UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE MATEMÁTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Uma introdução aos grandes desvios

Dissertação de Mestrado

Gustavo Henrique Müller

Porto Alegre, 10 de novembro de 2016

Dissertação submetida por Gustavo Henrique Müller[1] como requisito parcial para a obtenção do grau de Mestre em Matemática pelo Programa de Pós-Graduação em Matemática do Instituto de Matemática da Universidade Federal do Rio Grande do Sul.

Professora Orientadora:
Dra. Adriana Neumann de Oliveira

Banca Examinadora:
Dra. Adriana Neumann de Oliveira (PPGMAT-UFRGS)
Dra. Ana Patrícia Carvalho Gonçalves (IST-ULisboa)
Dr. Ricardo Misturini (PPGMAT-UFRGS)
Dra. Susana Frómeta Fernández (UFRGS)

Data da apresentação: 13 de outubro de 2016

[1] Bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES

Resumo:
Nesta dissertação de mestrado, vamos apresentar uma prova para os grandes desvios para variáveis aleatórias independentes e identicamente distribuídas com todos os momentos finitos e para a medida empírica de cadeias de Markov com espaço de estados finito e tempo discreto. Além disso, abordaremos os teoremas de Sanov e Gärtner-Ellis.

Palavras-Chave: Grandes desvios. Teorema de Cramér-Chernoff. Cadeias de Markov. Teorema de Sanov. Teorema de Gärtner-Ellis.

Abstract:
In this master thesis it is presented a proof of the large deviations for independent and identically distributed random variables with all finite moments and for the empirical measure of Markov chains with finite state space and with discrete time. Moreover, we address the theorems of Sanov and of Gartner-Ellis.

Key Words: Large Deviations. Theorem of Cramér-Chernoff. Markov chains. Theorem of Sanov. Theorem of Gärtner-Ellis.

A todos aqueles que torcem pelo meu sucesso

AGRADECIMENTOS

Gostaria inicialmente de agradecer aos meus pais todo apoio e incentivo recebido. Agradeço também aos demais familiares e amigos, que sempre acreditaram em mim e, com certeza, fazem parte dessa conquista.

Agradeço aos meus professores pelos conhecimentos adquiridos durante todos esses anos. Se não fosse por vocês, eu não teria chegado até aqui. Agradeço, em especial, a Adriana Neumann de Oliveira, minha orientadora, por toda a paciência e atenção, desde quando eu a procurei.

Agradeço aos membros da banca examinadora todas as correções e sugestões, que foram muito importantes para melhorar o texto. Qualquer incorreção que permaneça é de minha inteira responsabilidade.

Agradeço aos professores e funcionários do Instituto de Matemática da UFGRS o suporte. Agradeço a CAPES pelo apoio financeiro.

SUMÁRIO

INTRODUÇÃO

O objetivo principal desta dissertação é estudar grandes desvios em alguns contextos simples, para mostrar que as demonstrações de grandes desvios são permeadas pelas mesmas ideias e todas seguem a mesma linha de raciocínio. Um bom entendimento dessas ideias nos permite aplicá-las em contextos mais sofisticados, onde, pelas tecnicalidades inerentes a tais contextos, elas ficam mascaradas.

Vamos começar nossa jornada em um território familiar. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuidas (i.i.d. - veja Definição A.8) definidas no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e tomando valores em $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, onde $\mathcal{B}(\mathbb{R})$ denota a sigma-álgebra de Borel em $\mathbb{R}$. Escrevendo $\mathbb{E}$ para denotar a esperança com respeito a probabilidade $\mathbb{P}$, sejam $\mathbb{E}[X_1] = a \in \mathbb{R}$, $Var(X_1) = \mathbb{E}[X_1^2] - a^2 = \sigma^2 \in (0, \infty)$ e $S_n = X_1 + \cdots + X_n$ $(n \in \mathbb{N})$ as somas parciais. Nos livros clássicos de teoria da probabilidade, podem ser encontrados dois teoremas fundamentais tratando deste tipo de sequencias:

Lei Forte dos Grandes Números (LGN)[2]

$$\mathbb{P}\left[\frac{S_n}{n} \to a,\ quando\ n \to \infty\right] = 1.$$

Teorema Central do Limite (TCL)[3]

$$\frac{1}{\sigma\sqrt{n}}(S_n - an) \xrightarrow{d} N(0,1),\ \ quando\ n \to \infty,$$

onde $N(0,1)$ é a variável aleatória com distribuição normal, média 0 e variância 1 e $\xrightarrow{d}$ denota a convergência em distribuição (veja Definição A.25).

Enquanto a LGN trata da convergência de $\frac{S_n}{n}$ para a, o TCL quantifica a probabilidade da diferença entre S_n e na ser da ordem de $\sqrt{n}$. Desvios desse tamanho são chamados de "normais". Nesta dissertação, vamos tratar de desvios de ordem n. Pela LGN, sabemos que a probabilidade do evento

$$\left[\frac{S_n - an}{n} \geqslant \delta\right],\ \text{para}\ \delta > 0, \tag{1}$$

converge a zero, quando $n \to \infty$. O nosso objetivo é quantificar a taxa de convergência das probabilidades de eventos como este acima. Desvios deste tamanho são chamados "grandes".

Vamos ver que o princípio de grandes desvios, assunto dessa dissertação, estuda uma taxa de convergência EXPONENCIAL. Falando informalmente, estamos buscando uma função taxa $I \geqslant 0$ tal que a probabilidade do evento (1) seja da ordem de $e^{-nI(a+\delta)}$.

---

[2] A demonstração deste teorema pode ser encontrada em Durrett (2013)
[3] A demonstração deste teorema pode ser encontrada em Durrett (2013)

Para ser mais preciso sobre o que estuda os grandes desvios iremos apresentar logo a sua definição, que vamos utilizar durante praticamente todo o texto.

Definição. [4] *[Princípio de Grandes Desvios, em inglês: Large Deviations Principle - LDP] Seja $\{\mathcal{Z}_n\}$ uma sequência de variáveis aleatórias (veja Definição A.6) definidas no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e tomando valores num espaço polonês (espaço métrico completo e separável) R. Considere $\mathcal{B}(R)$ a σ-álgebra de Borel de R e $I : R \to [0, +\infty]$ uma função semicontinua inferiormente (veja Definição A.1).*

*Dizemos que $\{\mathcal{Z}_n\}$ satisfaz **o LDP com função taxa** I, se as seguintes condições são satisfeitas:*

(i) Cota superior: Seja $F \in \mathcal{B}(R)$ um fechado. Então,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\mathcal{Z}_n \in F\right] \leqslant - \inf_{x \in F} I(x); \tag{2}$$

(ii) Cota inferior: Seja $O \in \mathcal{B}(R)$ um aberto. Então,

$$\underline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\mathcal{Z}_n \in O\right] \geqslant - \inf_{x \in O} I(x).$$

Repare que, na definição acima, $\mathcal{Z}_n$ faz o papel de $\frac{S_n}{n}$ no contexto de variáveis aleatórias i.i.d. mencionado no início dessa introdução.

Como podemos ver pela definição acima, a demonstração de que uma sequência de variáveis aleatórias satisfaz o princípio de grandes desvios consiste em duas partes: a cota superior e a cota inferior. Em cada uma delas, precisamos limitar (superiormente ou inferiormente) a probabilidade destas variáveis aleatórias pertencerem a um determinado conjunto. Para obter este tipo de estimativas, começamos pela cota superior criando uma perturbação (exponencial) na probabilidade que queremos cotar e, em seguida, minimizamos sobre todas as possíveis perturbações. Em geral, fazemos esse processo utilizando conjuntos compactos e depois estendemos para os fechados através da Proposição A.19. Desta maneira, chegamos a uma expressão variacional para o funcional taxa. Entretanto, na cota inferior, precisamos ser mais precisos. Para isso, devemos entender algumas características do funcional taxa obtido na cota superior, pois precisamos utilizar uma perturbação específica: aquela que atinge a expressão variacional obtida na cota superior. Nesta parte, geralmente, precisamos que o modelo estudado satisfaça algumas hipóteses adicionais.

Esta técnica de demonstração foi introduzida por Donsker e Varadhan numa série de artigos: (DONSKER; VARADHAN, 1975a), (DONSKER; VARADHAN, 1975b), (DONSKER; VARADHAN, 1975c), (DONSKER; VARADHAN, 1976), (DONSKER; VARADHAN, 1989), entre outros. Em (VARADHAN, 1984) e (VARADHAN, 2008), o

4 Para uma definição de LDP ainda mais geral, veja a definição de LDP para sequências de medidas em A.32.

autor apresenta esta técnica aplicada em situações mais simples. A ideia dessa dissertação surgiu durante o estudo do artigo (VARADHAN, 2008). Porém, aqui temos como objetivo apresentar estas ideias para leitores que não sejam especialistas no assunto. Supomos apenas que o leitor possua conhecimentos básicos de análise, álgebra linear, medida e probabilidade. Por isso, apresentamos as demonstrações com bastante riqueza de detalhes.

Este texto é organizado da seguinte forma: Iniciamos esta dissertação provando grandes desvios para variáveis aleatórias i.i.d. No Capítulo 3, provamos grandes desvios para a sequência de medidas empíricas associadas a cadeias de Markov com tempo discreto e com espaço de estados finito (veja Definição 2.4). Antes disso, com a intenção de ter um texto completo, o Capítulo 2 foi dedicado aos conceitos básicos de cadeias de Markov, que serão importantes no Capítulo 3. No Capítulo 4, fizemos algumas considerações sobre outros resultados importantes no estudo de grandes desvios, que não foram abordados anteriormente. O Apêndice concentra o restante da parte teórica que é utilizada nas provas de grandes desvios. Ele contém os resultados de várias áreas da matemática e é dividido em assuntos de Análise, Probabilidade Básica e Convergência de Medidas.

A teoria apresentada nos Capítulos 1, 3 e 4 é baseada em (VARADHAN, 2008), (HOLLANDER, 2008), (OLIVIERI; VARES, 2005), (KIPNIS; LANDIM, 2013) e (STROOCK, 1984). O Capítulo 2 foi adaptado de (MÜLLER, 2015) e nele foram utilizados (LEVIN; PERES; WILMER, 2009), (HORN; JOHNSON, 2012) e (ROUSSEAU et al., 2015). As demonstrações apresentadas no Capítulo 1 não seguem as demonstrações clássicas apresentadas em (HOLLANDER, 2008) e (OLIVIERI; VARES, 2005). Como o objetivo deste texto é apresentar, de uma forma mais simples, as teorias de demonstração de LDP introduzidas por Donsker e Varadhan nos artigos mencionados anteriormente, o que fizemos foi aplicá-las, por exemplo, ao caso i.i.d. Um leitor mais experiente pode notar que, ao fazer isso, obtivemos uma demonstração muito parecida à do Teorema de Gärtner-Ellis (que é apresentada no Capítulo 4), mas sem as complicações técnicas deste teorema, pois estamos num contexto bem mais simples.

1 GRANDES DESVIOS PARA VARIÁVEIS ALEATÓRIAS I.I.D.

Neste capítulo, consideraremos $\{X_n\}$ uma sequência de variáveis aleatórias independentes e identicamente distribuidas (i.i.d. - veja Definição A.8) com espaço de estados $\mathbb{R}$. Como vamos seguir com esse estudo para casos mais gerais, algumas partes deste capítulo estão postas num contexto mais abrangente.

Para entender o motivo da função I, que aparece na definição do LDP, ser chamada de função taxa, considere $\{X_n\}$ um sequência de variáveis aleatórias i.i.d. com distribuição normal de média 0 e variância 1, ou seja, $X_1 \sim N(0,1)$. Seja $S_n = X_1 + \cdots + X_n$ a sua soma parcial. Sabemos, pela Lei dos Grandes Números, que se $m > 0$, então $\mathbb{P}\left[\frac{S_n}{n} \in [m, +\infty)\right]$ converge para 0, mas gostaríamos de saber qual a velocidade desta convergência. Para isso, utilizando a desigualdade de Chebyshev (Teorema A.12) com $A = [mn, +\infty)$, $X = S_n$ e $\psi(x) = e^{\lambda x}$, para algum $\lambda > 0$, temos que

$$\mathbb{P}\left[\frac{S_n}{n} \in [m, +\infty)\right] = \mathbb{P}[S_n \geqslant mn] \leqslant \frac{\mathbb{E}[e^{\lambda S_n}]}{e^{\lambda mn}}.$$

Como $\{X_n\}$ é uma sequência de variáveis aleatórias i.i.d, vale que

$$\mathbb{E}[e^{\lambda S_n}] = \mathbb{E}[e^{\lambda(X_1+\cdots+X_n)}] = \mathbb{E}[e^{\lambda X_1} \cdots e^{\lambda X_n}] = \mathbb{E}[e^{\lambda X_1}] \cdots \mathbb{E}[e^{\lambda X_n}] = \mathbb{E}[e^{\lambda X_1}]^n. \tag{1.1}$$

Denote a transformada de Laplace de X_1 por

$$\varphi(\lambda) = \mathbb{E}[e^{\lambda X_1}]. \tag{1.2}$$

Por (1.1) e com a notação introduzida em (1.2), temos que $\mathbb{E}[e^{\lambda S_n}] = \mathbb{E}[e^{\lambda X_1}]^n = \varphi(\lambda)^n$. E já que $X_1 \sim N(0,1)$, por (1.21), temos $\mathbb{E}[e^{\lambda S_n}] = e^{\frac{\lambda^2}{2}n}$. Assim,

$$\mathbb{P}\left[\frac{S_n}{n} \in [m, +\infty)\right] \leqslant e^{-n\left(\lambda m - \frac{\lambda^2}{2}\right)} \to 0,$$

se $\lambda m - \frac{\lambda^2}{2} > 0$. Para isso, basta escolher $\lambda < 2m$.

Portanto, temos um decaimento exponencial da probabilidade de $\left[\frac{S_n}{n} \in [m, +\infty)\right]$ e a taxa de decaimento é dada por $\lambda m - \frac{\lambda^2}{2}$. Se olharmos para $\frac{1}{n}\log \mathbb{P}\left[\frac{S_n}{n} \in [m, +\infty)\right]$, vemos que $\lambda m - \frac{\lambda^2}{2}$ está fazendo o papel da função I em (2). Por isso, I é chamada de função taxa.

Como podem notar, a função I herdará características do modelo estudado, pois a função $\frac{\lambda^2}{2}$ que apareceu acima é $\log \mathbb{E}[e^{\lambda X_1}]$. É claro que a função taxa I não é exatamente $\lambda m - \frac{\lambda^2}{2}$, pois dependeria da escolha de λ e não estaria bem definida. No texto que segue trabalharemos um pouco mais para encontrar a função taxa I.

1.1 O Teorema de Cramér

Agora, consideraremos o caso em que $\{X_n\}$ é uma sequência de variáveis aleatórias i.i.d integráveis tomando valores em $\mathbb{R}$ e, como antes, $S_n = X_1 + \cdots + X_n$. Vamos então começar encontrando uma cota superior para o caso $F = [m, +\infty)$. Neste caso, vamos inicialmente fazer uma estimativa para $\mathbb{P}\left[\frac{S_n}{n} \in [m, +\infty)\right]$ usando basicamente a mesma ideia do exemplo acima, aplicada ao caso mais geral. Note que

$$\mathbb{P}\left[\frac{S_n}{n} \in [m, +\infty)\right] = \mathbb{E}\left[\mathbb{1}_{\left[\frac{S_n}{n} \in [m,+\infty)\right]}\right],$$

onde $\mathbb{E}$ denota a esperança com respeito a probabilidade $\mathbb{P}$. Multiplicando e dividindo por $e^{\lambda S_n}$, para qualquer $\lambda \in \mathbb{R}$, temos

$$\mathbb{P}\left[\frac{S_n}{n} \in [m, +\infty)\right] = \mathbb{E}\left[\mathbb{1}_{\left[\frac{S_n}{n} \in [m,+\infty)\right]} e^{-\lambda S_n} e^{\lambda S_n}\right].$$

Se $\frac{S_n}{n} \in [m, +\infty)$, vale que $S_n \geqslant mn$. Então, para $\lambda \geqslant 0$, temos que $\lambda S_n \geqslant \lambda mn$ e, portanto, $e^{-\lambda S_n} \leqslant e^{-\lambda mn}$. Logo,

$$\mathbb{P}\left[\frac{S_n}{n} \in [m, +\infty)\right] = \mathbb{E}\left[\mathbb{1}_{\left[\frac{S_n}{n} \in [m,+\infty)\right]} e^{-\lambda S_n} e^{\lambda S_n}\right] \leqslant e^{-\lambda mn}\mathbb{E}[e^{\lambda S_n}] = e^{-\lambda mn}\varphi(\lambda)^n.$$

Então,

$$\frac{1}{n}\log \mathbb{P}\left[\frac{S_n}{n} \in [m, +\infty)\right] \leqslant \frac{1}{n}\log\left(e^{-\lambda mn}\varphi(\lambda)^n\right) = -\lambda m + \log\varphi(\lambda).$$

Defina

$$J_\theta(x) = \theta x - \log\varphi(\theta). \tag{1.3}$$

Portanto,

$$\overline{\lim_{n\to\infty}} \frac{1}{n}\log \mathbb{P}\left[\frac{S_n}{n} \in [m, +\infty)\right] \leqslant -J_\lambda(m), \text{ para qualquer } \lambda \geqslant 0. \tag{1.4}$$

Para o caso $F = (-\infty, M]$, utilizando um raciocínio análogo, concluímos que

$$\overline{\lim_{n\to\infty}} \frac{1}{n}\log \mathbb{P}\left[\frac{S_n}{n} \in (-\infty, M]\right] \leqslant -J_\alpha(M), \text{ para qualquer } \alpha \leqslant 0. \tag{1.5}$$

Teorema 1.1. *Seja $\{X_n\}$ uma sequência de variáveis aleatórias i.i.d. integráveis tomando valores em $\mathbb{R}$. Sejam $\varphi(\lambda) = \mathbb{E}[e^{\lambda X_1}]$, para $\lambda \in \mathbb{R}$ e*

$$I(x) = \sup_{\lambda\in\mathbb{R}} J_\lambda(x) = \sup_{\lambda\in\mathbb{R}}(\lambda x - \log\varphi(\lambda)), \tag{1.6}$$

para $x \in \mathbb{R}$.

Suponhamos que as seguintes hipóteses sejam cumpridas:

Hipótese (1). $\Phi^* := \{\lambda \in \mathbb{R} : \varphi(\lambda) < +\infty\}$ é um conjunto aberto.

Hipótese (2). Para cada $x \in \mathbb{R}$, $I(x) = +\infty$ ou $\exists \lambda_x \in \mathbb{R}$ tal que $I(x) = \lambda_x x - \log\varphi(\lambda_x)$.

Então, $\left\{\frac{1}{n}\sum_{k=1}^{n} X_k\right\}$ satisfaz o LDP com função taxa I definida em (1.6).

Observação 1. *Repare que* $0 \in \Phi^*$*, pois* $\varphi(0) = 1 < +\infty$*. Além disso, para* $\tilde{\lambda} \in (\Phi^*)^{\complement}$*, temos que* $\varphi(\tilde{\lambda}) = +\infty$ *e, portanto,* $J_{\tilde{\lambda}}(x) \equiv -\infty$*. Como* $\sup_{\lambda \in \mathbb{R}} J_\lambda(x) \geqslant J_0(x) = 0$*, podemos descartar todos os* $\tilde{\lambda} \in (\Phi^*)^{\complement}$ *e concluir que o supremo na definição de* I *em* (1.6) *pode ser calculado apenas sobre os* $\lambda \in \Phi^*$*, ou seja,*

$$I(x) := \sup_{\lambda \in \mathbb{R}} J_\lambda(x) = \sup_{\lambda \in \Phi^*} J_\lambda(x).$$

Observação 2. *O que a equação* (1.6) *diz é que a função* I *é a transformada de Legendre de* $\log \varphi(\lambda)$*. Esta relação voltará a aparece no Capítulo 4 quando apresentarmos o Teorema de Gärtner-Ellis.*

Demonstração. Dividiremos a demonstração deste teorema em três partes. Na primeira parte, provaremos que a função I satisfaz as hipóteses para ser função taxa de um LDP (não negativa e semicontínua inferiormente). Em seguida, mostraremos a cota superior de duas maneiras diferentes. Por último, mostraremos a cota inferior.

(i) Note que $\varphi(0) = \mathbb{E}[e^{0X_1}] = 1$ e, portanto,

$$I(x) = \sup_{\lambda \in \Phi^*} (\lambda x - \log \varphi(\lambda)) \geqslant 0x - \log \varphi(0) = 0.$$

Logo, a função I realmente toma valores no intervalo $[0, +\infty)$. O fato de I ser uma função semicontínua inferiormente segue diretamente do Lema A.2, já que todas as J_λ são contínuas em Φ^* (veja Lema A.24). Então, I satisfaz as hipóteses do LDP e é uma candidata à função taxa.

(ii) Provaremos a cota superior de duas maneiras diferentes. A primeira delas é mais simples, mas só funciona no caso particular do conjunto R (que aparece na definição de LDP) ser $\mathbb{R}$. A segunda é mais geral, porém requer argumentos mais complexos.

Prova 1: Seja $I^+(m) = \sup_{\lambda \geqslant 0} J_\lambda(m)$, onde $J_\lambda(m)$ foi definido em (1.3). Minimizando em $\lambda \geqslant 0$ a inequação (1.4), obtemos

$$\overline{\lim_{n \to \infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in [m, +\infty)\right] \leqslant -I^+(m). \tag{1.7}$$

Analogamente, minimizando em $\alpha \leqslant 0$ a inequação (1.5), concluímos que

$$\overline{\lim_{n \to \infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in (-\infty, M]\right] \leqslant -I^-(M), \tag{1.8}$$

com $I^-(M) = \sup_{\alpha \leqslant 0} J_\alpha(M)$.

Seja $a = \mathbb{E}[X_1]$. Utilizando a desigualdade de Jensen (Teorema A.10) com $\psi(x) = e^{\theta x}$, para qualquer $\theta \in \mathbb{R}$, temos que $\varphi(\theta) = \mathbb{E}[e^{\theta X_1}] \geqslant e^{\theta \mathbb{E}[X_1]} = e^{\theta a}$. Logo, para qualquer $\theta \in \mathbb{R}$, $J_\theta(a) \leqslant \theta a - \log e^{\theta a} = 0$. Portanto, $I^-(a) \leqslant 0$ e $I^+(a) \leqslant 0$. Por outro lado, utilizando o mesmo raciocínio do item anterior, podemos mostrar que $I^-(x) \geqslant 0$ e $I^+(x) \geqslant 0$, para qualquer $x \in \mathbb{R}$. Logo, $I^-(a) = 0 = I^+(a)$.

Para $b \leqslant c$, temos que $\lambda b \leqslant \lambda c$ para qualquer $\lambda \geqslant 0$. Então,

$$I^+(b) = \sup_{\lambda \geqslant 0}(\lambda b - \log \varphi(\lambda)) \leqslant \sup_{\lambda \geqslant 0}(\lambda c - \log \varphi(\lambda)) = I^+(c).$$

Então, I^+ é uma função não decrescente. Além disso, para $d \leqslant e$, vale que $\lambda d \geqslant \lambda e$ para qualquer $\lambda \leqslant 0$. Portanto, $I^-(d) \geqslant I^-(e)$, ou seja, I^- é não crescente.

Seja F um conjunto fechado. Se $a \in F$, pela Lei dos Grandes Números, concluímos que $\mathbb{P}\left[\frac{S_n}{n} \in F\right] \to 1$. Além disso, como $a \in F$, $\inf_{x \in F} I(x) \leqslant I(a) = 0$. Então,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in F\right] = 0 \leqslant -\inf_{x\in F} I(x),$$

ou seja, a condição de cota superior é satisfeita.

