$ q.t.p.

Agora, supondo que $a, b \in \mathbb{Q}$, observe que

$$A_{a,b} = \{\limsup X_n \geq b > a \geq \liminf X_n\}$$

tem medida nula, pois se não, haveria um conjunto de medida positiva com infinitos cruzamentos do intervalo (a, b), contrariando o fato de que $U < \infty$.

É claro que, então

$$A = \bigcup_{a,b\in\mathbb{Q},a<b} A_{a,b}$$

tem medida nula. Em particular

$$\{\limsup X_n > \liminf X_n\} = \{\limsup X_n = \liminf X_n\}^c$$

tem medida nula, isto é, (X_n) converge para uma v.a. X q.t.p.

Para finalizar, pelo lema de Fatou,

$$\mathbb{E}(|X|) \leq \liminf \mathbb{E}(|X_n|) < \infty,$$

provando então que $\mathbb{E}|X| < \infty$. ■

Note que no teorema de convergência 5.2.26 pedimos que $(X_n)_n$ seja um submartingale, porém em 5.3.2 pedimos que $(X)n)_n$ seja um supermartingale. É claro que tomando $(Y_n)_n = (-X_n)_n$ conseguimos mostrar que tanto faz usarmos submartingales ou supermartingales.
O único motivo para essas hipóteses serem diferentes é que utilizamos fortemente a desigualdade dos cruzamentos e a estratégia de prova das duas versões é levemente diferente, já que no primeiro caso conseguimos um resultado mais geral.

5.3.1 Decomposição de Doob

Teorema 5.3.3 (Decomposição de Doob)**.** *Seja (X_n) um processo adaptado à $(\mathcal{F}_n)$ com $(X_n) \subset L^1$. Então (X_n) pode ser escrito unicamente como*

$$X_n = X_0 + M_n + H_n,$$

onde M_n é um Martingale nulo em $n = 0$, H_n é uma sequência previsível nula em $n = 0$. Além disso, (X_n) é um submartingale se, e só se, H_n é crescente.

Demonstração. Se a decomposição existir, então

$$\mathbb{E}(X_n|\mathcal{F}_{n-1}) = X_0 + \mathbb{E}(M_n|\mathcal{F}_{n-1}) + \mathbb{E}(H_n|\mathcal{F}_{n-1}) = X_0 + M_{n-1} + H_n.$$

Como

$$M_{n-1} = X_{n-1} - H_{n-1} - X_0,$$

temos que

$$\mathbb{E}(X_n|\mathcal{F}_{n-1}) = X_{n-1} - H_{n-1} + H_n.$$

Ou seja, temos duas relações:

1. $H_n - H_{n-1} = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1}$,
2. $M_n = X_n - H_n$.

Definindo então $H_0 = M_0 = 0$, o item 1 nos dá uma receita para construir H_n unicamente, a menos de versões equivalentes da mesma v.a.. Note que como

$$H_n = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} + H_{n-1},$$

indutivamente, conseguimos mostrar que $H_n \in \mathcal{F}_{n-1}$. Além disso, pelo item 1, é fácil ver que (H_n) é crescente se, e só se, (X_n) é submartingale.

Uma vez construído a sequência previsível (H_n), conseguimos construir M_n unicamente pelo item 2.

Note que M_n é um Martingale, já que

$$\begin{aligned}
\mathbb{E}(M_n|F_{n-1}) &= \mathbb{E}(X_n|F_{n-1}) - \mathbb{E}(H_n|F_{n-1}) \\
&= \mathbb{E}(X_n|F_{n-1}) - H_n \\
&\underset{\text{Item 1.}}{=} X_{n-1} - H_{n-1} \\
&\underset{\text{Item 2.}}{=} M_{n-1}.
\end{aligned}$$

■

5.4 Desigualdade de Doob e Convergência L^p

Teorema 5.4.1. *Se (X_n) é um submartingale e N é um tempo de parada limitado por k, então*

$$\mathbb{E}X_0 \leq \mathbb{E}X_N \leq EX_k$$

Observação 5.4.2. Intuitivamente, é claro que esse resultado deve ser verdadeiro.

Demonstração. Já vimos que $(X_{N\wedge n})$ também é um submartingale, logo

$$\mathbb{E}X_0 = \mathbb{E}X_{N\wedge 0} \leq \mathbb{E}X_{N\wedge k} = \mathbb{E}X_N,$$

provando a primeira desigualdade.