Para o caso $a \notin F$, reescrevemos

$$I(y) = \sup_{\theta \in \mathbb{R}}(\theta y - \log \varphi(\theta)) = \max\left\{\sup_{\alpha \leqslant 0} J_\alpha(y), \sup_{\lambda \geqslant 0} J_\lambda(y)\right\} = \max\left\{I^-(y), I^+(y)\right\}.$$

Logo, considerando que I^- é não crescente, I^+ é não descrescente e $I^-(a) = 0 = I^+(a)$, concluímos que

$$I(y) = I^-(y)\mathbb{1}_{(-\infty,a)}(y) + I^+(y)\mathbb{1}_{[a,+\infty)}(y). \tag{1.9}$$

Como $F^\complement$ é aberto e $a \in F^\complement$, seja $I_F = (y_1, y_2)$ o maior intervalo aberto tal que $a \in I_F$ e $I_F \cap F = \emptyset$. Como $I_F \cap F = \emptyset$, temos que $F \subseteq {I_F}^\complement = (-\infty, y_1] \cup [y_2, +\infty)$. Então, denotando $F^- = F \cap (-\infty, y_1]$ e $F^+ = F \cap [y_2, +\infty)$, temos que $F = F^- \cup F^+$. Assim, podemos escrever

$$\inf_{y \in F} I(y) = \min\left\{\inf_{y \in F^-} I(y), \inf_{y \in F^+} I(y)\right\}. \tag{1.10}$$

Considerando que, para quaisquer $x \in F^+$ e $z \in F^-$, vale que $\mathbb{1}_{(-\infty,a)}(x) = 0$, $\mathbb{1}_{(-\infty,a)}(z) = 1$, $\mathbb{1}_{[a,+\infty)}(x) = 1$ e $\mathbb{1}_{[a,+\infty)}(z) = 0$, segue de (1.9) e (1.10) que

$$\inf_{y \in F} I(y) = \min\left\{\inf_{y \in F^-} I^-(y), \inf_{y \in F^+} I^+(y)\right\} = \min\left\{I^-(y_1), I^+(y_2)\right\}, \tag{1.11}$$

onde a última igualdade segue do fato de I^- ser não crescente e I^+ ser não decrescente. Como $F \subseteq (-\infty, y_1] \cup [y_2, +\infty)$, sabemos que

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in F\right] \leqslant \overline{\lim_{n\to\infty}} \frac{1}{n} \log\left(\mathbb{P}\left[\frac{S_n}{n} \in (-\infty, y_1]\right] + \mathbb{P}\left[\frac{S_n}{n} \in [y_2, +\infty)\right]\right).$$

Aplicando o Lema A.3, com $a_n = \mathbb{P}\left[\frac{S_n}{n} \in (-\infty, y_1]\right]$ e $b_n = \mathbb{P}\left[\frac{S_n}{n} \in [y_2, +\infty)\right]$, obtemos

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in F\right] \leqslant \max\left\{\overline{\lim_{n\to\infty}} \frac{1}{n} \log a_n, \overline{\lim_{n\to\infty}} \frac{1}{n} \log b_n\right\}.$$

Utilizando as desigualdades (1.7) e (1.8) e a igualdade (1.11), concluímos que

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in F\right] \leqslant \max\{-I^-(y_1), -I^+(y_2)\} = -\inf_{y\in F} I(y),$$

provando a condição de cota superior do LDP.

Prova 2: Começaremos encontrando cotas superiores quando $F = K$, onde K é um conjunto compacto em $\mathbb{R}$. Dado $\lambda \in \Phi^* = \{\lambda \in \mathbb{R} : \varphi(\lambda) < +\infty\}$, podemos escrever

$$\mathbb{P}\left[\frac{S_n}{n} \in K\right] = \mathbb{E}\left[\mathbb{1}_{\left[\frac{S_n}{n}\in K\right]} e^{-\lambda S_n} e^{\lambda S_n}\right].$$

Como anteriormente, queremos estimar essa probabilidade. Para isso, observe que

$$\mathbb{1}_{\left[\frac{S_n}{n}\in K\right]} e^{-\lambda S_n} \leqslant \sup_{\frac{x}{n}\in K} e^{-\lambda x} = \sup_{y\in K} e^{-\lambda n y} = e^{\sup\limits_{y\in K}(-\lambda n y)} = e^{-n \inf\limits_{y\in K} \lambda y}.$$

Vale ressaltar que o termo λy acima é um produto de números reais, pois estamos trabalhando com $E = \mathbb{R}$. Porém, como já dito anteriormente, essa demonstração é mais geral e pode ser estendida para outros tipos de conjunto E. Por exemplo, se $E = \mathbb{R}^d$, teríamos $\lambda \in \mathbb{R}^d$ e o termo λy seria um produto interno em $\mathbb{R}^d$. Seja $\mathcal{M}(A)$ o conjunto de todas as probabilidades sobre o espaço A. Se $E = \mathcal{M}(A)$, poderíamos generalizar λy como $\int y d\lambda$, onde y seria uma função e λ seria uma probabilidade sobre A.

Continuando os cálculos para o caso $E = \mathbb{R}$, note que, se $\lambda \geqslant 0$,

$$\inf_{y\in K} \lambda y = \lambda \inf_{y\in K} y$$

e, por outro lado, se $\lambda < 0$,

$$\inf_{y\in K} \lambda y = \lambda \sup_{y\in K} y.$$

Então, para $\inf\limits_{y\in K} \lambda y$ ser finito, precisamos que $\inf\limits_{y\in K} y$ e $\sup\limits_{y\in K} y$ sejam ambos finitos, o que é verdade, pois K é compacto. Logo,

$$\mathbb{P}\left[\frac{S_n}{n} \in K\right] = \mathbb{E}\left[\mathbb{1}_{\left[\frac{S_n}{n}\in K\right]} e^{-\lambda S_n} e^{\lambda S_n}\right] \leqslant e^{-n \inf\limits_{x\in K} \lambda x} \underbrace{\mathbb{E}[e^{\lambda S_n}]}_{\mathbb{E}[e^{\lambda X_1}]^n} = e^{-n \inf\limits_{x\in K} \lambda x} \varphi(\lambda)^n.$$

Então,

$$\frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in K\right] \leqslant -\inf_{x\in K} \lambda x + \log \varphi(\lambda).$$

Portanto, para quaisquer $\lambda \in \Phi^*$ e $n \in \mathbb{N}$,

$$\frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in K\right] \leqslant -\left(\inf_{x\in K} \lambda x - \log \varphi(\lambda)\right) = -\inf_{x\in K} (\lambda x - \log \varphi(\lambda)).$$

Pela definição de J_λ em (1.3) segue que, para quaisquer $\lambda \in \Phi^*$ e $n \in \mathbb{N}$,

$$\frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in K\right] \leqslant -\inf_{x\in K} J_\lambda(x).$$

Logo, para qualquer $\lambda \in \Phi^*$,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in K\right] \leqslant -\inf_{x\in K} J_\lambda(x). \tag{1.12}$$

Queremos então generalizar essa condição para qualquer fechado. Para isso, vamos mostrar que a sequência $\left\{\frac{S_n}{n}\right\}$ é exponencialmente rígida (ver Definição A.18) e usar a Proposição A.19. Dado $a > 0$, seja $K_a = [-j, j]$, onde $j = \max\{a + \log\varphi(-1), a + \log\varphi(1)\}$. Então,

$$\mathbb{P}\left[\frac{S_n}{n} \notin K_a\right] = \mathbb{P}\left[\frac{S_n}{n} \notin [-j, j]\right] = \mathbb{P}\left[\frac{S_n}{n} \in (-\infty, -j) \cup (j, +\infty)\right].$$

Como $\mathbb{P}$ é uma probabilidade, podemos reescrever a equação acima como

$$\mathbb{P}\left[\frac{S_n}{n} \notin K_a\right] \leqslant \mathbb{P}\left[\frac{S_n}{n} \in (-\infty, -j]\right] + \mathbb{P}\left[\frac{S_n}{n} \in [j, +\infty)\right].$$

Portanto,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \notin K_a\right] \leqslant \overline{\lim_{n\to\infty}} \frac{1}{n} \log\left(\mathbb{P}\left[\frac{S_n}{n} \in (-\infty, -j]\right] + \mathbb{P}\left[\frac{S_n}{n} \in [j, +\infty)\right]\right).$$

Aplicando o Lema A.3, com $a_n = \mathbb{P}\left[\frac{S_n}{n} \in (-\infty, -j]\right]$, $b_n = \mathbb{P}\left[\frac{S_n}{n} \in [j, +\infty)\right]$ e $c_n = n$, obtemos

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \notin K_a\right] \leqslant \max\left\{\overline{\lim_{n\to\infty}} \frac{1}{n} \log a_n, \overline{\lim_{n\to\infty}} \frac{1}{n} \log b_n\right\}.$$

Considerando as igualdades (1.4) e (1.5) para $\lambda = 1$, $m = j$, $\alpha = -1$ e $M = -j$ e substituindo na desigualdade acima, concluímos que

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \notin K_a\right] \leqslant \max\{-j + \log\varphi(-1), -j + \log\varphi(1)\}.$$

Lembre que $j = \max\{a + \log\varphi(-1), a + \log\varphi(1)\}$. Portanto, $-j \leqslant -a - \log\varphi(-1)$ e $-j \leqslant -a - \log\varphi(1)$. Consequentemente,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \notin K_a\right] \leqslant \max\{-a - \log\varphi(-1) + \log\varphi(-1), -a - \log\varphi(1) + \log\varphi(1)\}.$$

Então,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \notin K_a\right] \leqslant \max\{-a, -a\} = -a.$$

Logo, $\{X_n\}$ é exponencialmente rígida.

A generalização de (1.12) segue do fato de $\{X_n\}$ ser exponencialmente rígida e da Proposição A.19, considerando $f(x) = J_\lambda(x)$, ou seja,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in F\right] \leqslant -\inf_{x\in F} J_\lambda(x), \tag{1.13}$$

para qualquer F fechado.

Sabemos que, para qualquer O aberto, $\mathbb{P}\left[\frac{S_n}{n} \in O\right] \leqslant \mathbb{P}\left[\frac{S_n}{n} \in \overline{O}\right]$, onde $\overline{O}$ denota o fecho do conjunto O. Então,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in O\right] \leqslant \overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in \overline{O}\right].$$

Como $\overline{O}$ é fechado, utilizando (1.13), temos que

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in O\right] \leqslant - \inf_{x\in\overline{O}} J_\lambda(x) = \sup_{x\in\overline{O}} \widetilde{J}_\lambda(x),$$

onde $\widetilde{J}_\lambda(x) = -J_\lambda(x)$. Utilizando o fato de todas $\widetilde{J}_\lambda$ serem contínuas, segue que

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in O\right] \leqslant \sup_{x\in O} \widetilde{J}_\lambda(x).$$

Minimizando em λ a inequação acima, obtemos

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in O\right] \leqslant \inf_{\lambda\in\Phi^*} \sup_{x\in O} \widetilde{J}_\lambda(x).$$

Assim, como as $\widetilde{J}_\lambda$ são contínuas, podemos aplicar o Lema A.21 e concluir, para K compacto, que

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in K\right] \leqslant \sup_{x\in K} \inf_{\lambda\in\Phi^*} \widetilde{J}_\lambda(x) = - \inf_{x\in K} \sup_{\lambda\in\Phi^*} J_\lambda(x) = - \inf_{x\in K} I(x),$$

onde $I(x) = \sup\limits_{\lambda\in\Phi^*} J_\lambda(x)$, pela observação anterior. Aplicando a Proposição A.19 para $f = I$, mostramos que vale a cota superior do LDP.

(iii) Agora vamos mostrar a cota inferior. Para isso, seja O um conjunto aberto. Definimos $I^* = \{x \in \mathbb{R} : I(x) < +\infty\}$ e $O^* = O \cap I^*$.

Se $O^* = \emptyset$, então $I(x) = +\infty$, para todo $x \in O$. Assim, $\inf\limits_{x\in O} I(x) = +\infty$ e, portanto, vale a cota inferior.

Se $O^* \neq \emptyset$, tomamos $y \in O^*$. Em particular, $y \in O$. Como O é aberto, existe $\delta_0 > 0$ tal que $(y - \delta_0, y + \delta_0) \subseteq O$. Seja $\delta > 0$ tal que $\delta \leqslant \delta_0$. Note que $(y - \delta, y + \delta) \subseteq O$. Por outro lado, $y \in I^*$, ou seja, $I(y) < +\infty$. Pela *Hipótese (2)*, existe λ_y tal que $I(y) = \lambda_y y - \log\varphi(\lambda_y)$, ou seja, $\sup\limits_{\lambda\in\mathbb{R}}(\lambda y - \log\varphi(\lambda))$ é atingido em $\lambda = \lambda_y \in \mathbb{R}$. Logo, λ_y maximiza a função $\lambda y - \log\varphi(\lambda)$ e, portanto,

$$0 = y - \frac{\varphi'(\lambda_y)}{\varphi(\lambda_y)} = y - \frac{\mathbb{E}[X_1 e^{\lambda_y X_1}]}{\mathbb{E}[e^{\lambda_y X_1}]}. \tag{1.14}$$

Além disso,

$$\mathbb{P}\left[\frac{S_n}{n} \in O\right] = \mathbb{E}\left[\mathbb{1}_{\left[\frac{S_n}{n}\in O\right]} e^{-\lambda_y S_n} e^{\lambda_y S_n}\right] \geqslant \mathbb{E}\left[\mathbb{1}_{\left[\frac{S_n}{n}\in(y-\delta,y+\delta)\right]} e^{-\lambda_y S_n} e^{\lambda_y S_n}\right].$$

Observe que se $\frac{S_n}{n} \in (y - \delta, y + \delta)$, então $e^{-\lambda_y S_n} \geqslant \inf\limits_{\frac{z}{n}\in(y-\delta,y+\delta)} e^{-\lambda_y z}$. E como

$$\inf_{\frac{z}{n}\in(y-\delta,y+\delta)} e^{-\lambda_y z} = \inf_{x\in(y-\delta,y+\delta)} e^{-\lambda_y n x} = \exp\left(-n \sup_{x\in(y-\delta,y+\delta)} \lambda_y x\right),$$

temos que

$$\begin{aligned}\mathbb{1}_{\left[\frac{S_n}{n}\in(y-\delta,y+\delta)\right]}e^{-\lambda_y S_n} &\geqslant \mathbb{1}_{\left[\frac{S_n}{n}\in(y-\delta,y+\delta)\right]} \inf_{\frac{z}{n}\in(y-\delta,y+\delta)} e^{-\lambda_y z}\\ &\geqslant \mathbb{1}_{\left[\frac{S_n}{n}\in(y-\delta,y+\delta)\right]} \exp\left(-n \sup_{x\in(y-\delta,y+\delta)} \lambda_y x\right).\end{aligned}$$

Pelo mesmo raciocínio utilizado quando fizemos a cota superior, concluímos que

$$\sup_{x\in(y-\delta,y+\delta)} \lambda_y x = \begin{cases} \lambda_y \sup\limits_{x\in(y-\delta,y+\delta)} x = \lambda_y(y+\delta), & se \quad \lambda_y \geqslant 0\\ \lambda_y \inf\limits_{x\in(y-\delta,y+\delta)} x = \lambda_y(y-\delta), & se \quad \lambda_y < 0\end{cases}.$$

Logo,

$$\sup_{x\in(y-\delta,y+\delta)} \lambda_y x = \lambda_y y + |\lambda_y|\delta.$$

Assim,

$$\mathbb{P}\left[\frac{S_n}{n}\in O\right] \geqslant e^{-n\lambda_y y}e^{-n|\lambda_y|\delta}\mathbb{E}[e^{\lambda_y S_n}]\frac{\mathbb{E}\left[\mathbb{1}_{\left[\frac{S_n}{n}\in(y-\delta,y+\delta)\right]}e^{\lambda_y S_n}\right]}{\mathbb{E}[e^{\lambda_y S_n}]}. \tag{1.15}$$

Definimos a medida μ_{λ_y} como $\mu_{\lambda_y}(A) = \dfrac{\mathbb{E}\left[\mathbb{1}_A(X_1)e^{\lambda_y X_1}\right]}{\mathbb{E}[e^{\lambda_y X_1}]}$, para qualquer $A\in\mathcal{B}(\mathbb{R})$. Pelo Corolário A.17, existe uma sequência $\{Y_n\}$ de variáveis aleatórias i.i.d com distribuição dada por μ_{λ_y}. Seja $\widetilde{\mathbb{P}}_{\lambda_y}$ a distribuição conjunta da sequência $\{Y_i\}$. Assim, usando (1.14), concluímos que

$$\mathbb{E}_{\lambda_y}[Y_1] = \int y\mu_{\lambda_y}(dy) = \frac{\mathbb{E}[X_1 e^{\lambda_y X_1}]}{\mathbb{E}[e^{\lambda_y X_1}]} = y,$$

onde $\mathbb{E}_{\lambda_y}$ denota a esperança com respeito a probabilidade μ_{λ_y}. Além disso,

$$\begin{aligned}\widetilde{\mathbb{P}}_{\lambda_y}\left[\left|\frac{1}{n}\sum_{i=1}^n Y_i - y\right| < \delta\right] &= \int\cdots\int \mathbb{1}_{\left[\left|\frac{1}{n}\sum_{i=1}^n y_i - y\right|<\delta\right]}\mu_{\lambda_y}(dy_1)\cdots\mu_{\lambda_y}(dy_n)\\ &= \frac{\mathbb{E}\left[\mathbb{1}_{\left[\left|\frac{1}{n}\sum_{i=1}^n X_i - y\right|<\delta\right]}\exp\left(\lambda_y\sum_{i=1}^n X_i\right)\right]}{\mathbb{E}\left[\exp\left(\lambda_y\sum_{i=1}^n X_i\right)\right]} = \frac{\mathbb{E}\left[\mathbb{1}_{\left[\frac{S_n}{n}\in(y-\delta,y+\delta)\right]}e^{\lambda_y S_n}\right]}{\mathbb{E}[e^{\lambda_y S_n}]}.\end{aligned}$$

Pela Lei dos Grandes Números, $\widetilde{\mathbb{P}}_{\lambda_y}\left[\left|\frac{1}{n}\sum_{k=1}^n Y_k - y\right| < \delta\right] \to 1$, quando $n\to\infty$. Então, quando $n\to\infty$,

$$\log\left(\frac{\mathbb{E}\left[\mathbb{1}_{\left[\frac{S_n}{n}\in(y-\delta,y+\delta)\right]}e^{\lambda_y S_n}\right]}{\mathbb{E}[e^{\lambda_y S_n}]}\right) \to 0.$$

Assim, segue de (1.15) que

$$\varliminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left[\frac{S_n}{n}\in O\right] \geqslant -\lambda_y y - |\lambda_y|\delta + \log\varphi(\lambda_y).$$

Fazendo $\delta \to 0$, obtemos que

$$\varliminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in O\right] \geqslant -\lambda_y y + \log\varphi(\lambda_y) = -I(y). \tag{1.16}$$

Como a inequação anterior vale para qualquer $y \in O^*$, temos que

$$\varliminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in O\right] \geqslant \sup_{y\in O^*} -I(y) = -\inf_{y\in O^*} I(y).$$

Note que $I(z) = +\infty$, para qualquer $z \in O\backslash O^*$, ou seja, $\inf\limits_{y\in O^*} I(y) = \inf\limits_{y\in O} I(y)$. Logo,

$$\varliminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in O\right] \geqslant -\inf_{y\in O} I(y).$$

□

Na prática, quando queremos aplicar este Teorema, muitas vezes temos dificuldades em provar que certa sequência $\{X_n\}$ satisfaz a *Hipótese (2)*. Em alguns casos específicos, podemos facilitar este trabalho através do Corolário abaixo, que nos dá uma outra hipótese mais forte.

Corolário 1.2. *O Teorema anterior ainda é válido se substituirmos a Hipótese (2) por: Hipótese (2').* $\lim\limits_{|\lambda|\to+\infty} \dfrac{\log\varphi(\lambda)}{|\lambda|} = +\infty$.

Demonstração. Fixado $x \in \mathbb{R}$. Pela *Hipótese (2')*, temos que existe $\widetilde{\lambda}_x \geqslant 0$ tal que

$$\frac{\log\varphi(\lambda)}{|\lambda|} \geqslant |x|, \tag{1.17}$$

para qualquer λ com $|\lambda| > \widetilde{\lambda}_x$. Vamos separar em 2 casos: $x \geqslant 0$ e $x < 0$.

<u>Caso 1:</u> $x \geqslant 0$

Se $\lambda > \widetilde{\lambda}_x$, reescrevemos (1.17) como $\dfrac{\log\varphi(\lambda)}{\lambda} \geqslant x$. Assim, $\lambda x - \log\varphi(\lambda) \leqslant 0$. Por outro lado, se $\lambda < -\widetilde{\lambda}_x$, podemos reescrever (1.17) como $\dfrac{\log\varphi(\lambda)}{-\lambda} \geqslant x$. Então,

$$-\log\varphi(\lambda) \leqslant \lambda x. \tag{1.18}$$

Somando λx nos dois lados da desigualdade (1.18), obtemos que

$$\lambda x - \log\varphi(\lambda) \leqslant 2\lambda x \leqslant 0.$$

Portanto,

$$\sup_{\lambda<-\widetilde{\lambda}_x} [\lambda x - \log\varphi(\lambda)] \leqslant 0,$$

$$\sup_{\lambda>\widetilde{\lambda}_x} [\lambda x - \log\varphi(\lambda)] \leqslant 0.$$

Na demonstração do Teorema anterior, provamos que $I(x) = \sup_{\lambda \in \mathbb{R}} [\lambda x - \log \varphi(\lambda)] \geqslant 0$. Através das desigualdades obtidas acima, concluímos que

$$I(x) = \sup_{\lambda \in [-\widetilde{\lambda}_x, \widetilde{\lambda}_x]} [\lambda x - \log \varphi(\lambda)].$$

Como a função $\lambda x - \log \varphi(\lambda)$ é contínua em λ, existe algum $\lambda_x \in [-\widetilde{\lambda}_x, \widetilde{\lambda}_x]$ para o qual o supremo acima é atingido e, portanto,

$$I(x) = \lambda_x x - \log \varphi(\lambda_x).$$

<u>Caso 2:</u> $x < 0$

Neste caso, reescrevemos (1.17) como $\dfrac{\log \varphi(\lambda)}{|\lambda|} \geqslant -x$. Para $\lambda < 0$, basta repetir o argumento utilizado no caso anterior quando tínhamos $\lambda \geqslant 0$. Para $\lambda \geqslant 0$, repetimos o argumento utilizado anteriormente quando $\lambda < 0$.

Logo, para qualquer $x \in \mathbb{R}$, existe λ_x tal que

$$I(x) = \lambda_x x - \log \varphi(\lambda_x).$$

Disto segue que a *Hipótese (2)* do Teorema anterior é satisfeita e, portanto, vale o LDP. $\square$

A seguir, trataremos do Teorema de Cramér, que nos permite obter a mesma conclusão do Teorema 1.1 apenas com a hipótese de $\varphi(\lambda) < +\infty$, para todo $\lambda \in \mathbb{R}$. Antes, porém, vamos provar dois lemas que serão necessários na demonstração dele.

Lema 1.3. *Com as notações introduzidas anteriormente, suponhamos que $\varphi(\lambda) < +\infty$, para todo $\lambda \in \mathbb{R}$. Denotamos por μ a distribuição de X_1 e $a = \mathbb{E}[X_1]$. Fixado $x \in \mathbb{R}$ para o qual $I(x) < +\infty$ e $\nexists \lambda_x$ tal que $I(x) = \lambda_x x - \log \varphi(\lambda_x)$. Então, $\mu(x, +\infty) = 0$, se $x > a$ e $\mu(-\infty, x) = 0$, se $x < a$ e, em ambos os casos, $\mu(\{x\}) = e^{-I(x)}$.*

Demonstração. Como $I(x) = \sup_{\lambda \in \mathbb{R}} (\lambda x - \log \varphi(\lambda))$, existe uma sequência $\{\tilde{\lambda}_n\} \subseteq \mathbb{R}$ tal que $\tilde{\lambda}_n x - \log \varphi(\tilde{\lambda}_n) \to I(x)$, quando $n \to \infty$.

Suponhamos que $\{\tilde{\lambda}_n\}$ possui um ponto de acumulação $\tilde{\lambda} \in \mathbb{R}$. Então existe uma subsequência $\{\tilde{\lambda}_{n_k}\}$ tal que $\tilde{\lambda}_{n_k} x - \log \varphi(\tilde{\lambda}_{n_k}) \to \tilde{\lambda} x - \log \varphi(\tilde{\lambda})$, quando $k \to \infty$, pois a função $\lambda \mapsto J_\lambda(x)$ é contínua em λ. Disto segue que $I(x) = \tilde{\lambda} x - \log \varphi(\tilde{\lambda})$, o que contraria a hipótese sobre x. Logo, os pontos de acumulação de $\{\tilde{\lambda}_n\}$ são $+\infty$ ou $-\infty$.

Pela desigualdade de Jensen, temos que $\log \varphi(\tilde{\lambda}_n) = \log \mathbb{E}[e^{\tilde{\lambda}_n X_1}] \geqslant \tilde{\lambda}_n \mathbb{E}[X_1] = \tilde{\lambda}_n a$. Então, quando $n \to \infty$,

$$\tilde{\lambda}_n (x - a) = \tilde{\lambda}_n x - \tilde{\lambda}_n a \geqslant \tilde{\lambda}_n x - \log \varphi(\tilde{\lambda}_n) \to I(x) \geqslant 0.$$

Suponhamos que $x > a$. Se $\tilde{\lambda}_n \to -\infty$, então $\tilde{\lambda}_n x - \log \varphi(\tilde{\lambda}_n) \leqslant \tilde{\lambda}_n (x - a) \to -\infty$ e, portanto, $I(x) = -\infty$, o que é um absurdo, já que I é positiva. Portanto, obtemos que $\tilde{\lambda}_n \to +\infty$.