Agora considere a sequência previsível e limitada $H_n = 1_{N\leq n-1}$, isto é, apostamos só quando $n > N$. Temos então que

$$(H, X)_n = X_n - X_{N\wedge n}$$

é um submartingale e portanto

$$= \mathbb{E}(X_k) - \mathbb{E}(X_N) = \mathbb{E}(X, H)_k \geq |E(X, H)_0 = 0$$

■

Teorema 5.4.3 (Desigualdade de Doob)**.** *Sejam (X_n) um submartingale, $\lambda > 0$ e*

$$X'_n = \max_{0\leq m\leq n} X_n^+.$$

Então

$$\mathbb{P}(X'_n \geq \lambda) \leq \frac{\mathbb{E}(X_n 1_{X'_n\geq\lambda})}{\lambda} \leq \mathbb{E}X_n^+.$$

Demonstração. Defina o tempo de parada

$$N = \inf\{m : X_m \geq \lambda \text{ ou } m = n\},$$

isto é, N marca o menor tempo em que o processo fica maior que λ até o processo acabar no tempo n. Como N é limitado por n, temos pelo teorema anterior que

$$\mathbb{E}(X_N) \leq \mathbb{E}(X_n).$$

Note que no evento $\{X'_n \geq \lambda\}$ temos que $X_N \geq \lambda$ e então

$$\lambda\mathbb{P}(X'_n \geq \lambda) \leq \mathbb{E}(X_N 1_{X'_n\geq\lambda}).$$

Além disso, temos que

$$\begin{aligned}\mathbb{E}(X_N) &= \mathbb{E}(X_N 1_{X'_n\geq\lambda}) + \mathbb{E}(X_N 1_{X'_n<\lambda})\\ &= \mathbb{E}(X_N 1_{X'_n\geq\lambda}) + \mathbb{E}(X_n 1_{X'_n<\lambda})\\ &= \mathbb{E}(X_N 1_{X'_n\geq\lambda}) + \mathbb{E}(X_n) - \mathbb{E}(X_n 1_{X'_n\geq\lambda})\end{aligned}$$

e portanto, como $\mathbb{E}(X_N) \leq \mathbb{E}(X_n)$

$$\mathbb{E}(X_N 1_{X_n' \geq \lambda}) \leq \mathbb{E}(X_n 1_{X_n' \geq \lambda}).$$

Com isso temos que

$$\lambda \mathbb{P}(X_n' \geq \lambda) \leq \mathbb{E}(X_n 1_{X_n' \geq \lambda}).$$

E a segunda desigualdade do enunciado é trivial. ■

Exemplo 5.4.4 (Desigualdade Maximal de Kolmogorov)**.** Vamos apresentar uma prova alternativa para a Desigualdade Maximal de Kolmogorov 4.3.3.

Considere então (X_n) v.a.'s independentes com média 0. Sabemos que (S_n) é um martingale, e portanto (S_n^2) é um submartingale já que $x \mapsto x^2$ é convexa. Temos então pelo teorema anterior que, tomando $\lambda = x^2$

$$\mathbb{P}\left(\max_{0 \leq m \leq n} |S_m| \geq x\right) \leq \frac{\mathbb{V}\mathrm{ar}(S_n)}{x^2}.$$

5.5 Exercícios Resolvidos

Exercício 5.5.1 ([Dur10])**.** Prove a Desigualdade de Chebyshev no caso condicional, isto é,

$$\mathbb{P}(|X| > a|\mathcal{F}) \leq \frac{\mathbb{E}(X^2|\mathcal{F})}{a^2}, \ a > 0.$$

Demonstração. Basta notar que

$$a1_{|X|>a} \leq |X|.$$

∎

Exercício 5.5.2 ([Dur10])**.** Sejam X, Y v.a.'s com $\mathbb{E}(Y|\mathcal{G}) = X$ e $\mathbb{E}Y^2 = \mathbb{E}X^2$. Mostre que $X = Y$ q.t.p..

Demonstração. Como $\mathbb{E}(Y|\mathcal{G})) = X$ é claro que $X \in \mathcal{G}$. Então

$$\mathbb{E}(X^2) = \mathbb{E}(X\mathbb{E}(Y|\mathcal{G})) = \mathbb{E}(\mathbb{E}(XY|\mathcal{G})) = \mathbb{E}(XY).$$

Mas então

$$\mathbb{E}((X-Y)^2) = \mathbb{E}(X^2 - 2XY + Y^2) = 0.$$

∎

Capítulo 6

Probabilidade em Espaços Métricos e Movimento Browniano

6.1 Convergência

Definição 6.1.1 (Convergência Fraca)**.** Seja $(\Omega, \mathcal{B})$ um espaço de probabilidade onde Ω é um espaço métrico com distância d e $\mathcal{B}$ é a σ-álgebra de Borel gerada por d. Dado uma sequência de probabilidades μ_n em $(\Omega, \mathcal{B})$, dizemos que μ_n converge fracamente para μ se

$$\int f(x)d\mu_n(x) \to \int f(x)d\mu(x),$$

para toda função f contínua e limitada. Simbolizamos essa convergência como $\mu_n \Rightarrow \mu$.
Da mesma Forma, se (X_n) é uma sequência de v.a.'s, dizemos que X_n converge fracamente para X (ou que converge em lei) se

$$\mathbb{E}f(X_n) \to \mathbb{E}f(X),$$

para toda função f contínua e limitada. Simbolizamos essa convergência como $X_n \Rightarrow X$.

6.2 Exercícios Resolvidos

Parte I

Apêndice

Apêndice A

A.1 Fórmula de Stirling

As Principais referências usadas foram:

1. [Tao12]
2. [Spe14]
3. [Con16]

O nosso objetivo é mostrar a seguinte identidade assintótica:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \tag{A.1.1}$$

Lema A.1.1. *Temos que*

$$n! = \int_0^{+\infty} t^n e^{-t} dt.$$

Demonstração. Integrando por partes uma vez, temos que

$$\int_0^{+\infty} t^n e^{-t} dt = \left(e^{-t} t^n\right)\Big|_0^{\infty} + n\int_0^{+\infty} t^{n-1} e^{-t} dt.$$

Então o resultado sai por indução. ∎

Lema A.1.2. *A função*

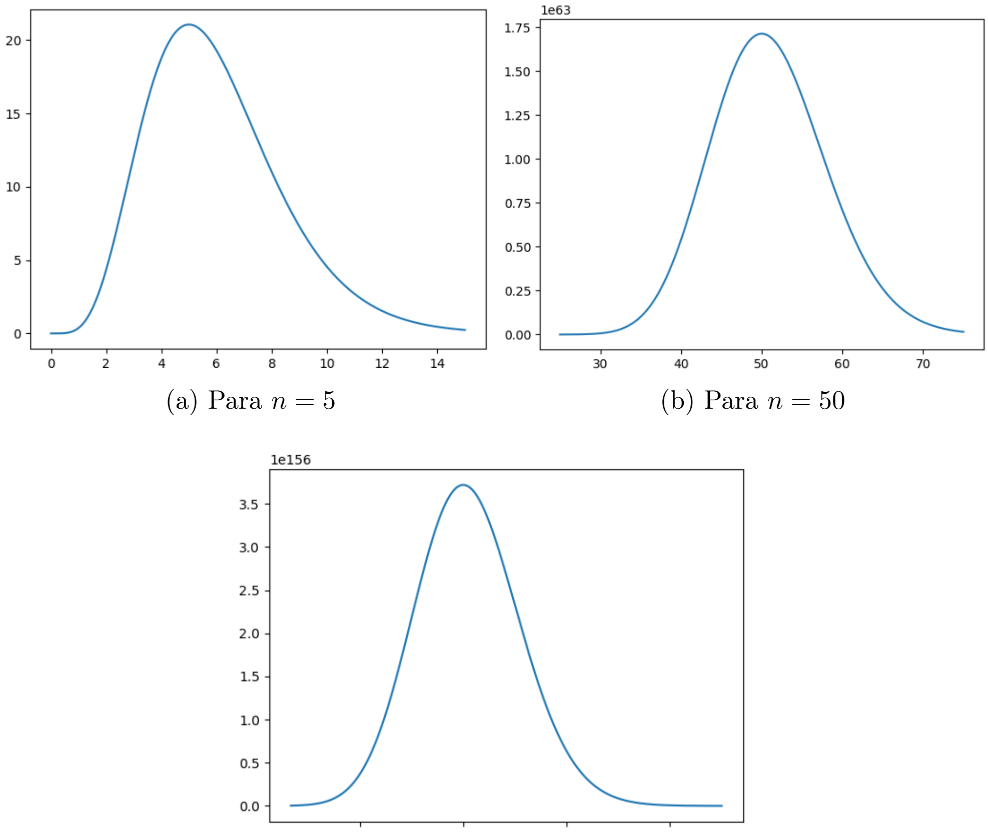
$$f(t) = t^n e^{-t}$$

atinge máximo em $t = n$.