Como $\tilde{\lambda}_n x - \log \varphi(\tilde{\lambda}_n) \to I(x)$, concluímos que $e^{-\tilde{\lambda}_n x}\varphi(\tilde{\lambda}_n) \to e^{-I(x)}$. Além disso, usando que $e^{\tilde{\lambda}_n(y-x)} \to 0$, quando $y < x$ e o fato de $\varphi(\lambda) < +\infty$, para todo $\lambda \in \mathbb{R}$, concluímos que a função $y \mapsto e^{\tilde{\lambda}_n(y-x)}$ é integrável e

$$\int_{(-\infty,x)} e^{\tilde{\lambda}_n(y-x)} d\mu(y) \to 0.$$

Além disso, podemos reescrever

$$\int_{[x,+\infty)} e^{\tilde{\lambda}_n(y-x)} d\mu(y) = \int_{\mathbb{R}} e^{\tilde{\lambda}_n(y-x)} d\mu(y) - \int_{(-\infty,x)} e^{\tilde{\lambda}_n(y-x)} d\mu(y).$$

Repare que o primeiro termo da soma acima pode ser reescrito como $e^{-\tilde{\lambda}_n x}\varphi(\tilde{\lambda}_n)$, que converge a $e^{-I(x)}$, conforme provamos antes. Ainda, o segundo termo da soma converge a 0. Portanto,

$$\int_{[x,+\infty)} e^{\tilde{\lambda}_n(y-x)} d\mu(y) \to e^{-I(x)} < +\infty.$$

Note que

$$\int_{[x,+\infty)} e^{\tilde{\lambda}_n(y-x)} d\mu(y) = \int_{\{x\}} e^{\tilde{\lambda}_n(y-x)} d\mu(y) + \int_{(x,+\infty)} e^{\tilde{\lambda}_n(y-x)} d\mu(y).$$

Observe que $e^{\tilde{\lambda}_n(y-x)} = 1$, se $y = x$, ou seja, a primeira parcela da soma acima vale $\mu(\{x\})$. Por outro lado, $e^{\tilde{\lambda}_n(y-x)} \to +\infty$, se $y > x$ e, portanto, $\mu(x, +\infty) = 0$, já que a integral converge para um valor finito.

Para o caso $x < a$ basta repetir esse mesmo argumento. Já a conclusão de que $\mu(\{x\}) = e^{-I(x)}$ segue do fato de

$$e^{-I(x)} = \lim_{n\to\infty} \int_{\{x\}} e^{\tilde{\lambda}_n(y-x)} d\mu(y) = \mu(\{x\}),$$

pelo que provamos antes. □

Lema 1.4. *Se $x \in O$ e vale que $\mu(\{x\}) = e^{-I(x)}$, então*

$$\frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in O\right] \geqslant -I(x).$$

Demonstração. Como $x \in O$, temos que

$$\mathbb{P}\left[\frac{S_n}{n} \in O\right] \geqslant \mathbb{P}\left[\frac{S_n}{n} = x\right] = \int \cdots \int \mathbb{1}_{\{y_1+\cdots+y_n=nx\}} d\mu(y_1)\cdots d\mu(y_n).$$

Repare que $\{y_1 = x, \ldots, y_n = x\} \subseteq \{y_1 + \cdots + y_n = nx\}$. Portanto, segue da desiguadade acima que

$$\mathbb{P}\left[\frac{S_n}{n} \in O\right] \geqslant \int \cdots \int \mathbb{1}_{\{y_1=x,\ldots,y_n=x\}} d\mu(y_1)\cdots d\mu(y_n) \geqslant \int \mathbb{1}_{y_1=x} d\mu(y_1)\cdots \int \mathbb{1}_{y_n=x} d\mu(y_n).$$

Note que cada parcela do produto acima vale $\mu(\{x\})$. Logo,

$$\mathbb{P}\left[\frac{S_n}{n} \in O\right] \geqslant \mu(\{x\})^n = e^{-nI(x)},$$

pela hipótese. Assim, concluímos que

$$\frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in O\right] \geqslant \frac{1}{n} \log e^{-nI(x)} = -I(x).$$

□

Teorema 1.5 (Teorema de Cramér)**.** [1] *Seja $\{X_n\}$ uma sequência de variáveis aleatórias i.i.d. tomando valores em $\mathbb{R}$. Suponhamos que $\varphi(\lambda) = \mathbb{E}[e^{\lambda X_1}] < +\infty$, para todo $\lambda \in \mathbb{R}$. Então, $\{\frac{1}{n}\sum\limits_{k=1}^{n} X_k\}$ satisfaz o LDP com função taxa I definida em* (1.6)*.*

Demonstração. Pela Proposição A.22, considerando $k = 1$, concluímos que X_1 é integrável. Seja $a = \mathbb{E}[X_1]$. A demonstração da cota superior é a mesma feita no Teorema 1.1, já que ela utiliza apenas o fato de X_1 ser integrável. Para a cota inferior, vamos tomar O um aberto, fixar $x \in O$ e provar que

$$\varliminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in O\right] \geqslant -I(x). \tag{1.19}$$

Essa parte será dividida em vários casos:

1. $I(x) = +\infty \Rightarrow$ não há o que provar;

2. $I(x) < +\infty$:

   2.1 $x = a \Rightarrow$ Neste caso, $I(a) = 0$. Além disso, pela Lei dos Grandes Números, temos que $\mathbb{P}\left[\frac{S_n}{n} \in O\right] \to 1$ e, portanto, vale (1.19);

   2.2 $x \neq a$:

   2.2.1 Existe λ_x tal que $I(x) = \lambda_x x - \log \varphi(\lambda_x) \Rightarrow$ Seguimos a prova do Teorema 1.1 até (1.16);

   2.2.2 Não existe λ_x tal que $I(x) = \lambda_x x - \log \varphi(\lambda_x) \Rightarrow$ Pelo Lema 1.3, segue que $\mu(\{x\}) = e^{-I(x)}$. Então, a desigualdade desejada é consequência direta do Lema 1.4.

Concluímos que, para qualquer $x \in O$, vale (1.19). Para concluir a demonstração da cota inferior, basta maximizar esta desigualdade em $x \in O$. $\square$

1.2 Exemplos

Agora, vamos encontrar a função taxa I para algumas das distribuições de probabilidades mais conhecidas que satisfazem as hipóteses de algum dos resultados anteriores.

Exemplo 1.1. *Sejam $\{X_n\}$ variáveis aleatórias i.i.d com distribuição de Bernoulli simétrica, isto é, $\mathbb{P}[X_1 = 0] = \mathbb{P}[X_1 = 1] = \dfrac{1}{2}$. Seja $S_n = \sum\limits_{i=1}^{n} X_i$. Vamos encontrar a função taxa I para esta sequência.*

Lembre que em (1.6) tínhamos definido $I(x) = \sup\limits_{\lambda\in\mathbb{R}}(\lambda x - \log\varphi(\lambda))$, onde $\varphi(\lambda) = \mathbb{E}[e^{\lambda X_1}]$. Para a sequência definida neste exemplo, temos

$$\varphi(\lambda) = \sum_{x\in\{0,1\}} e^{\lambda x}\mathbb{P}[X_1 = x] = e^0\mathbb{P}[X_1 = 0] + e^{\lambda}\mathbb{P}[X_1 = 1] = \frac{1}{2}(1 + e^{\lambda}).$$

[1] Adaptado de Stroock (1984)

Substituindo em $I(x)$, obtemos

$$I(x) = \sup_{\lambda\in\mathbb{R}} \Big(\lambda x - \log\frac{1}{2}(1+e^{\lambda})\Big) = \sup_{\lambda\in\mathbb{R}} \Big(\lambda x - \log\frac{1}{2} - \log(1+e^{\lambda})\Big).$$

Podemos reescrever a equação acima como

$$I(x) = \log 2 + \sup_{\lambda\in\mathbb{R}}(\lambda x - \log(1+e^{\lambda})). \tag{1.20}$$

Para $x < 0$, fazendo $\lambda \to -\infty$, concluímos que $\lambda x \to +\infty$ e $\log(1+e^{\lambda}) \to 0$. Logo, $\lambda x - \log(1+e^{\lambda})$ não é limitado superiormente e, portanto, $I(x) = +\infty$. Além disso, podemos reescrever (1.20) como

$$I(x) = \log 2 + \sup_{\lambda\in\mathbb{R}}\left[\log\left(\frac{e^{\lambda x}}{1+e^{\lambda}}\right)\right] = \log 2 + \sup_{\lambda\in\mathbb{R}}\left[\log\left(\frac{e^{\lambda x-\lambda}}{e^{-\lambda}+1}\right)\right],$$

e assim, fazendo uma substituição $-\lambda \mapsto \alpha$, concluímos que

$$I(x) = \log 2 + \sup_{\alpha\in\mathbb{R}}\left[\log\left(\frac{e^{\alpha(1-x)}}{1+e^{\alpha}}\right)\right] = I(1-x).$$

Logo, se $x > 1$, então $I(x) = I(1-x) = +\infty$, pois $1-x<0$.

Falta encontrar o valor de I no intervalo $[0,1]$. Vamos começar pelos extremos do intervalo. Note que $I(1) = I(1-1) = I(0) = \log 2 + \sup_{\lambda\in\mathbb{R}}(-\log(1+e^{\lambda}))$. Como $\log(1+e^{\lambda})$ é crescente, sabemos que $-\log(1+e^{\lambda})$ é decrescente. Logo,

$$I(1) = I(0) = \log 2 + \lim_{\lambda\to-\infty}(-\log(1+e^{\lambda})) = \log 2.$$

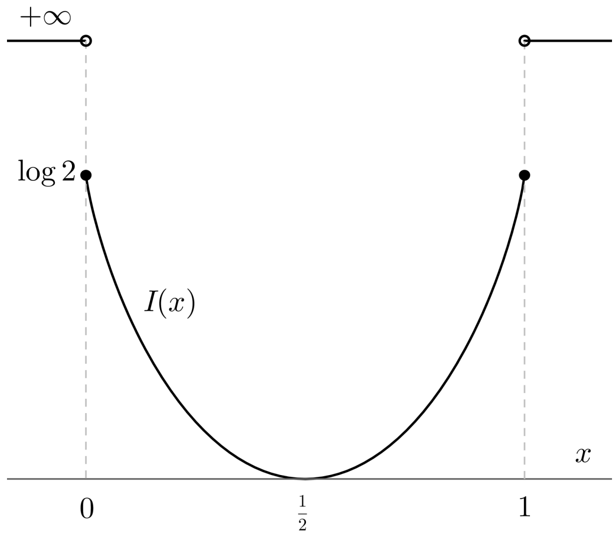
Por fim, basta encontrar o valor de I no intervalo aberto $(0,1)$. Para isso, seja $f_x(\lambda) = \lambda x - \log(1+e^{\lambda}) = \log\left(\dfrac{e^{\lambda x}}{1+e^{\lambda}}\right)$. Assim, $I(x) = \log 2 + \sup_{\lambda\in\mathbb{R}} f_x(\lambda)$. Queremos encontrar um ponto de máximo para esta f_x, ou seja, procuramos um λ_0 tal que $f_x'(\lambda_0) = 0$ e $f_x''(\lambda_0) < 0$. Note que $f_x'(\lambda) = \dfrac{x + xe^{\lambda} - e^{\lambda}}{1+e^{\lambda}}$ e $f_x''(\lambda) = \dfrac{-e^{\lambda}}{(1+e^{\lambda})^2} < 0$. Portanto, f_x possui um máximo em $\lambda_0 = \log\left(\dfrac{x}{1-x}\right) = \log x - \log(1-x)$. Então,

$$I(x) = \log 2 + \sup_{\lambda\in\mathbb{R}} f_x(\lambda) = \log 2 + f_x(\lambda_0) = \log 2 + x\log x + (1-x)\log(1-x).$$

Logo,

$$I(x) = \begin{cases} \log 2 + x\log x + (1-x)\log(1-x), & \text{se } x \in [0,1], \\ +\infty, & \text{caso contrário.} \end{cases}$$

Figura 1 – Função I para distribuição de Bernoulli simétrica



Exemplo 1.2. *Vamos descobrir I para a sequência $\{X_n\}$ de variáveis aleatórias i.i.d com distribuição de Poisson de parâmetro $\beta > 0$, ou seja, $X_n \sim Poisson(\beta)$.*

Como $X_1 \sim Poisson(\beta)$, temos que $\mathbb{P}[X_1 = k] = \dfrac{e^{-\beta}\beta^k}{k!}$, para qualquer $k \in \mathbb{N}$. Logo,

$$\varphi(\lambda) = \mathbb{E}[e^{\lambda X_1}] = \sum_{k=0}^{\infty} \frac{e^{\lambda k} e^{-\beta}\beta^k}{k!} = e^{-\beta} \sum_{k=0}^{\infty} \frac{(\beta e^{\lambda})^k}{k!} = e^{-\beta} e^{\beta e^{\lambda}} = e^{\beta(e^{\lambda}-1)}.$$

Substituindo em (1.6),

$$I(x) = \sup_{\lambda \in \mathbb{R}} [\lambda x - \log e^{\beta(e^{\lambda}-1)}] = \sup_{\lambda \in \mathbb{R}} [\lambda x - \beta(e^{\lambda} - 1)].$$

Se $x < 0$, fazendo $\lambda \to -\infty$, temos que $\lambda x - \beta(e^{\lambda} - 1) \to +\infty$. Portanto, $I(x) = +\infty$.

Para $x = 0$, $I(x) = \sup_{\lambda \in \mathbb{R}} [-\beta(e^{\lambda} - 1)] = \beta \sup_{\lambda \in \mathbb{R}} [-e^{\lambda} + 1]$. Como $-e^{\lambda} + 1$ é decrescente, vale que $\sup_{\lambda \in \mathbb{R}} [-e^{\lambda} + 1] = \lim_{\lambda \to -\infty} -e^{\lambda} + 1 = 1$. Logo, $I(x) = \beta$.

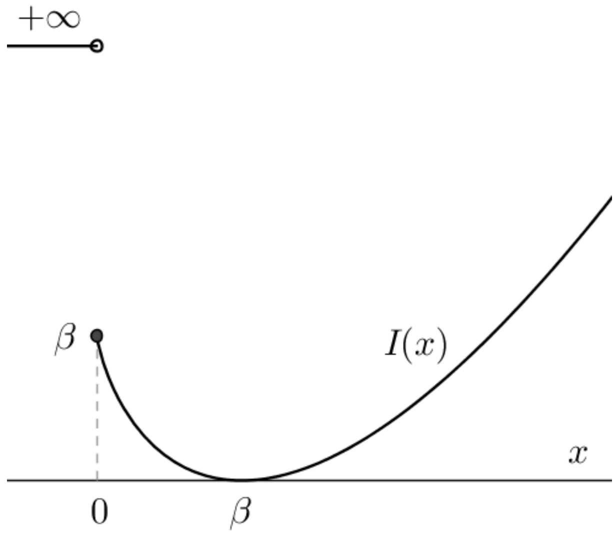
Se $x > 0$, definimos $f_x(\lambda) = \lambda x - \beta(e^{\lambda} - 1)$. Assim, $f_x'(\lambda) = x - \beta e^{\lambda}$ se anula em $\lambda_0 = \log\left(\dfrac{x}{\beta}\right)$ e $f_x''(\lambda) = -\beta e^{\lambda} < 0$. Logo, λ_0 é um máximo para f_x. Portanto,

$$I(x) = \sup_{\lambda \in \mathbb{R}} f_x(\lambda) = f_x(\lambda_0) = x \log\left(\frac{x}{\beta}\right) - x + \beta, \text{ se } x > 0.$$

Logo,

$$I(x) = \begin{cases} x \log\left(\dfrac{x}{\beta}\right) - x + \beta, & \text{se } x \geqslant 0, \\ +\infty, & \text{se } x < 0. \end{cases}$$

Figura 2 – Função I para distribuição de Poisson



Repare que $\beta = \mathbb{E}[X_1]$ e, como deveria ser, $I(\beta) = 0$.

Exemplo 1.3. *Vamos procurar a função taxa I para a sequência $\{X_n\}$ de variáveis aleatórias i.i.d tais que $X_n \sim N(\mu, \sigma^2)$.*

A função de densidade da variável aleatória X_1 é $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$, onde μ e σ estão fixados. Assim,

$$\varphi(\lambda) = \mathbb{E}[e^{\lambda X_1}] = \int_{\mathbb{R}} \frac{e^{\lambda x} e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} dx.$$

Fazendo uma substituição $x \mapsto (y+\mu)$, podemos reescrever a equação acima como

$$\varphi(\lambda) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\lambda(y+\mu) - \frac{y^2}{2\sigma^2}} dy = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\lambda\mu} e^{\frac{2\sigma^2\lambda y - y^2}{2\sigma^2}} dy.$$

Podemos reescrever $2\sigma^2\lambda y - y^2$ como $-(y-\sigma^2\lambda)^2 + \sigma^4\lambda^2$. Então,

$$\varphi(\lambda) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\lambda\mu} e^{\frac{-(y-\sigma^2\lambda)^2}{2\sigma^2}} e^{\frac{\sigma^2\lambda^2}{2}} dy.$$

Portanto,

$$\varphi(\lambda) = \frac{e^{\lambda\mu + \frac{\sigma^2\lambda^2}{2}}}{\sigma\sqrt{2\pi}} \underbrace{\int_{-\infty}^{+\infty} e^{\frac{-(y-\sigma^2\lambda)^2}{2\sigma^2}} dy}_{\sigma\sqrt{2\pi}} = e^{\lambda\mu + \frac{\sigma^2\lambda^2}{2}}. \tag{1.21}$$

Substituindo em (1.6),

$$I(x) = \sup_{\lambda\in\mathbb{R}} \left[\lambda x - \log e^{\lambda\mu + \frac{\sigma^2\lambda^2}{2}}\right] = \sup_{\lambda\in\mathbb{R}} \left[\lambda x - \lambda\mu - \frac{\sigma^2\lambda^2}{2}\right] = \sup_{\lambda\in\mathbb{R}} \left[\lambda(x-\mu) - \frac{\sigma^2\lambda^2}{2}\right].$$

Seja $f_x(\lambda) = \lambda(x-\mu) - \dfrac{\sigma^2\lambda^2}{2}$. Assim, $f_x'(\lambda) = (x-\mu) - \sigma^2\lambda$ e $f_x''(\lambda) = -\sigma^2$. Tomando $\lambda_0 = \dfrac{x-\mu}{\sigma^2}$, temos que $f_x'(\lambda_0) = 0$ e $f_x''(\lambda_0) < 0$, ou seja, f_x assume um máximo em λ_0.

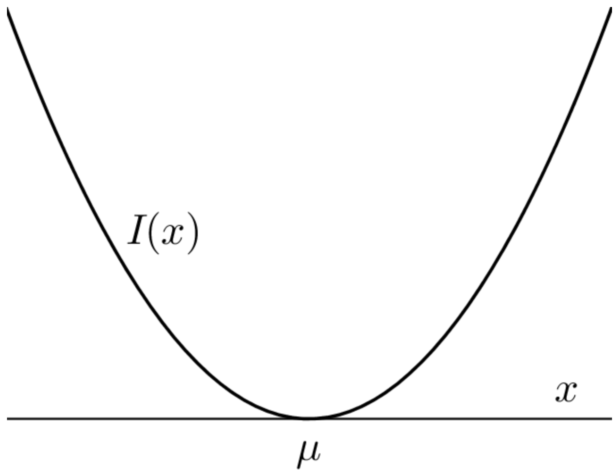
Então,

$$I(x) = f_x(\lambda_0) = (x-\mu)\left(\frac{x-\mu}{\sigma^2}\right) - \frac{\sigma^2}{2}\left(\frac{x-\mu}{\sigma^2}\right)^2 = \frac{(x-\mu)^2}{\sigma^2} - \frac{(x-\mu)^2}{2\sigma^2}.$$

Logo,

$$I(x) = \frac{(x-\mu)^2}{2\sigma^2}, \text{ para qualquer } x \in \mathbb{R}.$$

Figura 3 – Função I para distribuição $N(\mu, \sigma^2)$



Exemplo 1.4. *Vamos calcular a função I para a sequência $\{X_n\}$ de variáveis aleatórias i.i.d com distribuição exponencial de parâmetro β, ou seja, $X_n \sim exp(\beta)$.*

A função densidade de uma variável aleatória com distribuição exponencial de parâmetro β é $f(x;\beta) = \beta e^{-\beta x}\mathbb{1}_{[0,+\infty)}(x)$. Assim,

$$\varphi(\lambda) = \mathbb{E}[e^{\lambda X_1}] = \int_0^{+\infty} e^{\lambda x}\beta e^{-\beta x}dx = \beta\int_0^{+\infty} e^{(\lambda-\beta)x}dx.$$

Se $\lambda < \beta$, então $\lambda - \beta < 0$. Assim,

$$\varphi(\lambda) = \beta\left[\frac{e^{(\lambda-\beta)x}}{\lambda-\beta}\right]_{x=0}^{+\infty} = \beta\left[0 - \frac{1}{\lambda-\beta}\right] = \frac{\beta}{\beta-\lambda}.$$

Se $\lambda \geqslant \beta$, então $\varphi(\lambda) = +\infty$. Logo,

$$I(x) = \sup_{\lambda\in\mathbb{R}}[\lambda x - \log\varphi(\lambda)] = \sup_{\lambda<\beta}\left[\lambda x - \log\left(\frac{\beta}{\beta-\lambda}\right)\right] = -\log\beta + \sup_{\lambda<\beta}[\lambda x + \log(\beta-\lambda)].$$

Para $x \leqslant 0$, fazendo $\lambda \to -\infty$, concluímos que $\lambda x + \log(\beta - \lambda) \to +\infty$. Portanto, $I(x) = +\infty$.

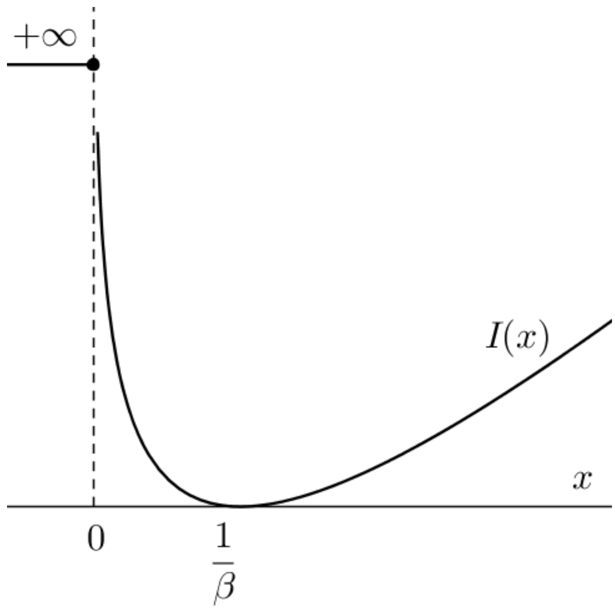
Para $x > 0$, definimos $f_x(\lambda) = \lambda x + \log(\beta-\lambda)$. Assim, $f'_x(\lambda) = x - \frac{1}{\beta-\lambda}$ se anula em $\lambda_0 = \beta - \frac{1}{x}$ e $f''_x(\lambda) = \frac{-1}{(\beta-\lambda)^2} < 0$. Logo, λ_0 é um máximo para f_x. Portanto, para $x > 0$,

$$I(x) = -\log\beta + \sup_{\lambda<\beta} f_x(\lambda) = -\log\beta + f_x(\lambda_0) = -\log\beta + \beta x - 1 - \log x = \beta x - \log(\beta x) - 1.$$

Logo,

$$I(x) = \begin{cases} \beta x - \log(\beta x) - 1, & \text{se } x > 0, \\ +\infty, & \text{se } x \leqslant 0. \end{cases}$$

Figura 4 – Função I para distribuição $exp(\beta)$



2 CADEIAS DE MARKOV COM ESPAÇO DE ESTADOS FINITO

Este capítulo contém um estudo acerca das cadeias de Markov e serve como base para o próximo capítulo. Para facilitar o entendimento e aliviar as notações, esse estudo será realizado no caso em que o espaço de estados é finito. Visando uma melhor compreensão dos conceitos aqui trabalhados, é importante que o leitor esteja familiarizado com certos termos e definições de Teoria da Probabilidade. Na Seção A.2 estão alguns dos mais importantes.

2.1 Conceitos Básicos

Durante todo este capítulo, consideraremos o conjunto $E = \{z_1, z_2, \dots, z_d\}$ como o espaço de estados da cadeia de Markov e o conjunto $\mathcal{E}$ como o conjunto das partes de E. Vale ainda ressaltar que $(E, \mathcal{E})$ é um espaço mensurável.

Definição 2.1. *A função $p : E \times \mathcal{E} \to [0,1]$ é uma probabilidade de transição se:*
(i) Fixado $x \in E$, a função $A \in \mathcal{E} \mapsto p(x, A)$ é uma medida de probabilidade em $(E, \mathcal{E})$.
(ii) Fixado $A \in \mathcal{E}$, a função $x \in E \mapsto p(x, A)$ é mensurável.

Notação. *Para simplificar a notação, vamos denotar $p(z_i, \{z_j\})$ por $p(z_i, z_j)$ ou por p_{ij}.*

Definição 2.2. *$P = (p_{ij})_{d \times d}$ é chamada matriz de transição ou matriz estocástica.*

Vale observar que todas as entradas da matriz P são não negativas, visto que cada uma delas é uma probabilidade.

Lema 2.3. *Para qualquer $x \in E$, $\sum\limits_{i=1}^{d} p(x, z_i) = 1$.*

Demonstração. Dado $x \in E$. Como p é uma probabilidade de transição, utilizando o item *(i)* da Definição 2.1, concluímos que $\sum\limits_{i=1}^{d} p(x, z_i) = p\left(x, \bigcup\limits_{i=1}^{d} \{z_i\}\right)$. Mas, $\bigcup\limits_{i=1}^{d} \{z_i\} = E$, logo, $\sum\limits_{i=1}^{d} p(x, z_i) = p(x, E) = 1$, já que $p(x, \cdot)$ é uma medida de probabilidade. □

Pelo lema acima, a matriz estocástica P também é chamada de matriz linha soma 1, pois a soma dos elementos de qualquer uma das suas linhas é $\sum\limits_{j=1}^{d} p_{ij} = \sum\limits_{j=1}^{d} p(z_i, z_j) = 1$.