Abaixo, exibimos o gráfico da função

$$f(t) = t^n e^{-t},$$

para diversos valores de n. Note que todos os gráficos, possuem o mesmo aspecto de sino quando observados em torno do máximo de $f(t)$.



(a) Para $n = 5$

(b) Para $n = 50$

(c) Para $n = 100$

Figura A.1: Gráfico de $f(t) = t^n e^{-t}$.

Façamos então uma mudança de coordenadas que leve em consideração esse fato, isto é, vamos, vamos integrar com respeito á $t = s + n$:

$$
\begin{aligned}
n! &= \int_0^{+\infty} t^n e^{-t} dt \\
&= \int_{-n}^{+\infty} (s+n)^n e^{-(s+n)} ds \\
&= \left(\frac{n}{e}\right)^n \int_{-n}^{+\infty} \left(1 + \frac{s}{n}\right)^n e^{-s} ds \\
&= \left(\frac{n}{e}\right)^n \int_{-n}^{+\infty} \exp\left(-s + n \ln\left(1 + \frac{s}{n}\right)\right) ds
\end{aligned}
$$

Intuição: Como para $|s| < n$,

$$n \ln\left(\frac{s}{n} + 1\right) = s - \frac{s^2}{2n} + \ldots,$$

fazendo a mudança de coordenadas $s = t\sqrt{n}$:

$$\begin{aligned} n! &= \left(\frac{n}{e}\right)^n \int_{-n}^{+\infty} \exp\left(-s + n\ln\left(1 + \frac{s}{n}\right)\right) ds \\ &\approx \left(\frac{n}{e}\right)^n \int_{-n}^{+\infty} \exp\left(-\frac{s^2}{2n}\right) ds. \\ &= \left(\frac{n}{e}\right)^n \sqrt{n} \int_{-\sqrt{n}}^{+\infty} \exp\left(-\frac{t^2}{2}\right) dt. \\ &\sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \end{aligned}$$

Para formalizar a intuição acima, considere a mudança de variáveis $s = t\sqrt{n}$. Então

$$n! = \left(\frac{n}{e}\right)^n \sqrt{n} \int_{-\sqrt{n}}^{+\infty} \exp\left(-t\sqrt{n} + n\ln\left(1 + \frac{t}{\sqrt{n}}\right)\right) dt.$$

Lema A.1.3. *Temos que*

$$\exp\left(-t\sqrt{n} + n\ln\left(1 + \frac{t}{\sqrt{n}}\right)\right) \xrightarrow{n\to\infty} e^{-t^2/2}.$$

Demonstração. Como t é fixo, basta estudarmos o que acontece quando n é bem maior que t. Suponha então que $n > t^2$, isto é, $|t/\sqrt{n}| < 1$.

Então,

$$\begin{aligned} n\ln\left(\frac{t}{\sqrt{n}} + 1\right) - \sqrt{n}t &= n\left(-\frac{t^2}{2n} + O((t/\sqrt{n})^3)\right) \\ &= -\frac{t^2}{2} + O(t^3/\sqrt{n}) \xrightarrow{n\to\infty} -t^2/2. \end{aligned}$$

∎

Defina,

$$f_n(t) := 1_{[-\sqrt{n},+\infty)} \exp\left(-t\sqrt{n} + n\ln\left(1 + \frac{t}{\sqrt{n}}\right)\right).$$

Então só o que nos falta mostrar é que

$$\lim_n \int f_n(t)dt = \int \lim_n f_n(t)dt = \int e^{-t^2/2}dt.$$

Lema A.1.4. *Existe uma função mensurável g com módulo integrável, tal que $|f_n| < g$. Portanto, pelo teorema da convergência dominada,*

$$\int_{-\sqrt{n}}^{+\infty} \exp\left(-t\sqrt{n} + n\ln\left(1 + \frac{t}{\sqrt{n}}\right)\right) dt \xrightarrow{n\to\infty} \int e^{-t^2/2}dt = \sqrt{2\pi}.$$

Demonstração. Já vimos que

$$\log f_n(t) = -t^2/2 + O(t^3/\sqrt{n}),$$

note que se $t < 0$, intuitivamente

$$\log f_n(t) = -t^2/2 + O(t^3/\sqrt{n}) < -t^2/2,$$

já que $O(t^3/\sqrt{n}) < 0$.

Se $t > 0$, intuitivamente, temos que

$$-\frac{t^2}{2} + \frac{t^3}{\sqrt{1}} > \frac{t^2}{2} + \frac{t^3}{\sqrt{2}} > \cdots > \frac{t^2}{2} + \frac{t^3}{\sqrt{n}} > \ldots,$$

isto é, $f_n \leq f_1 = (1+t)e^{-t}$.

Defina g como

$$g(t) = \begin{cases} e^{-t^2/2} & \text{se } -\sqrt{n} \leq t < 0, \\ (1+t)e^{-t} & \text{se } t \geq 0. \end{cases}.$$

De fato, não é difícil mostrar que $|f_n| \leq g$ e que g é integrável (usar técnicas simples de cálculo), portanto, pelo teorema da convergência dominada, como $f_n \xrightarrow{n\to\infty} e^{-t^2/2}$, temos o resultado. ∎

A.2 Expansão de Padé

As principais referências usadas foram:

1. [MF98].
2. [HBA+13].

Quando provamos a desigualdade de Bernstein, utilizamos o seguinte fato:

$$(1+x)\log(1+x) - x \geq \frac{x^2}{2(1+x/3)},$$

provar que essa desigualdade vale é um exercício normal de cálculo, a questão fundamental é: como alguém pode deduzir tal desigualdade?

Note que, assim como numa expansão de Taylor, estamos aproximando

$$f(x) = (1+x)\log(1+x) - x$$

por funções mais simples, isto é, quocientes de polinômios.

Em geral, suponha que

$$f(x) = \sum_{i=0}^{\infty} a_i x^i,$$

e que queremos escrever [1]

$$f(x)q(x) = p(x) + z(x),$$

onde $p(x)$ é um polinômio de grau n, $q(x)$ é um polinômio de grau m e as derivadas de f e $r(x) := p(x)/q(x)$ coincidam em 0 até o grau $n+m$.

Note que, pela condição das derivadas coincidirem em 0 até $n+m$, é fácil ver que $z(x)$ tem que ser da forma:

$$z(x) = \sum_{i=n+m+1}^{\infty} z_i x^i.$$

Além disso, como estamos expandindo f em torno do 0, queremos que o $q(0) \neq 0$, logo, podemos supor que $q_0 = 1$. Portanto:

$$\left(\sum_{i=0}^{\infty} a_i x^i\right)\left(1 + \sum_{i=1}^{m} q_i x^i\right) - \sum_{i=0}^{n} p_i x^i = \sum_{i=n+m+1}^{\infty} z_i x^i.$$

Dessa forma, conseguimos um sistema linear de $n+m+1$ incógnitas que podemos resolver para encontrarmos os coeficientes q_i, p_i. Abaixo, exibimos as 4 primeiras

[1] ou mais informalmente $f(x) \approx \dfrac{p(x)}{q(x)}$.

equações lineares, que vamos chamar de *equações de Padé*:

$$\begin{aligned} a_0 - p_0 &= 0 \\ q_1 a_0 + a_1 - p_1 &= 0 \\ q_2 a_0 + q_1 a_1 + a_2 - p_2 &= 0 \\ q_3 a_0 + q_2 a_1 + q_1 a_2 + a_3 - p_3 &= 0. \end{aligned}$$