Vamos considerar $\Omega = \{\omega = (\omega_0, \omega_1, \omega_2, \dots) : \omega_i \in E, i \in \{0, 1, 2, \dots\}\} = E^{\mathbb{N}}$. Seja $\{X_n\}$ uma sequência de funções $X_n : \Omega \to E$ mensuráveis tais que $X_n(\omega) = \omega_n$, isto é, X_n é a projeção na n-ésima coordenada.

Definição 2.4. *Uma sequência de variáveis aleatórias $\{X_n\}$ é chamada de cadeia de Markov com probabilidade de transição p se, para qualquer B em $\mathcal{E}$, vale a igualdade*

$$\mathbb{P}[X_{n+1} \in B|X_0 = x_0, \ldots, X_n = x_n] = \mathbb{P}[X_{n+1} \in B|X_n = x_n] = p(x_n, B), \tag{2.1}$$

ou seja, a probabilidade da cadeia estar em um certo conjunto de estados B no tempo $n+1$ depende exclusivamente do seu valor no momento imediatamente anterior, X_n, ignorando todos os tempos anteriores: $X_{n-1}, X_{n-2}, \ldots, X_0$.

Note que $p_{ij} = p(z_i, z_j) = \mathbb{P}[X_{n+1} = z_j|X_n = z_i]$, ou seja, p_{ij} é a probabilidade da cadeia passar do estado i para o estado j no tempo imediatamente seguinte.

Exemplo 2.1. [1] *Um sapo vive em uma lagoa com duas folhas, leste (l) e oeste (o). Certo dia ele encontrou duas moedas no fundo da lagoa e colocou uma em cada folha. Desde então, toda manhã, o sapo decide se irá saltar jogando uma moeda da seguinte forma: se a moeda cair cara, o sapo pula para a outra folha e se cair coroa, ele permanece onde está.*

Sejam $E = \{l, o\}$ e $(X_0, X_1, \ldots)$ a sequência de folhas ocupadas pelo sapo ao longo dos dias. Sobre as moedas, nós não iremos assumir que elas são necessariamente justas. Suponhamos que a moeda leste tenha probabilidade p de cair cara, enquanto a moeda oeste tem probabilidade q de cair cara. Neste caso, obedecendo as regras para o sapo pular, obtemos que:

$$p(l,l) = probabilidade\ de\ estar\ na\ folha\ leste\ e\ permanecer\ na\ folha\ leste = 1-p$$

$$p(l,o) = probabilidade\ de\ estar\ na\ folha\ leste\ e\ pular\ para\ a\ folha\ oeste = p$$

$$p(o,l) = probabilidade\ de\ estar\ na\ folha\ oeste\ e\ pular\ para\ a\ folha\ leste = q$$

$$p(o,o) = probabilidade\ de\ estar\ na\ folha\ oeste\ e\ permanecer\ na\ folha\ oeste = 1-q$$

Podemos então notar que a probabilidade do sapo estar em uma determinada folha, depende exclusivamente da posição dele no dia imediatamente anterior. Portanto, apesar de nem termos mencionado o espaço de probabilidade $(\Omega, \mathfrak{F}, \mathbb{P})$, temos que a sequência $\{X_n\}$ é uma cadeia de Markov com matriz de transição

$$P = \begin{bmatrix} p(l,l) & p(l,o) \\ p(o,l) & p(o,o) \end{bmatrix} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}.$$

Mas, alguém poderia se perguntar: "Quem é $(\Omega, \mathfrak{F}, \mathbb{P})$?", porque esta pessoa poderia querer saber qual é a probabilidade de no segundo salto o sapo estar na folha leste, em símbolos, $\mathbb{P}[X_2 = l]$.

Podemos tentar responder isso usando um argumento que irá supor a existência de $(\Omega, \mathfrak{F}, \mathbb{P})$. Assim, usando probabilidade condicional, temos:

$$\mathbb{P}[X_2 = l] = \mathbb{P}[X_2 = l|X_1 = l]\mathbb{P}[X_1 = l] + \mathbb{P}[X_2 = l|X_1 = o]\mathbb{P}[X_1 = o].$$

[1] Exemplo adaptado de Levin, Peres e Wilmer (2009, p. 3)

Considerando as definições de $p(l,l)$ e $p(o,l)$ e usando novamente probabilidade condicional, concluímos:

$$\begin{aligned}\mathbb{P}[X_2 = l] &= p(l,l)\left\{\mathbb{P}[X_1 = l|X_0 = l]\mathbb{P}[X_0 = l] + \mathbb{P}[X_1 = l|X_0 = o]\mathbb{P}[X_0 = o]\right\} \\ &+ p(o,l)\left\{\mathbb{P}[X_1 = o|X_0 = l]\mathbb{P}[X_0 = l] + \mathbb{P}[X_1 = o|X_0 = o]\mathbb{P}[X_0 = o]\right\}.\end{aligned}$$

Utilizando as definições de $p(l,l)$, $p(o,l)$, $p(l,o)$ e $p(o,o)$, temos:

$$\begin{aligned}\mathbb{P}[X_2 = l] &= p(l,l)\left\{p(l,l)\mathbb{P}[X_0 = l] + p(o,l)\mathbb{P}[X_0 = o]\right\} \\ &+ p(o,l)\left\{p(l,o)\mathbb{P}[X_0 = l] + p(o,o)\mathbb{P}[X_0 = o]\right\}.\end{aligned}$$

Para finalizar este cálculo, temos que definir $\mathbb{P}[X_0 = l]$ e $\mathbb{P}[X_0 = o]$, ou seja, as probabilidades do sapo iniciar na folha leste e oeste. Vamos supor que ele inicia na folha leste, ou seja, $\mathbb{P}[X_0 = l] = 1$. Assim,

$$\mathbb{P}[X_2 = l] = p(l,l)p(l,l) + p(l,o)p(o,l) = \sum_{z\in\{l,o\}} p(l,z)p(z,l).$$

Este exemplo, apesar de simples, irá nos dar a intuição do que devemos fazer para responder a seguinte pergunta: "dada uma matriz de transição P, será que existe um espaço de probabilidade $(\Omega, \mathfrak{F}, \mathbb{P})$ e uma sequência de variáveis aleatórias $\{X_n\}$ que satisfaça a condição (2.1)?".

Para isso, vamos fazer a seguinte construção:

Definição 2.5. *Inicialmente, definimos os seguintes conjuntos:*
(i) $E^n = E \times E \times \ldots \times E$ *o produto cartesiano de* n *espaços de estados* E*;*
(ii) $\mathcal{E}^n = \mathcal{E} \times \mathcal{E} \times \ldots \times \mathcal{E}$ *o produto cartesiano de* n σ*-álgebras* $\mathcal{E}$*;*
(iii) $E^{\mathbb{N}}$ *o conjunto das sequências de elementos de* E*;*
(iv) $\mathcal{E}^{\mathbb{N}}$ *o conjunto das sequências de elementos de* $\mathcal{E}$*;*

Definição 2.6. *Para cada* $x_0 \in E$*, definimos* $v_0(x_0)$ *como a probabilidade da cadeia de Markov começar em* x_0*. Vamos denotar por* $v_0 = (v_0(z_1), \ldots, v_0(z_d))$ *a distribuição inicial da cadeia de Markov.*

A distribuição inicial da cadeia de Markov, pode ser qualquer vetor definido conforme dito acima. Porém, na maioria dos casos, sabe-se o estado em que a cadeia inicia, ou então, o resultado desejado independe da distribuição inicial. Nestes casos, a definição abaixo é bastante útil.

Definição 2.7. *Denotamos por* δ_x *a distribuição inicial da cadeia de Markov que inicia no estado* x*, ou seja, dado* $y \in E$*, a probabilidade da cadeia iniciar em* y *é:*

$$\delta_x(y) = \begin{cases} 1, & se \quad y = x \\ 0, & se \quad y \neq x \end{cases}.$$

Note que, no Exemplo 2.1, definimos $v_0 = \delta_l$.

Definição 2.8. *Em cada* $(E^n, \mathcal{E}^n)$, *definimos* μ_n *por:*

$$\mu_n(A_0 \times A_1 \times \cdots \times A_{n-1}) = \sum_{x_0 \in A_0} \sum_{x_1 \in A_1} \cdots \sum_{x_{n-1} \in A_{n-1}} v_0(x_0) p(x_0, x_1) p(x_1, x_2) \ldots p(x_{n-2}, x_{n-1}).$$

Proposição 2.9. *Cada* μ_n *é uma probabilidade em* $(E^n, \mathcal{E}^n)$.

Demonstração. Vamos dividir essa prova em quatro partes: (i) $\mu_n(\emptyset^n) = 0$; (ii) $\mu_n(E^n) = 1$; (iii) $\mu_n(B) = \mu_n(B_0 \times B_1 \times \cdots \times B_{n-1}) \geqslant 0$, para qualquer $B \in \mathcal{E}^n$; (iv) μ_n é σ-aditiva.

(i) Vale trivialmente, pois se trata de uma soma vazia;

(ii) Reesrevemos

$$\mu_n(E^n) = \sum_{x_0 \in E} \cdots \sum_{x_{n-2} \in E} v_0(x_0) p(x_0, x_1) \ldots p(x_{n-3}, x_{n-2}) \underbrace{\sum_{x_{n-1} \in E} p(x_{n-2}, x_{n-1})}_{1}.$$

Seguimos este procedimento até concluirmos que

$$\mu_n(E^n) = \sum_{x_0 \in E} \sum_{x_1 \in E} v_0(x_0) p(x_0, x_1) = \sum_{x_0 \in E} v_0(x_0) \sum_{x_1 \in E} p(x_0, x_1) = \sum_{x_0 \in E} v_0(x_0) = 1$$

(iii) Utilizando o mesmo procedimento do item anterior reescrevemos

$$\mu_n(B_0 \times B_1 \times \cdots \times B_{n-1}) = \sum_{x_0 \in B_0} v_0(x_0) \sum_{x_1 \in B_1} p(x_0, x_1) \cdots \sum_{x_{n-1} \in B_{n-1}} p(x_{n-2}, x_{n-1}).$$

Como p é probabilidade, olhando cada soma dessas da direita para a esquerda, concluímos que cada uma delas está entre 0 e 1. Portanto, temos que $0 \leqslant \mu_n(B_0 \times B_2 \times \ldots \times B_{n-1}) \leqslant 1$, para qualquer $B = B_0 \times B_1 \times \cdots \times B_{n-1} \in \mathcal{E}^n$;

(iv) Sejam $\{U_i\} \subseteq \mathcal{E}^n$ disjuntos e $U = \bigcup_{i=0}^{\infty} U_i$. Note que podemos reescrever

$$\mu_n(U) = \sum_{(x_0, \ldots, x_{n-1}) \in U} v_0(x_0) p(x_0, x_1) p(x_1, x_2) \ldots p(x_{n-2}, x_{n-1}).$$

Como $U = \bigcup_{i=1}^{\infty} U_i$ é uma união disjunta, podemos decompor

$$\mu_n(U) = \sum_{i=1}^{\infty} \sum_{(x_0, \ldots, x_{n-1}) \in U_i} v_0(x_0) p(x_0, x_1) p(x_1, x_2) \ldots p(x_{n-2}, x_{n-1}) = \sum_{i=1}^{\infty} \mu_n(U_i).$$

Logo, μ_n é σ-aditiva. □

Definição 2.10. *Seja* $\psi_n : E^n \to E^{n-1}$ *a projeção nas* $n-1$ *primeiras coordenadas, ou seja,* $\psi_n(X_0, X_1, \ldots, X_{n-1}) = (X_0, X_1, \ldots, X_{n-2})$.

Proposição 2.11. *Para todo* $\Gamma \subseteq E^{n-1}$, *vale que, para todo* $n \geqslant 1$,

$$(\mu_n \circ \psi_n^{-1})(\Gamma) = \mu_n(\psi_n^{-1}(\Gamma)) = \mu_{n-1}(\Gamma),$$

onde $\varphi_n^{-1}(\Gamma)$ *é a pré-imagem de* Γ *pela função* ψ_n, *ou seja,* $\psi_n^{-1}(\Gamma) = \{y \in E^n : \psi_n(y) \in \Gamma\}$.

Demonstração. Inicialmente, aplicamos $\mu_n \circ \psi_n^{-1}$ em um conjuto qualquer $A_0 \times \ldots \times A_{n-2}$ de E^{n-1}. Note que a imagem inversa do conjunto $A_0 \times \ldots \times A_{n-2}$ pela função ψ_n é $A_0 \times \ldots \times A_{n-2} \times E$. Temos que $\mu_n(\psi_n^{-1}(A_0 \times \ldots \times A_{n-2})) = \mu_n(A_0 \times \ldots \times A_{n-2} \times E)$. Usando a definição de μ_n, podemos concluir que

$$\mu_n(\psi_n^{-1}(A_0 \times \ldots \times A_{n-2})) = \sum_{x_0 \in A_0} \ldots \sum_{x_{n-1} \in E} v_0(x_0)p(x_0,x_1)\ldots p(x_{n-3},x_{n-2})p(x_{n-2},x_{n-1}).$$

Como apenas o último fator do produto depende de x_{n-1}, podemos reescrever o segundo termo da igualdade acima como

$$\sum_{x_0 \in A_0} \ldots \sum_{x_{n-2} \in A_{n-2}} v_0(x_0)p(x_0,x_1)\ldots p(x_{n-3},x_{n-2}) \underbrace{\sum_{x_{n-1} \in E} p(x_{n-2},x_{n-1})}_{1}.$$

Portanto, obtemos que

$$\mu_n(\psi_n^{-1}(A_0 \times \ldots \times A_{n-2})) = \sum_{x_0 \in A_0} \ldots \sum_{x_{n-2} \in A_{n-2}} v_0(x_0)p(x_0,x_1)\ldots p(x_{n-3},x_{n-2}).$$

Desta última, segue que $\mu_n(\psi_n^{-1}(A_0 \times \ldots \times A_{n-2})) = \mu_{n-1}(A_0 \times \ldots \times A_{n-2})$. $\square$

Como as μ_n que construímos satisfazem as hipóteses do Teorema da extensão de Kolmogorov (Teorema A.16), concluímos que existe uma única medida de probabilidade $\mathbb{P}$ em $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$ com

$$\mathbb{P}(w \in E^{\mathbb{N}} : \ w_i \in A_i, 0 \leqslant i \leqslant n-1) = \mu_n(A_0 \times \ldots \times A_{n-1}). \tag{2.2}$$

Com isso, vimos que, para cada cadeia de Markov com espaços de estados finito existe uma única matriz de transição P, bem como, para cada matriz de transição existe uma única cadeia de Markov. Portanto, essa é uma relação biunívoca. Dizemos que a probabilidade $\mathbb{P}$ definida em (2.2) é induzida pela cadeia de Markov e pela medida inicial v_0.

Voltando ao Exemplo 2.1, quando calculamos que $\mathbb{P}[X_2 = l] = \sum_{z \in E} p(l,z)p(z,l)$, se já soubéssemos desta construção poderíamos ter calculado $\mathbb{P}[X_2 = l]$ da seguinte forma:

$$\begin{aligned}\mathbb{P}[X_2 = l] &= \mathbb{P}[\omega = (\omega_0,\omega_1,\omega_2,\ldots); \omega_0 \in E, \omega_1 \in E, \omega_2 = l] = \mu_n(E \times E \times \{l\}) \\ &= \sum_{x_0 \in E}\sum_{x_1 \in E}\sum_{x_2 \in \{l\}} v_0(x_0)p(x_0,x_1)p(x_1,x_2) = \sum_{x_0 \in E}\sum_{x_1 \in E} v_0(x_0)p(x_0,x_1)p(x_1,l)\end{aligned}$$

Lembre que tínhamos escolhido que o sapo iniciava na folha leste, ou seja, a distribuição inicial era $v_0 = \delta_l$. Assim,

$$\mathbb{P}[X_2 = l] = \sum_{x_1 \in E} p(l,x_1)p(x_1,l)$$

Definição 2.12. *Seja $p_{ij}^n = p^n(z_i,z_j) = \mathbb{P}[X_n = z_j | X_0 = z_i]$ a probabilidade de uma cadeia de Markov sair do estado i e estar no estado j em n passos, onde p_{ij}^n é a (i,j)-ésima entrada da matriz P^n.*

Lema 2.13 (Equação de Chapman-Kolmogorov)**.**

$$p^{k+n}(x_0, z) = \sum_{y \in E} p^k(x_0, y)p^n(y, z).$$

Demonstração. Sabemos, da Álgebra Linear, que $P^{k+n} = P^n P^k$. Então, a probabilidade de uma cadeia passar do estado x_0 para o estado z em um tempo $k+n$, representada pela respectiva entrada na matriz P^{k+n} (linha correspondente a z e coluna correspondente a x_0) é igual a esta mesma entrada na matriz $P^n P^k$. Esta entrada, por sua vez, pode ser representada, utilizando multiplicação de matrizes, pelo produto da linha correspondente a z em P^n pela coluna correspondente a x_0 em P^k, ou seja, por $\sum_{y \in E} p^n(y, z)p^k(x_0, y)$. Logo, concluímos que $p^{k+n}(x_0, z) = \sum_{y \in E} p^k(x_0, y)p^n(y, z)$. □

Definição 2.14. *Definimos* $v_n(x_0) = (p^n(x_0, z_1), p^n(x_0, z_2), \ldots, p^n(x_0, z_d))$ *para* $n \geqslant 1$.

Observe que o vetor $v_n(x_0)$ definido acima guarda toda a informação das probabilidades da cadeia partir de x_0 e chegar em cada um dos $z_i, 1 \leqslant i \leqslant d$, em n passos.

Notação. *Chamamos de*

$$v_n^i(x_0) = p^n(x_0, z_i) \tag{2.3}$$

à i-ésima coordenada de $v_n(x_0)$.

Proposição 2.15. $v_n(x_0) = \delta_{x_0} P^n$, *onde* δ_{x_0} *é a distribuição inicial da cadeia.*

Demonstração. Utilizando a equação de Chapman-Kolmogorov (Lema 2.13) e reescrevendo os termos conforme as notações introduzidas, podemos concluir que, dado qualquer $j \in \{1, \ldots, d\}$,

$$v_{n+1}^j(x_0) = p^{n+1}(x_0, z_j) = \sum_{i=1}^{d} p^n(x_0, z_i)p(z_i, z_j) = \sum_{i=1}^{d} v_n^i(x_0)p_{ij}.$$

Então, $v_{n+1}^j(x_0) = v_n(x_0)P_j$, com P_j sendo a j-ésima coluna de P. Portanto, $v_{n+1}(x_0) = v_n(x_0)P$. Logo, $v_n(x_0) = v_{n-1}(x_0)P = v_{n-2}(x_0)P^2 = \cdots = v_1(x_0)P^{n-1} = \delta_{x_0}P^n$. □

Voltando ao Exemplo 2.1, lembre que o sapo começava na folha *leste* e queríamos calcular $\mathbb{P}[X_2 = l]$. Com as notações fixadas anteriormente, temos que $v_0 = \delta_l = (1, 0)$. Queremos então calcular $v_2^2(l) = p^2(l, l)$. Pela proposição anterior, obtemos que

$$v_2(l) = \delta_l P^2 = \delta_l \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} (1-p)^2 + pq & p(1-p) + p(1-q) \\ (1-p)q + (1-q)q & pq + (1-q)^2 \end{bmatrix}.$$

Assim,

$$v_2(l) = \begin{bmatrix} (1-p)^2 + pq & p(1-p) + p(1-q) \end{bmatrix}.$$

Logo, $\mathbb{P}[X_2 = l] = (1-p)^2 + pq = p(l, l)p(l, l) + p(l, o)p(o, l)$, como já havíamos encontrado anteriormente na forma de somatório.

Podemos ainda ampliar este conceito ao caso geral em que a distribuição inicial v_0 não é necessariamente da forma δ_{x_0} para algum $x_0 \in E$.

Teorema 2.16. *Seja $\{X_n\}$ uma cadeia de Markov com matriz de transição P e distribuição inicial v_0. Então, $v_0P^n = v_n$, onde $v_n = (\mathbb{P}[X_n = z_1], \mathbb{P}[X_n = z_2], \ldots, \mathbb{P}[X_n = z_d])$.*

Demonstração. Observe que podemos escrever $v_0 = \sum_{i=1}^{d} v_0(z_i)\delta_{z_i}$. Por linearidade, temos

$$v_0P^n = \left(\sum_{i=1}^{d} v_0(z_i)\delta_{z_i}\right) P^n = \sum_{i=1}^{d} v_0(z_i)\delta_{z_i}P^n = \sum_{i=1}^{d} v_0(z_i)v_n(z_i).$$

Mas, usando que $v_n(z_i) = (\mathbb{P}[X_n = z_1|X_0 = z_i], \ldots, \mathbb{P}[X_n = z_d|X_0 = z_i])$ e ainda $v_0(z_i) = \mathbb{P}[X_0 = z_i]$, obtemos:

$$\begin{aligned} v_0P^n &= \sum_{i=1}^{d} \Big(\mathbb{P}[X_n = z_1|X_0 = z_i]\mathbb{P}[X_0 = z_i], \ldots, \mathbb{P}[X_n = z_d|X_0 = z_i]\mathbb{P}[X_0 = z_i]\Big) \\ &= \left(\sum_{i=1}^{d} \mathbb{P}[X_n = z_1|X_0 = z_i]\mathbb{P}[X_0 = z_i], \ldots, \sum_{i=1}^{d} \mathbb{P}[X_n = z_d|X_0 = z_i]\mathbb{P}[X_0 = z_i]\right). \end{aligned}$$

Note que cada coordenada de v_0P^n é então da forma $\sum_{i=1}^{d} \mathbb{P}[X_n = z_j|X_0 = z_i]\mathbb{P}[X_0 = z_i]$. Definindo $A_i = \{\omega \in \Omega : X_0(\omega) = z_i\}$, para $1 \leqslant i \leqslant d$, temos que $\{A_1, \ldots, A_n\}$ é partição de Ω. Podemos supor, sem perda de generalidade, que $\mathbb{P}[A_i] > 0$, para qualquer i, pois, caso contrário, bastaria retirar os que tem probabilidade nula. Então, pela lei da probabilidade total (Proposição A.9),

$$\sum_{i=1}^{d} \mathbb{P}[X_n = z_j|X_0 = z_i]\mathbb{P}[X_0 = z_i] = \sum_{i=1}^{d} \mathbb{P}[X_n = z_j|A_i]\mathbb{P}[A_i] = \mathbb{P}[X_n = z_j].$$

Logo, $v_0P^n = v_n$. □

Lema 2.17. *Para todo $n \in \mathbb{N}$, $\sum_{i=1}^{d} v_n^i(x_0) = 1$, onde v_n^i foi definido em (2.3).*

Demonstração. Seja $n \in \mathbb{N}$. Utilizando a definição de $v_n(x_0)$ e $p^n(i,j)$, podemos escrever

$$\sum_{i=1}^{d} v_n^i(x_0) = \sum_{i=1}^{d} p^n(x_0, z_i) = \sum_{i=1}^{d} \mathbb{P}[X_n = z_i|X_0 = x_0].$$

Utilizando o fato de $\mathbb{P}$ ser probabilidade, obtemos que

$$\sum_{i=1}^{d} v_n^i(x_0) = \mathbb{P}\Big[x_n \in \bigcup_{i=1}^{d}\{z_i\}\Big| X_0 = x_0\Big].$$

Mas $\bigcup_{i=1}^{d}\{z_i\} = E$, logo,

$$\sum_{i=1}^{d} v_n^i(x_0) = \mathbb{P}[x_n \in E|X_0 = x_0] = p(x_0, E) = 1.$$

□

Quando estudamos cadeias de Markov, geralmente estamos interessados em fazer alguma previsão sobre a posição desta cadeia após um longo período de tempo. Em outras palavras, procuramos algum vetor π, que chamamos de vetor estacionário, tal que $v_n \to \pi$ quando $n \to \infty$, independentemente de qual seja a distribuição inicial v_0.

2.2 Vetores Estacionários

Esta seção é dedicada ao estudo dos vetores estacionários das cadeias de Markov, que são os vetores π não negativos, com coordenadas somando 1 e satisfazendo $\pi = \pi P$. Note que, se existir tal π, ele é um autovetor à esquerda de P associado ao autovalor 1.

Proposição 2.18. *Qualquer matriz de transição P possui pelo menos um autovalor, o 1.*

Demonstração. Basta ver que a soma dos elementos de cada uma das linhas de $P-I$ é zero. Portanto, as colunas de $P-I$ são linearmente dependentes e segue que $det(P-I)=0$. Logo, 1 é autovalor de P. $\square$

Vimos então que qualquer matriz de transição possui um autovalor 1. Mas isso não nos ajuda muito, pois como P é matriz linha soma 1, o vetor coluna $f = (1, \ldots, 1)$ é um autovetor à direita de P associado ao autovalor 1. O vetor π que queremos deve ser não negativo e suas coordenadas devem somar 1, além de ser autovetor à esquerda. Vamos verificar isto agora.