Com as quatro equações acima, já somos capazes de aproximar $f(x)$ por funções da forma $p(x)/q(x)$ com p de grau 2 e q de grau 1. Por exemplo, considere novamente

$$f(x) = (1+x)\log(1+x) - x,$$

é fácil[2] ver que

$$f(x) = 0 + 0x + x^2/2 - x^3/6 + \cdots,$$

e então, substituindo os coeficientes a_i de f nas equações de Padé, temos que

$$\begin{aligned} 0 - p_0 = 0 &\longrightarrow p_0 = 0 \\ 0q_1 + 0 - p_1 = 0 &\longrightarrow p_1 = 0 \\ 0q_2 + 0q_1 + 1/2 - p_2 = 0 &\longrightarrow p_2 = 1/2 \\ 0q_3 + 0q_2 + q_1/2 - 1/6 - 0 = 0 &\longrightarrow q_1 = 1/3, \end{aligned}$$

e podemos concluir que

$$f(x) \approx \frac{x^2/2}{1 + x/3},$$

que nos leva exatamente à desigualdade que precisamos, como ilustrado abaixo [3]

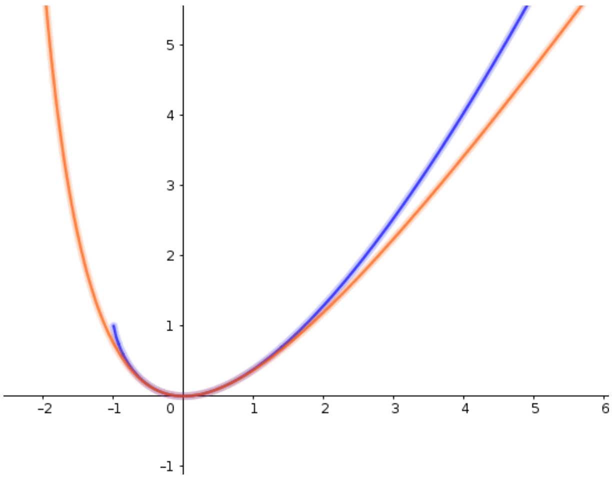


Figura A.2: Em azul temos $(1+x)\log(1+x) - x$ e em vermelho $\dfrac{x^2/2}{1+x/3}$.

Façamos mais um exemplo, dessa vez para

$$g(x) = 1 + x - \sqrt{1+2x} = 0 + 0x + x^2/2 - x^3/2 + \cdots,$$

[2]Wolfram =)

[3]Imagens geradas no Geogebra [HBA+13].

e então resolvendo as equações de Padé, temos que

$$g(x) \approx \frac{x^2/2}{1+x},$$

e é possível mostrar que podemos trocar $\approx$ por $\geq$.

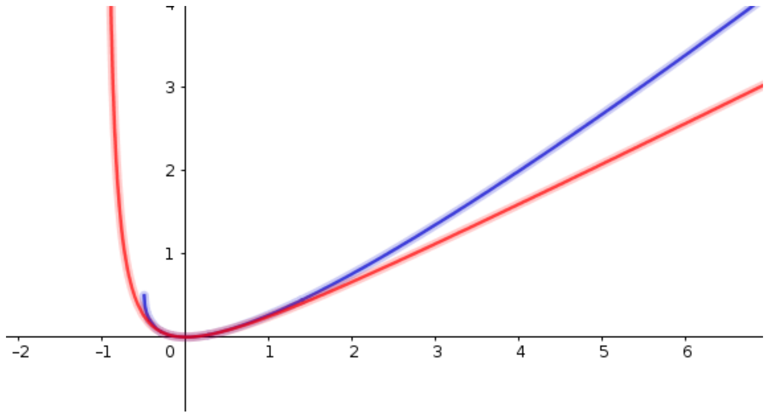


Figura A.3: Em azul temos $1 + x - \sqrt{1+2x}$ e em vermelho $\dfrac{x^2/2}{1+x}$.

Para finalizar, já sabemos que

$$\frac{x}{x+1} \leq \log(1+x), \qquad x > -1,$$

mas então, para $x \in (0,1)$, temos que

$$-\log(1-x) - x \leq \frac{x}{-x+1} - x = \frac{x^2}{1-x}.$$

Utilizando a aproximação de Padé, conseguimos encontrar um bound com uma constante melhor, já que como

$$-\log(1-x) - x = x^2/2 + x^3/3 + \cdots,$$

então por Padé,

$$-\log(1-x) - x \approx \frac{x^2}{1 - 2x/3},$$

e é possível provar que podemos trocar $\approx$ por $\leq$, se $x \in (0,1)$.

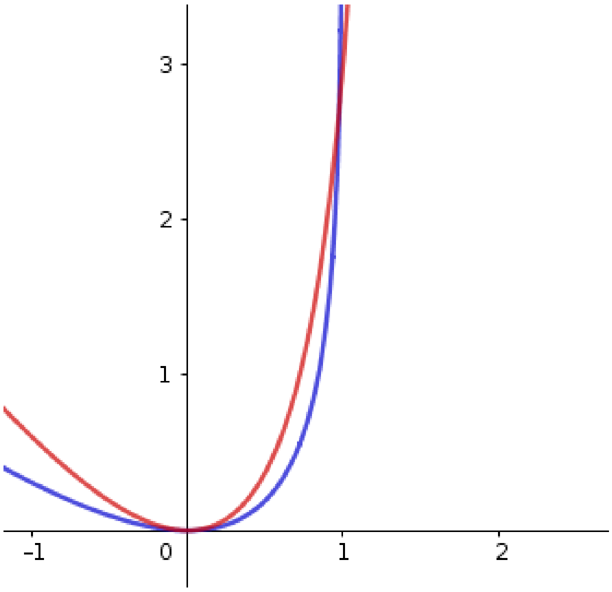


Figura A.4: Em azul temos $-\log(1-x)-x$ e em vermelho $\dfrac{x^2}{1-2x/3}$.

Referências Bibliográficas

[AS16] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley Publishing, 4th edition, 2016.

[Bil99] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.

[Chu01] K.L. Chung. *A Course in Probability Theory*. Elsevier Science, 2001.

[Con16] Keith Conrad. Stirling's formula. 2016.

[CY12] Loïc Chaumont and Marc Yor. *Exercises in Probability: A Guided Tour from Measure Theory to Random Processes, via Conditioning*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2 edition, 2012.

[Dur96] R. Durrett. *Stochastic Calculus: A Practical Introduction*. Probability and Stochastics Series. Taylor & Francis, 1996.

[Dur10] Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, New York, NY, USA, 4th edition, 2010.

[GS01] G. Grimmett and D. Stirzaker. *One Thousand Exercises in Probability*. One Thousand Exercises in Probability. OUP Oxford, 2001.

[Gut12] A. Gut. *Probability: A Graduate Course*. Springer Texts in Statistics. Springer New York, 2012.

[HBA+13] M. Hohenwarter, M. Borcherds, G. Ancsin, B. Bencze, M. Blossier, A. Delobelle, C. Denizet, J. Éliás, Á Fekete, L. Gál, Z. Konečný, Z. Kovács, S. Lizelfelner, B. Parisse, and G. Sturr. GeoGebra 4.4, December 2013. http://www.geogebra.org.

[Kee10] R.W. Keener. *Theoretical Statistics: Topics for a Core Course*. Springer Texts in Statistics. Springer New York, 2010.

[Kni09] O. Knill. *Probability Theory and Stochastic Processes with Applications*. Overseas Press, 2009.

[Kol06] Vladimir Koltchinskii. Local rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.*, 34(6):2593–2656, 12 2006.

[MF98] John H. Mathews and Kurtis D. Fink. *Numerical Methods Using MATLAB*. Simon & Schuster, Inc., New York, NY, USA, 3rd edition, 1998.

[MRT12] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. The MIT Press, 2012.

[Sha03] J. Shao. *Mathematical Statistics*. Springer Texts in Statistics. Springer, 2003.

[Sim15] B. Simon. *Real Analysis*. A comprehensive course in analysis. American Mathematical Society, 2015.

[Spe14] Joel Spencer. *Asymptopia*. American Mathematical Society, 2014.

[Sta] Statlect. Digital textbook on probability and statistics. https://www.statlect.com/. [Online; accessed 20/02/2019].

[Sto14] J.M. Stoyanov. *Counterexamples in Probability: Third Edition*. Dover Books on Mathematics. Dover Publications, 2014.

[Str10] D.W. Stroock. *Probability Theory: An Analytic View*. Cambridge University Press, 2010.

[Tao12] T. Tao. *Topics in Random Matrix Theory*. Graduate studies in mathematics. American Mathematical Soc., 2012.

[vdV00] A.W. van der Vaart. *Asymptotic Statistics*. Asymptotic Statistics. Cambridge University Press, 2000.

[Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

[Was10] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Publishing Company, Incorporated, 2010.

[Wil91] David Williams. *Probability with Martingales*. Cambridge University Press, 1991.

Índice Remissivo