Definição 2.19. *Seja A uma matriz $d \times d$ com entradas a_{ij} tais que $a_{ij} \geqslant 0$, para quaisquer $i, j \in \{1, \ldots, d\}$. Definimos Λ_A como o conjunto de todos os $\lambda \in \mathbb{R}$ para os quais existe um vetor $x = (x_1, \ldots, x_d)$ tal que $\sum_{j=1}^{d} x_j = 1$, $x_j \geqslant 0$ para todo $j \in \{1, \ldots, d\}$ e $xA \geqslant \lambda x$ (cada coordenada de xA é maior ou igual que a respectiva coordenada de λx). Definimos ainda $\lambda_A = \sup_{\lambda \in \Lambda_A} \lambda$.*

Note que $0 \in \Lambda_A$ e, portanto, o conjunto é não vazio. Além disso, para qualquer $\lambda \in \Lambda_A$, existe um x que satisfaz $\sum_{i=1}^{d} x_i = 1$, $x_i \geqslant 0$ e $xA \geqslant \lambda x$. Portanto, pelo menos uma das entradas, que chamaremos de x_j, satisfaz $x_j \geqslant \frac{1}{d}$. Chamando de M a maior entrada de A, vale que $(xA)_j = \sum_{i=1}^{d} a_{ij} x_i \leqslant \sum_{i=1}^{d} M x_i = M$, para qualquer $j \in \{1, \ldots, d\}$. Logo, $M \geqslant (xA)_j \geqslant \lambda x_j \geqslant \lambda \frac{1}{d}$. Então, $\lambda_A = \sup_{\lambda \in \Lambda_A} \lambda \leqslant Md$. Logo, λ_A está bem definido.

Teorema 2.20 (Teorema de Frobenius)**.** *Seja λ_A como definido acima para A uma matriz tal que $a_{ij} \geqslant 0$ para quaisquer $i, j \in \{1, \ldots, d\}$ e existe pelo menos um a_{ij} não nulo. Então,*

(i) λ_A é autovalor de A e é possível escolher um autovetor à esquerda associado a λ_A com todas as coordenadas não negativas, sendo pelo menos uma delas não nula;

(ii) Se λ é autovalor de A, então $|\lambda| \leqslant \lambda_A$.

Demonstração. Caso 1: Inicialmente, vamos provar que se A possui todas as entradas positivas, então existe um x^0 com todas as entradas positivas tal que $x^0 A = \lambda_A x^0$. Para isso, vamos considerar uma sequência $\{\lambda_i < \lambda_A\}$ de elementos de Λ_A que converge a

λ_A. Pela forma como Λ_A foi definido, existem vetores $x^{(i)}$, $i \in \mathbb{N}$, tais que $x^{(i)} \geqslant 0$, $x^{(i)}A \geqslant \lambda_i x^{(i)}$ e

$$\sum_{j=1}^{d} x_j^{(i)} = 1.$$

Como o conjunto dos vetores cujas entradas são positivas e somam 1 é um conjunto compacto e todos os $x^{(i)}$ pertencem a ele, podemos escolher uma subsequência convergente $\{x^{(n_i)}\}$ com $n_1 < n_2 < \cdots$. Seja x^0 o limite dessa subsequência. Note que x^0 também satisfaz $x^0 \geqslant 0$, $x^0 A \geqslant \lambda_A x^0$ e $\sum_j x_j^0 = 1$.

Queremos provar que $x^0 A = \lambda_A x^0$. Para isso, suponhamos que alguma entrada de $x^0 A$ é estritamente maior que a respectiva entrada de $\lambda_A x^0$, ou seja, existe k tal que $(x^0 A)_k > \lambda_A x_k^0$. Como A possui todas as entradas positivas, definindo $y^0 = x^0 A$ e denotando a base canônica de $\mathbb{R}^d$ por $\{e_1, \ldots, e_d\}$, obtemos que

$$(y^0 A)_j = \sum_{i=1}^{d} y_i^0 a_{ij} > \sum_{i=1}^{d} \lambda_A x_i^0 a_{ij} = \lambda_A (x^0 A)_j = \lambda_A y_j^0,$$

isto é, $y^0 A$ possui todas as entradas maiores que a respectiva entrada de $\lambda_A y^0$. Portanto, existe $\varepsilon > 0$ tal que $y^0 A$ possui todas as entradas maiores que a respectiva entrada de $(\lambda_A + \varepsilon) y^0$. Normalizando y^0 de modo que $\sum_j y_j^0 = 1$, concluímos que $\lambda_A + \varepsilon \in \Lambda_A$. Mas isso é um absurdo, pois λ_A é o supremo. Logo, $x^0 A = \lambda_A x^0$.

Precisamos ainda mostrar que todas as entradas de x^0 são positivas. Como A possui todas as entradas positivas e x^0 possui pelo menos uma entrada não nula, pelo mesmo raciocínio anterior, concluímos que $x^0 A$ possui todas as entradas estritamente maiores que zero. Pela igualdade que provamos acima, $\lambda_A x^0 = x^0 A$ possui todas as entradas positivas. Então, x^0 possui todas as entradas positivas.

Caso geral: Vamos agora de fato provar o item (i) do Teorema. Denotamos por B a matriz $n \times n$ com todas as entradas iguais a 1. Note que, se $x \geqslant 0$ tem pelo menos uma entrada não nula, $xB \geqslant x$. Se A é uma matriz com todas as entradas positivas e pelo menos uma delas não nula, temos que $A + \delta B$ é uma matriz positiva, para qualquer $\delta > 0$. Sejam $x \geqslant 0$ com pelo menos uma entrada não nula satisfazendo $\sum_j x_j = 1$ e $\{\delta_i\}$ uma sequência decrescente de números reais convergindo a 0. Aplicando o caso 1 para $A + \delta_i B$, é possível encontrar $x(\delta_i)$ com todas as entradas positivas satisfazendo $\sum_j x_j(\delta_i) = 1$ e $x(\delta_i)(A + \delta_i B) = \lambda_{(A+\delta_i B)} x(\delta_i)$.

Analogamente ao caso 1, existe uma subsequência $\{\delta_{n_i}\}$ tal que $x(\delta_{n_i})$ converge a um ponto de acumulação x^0. Seja λ_0 o limite dos $\lambda_{(A+\delta_i B)}$. Note que δ_i decresce para 0 e $\lambda_{(A+\delta_i B)}$ é crescente. Logo, $\lambda_0 \geqslant \lambda_A$. Tomando o limite em ambos os lados da igualdade $x(\delta_{n_i})(A + \delta_{n_i} B) = \lambda_{(A+\delta_{n_i} B)} x(\delta_{n_i})$, segue que $x^0 A = \lambda_0 x^0$ e, portanto, $\lambda_0 \in \Lambda_A$. Logo, $\lambda_0 \leqslant \lambda_A$. Assim, $\lambda_0 = \lambda_A$ e a prova do item *(i)* está completa.

Seja $\lambda \neq \lambda_A$ outro autovalor de A e y o respectivo autovetor não nulo associado à esquerda. Então, podemos dizer que

$$(yA)_j = \sum_{i=1}^{d} a_{ij} y_i = \lambda y_j.$$

Tomando o módulo em ambos os lados da igualdade acima, temos que

$$|\lambda||y_j| = \left|\sum_{i=1}^{d} a_{ij} y_i\right| \leqslant \sum_{i=1}^{d} a_{ij}|y_i|.$$

Segue da desigualdade acima que $|y|A \geqslant |\lambda||y|$, onde $|y| = (|y_1|, \dots, |y_d|)$. Normalizando $|y|$ de forma adequada, concluímos que $|\lambda| \in \Lambda$. Pela definição de λ_A, segue que $|\lambda| \leqslant \lambda_A$. $\square$

Corolário 2.21. *Para qualquer matriz A com entradas não negativas e pelo menos uma delas não nula vale que $\lambda_A = \lambda_{A^\mathsf{T}}$.*

Demonstração. Pelo item *(i)* do Teorema 2.20, λ_A é autovalor de A, enquanto λ_{A^T} é autovalor de A^T. Como os autovalores de A e A^T são iguais, temos que λ_A e λ_{A^T} são autovalores de A e A^T simultaneamente. Aplicando o item *(ii)* do Teorema 2.20 para A, temos $\lambda_{A^\mathsf{T}} \leqslant \lambda_A$. Analogamente para A^T obtemos $\lambda_A \leqslant \lambda_{A^\mathsf{T}}$. Logo, $\lambda_A = \lambda_{A^\mathsf{T}}$. $\square$

Corolário 2.22. *Se P é matriz de transição de uma cadeia de Markov, então $\lambda_P = 1$.*

Demonstração. Veja que P satisfaz as hipóteses do Teorema 2.20. Então, $Q = P^\mathsf{T}$ satisfaz as mesmas hipóteses. Logo, pelo item *(i)* deste teorema, vale que existe u autovetor à esquerda associado a λ_Q com $u_i \geqslant 0$ para $i \in \{1, \dots, d\}$ e $u_i > 0$ para algum i. Portanto, $\sum_{i=1}^{d} u_i = C > 0$. Obtemos então que $v = \frac{u}{C}$ satisfaz: $\sum_{i=1}^{d} v_i = \frac{1}{C}\sum_{i=1}^{d} u_i = 1$; $v_i \geqslant 0$ para qualquer $i \in \{1, \dots, d\}$ e $vQ = \lambda_Q v$. Seja v_k a maior coordenada de v. Note que $v_k > 0$ e satisfaz

$$\lambda_Q v_k = (vQ)_k = \sum_{j=1}^{d} v_j q_{jk} \leqslant \sum_{j=1}^{d} v_k q_{jk} = v_k \sum_{j=1}^{d} p_{kj} = v_k.$$

Logo, $\lambda_Q \leqslant 1$.

Por outro lado, sabemos pela Proposição 2.18 que 1 é autovalor de P. Então, 1 também é autovalor de Q. Logo, pelo item *(ii)* do Teorema 2.20 vale que $1 \leqslant \lambda_Q$.

Portanto, $\lambda_Q = 1$. Logo, $\lambda_P = 1$. $\square$

Segue do Corolário 2.22 que $|\lambda| \leqslant 1$ para qualquer λ autovalor de uma matriz de transição. Além disso, durante a demonstração desse corolário, mostramos que existe um autovetor π associado a 1 com todas as entradas não negativas satisfazendo $\sum_{i=1}^{d} \pi_i = 1$.

Será que, dada qualquer distribuição inicial v_0, vale que $v_n \to \pi$ quando $n \to \infty$, onde π independe da distribuição inicial v_0? Isto não é verdade para qualquer matriz de transição, como podemos ver pelos seguintes exemplos:

Exemplo 2.2. *Seja* $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ *matriz de transição de uma cadeia de Markov com distribuição inicial* v_0*. Então, existe* π *distribuição estacionária, mas não vale que* $v_n \to \pi$ *quando* $n \to \infty$*.*

Veja que $\pi = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$, pois satisfaz $\pi P = \pi$. Note ainda que $e_1 P = e_2$, $e_2 P = e_3$ e $e_3 P = e_1$. Portanto, tomando $v_0 = e_1$, temos que

$$v_n = \begin{cases} e_1, & se \quad n \in \{0, 3, 6, 9, \dots\} \\ e_2, & se \quad n \in \{1, 4, 7, 10, \dots\} \\ e_3 & se \quad n \in \{2, 5, 8, 11, \dots\} \end{cases}$$

e, portanto, $v_n \nrightarrow \pi$.

Exemplo 2.3. *Seja* P *uma matriz que possui o formato* $P = \begin{bmatrix} P_1 & 0_{k \times n} \\ 0_{n \times k} & P_2 \end{bmatrix}$*, onde* P_1 *e* P_2 *são matrizes de transição* $k \times k$ *e* $n \times n$*, respectivamente, com* $n + k = d$ *e as demais entradas são todas nulas. Então* P *possui pelo menos dois autovetores linearmente independentes associados ao autovalor 1.*

De fato, sejam $\pi_1 = (\pi_{11}, \pi_{12}, \dots, \pi_{1k})$ e $\pi_2 = (\pi_{21}, \pi_{22}, \dots, \pi_{2n})$ autovetores associados ao autovalor 1 em P_1 e P_2, respectivamente. Note que $(\pi_{11}, \pi_{12}, \dots, \pi_{1k}, 0, \dots, 0)$ e $(0, \dots, 0, \pi_{21}, \pi_{22}, \dots, \pi_{2n})$ são dois autovetores linearmente independentes de P associados ao autovalor 1.

Dado $v_0 = (v_1, \dots, v_k, u_1, \dots, u_n)$, defina $A = \sum_{j=1}^{k} v_j$ e $B = \sum_{j=1}^{n} u_j$. Observe que as k primeiras coordenadas são aplicadas apenas em P_1, enquanto as restantes em P_2. Supondo que $v_n \to \pi_1$ em P_1 e $v_n \to \pi_2$ em P_2 para quaisquer distribuições iniciais, observe que $v_0 P^n = v_n \to (A\pi_{11}, \dots, A\pi_{1k}, B\pi_{21}, \dots, B\pi_{2n})$, que depende da distribuição inicial v_0.

Vimos então que, nestes exemplos, não temos $v_n \to \pi$ para qualquer v_0. Para garantir isto, precisamos de algumas hipóteses sobre P, presentes no teorema abaixo.

Teorema 2.23. *Seja* P *uma matriz de transição satisfazendo as seguintes hipóteses:*

(i) P *possui exatamente um autovalor* λ *com* $|\lambda| = 1$ *(pela Proposição 2.18, este autovalor é exatamente* $\lambda = 1$*);*

(ii) Dados u, v *autovetores associados a 1, temos que* $u = \alpha v$*,* $\alpha \in \mathbb{R}$ *(o subespaço associado ao autovalor 1 é uma reta);*

(iii) P *é diagonalizável.*

Então, existe um único π *satisfazendo* $\pi_i \geqslant 0$*,* $\pi_i = \sum_{j=1}^{d} \pi_j p_{ji}$ *e* $\sum_{i=1}^{d} \pi_i = 1$*.*

Além disso, dado qualquer v_0 *distribuição inicial, vale que* $P^n v_0 \to \pi$ *quando* $n \to \infty$*.*

Demonstração. Conforme já citado anteriormente, durante a demonstração do Corolário 2.22 mostramos que existe um vetor π associado ao autovalor 1 tal que $\pi_i \geqslant 0$ e $\pi_i = \sum_{j=1}^{d} \pi_j p_{ji}$.

Como π é autovetor associado a 1, vale que $\pi P = \pi$ e, portanto, $\pi_i = (\pi P)_i = \sum_{j=1}^{d} \pi_j p_{ji}$.

Para mostrar a segunda parte, vamos indexar os autovalores em ordem decrescente de módulos, repetindo cada autovalor quantas vezes ele for raiz do polinômio característico, ou seja, $|\lambda_1| \geqslant |\lambda_2| \geqslant \cdots \geqslant |\lambda_d|$. Sabemos que 1 é o autovalor com maior módulo, isto é, $\lambda_1 = 1$. Além disso, as hipóteses *(ii)* e *(iii)* nos dizem que 1 é raiz simples de $p(\lambda)$, pois o fato de P ser diagonalizável significa que existem d autovetores linearmente independentes de P, um associado a cada raiz do polinômio característico. Portanto, a primeira desigualdade é estrita, ou seja, $1 = \lambda_1 > |\lambda_2| \geqslant \cdots \geqslant |\lambda_d|$.

Seja u_i um autovetor associado a λ_i, para qualquer $i \in \{1, \ldots, d\}$. Podemos ainda supor, sem perda de generalidade, que $u_1 = \pi$. Pela hipótese *(iii)*, o conjunto dos u_i é uma base para o espaço em que P atua. Logo, dado v_0 distribuição inicial, podemos escrever $v_0 = \sum_{i=1}^{d} a_i u_i$, onde a_i são os coeficientes de v_0 nesta base.

Seja $u = (1, \ldots, 1)$. Para cada $i \in \{1, \ldots, d\}$, podemos simplificar $u_i P u^\mathsf{T}$ de duas maneiras diferentes:

$$u_i P u^\mathsf{T} = (u_i P) u^\mathsf{T} = \lambda_i u_i u^\mathsf{T}$$
$$u_i P u^\mathsf{T} = u_i (P u^\mathsf{T}) = u_i u^\mathsf{T}.$$

Disto seque que $u_i(\lambda_i - 1)u^\mathsf{T} = 0$ para qualquer $i \in \{1, \ldots, d\}$. Para $i \geqslant 2$, temos $\lambda_i \neq 1$ e, portanto, $0 = u_i u^\mathsf{T} = \sum_{j=1}^{d} (u_i)_j$. Assim,

$$1 = \sum_{j=1}^{d} (v_0)_j = \sum_{j=1}^{d} \sum_{i=1}^{d} a_i (u_i)_j = \sum_{i=1}^{d} a_i \sum_{j=1}^{d} (v_i)_j = a_1 \sum_{j=1}^{d} \pi_j + \sum_{i=2}^{d} a_i \underbrace{\sum_{j=1}^{d} (u_i)_j}_{0} = a_1 \underbrace{\sum_{j=1}^{d} \pi_j}_{1} = a_1.$$

Note que

$$v_0 P^n = \left(\sum_{j=1}^{d} a_j u_j \right) P^n = \sum_{j=1}^{d} a_j u_j P^n = \sum_{j=1}^{d} a_j \lambda_j^n u_j = \underbrace{a_1 \lambda_1^n u_1}_{\pi} + \sum_{j=2}^{d} \lambda_j^n a_j u_j = \pi + \sum_{j=2}^{d} \lambda_j^n a_j u_j.$$

Logo, fixado qualquer v_0, temos que $||v_0 P^n - \pi|| = \left|\left| \sum_{j=2}^{d} \lambda_j^n a_j u_j \right|\right| \leqslant \sum_{j=2}^{d} |\lambda_j|^n |a_j| ||u_j|| \to 0$ quando $n \to \infty$, pois $|\lambda_j| < 1$. Repare que a norma utilizada acima pode ser qualquer norma de $\mathbb{R}^d$.

Assim, $v_n = v_0 P^n \to \pi$ quando $n \to \infty$, para qualquer v_0. □

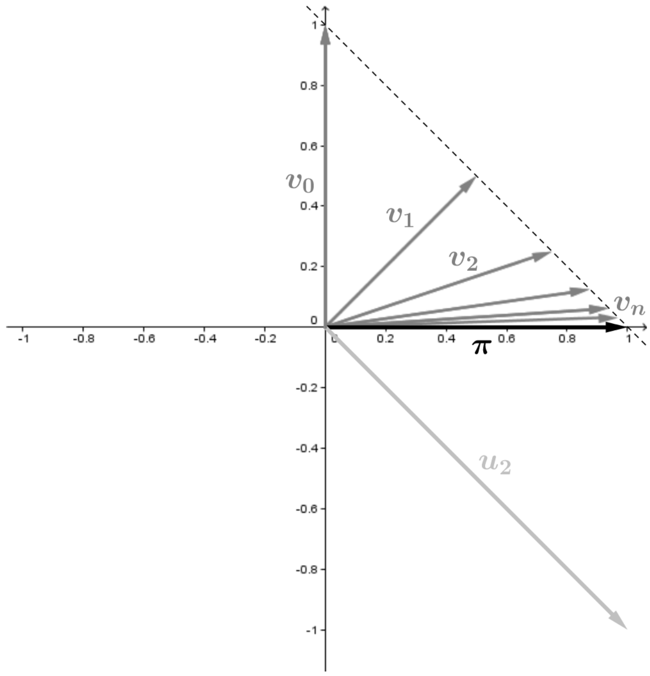
Exemplo 2.4. *No caso em que $d = 2$, é possível visualizar geometricamente que os vetores v_n se aproximam do vetor π. Para fazer isso, vamos considerar o caso em que $v_0 = (0, 1)$ e*

$$P = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Então, o polinômio característico é $p(\lambda) = (1-\lambda)\left(\frac{1}{2}-\lambda\right)$. Logo, os autovalores de P são $\lambda_1 = 1$ e $\lambda_2 = \frac{1}{2}$ e os respectivos autovetores associados são $u_1 = (1,0) = \pi$ e $u_2 = (1,-1)$. É fácil ver que, de fato, $\pi_i \geqslant 0$, $\pi_i = \sum_{j=1}^{d} \pi_j p_{ji}$ e $\sum_{j=1}^{d} \pi_i = 1$. Escrevendo $v_0 = \pi - u_2$, concluímos que

$$v_0 P^n = (\pi - u_2)P^n = \pi P^n + u_2 P^n = \pi - \lambda_2^n u_2 = \pi - \left(\frac{1}{2}\right)^n u_2.$$

Figura 5 – Visão geométrica dos vetores v_n



Observando geometricamente os vetores v_n na Figura 5, podemos perceber que eles estão se aproximando do vetor π.

Até aqui, vimos quais as hipóteses suficientes sobre P para se ter $v_0 P^n \to \pi$ quando $n \to \infty$ para qualquer v_0, no caso geral em que P é uma matriz de transição qualquer. A partir de agora, vamos focar nas matrizes de transição **positivas**, ou seja, matrizes P em que $p_{ij} > 0$ para quaisquer $i, j \in \{1, \ldots, d\}$. Nestes casos, o fato de $v_0 P^n \to \pi$ quando $n \to \infty$ é uma consequência direta do teorema abaixo.

Teorema 2.24 (Teorema de Perron)**.** *Seja A uma matriz positiva $d \times d$. Então,*

(i) $\lambda_A > 0$*;*

(ii) λ_A *é raiz simples de* $p(\lambda)$*;*

(iii) existe um único vetor x tal que $Ax^{\mathsf{T}} = \lambda_A x^{\mathsf{T}}$, $\sum_{i=1}^{d} x_i = 1$ e $x_i \geqslant 0$, para qualquer $i \in \{1, \ldots, d\}$;

(iv) existe um único vetor y tal que $yA = \lambda_A y$, $\sum_{i=1}^{d} x_i y_i = 1$ e $y_i \geqslant 0$, para qualquer $i \in \{1, \ldots, d\}$;

(v) $|\lambda| < \lambda_A$ *para qualquer* $\lambda \neq \lambda_A$ *autovalor de* A*;*

(vi) $(\lambda_A^{-1}A)^n \to x^\mathsf{T} y$ *quando* $n \to \infty$.

A demonstração deste teorema requer um estudo mais aprofundado de Álgebra Linear, mais precisamente, de matrizes positivas e das suas propriedades. Tal estudo não condiz com o foco deste trabaho e, por isso, esta demonstração não será feita aqui. Um maior estudo acerca deste assunto, contendo todos os passos para a demonstração deste teorema, pode ser encontrado em Horn e Johnson (2012, p. 524-528).

Corolário 2.25. *Seja* A *uma matriz positiva e* λ_A *como definido anteriormente. Denotamos por* $A(x,y)$ *a entrada de* A *cuja linha corresponde a* x *e coluna corresponde a* y*, com* $x, y \in E$*. Sejam* $B, C \in \mathcal{E}$ *e* μ, ν *duas medidas sobre* E*. Então, quando* $n \to \infty$*,*

$$\frac{1}{n}\log \sum_{x_0\in B}\sum_{y_0\in C} \nu(x_0)\mu(y_0)A^n(x_0,y_0) \to \log \lambda_A.$$

Demonstração. Podemos escrever

$$\frac{1}{n}\log \sum_{x_0\in B}\sum_{y_0\in C} \nu(x_0)\mu(y_0)A^n(x_0,y_0) = \frac{1}{n}\log \sum_{x_0\in B}\sum_{y_0\in C} \nu(x_0)\mu(y_0)A^n(x_0,y_0)(\lambda_A^{-1})^n\lambda_A^n,$$

ou ainda,

$$\frac{1}{n}\log \sum_{x_0\in B}\sum_{y_0\in C} \nu(x_0)\mu(y_0)A^n(x_0,y_0) = \frac{1}{n}\log \sum_{x_0\in B}\sum_{y_0\in C} \nu(x_0)\mu(y_0)(\lambda_A^{-1}A)^n(x_0,y_0)+\frac{1}{n}\log\lambda_A^n.$$

Por propriedade de logaritmos,

$$\frac{1}{n}\log \sum_{x_0\in B}\sum_{y_0\in C} \nu(x_0)\mu(y_0)A^n(x_0,y_0) = \frac{1}{n}\log \sum_{x_0\in B}\sum_{y_0\in C} \nu(x_0)\mu(y_0)(\lambda_A^{-1}A)^n(x_0,y_0)+\log\lambda_A.$$

Pelo Teorema de Perron, existem (e são únicos) vetores x e y, satisfazendo as condições descritas nos itens *(iii)* e *(iv)*, tais que $(\lambda_A^{-1}A)^n \to xy^\mathsf{T}$, quando $n \to \infty$. Assim,

$$\frac{1}{n}\log \sum_{x_0\in B}\sum_{y_0\in C} \nu(x_0)\mu(y_0)(\lambda_A^{-1}A)^n(x_0,y_0) \to 0,$$

quando $n \to \infty$ e, portanto,

$$\frac{1}{n}\log \sum_{x_0\in B}\sum_{y_0\in C} \nu(x_0)\mu(y_0)A^n(x_0,y_0) \to \log \lambda_A.$$

□

O próximo resultado é uma consequência interessante do que fizemos até aqui, mas isso não é usado no capítulo de grandes desvios para Cadeias de Markov.

Corolário 2.26. *Seja P uma matriz de transição positiva. Então, quando $n \to \infty$,*

$$P^n \to Q = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_d \\ \pi_1 & \pi_2 & \cdots & \pi_d \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_d \end{bmatrix}.$$

Demonstração. Seja $A = P^\mathsf{T}$. Pelos Corolários 2.21 e 2.22, $\lambda_A = 1$. Note que $x = \pi$ e $y = (1, \ldots, 1)$ satisfazem *(iii)* e *(iv)* do Teorema 2.24. Portanto, pelo item *(vi)* deste teorema, temos que

$$A^n \to \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_d \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} = Q^\mathsf{T}.$$

Logo, $P^n \to Q$. □

Com este corolário podemos mostrar que $v_0 P^n \to \pi$, para qualquer v_0. Basta notar que para qualquer distribuição inicial v_0 vale que $v_0 Q = \pi$, pois

$$\begin{aligned} v_0 Q &= \begin{bmatrix} (v_0)_1 & (v_0)_2 & \cdots & (v_0)_d \end{bmatrix} \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_d \\ \pi_1 & \pi_2 & \cdots & \pi_d \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_d \end{bmatrix} \\ &= \begin{bmatrix} \pi_1 \sum_{i=1}^{d} (v_0)_i & \pi_2 \sum_{i=1}^{d} (v_0)_i & \cdots & \pi_d \sum_{i=1}^{d} (v_0)_i \end{bmatrix} \\ &= \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_d \end{bmatrix} = \pi. \end{aligned}$$

Assim, para qualquer v_0, temos que $v_0 P^n \to v_0 Q = \pi$, quando $n \to \infty$, que é exatamente o que queríamos verificar.

Concluímos então que, se P é matriz de transição positiva, vale imediatamente que $v_0 P^n \to \pi$ para qualquer v_0, sem serem necessárias hipóteses adicionais.

Pondo este resultado em notação probabilistica, o que provamos é que se $\mathbb{P}$ é a probabilidade induzida pela cadeia de Markov $\{X_n\}$ e a medida inicial v_0, como em (2.2), temos que

$$\mathbb{P}[X_n = i] \to \pi(i).$$

Como comentamos na Definição de LDP, geralmente queremos estudar o comportamento das médias $\frac{1}{n} \sum_{k=1}^{n} X_k$.

Para cadeias de Markov, o que é estudado é $\frac{1}{n} \sum_{k=1}^{n} f(X_k)$, onde $f : E \to \mathbb{R}$ é uma função limitada, veja o Teorema 2.29. Para apresentarmos este resultado precisamos primeiro apresentar a definição de medida empírica, faremos isso num contexto geral na próxima seção e na seção seguinte vamos aplicar esta definição para entender a convergência de $\frac{1}{n} \sum_{k=1}^{n} f(X_k)$.

2.3 Medidas Empíricas

Se $\{Y_n\}$ é uma sequência de variáveis aleatórias definidas no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ e tomando valores em um espaço polonês $(E, \mathcal{E})$, para cada $\omega \in \Omega$, consideramos como medida empírica a seguinte medida aleatória, que associa a cada conjunto $A \in \mathcal{E}$ o número:

$$L_n(\omega, A) = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i(\omega)}(A), \tag{2.4}$$

onde δ_a é a medida Delta de Dirac, que dá massa 1 ao ponto a. Por simplicidade, vamos omitir ω da notação e denotar apenas por L_n. Note que essa medida empírica é uma probabilidade em E e ela pode ser reescrita da seguinte forma, para cada $A \in \mathcal{E}$,

$$L_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(Y_i). \tag{2.5}$$

L_n é a porção de vezes que a observação cai no conjunto A até o tempo n.

2.4 Teorema Ergódico

Definição 2.27. *Dizemos que uma cadeia de Markov com espaço de estados E e matriz de transição $P = (p_{ij})$ é irredutível se, dados $x, y \in E$, existe $m \in \mathbb{N}$ tal que $p_{ij}^m(x, y) > 0$, isto é, sempre é possível ir de um estado para outro em algum número de passos com probabilidade não nula.*

Definição 2.28. *Seja $\{X_n\}$ uma cadeia de Markov. Definimos o tempo de primeira passagem pelo estado x como uma variável aleatória T_i definida por*

$$T_i(\omega) = \inf\{n \geqslant 1 : X_n(\omega) = i\}.$$

Por convenção, consideraremos $\inf \emptyset = +\infty$.

Teorema 2.29 (Teorema ergódico). *Seja $\{X_n\}$ uma cadeia de Markov irredutível com espaço de estados E e probabilidade de transição p. Então,*

$$\mathbb{P}\left(L_n(\{i\}) \to \frac{1}{m_i}, \text{ quando } n \to \infty \right) = 1,$$

onde L_n foi definido em (2.5) *(usando X_n no lugar de Y_n) e $m_i = \mathbb{E}_i(T_i)$ é o tempo esperado de retorno ao estado i, onde $\mathbb{E}_i$ denota a esperança com respeito a probabilidade $\mathbb{P}_i$ definida por $\mathbb{P}_i(A) = \mathbb{P}[A|X_0 = i]$. Ainda, se $m_i < \infty$ para todo $i \in E$, então, para qualquer função $f : E \to \mathbb{R}$ limitada, temos*

$$\mathbb{P}\left(\frac{1}{n} \sum_{k=1}^{n} f(X_k) \to \sum_{i \in E} \pi_i f(i), \text{ quando } n \to \infty \right) = 1,$$

onde $(\pi_i : i \in E)$ é a única distribuição estacionária da cadeia.

A demonstração deste teorema pode ser encontrada em Norris (1998, p. 53-55).

Note que, no caso $E \subseteq \mathbb{R}$ finito, aplicando o teorema acima para $f : E \to E$ com $f(x) = x$, o que obtemos é

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k \to \sum_{i\in E} \pi_i i, \text{ quando } n \to \infty\right) = 1.$$

Observando isso, repare que este resultado é uma espécie de Lei dos Grandes Números.

3 GRANDES DESVIOS PARA CADEIAS DE MARKOV COM ESPAÇO DE ESTADOS FINITO

Neste capítulo, vamos provar o LDP para a sequência de medidas empíricas associadas a cadeias de Markov. Como vimos no Teorema 2.29, se a cadeia é irredutível, temos que a medida empírica L_n, definida em (2.4), converge para a medida estacionária da cadeia. Aqui vamos vamos obter a taxa exponencial desta convergência, utilizando uma abordagem semelhante àquela feita no teorema para variáveis aleatórias i.i.d.

Vamos estender o estudo de grandes desvios para o caso em que $\{X_n\}$ é uma cadeia de Markov com espaço de estados E, distribuição inicial v_0 e probabilidade de transição p. Por simplicidade, vamos assumir que $E = \{z_1, \ldots, z_d\}$ e $p(x, y) > 0$, para quaisquer $x, y \in E$. Seja

$$P = \begin{bmatrix} p(z_1, z_1) & \cdots & p(z_1, z_d) \\ \vdots & \ddots & \vdots \\ p(z_d, z_1) & \cdots & p(z_d, z_d) \end{bmatrix}$$

a matriz de transição desta cadeia.

Seja $x_0 \in E$. Supondo que $v_0 = \delta_{x_0}$, isto é, x_0 é, de fato, o estado inicial da cadeia de Markov, pelo Teorema da extensão de Kolmogorov (Teorema A.16), definimos $\mathbb{P}_{x_0}$ uma probabilidade em $E^{\mathbb{N}}$ tal que, em cada E^n,

$$\mathbb{P}_{x_0}[X_1 = x_1, \ldots, X_n = x_n] = p(x_0, x_1) \cdots p(x_{n-1}, x_n).$$

Note que

$$\mathbb{P}_{x_0}[X_n = x_n] = \sum_{x_1 \in E} \cdots \sum_{x_{n-1} \in E} p(x_0, x_1) \cdots p(x_{n-1}, x_n).$$

Fixado $n \in \mathbb{N}$, seja $L_n(\omega) = \frac{1}{n} \sum\limits_{i=1}^{n} \delta_{X_i(\omega)}$, como já definido em (2.4) a medida empírica em E. Vamos denotar por $\mathcal{M} = \mathcal{M}(E)$ o conjunto de todas as probabilidades em E. Definimos uma medida $Q_n^{x_0}$ em $\mathcal{M}$ tal que, para cada $A \subseteq \mathcal{M}$,

$$Q_n^{x_0}(A) = \mathbb{P}_{x_0}\left[\{\omega : L_n(\omega) \in A\}\right] := \mathbb{P}_{x_0}\left[L_n \in A\right].$$

Queremos estender esse conceito para cadeias de Markov com outros tipos de medida inicial. Assim, supondo que $v_0 = \mu$, vamos definir $\mathbb{P}_\mu$ por

$$Q_n(A) = \mathbb{P}_\mu\left[L_n \in A\right] = \int \mathbb{P}_{x_0}\left[L_n \in A\right] d\mu(x) = \sum_{x_0 \in E} Q_n^{x_0}(A)\mu(x_0).$$

Teorema 3.1. *Seja $\{X_n\}$ uma cadeia de Markov com espaço de estados E finito, distribuição inicial $\mu \in \mathcal{M}$ e medidas empíricas denotadas por L_n. Então, $\{L_n\}$ satisfaz o LDP com função taxa*

$$I(q) = \sup_V \left(\int V dq - \log \lambda_p(V) \right) = \sup_V \left(\sum_x V(x)q(x) - \log \lambda_p(V) \right)^1, \qquad (3.1)$$

onde o supremo acima é tomado sobre todas as funções $V : E \to \mathbb{R}$ e $\lambda_p(V)$ é o maior autovalor da matriz P^V, cujas entradas são $p^V(x,y) = p(x,y)e^{V(y)}$.

A demonstração desse teorema será feita nas seções abaixo. Antes disso, vamos verificar que a função I, definida em 3.1, é uma candidata a função taxa. Observe que a função I é semicontínua inferiormente, pois é supremo de funções contínuas. Ainda, tomando $V \equiv 0$, temos que $P^V = P$, ou seja, $\lambda_p(V) = 1$. Neste caso, $\int V dq - \log \lambda_p(V) = 0$ e, portanto, I é positiva. Logo, I satisfaz as hipóteses para ser função taxa de um LDP.

Vale ressaltar que, no estudo a seguir, vamos precisar aplicar alguns resultados de análise. A princípio, isso poderia ser um problema, pois tais resultados valem para os espaços $\mathbb{R}^k$ e aqui estamos trabalhando com $\mathcal{M} = \mathcal{M}(E)$, o conjunto das probabilidades sobre E. Mas, na verdade, isso não é problema algum, pois $\mathcal{M}$ é homeomorfo ao simplex de $\mathbb{R}^d$, que definimos como sendo

$$\Theta_d = \left\{ (y_1, \ldots, y_d) \in \mathbb{R}^d : 0 \leqslant y_i \leqslant 1, \text{ para todo } i \in \{1, \ldots, d\}, \text{ e } \sum_{i=1}^d y_i = 1 \right\}, \qquad (3.2)$$

através do homeomorfismo $\Phi : \mathcal{M} \to \Theta_d$ tal que $\Phi(\mu) = (\mu(z_1), \ldots, \mu(z_d))$. Assim sendo, $\mathcal{M}$ e Θ_d são iguais do ponto de vista topológico. Portanto, podemos simplesmente trabalhar com eles como se fossem o mesmo espaço.

3.1 Cota superior

Seja $V : E \to \mathbb{R}$ uma função qualquer e $x_0 \in E$. Então, podemos reescrever

$$Q_n^{x_0}(A) = \mathbb{E}_{x_0}\left[\mathbb{1}_{[L_n \in A]}\right] = \mathbb{E}_{x_0}\left[\exp\left(-\sum_{k=1}^n V(X_k)\right) \mathbb{1}_{[L_n \in A]} \exp\left(\sum_{k=1}^n V(X_k)\right)\right],$$

onde $\exp(x)$ denota e^x.

Note que , se $L_n \in A$,

$$\sum_{k=1}^n V(X_k) = \int_E V(x) n dL_n(x) \geqslant n \inf_{q \in A} \int_E V dq.$$

Portanto,

$$Q_n^{x_0}(A) \leqslant e^{-n \inf\limits_{q \in A} \int V dq} \mathbb{E}_{x_0}\left[\exp\left(\sum_{k=1}^n V(X_k)\right)\right]. \qquad (3.3)$$

[1] Repare que essa função taxa é independente da distribuição inicial μ da cadeia.

Lema 3.2. *Sejam V e P^V como definidas anteriormente. Então,*

$$\mathbb{E}_{x_0}\left[\exp\left(\sum_{k=1}^{n} V(X_k)\right)\right] = \sum_{y\in E}(P^V)^n(x_0,y),$$

onde $(P^V)^n(x_0,y)$ denota a respectiva entrada da matriz $(P^V)^n$.

Demonstração. Note que

$$\mathbb{E}_{x_0}\left[\exp\left(\sum_{k=1}^{n} V(X_k)\right)\right] = \sum_{y_1\in E}\cdots\sum_{y_n\in E}\exp\left(\sum_{k=1}^{n} V(y_k)\right)\mathbb{P}_{x_0}[X_1=y_1,\ldots,X_n=y_n].$$

Como os somatórios são todos finitos, podemos trocar a ordem deles. Além disso, pela a forma como $\mathbb{P}_{x_0}$ foi definida, concluímos que

$$\begin{aligned}\mathbb{E}_{x_0}\left[\exp\left(\sum_{k=1}^{n} V(X_k)\right)\right] &= \sum_{y_n\in E}\sum_{y_1\in E}\cdots\sum_{y_{n-1}\in E} e^{V(y_1)}\cdots e^{V(y_n)}p(x_0,y_1)\cdots p(y_{n-1},y_n)\\ &= \sum_{y_n\in E}\sum_{y_1\in E}\cdots\sum_{y_{n-1}\in E} p^V(x_0,y_1)p^V(y_1,y_2)\cdots p^V(y_{n-1},y_n).\end{aligned}$$

Seguindo o mesmo raciocínio da demonstração da Equação de Chapman-Kolmogorov (veja Lema 2.13), obtemos que

$$\mathbb{E}_{x_0}\left[\exp\left(\sum_{k=1}^{n} V(X_k)\right)\right] = \sum_{y_n\in E}(P^V)^n(x_0,y_n).$$

□

Aplicando o Lema acima, segue de (3.3) que

$$Q_n^{x_0}(A) \leqslant e^{-n\inf\limits_{q\in A}\int V\,dq}\sum_{y\in E}(P^V)^n(x_0,y).$$

Integrando a desigualdade acima com relação a medida inicial μ (do Teorema 3.1), temos

$$Q_n(A) = \sum_{x_0\in E} Q_n^{x_0}(A)\mu(x_0) \leqslant e^{-n\inf\limits_{q\in A}\int V\,dq}\sum_{y\in E}\sum_{x_0\in E}\mu(x_0)(P^V)^n(x_0,y).$$

Denotando por $\lambda_p(V)$ o maior autovalor de P^V e aplicando o Corolário 2.25 para a matriz P^V com $B=E$, $C=E$, $\mu\equiv 1$ e $\nu=\mu$, concluímos que

$$\overline{\lim_{n\to\infty}}\frac{1}{n}\log Q_n(A) \leqslant -\inf_{q\in A}\int V\,dq + \log\lambda_p(V).$$

Podemos reescrever

$$\overline{\lim_{n\to\infty}}\frac{1}{n}\log Q_n(A) \leqslant \sup_{q\in A}\left(-\int V\,dq + \log\lambda_p(V)\right).$$

Observe que a desigualdade acima vale para qualquer V. Minimizando sobre todas as $V: E\to\mathbb{R}$, obtemos

$$\overline{\lim_{n\to\infty}}\frac{1}{n}\log Q_n(A) \leqslant \inf_{V}\sup_{q\in A}\left(-\int V\,dq + \log\lambda_p(V)\right).$$

Pela forma como $Q_n(A)$ foi definida,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}_\mu [L_n \in A] \leqslant \inf_V \sup_{q\in A} \left(-\int V dq + \log \lambda_p(V) \right).$$

Note que, até aqui, estamos trabalhando com $A \subseteq \mathcal{M}$ qualquer. Suponhamos agora que A é aberto. Veja ainda que $-\int V dq + \log \lambda_p(V)$ é semicontínua superiormente em q, já que é contínua. Então, pelo Lema A.21, concluímos que, para qualquer $K \subseteq \mathcal{M}$ compacto,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}_\mu [L_n \in K] \leqslant \sup_{q\in K} \inf_V \left(-\int V dq + \log \lambda_p(V) \right).$$

O conjunto $\mathcal{M}(E)$ das probabilidades em $(E, \mathcal{E})$ é rígido (ver Definição A.26), pois E é finito (logo, compacto) e $q(E) = 1 > 1-\varepsilon$, para quaisquer $\varepsilon > 0$ e $q \in \mathcal{M}(E)$. Pelo Teorema de Prohorov (Teorema A.27), toda sequência de $\mathcal{M}(E)$ tem uma subsequência fracamente convergente, ou seja, $\mathcal{M}(E)$ é relativamente compacto. Portanto, $\overline{\mathcal{M}(E)} = \mathcal{M}(E)$ é compacto.

Sabemos que qualquer conjunto fechado em um espaço compacto é compacto. Logo, dado $F \subseteq \mathcal{M}(E)$ fechado, temos que F é compacto e, portanto,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}_\mu [L_n \in F] \leqslant \sup_{q\in F} \inf_V \left(-\int V dq + \log \lambda_p(V) \right).$$

Reescrevemos a desigualdade acima como

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}_\mu [L_n \in F] \leqslant -\inf_{q\in F} \sup_V \left(\int V dq - \log \lambda_p(V) \right).$$

Logo, pela forma como I foi definida em (3.1),

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}_\mu [L_n \in F] \leqslant -\inf_{q\in F} I(q),$$

concluindo a prova da cota superior.

3.2 Cota inferior

Seja O um aberto de $\mathcal{M} = \mathcal{M}(E)$. Denotamos $O^* = O \cap int(\mathcal{M})$, onde $int(\mathcal{M})$ denota o interior do conjunto $\mathcal{M}$, isto é, $int(\mathcal{M}) = \{\mu \in \mathcal{M} : \mu(x) > 0, \text{ para qualquer } x \in E\}$. Repare que essa noção de interior vem do fato de $\mathcal{M}$ ser homeomorfo a Θ_d.

Se $O = \emptyset$, por convenção, ínfimo sobre o conjunto vazio é $+\infty$ e, portanto, vale o LDP. Suponhamos que $O \neq \emptyset$. Como O é aberto, pela forma como O^* foi definido, podemos concluir que $O^* \neq \emptyset$. Tomamos $q \in O^*$. Então, $q \in O$ e, portanto, existe δ tal que $B_q = B(q, \delta) \subseteq O$. Repare que, devido ao homeomorfismo existente entre $\mathcal{M}$ e Θ_d, esta bola é a pré-imagem da bola $B(\Phi(q), \delta)$ em Θ_d.

Queremos escolher uma nova cadeia de Markov para a qual q é um vetor estacionário e a probabilidade de transição $\tilde{p}$ desta cadeia satisfaça $\tilde{p}(x, y) > 0$, para quaisquer $x, y \in E$. Note que, para que essa escolha faça sentido, precisamos garantir que exista pelo menos uma cadeia satisfazendo as seguintes condições:

(i) $\tilde{p}(x,y) > 0$, para quaisquer $x, y \in E$;

(ii) $\sum_{x\in E} q(x)\tilde{p}(x,y) = q(y)$, para todo $y \in E$.

Como tomamos $q \in O^*$, sabemos que $q(y) > 0$, para qualquer $y \in E$. Assim, se tomarmos $\tilde{p}(x,y) = q(y)$, teremos $\tilde{p}(x,y) > 0$, para quaisquer x, y, e

$$\sum_x q(x)\tilde{p}(x,y) = q(y)\sum_x q(x) = q(y),$$

ou seja, $\tilde{p}(x,y) = q(y)$ satisfaz as hipóteses *(i)* e *(ii)* acima. Portanto, podemos escolher uma nova cadeia de Markov para a qual q é um vetor estacionário e a probabilidade de transição $\tilde{p}$ desta cadeia satisfaça $\tilde{p}(x,y) > 0$, para quaisquer x, y.

Definimos $\tilde{\mathbb{P}}_q$ de modo que, para qualquer $H : E^{n+1} \to \mathbb{R}$,

$$\int H d\tilde{\mathbb{P}}_q = \sum_{x_0,x_1,\ldots,x_n} H(x_0,\ldots,x_n)q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-1},x_n). \tag{3.4}$$

Qual a relação entre $\mathbb{P}_\mu$ e $\tilde{\mathbb{P}}_q$? Para obter essa relação, note que, para qualquer $B_n \in \mathcal{E}^{n+1}$,

$$\mathbb{P}_\mu(B_n) = \sum_{(x_0,\ldots,x_n)\in B_n} \mu(x_0)p(x_0,x_1)\cdots p(x_{n-1},x_n).$$

Podemos reescrever a equação acima como

$$\mathbb{P}_\mu(B_n) = \sum_{(x_0,\ldots,x_n)\in B_n} \frac{\mu(x_0)p(x_0,x_1)\cdots p(x_{n-1},x_n)}{q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-1},x_n)} q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-1},x_n).$$

Denotando $V_n = V_n(x_0,\ldots,x_n) = -\log\left[\frac{q(x_0)}{\mu(x_0)}\right] - \sum_{i=1}^n \log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right]$, concluímos que

$$\mathbb{P}_\mu(B_n) = \sum_{(x_0,\ldots,x_n)\in B_n} e^{V_n} q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-1},x_n) = \int_{B_n} e^{V_n} d\tilde{\mathbb{P}}_q,$$

para qualquer B_n. Essa é a relação entre $\mathbb{P}_\mu$ e $\tilde{\mathbb{P}}_q$!

Definimos $A_n = \left[\frac{1}{n}\sum_{i=1}^n \delta_{X_i} \in B_q\right]$. Como $B_q \subseteq O$,

$$\varliminf_{n\to\infty} \frac{1}{n}\log\mathbb{P}_\mu\left[\frac{1}{n}\sum_{i=1}^n \delta_{X_i} \in O\right] \geqslant \varliminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\mu\left[\frac{1}{n}\sum_{i=1}^n \delta_{X_i} \in B_q\right] = \varliminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\mu(A_n).$$

Pela relação existente entre $\mathbb{P}_\mu$ e $\tilde{\mathbb{P}}_q$, obtemos que

$$\mathbb{P}_\mu(A_n) = \int_{A_n} e^{V_n} d\tilde{\mathbb{P}}_q = \left(\int_{A_n} e^{V_n} d\frac{\tilde{\mathbb{P}}_q}{\tilde{\mathbb{P}}_q(A_n)}\right)\tilde{\mathbb{P}}_q(A_n).$$

Aplicando a Desigualdade de Jensen para funções côncavas (Corolário A.11) e usando propriedade de logarítmos, concluímos que

$$\log\mathbb{P}_\mu(A_n) = \log\int_{A_n} e^{V_n} d\frac{\tilde{\mathbb{P}}_q}{\tilde{\mathbb{P}}_q(A_n)} + \log\tilde{\mathbb{P}}_q(A_n) \geqslant \frac{\int_{A_n} V_n d\tilde{\mathbb{P}}_q}{\tilde{\mathbb{P}}_q(A_n)} + \log\tilde{\mathbb{P}}_q(A_n).$$

Logo,

$$\varliminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\mu\left[\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}\in O\right]\geqslant\varliminf_{n\to\infty}\frac{\frac{1}{n}\int_{A_n}V_n d\tilde{\mathbb{P}}_q}{\tilde{\mathbb{P}}_q(A_n)}+\varliminf_{n\to\infty}\frac{1}{n}\log\tilde{\mathbb{P}}_q(A_n). \tag{3.5}$$

Queremos determinar quanto vale o lado direito da desigualdade acima. Para isso, vamos ver a seguir uma sequência de lemas que nos levarão ao resultado desejado.

Lema 3.3. *Sejam $\tilde{\mathbb{P}}_q$ e A_n como definidos anteriormente. Então, quando $n\to+\infty$,*

(i) $\tilde{\mathbb{P}}_q(A_n)\to 1$;

(ii) $\displaystyle\frac{1}{n}\int_{A_n^{\complement}}V_n d\tilde{\mathbb{P}}_q\to 0.$

Demonstração. Como escolhemos $\tilde{p}$ tal que $\tilde{p}(x,y)>0$, para quaisquer $x,y\in E$, temos que esta cadeia de Markov é irredutível (veja Definição 2.27). Sabemos ainda que q é estacionária para esta cadeia. Como $q(z_i)>0$, para qualquer $z_i\in E$, podemos concluir que a cadeia passa em algum momento em cada um dos estados possíveis, ou seja, se esta cadeia iniciar num estado z_i, sabemos que ela, em algum momento, terá passado por cada um dos estados $z_k\in E$, desde que tenhamos a observado por um tempo suficientemente grande. Em particular, temos que a cadeia retorna para o estado z_i.

O argumento acima nos dá uma ideia de que o tempo esperado de retorno ao estado z_i é finito. Portanto, segue do Teorema ergódico 2.29 que

$$\tilde{\mathbb{P}}_q\left(\frac{1}{n}\sum_{k=1}^{n}f(X_k)\to\sum_{z_i\in E}q(z_i)f(z_i),\ \text{quando } n\to\infty\right)=1. \tag{3.6}$$

Lembre que tínhamos definido $L_n=\frac{1}{n}\sum\limits_{j=1}^{n}\delta_{X_j}$. Além disso, note que, para qualquer f contínua e limitada,

$$\int f dL_n=\frac{1}{n}\sum_{j=1}^{n}f(X_j)$$

e

$$\int f dq=\sum_{z_i\in E}q(z_i)f(z_i).$$

Assim, segue de (3.6) que

$$\tilde{\mathbb{P}}_q\left(\int f dL_n\to\int f dq,\ \text{quando } n\to\infty\right)=1.$$

Por definição de convergência fraca de medidas (veja Definição A.25), concluímos que

$$\tilde{\mathbb{P}}_q(L_n\Rightarrow q)=1.$$

Como a convergência fraca de medidas é equivalente a convergência no simplex de $\mathbb{R}^d$ (veja Seção A.3), repare que, se $L_n\Rightarrow q$, então existe n_0 tal que $L_n\in B_q=B(q,\delta)$, para qualquer $n\geqslant n_0$. Disto segue que, quando $n\to\infty$,

$$\tilde{\mathbb{P}}_q[L_n\in B_q]\to 1,$$

concluindo a demonstração do primeiro item, pois definimos $A_n = [L_n \in B_q]$.

Vamos agora provar o item *(ii)* do lema. Pela forma como V_n foi definida, podemos reescrever

$$\frac{1}{n}\int_{A_n^\complement} V_n d\tilde{\mathbb{P}}_q = -\frac{1}{n}\int_{A_n^\complement} \log\left[\frac{q(x_0)}{\mu(x_0)}\right] d\tilde{\mathbb{P}}_q - \frac{1}{n}\int_{A_n^\complement} \sum_{i=1}^{n} \log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] d\tilde{\mathbb{P}}_q. \tag{3.7}$$

Note que, pelo item *(i)*, quando $n \to \infty$,

$$-\frac{1}{n}\int_{A_n^\complement} \log\left[\frac{q(x_0)}{\mu(x_0)}\right] d\tilde{\mathbb{P}}_q = -\frac{1}{n}\log\left[\frac{q(x_0)}{\mu(x_0)}\right]\tilde{\mathbb{P}}_q(A_n^\complement) \to 0.$$

Portanto, a primeira parcela da soma no lado direito da igualdade (3.7) converge a zero. Falta mostrar que a segunda também converge a zero. Para isso, note que p e $\tilde{p}$ são probabilidade de transição estritamente positivas e, portanto, existem constantes C_1 e C_2 tais que $C_1 \leqslant \tilde{p}(x,y) \leqslant 1$ e $C_2 \leqslant p(x,y) \leqslant 1$, para quaisquer $x, y \in E$. Então,

$$n\log C_1 = \sum_{i=1}^{n}\log C_1 \leqslant \sum_{i=1}^{n}\log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] \leqslant \sum_{i=1}^{n}\log\left(\frac{1}{C_2}\right) = n\log\left(\frac{1}{C_2}\right).$$

Logo,

$$\int_{A_n^\complement}\frac{1}{n}(n\log C_1)d\tilde{\mathbb{P}}_q \leqslant \int_{A_n^\complement}\frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] d\tilde{\mathbb{P}}_q \leqslant \int_{A_n^\complement}\frac{1}{n}(-n\log C_2)d\tilde{\mathbb{P}}_q.$$

Repare que, pelo item *(i)*, quando $n \to \infty$,

$$\int_{A_n^\complement}\frac{1}{n}n\log C_1 d\tilde{\mathbb{P}}_q = \log C_1\tilde{\mathbb{P}}_q(A_n^\complement) \to 0$$

e

$$\int_{A_n^\complement}\frac{1}{n}n(-\log C_2)d\tilde{\mathbb{P}}_q = -\log C_2\tilde{\mathbb{P}}_q(A_n^\complement) \to 0.$$

Então, quando $n \to 0$,

$$\int_{A_n^\complement}\frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] d\tilde{\mathbb{P}}_q \to 0.$$

Assim, quando $n \to \infty$,

$$\frac{1}{n}\int_{A_n^\complement} V_n d\tilde{\mathbb{P}}_q \to 0.$$

□

Lema 3.4. *Assumindo as definições anteriores, temos que*

$$\sum_{x_0,x_1} q(x_0)\tilde{p}(x_0,x_1)f(x_1) = \sum_{x_0} q(x_0)f(x_0),$$

para qualquer função f.

Demonstração. Como q é invariante, sabemos que

$$\sum_{x_0} q(x_0)\tilde{p}(x_0, x_1) = q(x_1).$$

Portanto, concluímos que

$$\sum_{x_0,x_1} q(x_0)\tilde{p}(x_0, x_1)f(x_1) = \sum_{x_1} f(x_1) \sum_{x_0} q(x_0)\tilde{p}(x_0, x_1) = \sum_{x_1} f(x_1)q(x_1).$$

□

Lema 3.5. *Utilizando as mesmas definições anteriores, temos que*

$$\varliminf_{n\to\infty} \frac{1}{n} \int_{A_n} V_n d\tilde{\mathbb{P}}_q = -\sum_{x,y} \tilde{p}(x, y) \log \left(\frac{\tilde{p}(x, y)}{p(x, y)}\right) q(x).$$

Demonstração. Segue de (3.4) que

$$\int_A V_n d\tilde{\mathbb{P}}_q = \sum_{(x_0,x_1,\ldots,x_n)\in A} V_n q(x_0)\tilde{p}(x_0, x_1)\cdots\tilde{p}(x_{n-1}, x_n).$$

Podemos ainda decompor

$$\frac{1}{n}\int_{A_n} V_n d\tilde{\mathbb{P}}_q = \frac{1}{n}\int V d\tilde{\mathbb{P}}_q - \frac{1}{n}\int_{A_n^{\complement}} V_n d\tilde{\mathbb{P}}_q.$$

Então,

$$\begin{aligned}\frac{1}{n}\int_{A_n} V_n d\tilde{\mathbb{P}}_q = &\frac{1}{n}\sum_{x_0,\ldots,x_n}\left(-\log\left[\frac{q(x_0)}{\mu(x_0)}\right] q(x_0)\tilde{p}(x_0, x_1)\cdots\tilde{p}(x_{n-1}, x_n)\right)\\ &+\frac{1}{n}\sum_{x_0,\ldots,x_n}\left(-\sum_{i=1}^{n}\log\left[\frac{\tilde{p}(x_{i-1}, x_i)}{p(x_{i-1}, x_i)}\right] q(x_0)\tilde{p}(x_0, x_1)\cdots\tilde{p}(x_{n-1}, x_n)\right)\\ &-\frac{1}{n}\int_{A_n^{\complement}} V_n d\tilde{\mathbb{P}}_q.\end{aligned} \tag{3.8}$$

Por simplicidade, vamos analisar cada uma das três parcelas da soma acima separadamente. Pelo item *(ii)* do Lema 3.3, sabemos que a última parcela da soma converge a zero quando $n \to \infty$. Vamos agora analisar a primeira parcela dessa soma. Podemos reescreve-la como

$$-\frac{1}{n}\sum_{x_0} q(x_0)\log\frac{q(x_0)}{\mu(x_0)}\sum_{x_1}\tilde{p}(x_0, x_1)\cdots\sum_{x_{n-1}}\tilde{p}(x_{n-2}, x_{n-1})\sum_{x_n}\tilde{p}(x_{n-1}, x_n).$$

Como $\tilde{p}$ é probabilidade de transição, vale que $\sum_y \tilde{p}(x, y) = 1$. Então, calculando o valor de cada somatório acima, da direita para a esquerda, nos sobra apenas

$$-\frac{1}{n}\sum_{x_0} q(x_0)\log\frac{q(x_0)}{\mu(x_0)} = -\frac{1}{n}H(q|\mu),$$

onde $H(q|\mu)$ é a entropia relativa[2] de q com relação à μ.

[2] Ver Kipnis e Landim (2013, p.338-340) para mais detalhes sobre entropia relativa

Logo, a primeira parcela da soma em (3.8) também converge a zero quando $n \to \infty$. Portanto, segue que

$$\lim_{n\to\infty} \frac{1}{n} \int_{A_n} V_n d\tilde{\mathbb{P}}_q = \lim_{n\to\infty} -\frac{1}{n} \sum_{x_0,\ldots,x_n} \sum_{i=1}^{n} \log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-1},x_n). \tag{3.9}$$

Podemos reescrever

$$\sum_{i=1}^{n} \log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] = \sum_{i=1}^{n-1} \log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] + \log\left[\frac{\tilde{p}(x_{n-1},x_n)}{p(x_{n-1},x_n)}\right].$$

Concluímos de (3.9) que

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \int_{A_n} V_n d\tilde{\mathbb{P}}_q = \lim_{n\to\infty} -\frac{1}{n} \Bigg(&\sum_{x_0,\ldots,x_n} \sum_{i=1}^{n-1} \log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-1},x_n) \\ &+ \sum_{x_0,\ldots,x_n} \log\left[\frac{\tilde{p}(x_{n-1},x_n)}{p(x_{n-1},x_n)}\right] q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-1},x_n)\Bigg).\end{aligned}$$

Reescrevendo a primeira parcela da soma do lado direito da equação anterior, obtemos

$$\sum_{x_0,\ldots,x_{n-1}} \sum_{i=1}^{n-1} \log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-2},x_{n-1}) \underbrace{\sum_{x_n} \tilde{p}(x_{n-1},x_n)}_{1}.$$

Logo,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \int_{A_n} V_n d\tilde{\mathbb{P}}_q = \lim_{n\to\infty} -\frac{1}{n} \Bigg(&\sum_{x_0,\ldots,x_{n-1}} \sum_{i=1}^{n-1} \log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-2},x_{n-1}) \\ &+ \sum_{x_0,\ldots,x_n} \log\left[\frac{\tilde{p}(x_{n-1},x_n)}{p(x_{n-1},x_n)}\right] q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-1},x_n)\Bigg).\end{aligned} \tag{3.10}$$

Podemos reescrever a segunda parcela da soma do lado direito dessa igualdade como

$$\sum_{x_0,\ldots,x_{n-1}} q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-2},x_{n-1}) \sum_{x_n} \tilde{p}(x_{n-1},x_n) \log\left[\frac{\tilde{p}(x_{n-1},x_n)}{p(x_{n-1},x_n)}\right], \tag{3.11}$$

Seja ψ função tal que

$$\psi(z) = \sum_y \log\left(\frac{\tilde{p}(z,y)}{p(z,y)}\right) \tilde{p}(z,y).$$

Então, reescrvemos (3.11) como

$$\sum_{x_0,\ldots,x_{n-1}} q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-2},x_{n-1})\psi(x_{n-1}).$$

Denotando

$$f(z) = \sum_{x_2,\ldots,x_{n-1}} \tilde{p}(z,x_2)\cdots\tilde{p}(x_{n-2},x_{n-1})\psi(x_{n-1})$$

e aplicando o Lema 3.4, obtemos

$$\sum_{x_0,\ldots,x_{n-1}} q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-2},x_{n-1})\psi(x_{n-1}) = \sum_{x_0,x_1} q(x_0)\tilde{p}(x_0,x_1)f(x_1) = \sum_{x_0} q(x_0)f(x_0).$$

ote que

$$\sum_{x_0} q(x_0)f(x_0) = \sum_{x_0} q(x_0) \sum_{x_2,\ldots,x_{n-1}} \tilde{p}(x_0,x_2)\cdots\tilde{p}(x_{n-2},x_{n-1})\psi(x_{n-1})$$

e, portanto,

$$\sum_{x_0,\ldots,x_{n-1}} q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-2},x_{n-1})\psi(x_{n-1}) = \sum_{x_0,x_2} q(x_0)\tilde{p}(x_0,x_2)g(x_2),$$

onde

$$g(z) = \sum_{x_3,\ldots,x_{n-1}} \tilde{p}(z,x_3)\cdots\tilde{p}(x_{n-2},x_{n-1})\psi(x_{n-1}).$$

Logo, aplicando novamente o Lema 3.4,

$$\sum_{x_0,\ldots,x_{n-1}} q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-2},x_{n-1})\psi(x_{n-1}) = \sum_{x_0} q(x_0)g(x_0).$$

Repetindo o argumento anterior diversas vezes, concluímos que

$$\sum_{x_0,\ldots,x_{n-1}} q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-2},x_{n-1})\psi(x_{n-1}) = \sum_{x_0} q(x_0)h(x_0),$$

onde

$$h(z) = \sum_{x_{n-1}} \tilde{p}(z,x_{n-1})\psi(x_{n-1}).$$

Note que, aplicando mais uma vez o Lema 3.4, obtemos

$$\sum_{x_0} q(x_0)g(x_0) = \sum_{x_0,x_{n-1}} q(x_0)\tilde{p}(x_0,x_{n-1})\psi(x_{n-1}) = \sum_{x_0} q(x_0)\psi(x_0).$$

Logo,

$$\sum_{x_0,\ldots,x_{n-1}} q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-2},x_{n-1})\psi(x_{n-1}) = \sum_{x_0} q(x_0)\psi(x_0).$$

Substituindo em (3.10), temos que

$$\varliminf_{n\to\infty} \frac{1}{n}\int_{A_n} V_n d\tilde{\mathbb{P}}_q = \varliminf_{n\to\infty} -\frac{1}{n}\left(\sum_{x_0,\ldots,x_{n-1}} \sum_{i=1}^{n-1} \log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-2},x_{n-1}) \right.$$
$$\left. + \sum_{x_0} q(x_0)\psi(x_0)\right).$$

Podemos reescrever a equação acima como

$$\varliminf_{n\to\infty} \frac{1}{n}\int_{A_n} V_n d\tilde{\mathbb{P}}_q = \varliminf_{n\to\infty} -\frac{1}{n}\left(\sum_{x_0,\ldots,x_{n-1}} \sum_{i=1}^{n-2} \log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-2},x_{n-1}) \right.$$
$$+ \sum_{x_0,\ldots,x_{n-1}} \log\left[\frac{\tilde{p}(x_{n-2},x_{n-1})}{p(x_{n-2},x_{n-1})}\right] q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-2},x_{n-1})$$
$$\left. + \sum_{x_0} q(x_0)\psi(x_0)\right).$$

Utilizando o mesmo argumento anterior, obtemos

$$\lim_{n\to\infty} \frac{1}{n}\int_{A_n} V_n d\tilde{\mathbb{P}}_q = \lim_{n\to\infty} -\frac{1}{n}\Bigg(\sum_{x_0,\ldots,x_{n-2}} \sum_{i=1}^{n-2} \log\left[\frac{\tilde{p}(x_{i-1},x_i)}{p(x_{i-1},x_i)}\right] q(x_0)\tilde{p}(x_0,x_1)\cdots\tilde{p}(x_{n-3},x_{n-2}) + 2\sum_{x_0} q(x_0)\psi(x_0)\Bigg).$$

Repetindo esse argumento várias vezes, concluímos que

$$\lim_{n\to\infty} \frac{1}{n}\int_{A_n} V_n d\tilde{\mathbb{P}}_q = \lim_{n\to\infty} -\frac{1}{n}\left(\sum_{x_0,x_1} \log\left[\frac{\tilde{p}(x_0,x_1)}{p(x_0,x_1)}\right] q(x_0)\tilde{p}(x_0,x_1) + (n-1)\sum_{x_0} q(x_0)\psi(x_0)\right).$$

Note que

$$\sum_{x_0,x_1} \log\left[\frac{\tilde{p}(x_0,x_1)}{p(x_0,x_1)}\right] q(x_0)\tilde{p}(x_0,x_1) = \sum_{x_0} q(x_0)\sum_{x_1}\log\left[\frac{\tilde{p}(x_0,x_1)}{p(x_0,x_1)}\right]\tilde{p}(x_0,x_1) = \sum_{x_0} q(x_0)\psi(x_0).$$

Logo,

$$\lim_{n\to\infty} \frac{1}{n}\int_{A_n} V_n d\tilde{\mathbb{P}}_q = \lim_{n\to\infty} -\frac{1}{n}\left(n\sum_{x_0} q(x_0)\psi(x_0)\right) = -\sum_{x_0} q(x_0)\psi(x_0).$$

Pela forma como ψ foi definida, temos que

$$\lim_{n\to\infty} \frac{1}{n}\int_{A_n} V_n d\tilde{\mathbb{P}}_q = -\sum_{x_0} q(x_0)\sum_{y}\log\left(\frac{\tilde{p}(x_0,y)}{p(x_0,y)}\right)\tilde{p}(x_0,y).$$

Então,

$$\lim_{n\to\infty} \frac{1}{n}\int_{A_n} V_n d\tilde{\mathbb{P}}_q = -\sum_{x,y} \tilde{p}(x,y)\log\left(\frac{\tilde{p}(x,y)}{p(x,y)}\right)q(x).$$

□

Pelo Lema 3.5 e pelo item *(i)* do Lema 3.3, segue de (3.5) que

$$\varliminf_{n\to\infty} \frac{1}{n}\log\mathbb{P}_\mu\left[\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}\in O\right] \geqslant -\sum_{x,y}\tilde{p}(x,y)\log\frac{\tilde{p}(x,y)}{p(x,y)}q(x).$$

Veja que a desigualdade acima vale para qualquer $\tilde{p}$ tal que $\sum\limits_{x} q(x)\tilde{p}(x,y) = q(y)$ e $\tilde{p}(x,y)>0$, para qualquer $x,y\in E$. Vamos denotar por $\tilde{\Pi}$ o conjunto de todas as $\tilde{p}$ com essas propriedades. Maximizando em $\tilde{p}\in\tilde{\Pi}$, obtemos

$$\varliminf_{n\to\infty} \frac{1}{n}\log\mathbb{P}_\mu\left[\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}\in O\right] \geqslant -\inf_{\tilde{p}\in\tilde{\Pi}}\sum_{x,y}\tilde{p}(x,y)\log\frac{\tilde{p}(x,y)}{p(x,y)}q(x).$$

Definindo $L_n(\omega) = \frac{1}{n}\sum\limits_{i=1}^{n}\delta_{X_i(\omega)}$ e

$$\tilde{I}(q) = \inf_{\tilde{p}\in\tilde{\Pi}}\sum_{x,y}\tilde{p}(x,y)\log\frac{\tilde{p}(x,y)}{p(x,y)}q(x), \tag{3.12}$$

concluímos que, para qualquer $q\in O^*$,

$$\varliminf_{n\to\infty} \frac{1}{n}\log\mathbb{P}_\mu\left[L_n\in O\right] \geqslant -\tilde{I}(q). \tag{3.13}$$

Repare que, para concluirmos a demonstração da cota inferior, precisamos que o lado direito da desigualdade seja $-\inf_{q\in O} I(q)$. Porém, no momento, maximizando em $q \in O^*$, poderíamos apenas obter $-\inf_{q\in O^*} \tilde{I}(q)$. Para consguirmos o que queremos, precisamos trabalhar mais um pouco antes de maximizar. Precisaremos agora da desigualdade fornecida pelo lema abaixo.

Lema 3.6. *Sejam I e $\tilde{I}$ como definidas em* (3.1) *e* (3.12)*, respectivamente. Então, para qualquer $q \in O^*$, $\tilde{I}(q) \leqslant I(q)$.*

Demonstração. Para simplificar a notação, vamos denotar $\sum_y q(y)p(y,x) = (qp)(x)$. Começaremos encontrando uma função f_q tal que $\tilde{p}(x,y) = p(x,y)e^{f_q(y)}$ e $q(x) = (q\tilde{p})(x)$. Queremos então encontrar uma função f_q que satisfaça a equação

$$q(x) = (q\tilde{p})(x) = \sum_y q(y)\tilde{p}(y,x) = \underbrace{\sum_y q(y)p(y,x)}_{(qp)(x)}\, e^{f_q(x)} = (qp)(x)e^{f_q(x)},$$

para qualquer $x \in E$. Assim,

$$f_q(x) = \log\left(\frac{q(x)}{(qp)(x)}\right).$$

Note que

$$\log\frac{\tilde{p}(x,y)}{p(x,y)} = \log\frac{p(x,y)e^{f_q(y)}}{p(x,y)} = f_q(y),$$

para qualquer $x \in E$. Portanto,

$$\sum_{x,y} q(x)\tilde{p}(x,y)\log\frac{\tilde{p}(x,y)}{p(x,y)} = \sum_y \underbrace{\sum_x q(x)\tilde{p}(x,y)}_{q(y)}\, f_q(y) = \sum_y q(y)f_q(y). \tag{3.14}$$

Agora, queremos encontrar qual o valor de $\log\lambda_p(f_q)$, onde $\lambda_p(f_q)$ é o autovalor maximal da matriz P^{f_q}. Aplicando o Corolário 2.25 para a matriz P^{f_q} com $B = E$, $C = \{x\}$ e $\mu \equiv 1$, concluímos que

$$\log\lambda_p(f_q) = \lim_{n\to\infty}\frac{1}{n}\log\sum_y v(y)(p^{f_q})^n(y,x) = \lim_{n\to\infty}\frac{1}{n}\log(\nu(\tilde{p})^n)(x),$$

para quaisquer $x \in E$ e ν medida sobre E.

Sabemos que $(qp^{f_q})(x) = (q\tilde{p})(x) = q(x)$, isto é, q é um autovetor a esquerda de P^{f_q} associado ao autovalor 1. Portanto, q é autovetor a esquerda de $(P^{f_q})^n$ associado ao autovalor 1, ou seja, $(q(\tilde{p})^n)(x) = q(x)$, para qualquer $n \in \mathbb{N}$. Tomando $\nu = q$ na igualdade anterior, obtemos

$$\log\lambda_p(f_q) = \lim_{n\to\infty}\frac{1}{n}\log\underbrace{(q(\tilde{p})^n)(x)}_{q(x)} = \lim_{n\to\infty}\frac{1}{n}\log q(x) = 0.$$

Assim, podemos subtrair $\log \lambda_p(f_q)$ no lado direito de (3.14) sem alterar a igualdade. Portanto, para esta $\tilde{p} = p^{f_q}$, vale

$$\sum_{x,y} q(x)\tilde{p}(x,y) \log \frac{\tilde{p}(x,y)}{p(x,y)} = \sum_y q(y) f_q(y) - \log \lambda_p(f_q).$$

Portanto,

$$\tilde{I}(q) = \inf_{\tilde{p} \in \tilde{\Pi}} \sum_{x,y} q(x)\tilde{p}(x,y) \log \frac{\tilde{p}(x,y)}{p(x,y)} \leqslant \sum_y q(y) f_q(y) - \log \lambda_p(f_q).$$

Logo, para qualquer $q \in O^*$,

$$\tilde{I}(q) \leqslant \sum_y q(y) f_q(y) - \log \lambda_p(f_q) \leqslant \sup_V \left(\sum_y q(y) V(y) - \log \lambda_p(V) \right) = I(q).$$

$\square$

Pelo Lema 3.6, segue de (3.13) que, para qualquer $q \in O^*$,

$$\varliminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}_\mu \left[L_n \in O \right] \geqslant -\tilde{I}(q) \geqslant -I(q).$$

Maximizando em $q \in O^*$ a desigualdade acima, obtemos

$$\varliminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}_\mu \left[L_n \in O \right] \geqslant - \inf_{q \in O^*} I(q). \tag{3.15}$$

Observe que estamos quase concluindo a prova da cota inferior do LDP. Para isso, precisamos apenas que o ínfimo no lado direito da desigualdade seja tomado em $q \in O$, ou seja, precisamos retirar a restrição de que q é positivo.

Vamos agora nos dedicar a provar que $\inf\limits_{q \in O^*} I(q) = \inf\limits_{q \in O} I(q)$. Começaremos provando outros dois lemas, que serão necessários na demonstração dessa igualdade.

Lema 3.7. *A função I, definida em (3.1), é convexa.*

Demonstração. Sejam $q_1, q_2 \in \mathcal{M}$ e $\lambda, \beta \in [0,1]$ com $\lambda + \beta = 1$. Vamos denotar

$$J_V(q) = \sum_y q(y) V(y) - \log \lambda_p(V).$$

Repare que a função J_V satisfaz $J_V(\lambda q_1 + \beta q_2) = \lambda J_V(q_1) + \beta J_V(q_2)$, pois $\lambda + \beta = 1$. Portanto,

$$I(\lambda q_1 + \beta q_2) = \sup_V J_V(\lambda q_1 + \beta q_2) = \sup_V \{ \lambda J_V(q_1) + \beta J_V(q_2) \}.$$

Como $\lambda \geqslant 0$ e $\beta \geqslant 0$, por propriedade do supremo segue que

$$I(\lambda q_1 + \beta q_2) \leqslant \lambda \sup_V J_V(q_1) + \beta \sup_V J_V(q_2) = \lambda I(q_1) + \beta I(q_2).$$

Logo, I é convexa. $\square$

Lema 3.8. *Sejam I e $I^{(1)}$ tais que, para qualquer $q \in \mathcal{M} = \mathcal{M}(E)$,*

$$I(q) = \sup_V \left(\sum_{x \in E} q(x)V(x) - \log \lambda_p(V) \right)$$

e

$$I^{(1)}(q) = \sup_{u>0} \sum_{y \in E} q(y) \log \left(\frac{u(y)}{(up)(y)} \right),$$

onde $u : E \to (0, +\infty)$ e $(up)(y) = \sum_{x \in E} u(x)p(x, y)$. Então, $I = I^{(1)}$.

Demonstração. Vamos começar mostrando que $I^{(1)}(q) \leqslant I(q)$, para qualquer $q \in \mathcal{M}$. Para isso, seja $u : E \to (0, +\infty)$. Vamos denotar

$$V_0(x) = \log \left(\frac{u(x)}{(up)(x)} \right).$$

Então,

$$\sum_{x \in E} q(x)V_0(x) = \sum_{x \in E} q(x) \log \frac{u(x)}{(up)(x)}.$$

Segue do Corolário 2.25 que, para quaisquer $x \in E$ e $v : E \to \mathbb{R}$,

$$\log \lambda_p(V) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{y \in E} v(y)(P^V)^n(x, y) = \lim_{n\to\infty} \frac{1}{n} \log(v(P^V)^n)(x).$$

Note que

$$(uP^{V_0})(x) = \sum_y u(y)p(y, x)e^{V_0(x)} = (up)(x)e^{V_0(x)} = (up)(x)\frac{u(x)}{(up)(x)} = u(x)$$

e, portanto, $(u(P^{V_0})^n)(x) = u(x)$. Logo,

$$\log \lambda_p(V_0) = \lim_{n\to\infty} \frac{1}{n} \log u(x) = 0.$$

Logo, para qualquer $u > 0$, temos que

$$\sum_{x \in E} q(x) \log \left(\frac{u(x)}{(up)(x)} \right) = \sum_{x \in E} q(x)V_0(x) = \sum_{x \in E} q(x)V_0(x) - \log \lambda_p(V_0) \leqslant I(q).$$

Maximizando a desigualdade acima para $u > 0$, concluímos que $I^{(1)} \leqslant I$.

Vamos agora mostrar a desigualdade oposta. Para isso, seja $V : E \to \mathbb{R}$. Pelo Teorema de Perron (Teorema 2.24), existe $u \in \mathbb{R}^d$ positivo tal que $uP^V = \lambda_p(V)u$. Como $\mathcal{M}$ é homeomorfo ao $\mathbb{R}^d$, podemos associar esse vetor a uma $u_p^V : E \to (0, +\infty)$ tal que, para qualquer $x \in E$,

$$(u_p^V P^V)(x) = \lambda_p(V)u_p^V(x).$$

Como $\sum_{x \in E} q(x) = 1$, podemos reescrever

$$\log \lambda_p(V) = \sum_{x \in E} q(x) \log \lambda_p(V) = \sum_{x \in E} q(x) \log \frac{(u_p^V P^V)(x)}{u_p^V(x)}.$$

Note que

$$(u_p^V P^V)(x) = \sum_y u(y)p^V(y,x) = \sum_y u(y)p(y,x)e^{V(x)} = (u_p^V P)(x)e^{V(x)}$$

e, portanto,

$$\log \lambda_p(V) = \sum_{x \in E} q(x) \log \frac{(u_p^V P)(x)e^{V(x)}}{u_p^V(x)}.$$

Por propriedade de logarítmos,

$$\log \lambda_p(V) = \sum_{x \in E} q(x) \log \frac{(u_p^V P)(x)}{u_p^V(x)} + \sum_{x \in E} q(x)V(x) = - \sum_{x \in E} q(x) \log \frac{u_p^V(x)}{(u_p^V P)(x)} + \sum_{x \in E} q(x)V(x).$$

Disto segue que

$$\sum_{x \in E} q(x) \log \frac{u_p^V(x)}{(u_p^V P)(x)} = \sum_{x \in E} q(x)V(x) - \log \lambda_p(V).$$

Maximizando em $u > 0$ a desigualdade acima, concluímos que

$$I^{(1)}(q) = \sup_{u>0} \sum_{y \in E} q(y) \log \left(\frac{u(y)}{(up)(y)} \right) \geqslant \sum_{x \in E} q(x) \log \frac{u_p^V(x)}{(u_p^V P)(x)} = \sum_{x \in E} q(x)V(x) - \log \lambda_p(V).$$

Repare que a desigualdade acima vale para qualquer V. Portanto,

$$I^{(1)}(q) \geqslant \sup_V \left(\sum_{x \in E} q(x)V(x) - \log \lambda_p(V) \right) = I(q).$$

Logo, $I^{(1)} \geqslant I$. Então, $I^{(1)} = I$. □

Visto os dois lemas anteriores, podemos agora seguir para o lema abaixo, que nos dá a igualdade que queremos.

Lema 3.9. *Com as notações introduzidas anteriormente, temos*

$$\inf_{q \in O^*} I(q) = \inf_{q \in O} I(q).$$

Demonstração. Lembre que $O^* = O \cap int(\mathcal{M})$. Assim, para mostrar que os ínfimos são iguais, vamos mostrar que, para cada $q \in O \backslash O^*$, temos $I(q) \geqslant \inf_{q \in O^*} I(q)$. Para isso, começamos fixando $q \in O \backslash O^*$. Pela forma como O^* foi definido, temos que $q \notin int(\mathcal{M})$, ou seja, existe algum i tal que $q(z_i) = 0$.

Vamos considerar $\mu = \sum_{i=1}^{d} \alpha_i \mathbb{1}_{q(z_i)=0} \delta_{z_i}$ com $\alpha_i > 0$ e $\sum_{i=1}^{d} \alpha_i \mathbb{1}_{q(z_i)=0} = 1$. Repare que esse μ é combinação convexa das deltas de Dirac dos pontos de E com medida q nula. Além disso, $\mu(z_k) = \alpha_k \mathbb{1}_{q(z_k)=0}$, para qualquer $k \in \{1, \ldots, d\}$. Defina

$$q_n = \left(1 - \frac{1}{n}\right) q + \frac{1}{n}\mu,$$

ou seja,

$$q_n(z_k) = \left(1 - \frac{1}{n}\right) q(z_k) + \frac{1}{n}\mu(z_k) = \begin{cases} \frac{1}{n}\alpha_k, & \text{se } q(z_k) = 0 \\ q(z_k)\left(1 - \frac{1}{n}\right), & \text{se } q(z_k) > 0 \end{cases}.$$

Logo, $q_n(z_i) > 0$, para qualquer $i \in \{1, \ldots, d\}$. Portanto, $q_n \in int(\mathcal{M})$. Podemos ainda escrever $q_n = q + \dfrac{1}{n}(\mu - q)$. Disto segue que $q_n \to q$ e, portanto, existe n_0 tal que, $q_n \in O$, para qualquer $n \geqslant n_0$, pois, caso contrário, q_n não convergiria a q, já que O é aberto.

Mostramos que, fixada $q \in O \backslash O^*$, existe uma sequência $\{q_n\} \subseteq O \cap int(\mathcal{M}) = O^*$ tal que $q_n \to q$. Como I é semicontínua inferiormente, segue que

$$I(q) \leqslant \varliminf_{n\to\infty} I(q_n). \tag{3.16}$$

Por outro lado, como I é convexa e $q_n = \left(1 - \dfrac{1}{n}\right) q + \dfrac{1}{n}\mu$, obtemos

$$I(q_n) \leqslant \left(1 - \frac{1}{n}\right) I(q) + \frac{1}{n} I(\mu). \tag{3.17}$$

Repare que

$$I(\mu) = I^{(1)}(\mu) = \sup_{u>0} \sum_y \mu(y) \log \frac{u(y)}{(up)(y)}.$$

Como $u > 0$ e $p > 0$, temos que, para todo i,

$$\frac{u(z_i)}{(up)(z_i)} = \frac{u(z_i)}{\sum\limits_{j=1}^{d} u(z_j)p(z_j, z_i)} \leqslant \frac{u(z_i)}{u(z_i)p(z_i, z_i)} = \frac{1}{p(z_i, z_i)}$$

e concluímos que

$$I(\mu) \leqslant \sum_{i=1}^{d} \mu(z_i) \log \frac{1}{p(z_i, z_i)} \leqslant \sum_{i=1}^{d} \mu(z_i) C = C < +\infty,$$

onde $C = \max\limits_{1 \leqslant i \leqslant d} \dfrac{1}{p(z_i, z_i)}$.

Assim, segue de (3.17) que

$$\varlimsup_{n\to\infty} I(q_n) \leqslant I(q). \tag{3.18}$$

De (3.16) e (3.18), concluímos que $\lim\limits_{n\to\infty} I(q_n) = I(q)$. Como $q_n \in O^*$, temos que $I(q_n) \geqslant \inf\limits_{q \in O^*} I(q)$ e, portanto, $I(q) \geqslant \inf\limits_{q\in O^*} I(q)$, para qualquer $q \in O \backslash O^*$. Então,

$$\inf_{q\in O} I(q) = \min\left\{ \inf_{q\in O^*} I(q), \inf_{q \in O\backslash O^*} I(q) \right\} = \inf_{q\in O^*} I(q).$$

$\square$

Pelo Lema 3.9, segue de (3.15) que

$$\varliminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu \left[L_n \in O\right] \geqslant -\inf_{q\in O} I(q),$$

concluindo a prova do Teorema 3.1.

4 CONSIDERAÇÕES GERAIS

Com o objetivo de ter um texto completo sobre os assuntos abordados nesta dissertação, introduzimos este capítulo para tratar de dois importantes teoremas relacionados as ideias apresentadas nos Capítulos 1 e 3.

O primeiro resultado que abordaremos é o Teorema de Sanov, que trata de LDP para $L_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k}$, onde $\{X_n\}$ são variáveis aleatórias i.i.d. Logo, o Teorema de Sanov é muito parecido com o teorema que provamos no Capítulo 3, onde $\{X_n\}$ era uma cadeia de Markov. Por isso, aqui apresentaremos apenas as suas principais ideias e não faremos a sua demonstração completa.

O outro resultado apresentado neste capítulo é o Teorema de Gärtner-Ellis. Este é um resultado super geral, que implica os resultados dos Capítulos 1 e 3. Porém, para obtermos a sua demonstração completa, precisamos de uma grande fundamentação teórica de Análise Convexa, o que poderia afastar um leitor que está começando a estudar grandes desvios das importantes ideias probabilísticas por trás de suas provas. Como o objetivo do texto é apresentar estas ideias, faremos aqui somente a demonstração de parte do Teorema de Gärtner-Ellis. Será usada toda a fundamentação teórica de Análise Convexa sem demonstrações.

4.1 Teorema de Sanov

Acrescentamos esta seção para tratar do Teorema de Sanov, que aborda os grandes desvios para variáveis aleatórias i.i.d utilizando medidas empíricas, de maneira semelhante à que foi utilizada para tratar de grandes desvios para Cadeias de Markov no Teorema 3.1. Inclusive, alguns dos passos da demonstração do Teorema de Sanov são os mesmos que fizemos na prova do Teorema 3.1.

Definição 4.1. *Sejam μ e ν probabilidades sobre um espaço polonês E. Definimos a entropia relativa de ν com respeito a μ como*

$$H(\nu|\mu)^1 = \begin{cases} \int h \log h d\mu, & se\ \nu \ll \mu\ e\ h = \frac{d\nu}{d\mu} \\ +\infty, & caso\ contr\acute{a}rio \end{cases}.$$

Lema 4.2. *Fixada uma probabilidade μ, a função $\nu \mapsto H(\nu|\mu)$ é convexa.*

Demonstração. Dados ν_1, ν_2 probabilidades e $\alpha, \beta \in [0,1]$ com $\alpha + \beta = 1$, note que

$$H(\alpha\nu_1 + \beta\nu_2|\mu) = \int \left(\alpha\frac{d\nu_1}{d\mu} + \beta\frac{d\nu_2}{d\mu}\right) \log\left(\alpha\frac{d\nu_1}{d\mu} + \beta\frac{d\nu_2}{d\mu}\right) d\mu.$$

[1] $H(\nu|\mu)$ também é chamada de informação de Kullback Leibler.

Como a função $f(x) = x \log x$ é convexa (pois $f''(x) = \frac{1}{x} > 0$, já que f está definida apenas para $x > 0$), temos que

$$H(\alpha\nu_1 + \beta\nu_2|\mu) \leqslant \alpha \int \frac{d\nu_1}{d\mu} \log\left(\frac{d\nu_1}{d\mu}\right) d\mu + \beta \int \frac{d\nu_2}{d\mu} \log\left(\frac{d\nu_2}{d\mu}\right) d\mu = \alpha H(\nu_1|\mu) + \beta H(\nu_2|\mu).$$

□

Seja E um espaço polonês. Vamos denotar por $\mathcal{M} = \mathcal{M}(E)$ o conjunto de todas as probabilidades sobre E, por $C_b(E)$ o conjunto das funções $V : E \to \mathbb{R}$ contínuas e limitadas e definir o produto interno

$$\langle V, \mu \rangle = \int V d\mu,$$

onde $V \in C_b(E)$ e $\mu \in \mathcal{M}$.

Teorema 4.3 (Teorema de Sanov)**.** *Sejam E um espaço polonês, μ uma probabilidade sobre E e $\{Y_n\}$ uma sequência de variáveis aleatórias i.i.d tomando valores em E com distribuição μ. Se denotarmos $L_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$, então*

(i) A sequência L_n satisfaz o LDP com função taxa I_μ definida, para cada $\nu \in \mathcal{M}$, por

$$I_\mu(\nu) = \sup_{V \in C_b(E)} \left\{ \langle V, \nu \rangle - \log\langle e^V, \mu \rangle \right\}; \tag{4.1}$$

(ii) Se C é uma união finita de abertos convexos em $\mathcal{M}$, então, quando $n \to \infty$,

$$\frac{1}{n} \log \mathbb{P}[L_n \in C] \to - \inf_{\nu \in C} I_\mu(\nu);$$

(iii) $I_\mu(\nu) = H(\nu|\mu)$, para toda $\nu \in \mathcal{M}$.

Como a intenção é apenas fazer algumas considerações sobre o Teorema de Sanov, não vamos provar os itens *(i)* e *(ii)*. Observe que, se E é finito, o item *(i)* é um caso particular do Teorema 3.1. A demonstração completa e mais informações acerca do Teorema de Sanov podem ser encontradas em Olivieri e Vares (2005, p. 56-63). Aqui, daremos apenas uma ideia da cota superior do item *(i)* no caso E espaço polonês e provaremos o item *(iii)*, para dar uma ideia sobre o fato da função taxa ser a entropia relativa. Vale ressaltar que essa é a grande vantagem do Teorema de Sanov, pois através da entropia relativa é possível dar uma cara explícita (e não apenas variacional) ao funcional taxa. Por exemplo, se E é enumerável, então a entropia relativa pode ser escrita como

$$H(\nu|\mu) = \sum_{x \in E} \nu(x) \log \frac{\nu(x)}{\mu(x)},$$

conforme Kipnis e Landim (2013, p. 338-339).

Inicialmente, vamos provar o lema abaixo, que nos permite trocar o conjunto no qual estamos tomando o supremo na definição de I_μ. Este lema será necessário na demonstração do item *(iii)* do Teorema de Sanov.

Lema 4.4. *Se μ e ν são probabilidades tais que $I_\mu(\nu) < +\infty$, vale*

$$I_\mu(\nu) := \sup_{V \in C_b(E)} \left\{ \langle V, \nu \rangle - \log \langle e^V, \mu \rangle \right\} = \sup_{V \in B_b(E)} \left\{ \langle V, \nu \rangle - \log \langle e^V, \mu \rangle \right\},$$

onde $C_b(E)$ é o conjunto das funções $V : E \to \mathbb{R}$ contínuas e limitadas e $B_b(E)$ denota o conjunto das funções $V : E \to \mathbb{R}$ limitadas e Borel-mensuráveis.

Demonstração. Fixadas as probabilidades ν e μ. Como $C_b(E) \subseteq B_b(E)$, pois toda função contínua é Borel-mensurável, temos que

$$I_\mu(\nu) = \sup_{V \in C_b(E)} \left\{ \langle V, \nu \rangle - \log \langle e^V, \mu \rangle \right\} \leqslant \sup_{V \in B_b(E)} \left\{ \langle V, \nu \rangle - \log \langle e^V, \mu \rangle \right\}.$$

Para a desigualdade oposta, seja $V \in B_b(E)$. Tome uma sequência $\{V_n\} \subseteq C_b(E)$ tal que $V_n \to V$ pontualmente em todo $z \in E$ e $V_n \leqslant V$. Lembrando que V é limitada e ν é probabilidade (e, portanto, V é integrável com respeito a ν), pelo Teorema da convergência dominada (Teorema A.15), temos que $\int V_n d\nu \to \int V d\nu$. Pelos mesmos motivos, $\int e^{V_n} d\mu \to \int e^V d\mu$.

Pelas convergências estabelecidas acima e mais o fato da função logarítmo ser contínua, obtemos

$$\langle V, \nu \rangle - \log \langle e^V, \mu \rangle = \int V d\nu - \log \int e^V d\mu = \lim_{n \to \infty} \left(\int V_n d\nu - \log \int e^{V_n} d\mu \right).$$

Como $V_n \in C_b(E)$, para todo $n \in \mathbb{N}$, segue da igualdade acima que

$$\langle V, \nu \rangle - \log \langle e^V, \mu \rangle \leqslant \sup_{V \in C_b(E)} \left\{ \int V d\nu - \log \int e^V d\mu \right\} = I_\mu(\nu).$$

Observe que esta desigualdade vale para qualquer $V \in B_b(E)$. Maximizando, obtemos

$$\sup_{V \in B_b(E)} \left\{ \langle V, \nu \rangle - \log \langle e^V, \mu \rangle \right\} \leqslant \sup_{V \in C_b(E)} \left\{ \int V d\nu - \log \int e^V d\mu \right\} = I_\mu(\nu).$$

□

Cota superior do item *(i)* do Teorema de Sanov:

Seja $V : E \to \mathbb{R}$ uma função qualquer e μ a distribuição das variáveis Y_i. Procedendo de maneira semelhante ao argumento da página 44, podemos observar que

$$\mathbb{P}[L_n \in A] = \mathbb{E}\left[\mathbb{1}_{[L_n \in A]} e^{-n\langle V, L_n \rangle} e^{n \langle V, L_n \rangle} \right] \leqslant e^{-n \inf_{q \in A} \langle V, q \rangle} \mathbb{E}\left[\exp\left(\sum_{k=1}^n V(Y_k) \right) \right],$$

pela forma como definimos L_n.

Usando que a sequência é i.i.d com distribuição μ, podemos reescrever a desigualdade acima como

$$\mathbb{P}[L_n \in A] \leqslant \left(e^{-\inf_{q \in A} \langle V, q \rangle} \int e^{V(x)} d\mu(x) \right)^n = \left(e^{-\inf_{q \in A} \langle V, q \rangle} \langle e^V, \mu \rangle \right)^n.$$

Disto segue que
$$\frac{1}{n}\log \mathbb{P}[L_n \in A] \leqslant -\inf_{q\in A}\left\{\langle V,q\rangle - \log\langle e^V,\mu\rangle\right\}.$$
Já que a desigualdade acima vale para qualquer V, podemos minimizar em V. Além disso, como apenas o lado esquerdo da desigualdade depende de n, temos

$$\overline{\lim_{n\to\infty}}\frac{1}{n}\log \mathbb{P}[L_n \in A] \leqslant \inf_{V}\sup_{q\in A}\Big(-\langle V,q\rangle + \log\langle e^V,\mu\rangle\Big).$$

Procedendo como no Capítulo 3 temos

$$\overline{\lim_{n\to\infty}}\frac{1}{n}\log \mathbb{P}[L_n \in A] \leqslant -\inf_{q\in F}\sup_{V}\Big(\langle V,q\rangle - \log\langle e^V,\mu\rangle\Big) = -\inf_{q\in F} I_\mu(q).$$

Com isso, verificamos a cota superior. É claro que isso não conclui a prova, porém nos permite ver que a função taxa dada é coerente e que a linha de raciocínio da demonstração segue o mesmo padrão das anteriores.

Item *(iii)* do Teorema de Sanov:

Vamos começar provando que $I_\mu(\nu) \leqslant H(\nu|\mu)$. Se $H(\nu|\mu) = +\infty$, não há o que provar. Assim, suponhamos que $H(\nu|\mu) < +\infty$. Então, $\nu \ll \mu$. Portanto, existe $\frac{d\nu}{d\mu} \geqslant 0$ (pois ν e μ são probabilidades) tal que

$$H(\nu|\mu) = \int \frac{d\nu}{d\mu}\log\frac{d\nu}{d\mu}d\mu = \int \log\frac{d\nu}{d\mu}d\nu = \left\langle \log\frac{d\nu}{d\mu},\nu\right\rangle.$$

Então,
$$\langle V,\nu\rangle - H(\nu|\mu) = \left\langle\left(V - \log\frac{d\nu}{d\mu}\right),\nu\right\rangle.$$
Tomando exponencial na igualdade acima,

$$\exp\left(\langle V,\nu\rangle - H(\nu|\mu)\right) = \exp\left\langle\left(V - \log\frac{d\nu}{d\mu}\right),\nu\right\rangle = \exp\int\left(V - \log\frac{d\nu}{d\mu}\right)d\nu.$$

Pela desigualdade de Jensen (Teorema A.10) e usando propriedades de exponenciais e logarítmos, temos

$$\exp\left(\langle V,\nu\rangle - H(\nu|\mu)\right) \leqslant \int \exp\left(V - \log\frac{d\nu}{d\mu}\right)d\nu = \int\frac{e^V}{\exp\left(\log\frac{d\nu}{d\mu}\right)}d\nu = \int e^V d\mu = \langle e^V,\mu\rangle.$$

Tomando logarítmo na desigualdade acima,

$$\langle V,\nu\rangle - H(\nu|\mu) \leqslant \log\langle e^V,\mu\rangle.$$

Reorganizando os termos da desigualdade anterior, obtemos, para qualquer função limitada $V : E \to \mathbb{R}$, a seguinte desigualdade:

$$\langle V,\nu\rangle - \log\langle e^V,\mu\rangle \leqslant H(\nu|\mu).$$

Maximizando a desigualdade acima em $V \in C_b(E)$ e observando a expressão de I_μ em (4.1), concluímos que $I_\mu(\nu) \leqslant H(\nu|\mu)$.

Vamos agora usar o Lema 4.4 para provar que $H(\nu|\mu) \leqslant I_\mu(\nu)$. Se $I_\mu(\nu) = +\infty$, não há o que provar. Assim, suponhamos que $I_\mu(\nu) < +\infty$. Precisamos, primeiramente, mostrar que $H(\nu|\mu) < +\infty$. Então, vamos começar provando que $\nu \ll \mu$.

Sejam $k \in \mathbb{N}$ e $A \in \mathcal{B}(E)$ tal que $\mu(A) = 0$. Tomando $V_0 = k\mathbb{1}_A$, temos que $V_0 \in B_b(E)$ e, portanto,

$$\int V_0 d\nu - \log \int e^{V_0} d\mu \leqslant \sup_{V \in B_b(E)} \left\{ \langle V, \nu \rangle - \log \langle e^V, \mu \rangle \right\} = I_\mu(\nu) < +\infty.$$

Então, $k \int \mathbb{1}_A d\nu - \log \int e^{k\mathbb{1}_A} d\mu \leqslant I_\mu(\nu) < +\infty$. Como

$$\log \int e^{k\mathbb{1}_A} d\mu = \log \int \left(e^k \mathbb{1}_A + \mathbb{1}_{A^\complement} \right) d\mu = \log \left(e^k \mu(A) + \mu(A^\complement) \right) = \log(1) = 0,$$

segue que $k \int \mathbb{1}_A d\nu \leqslant I_\mu(\nu) < +\infty$. Concluímos que, para qualquer $k \in \mathbb{N}$,

$$k\nu(A) \leqslant I_\mu(\nu) < +\infty.$$

Logo, $\nu(A) = 0$ e, portanto, $\nu \ll \mu$. Assim, existe $\frac{d\nu}{d\mu} \geqslant 0$, pois ν e μ são probabilidades.

Suponhamos que existam constantes c e C tais que $0 < c < C < +\infty$ e $c \leqslant \frac{d\nu}{d\mu} \leqslant C$. Neste caso, tomando $V = \log \frac{d\nu}{d\mu})$, concluímos que

$$\left\langle \log \frac{d\nu}{d\mu}, \nu \right\rangle - \log \left\langle e^{\log \frac{d\nu}{d\mu}}, \mu \right\rangle = \int \log \left(\frac{d\nu}{d\mu} \right) d\nu - \underbrace{\log \int \frac{d\nu}{d\mu} d\mu}_{0} = \underbrace{\int \log \left(\frac{d\nu}{d\mu} \right) \frac{d\nu}{d\mu} d\mu}_{H(\nu|\mu)} = H(\nu|\mu).$$

Repare que, pelo que vimos acima e observando a forma como I_μ foi definida, temos que $H(\nu|\mu) \leqslant I_\mu(\nu)$ no caso em que $0 < c \leqslant \frac{d\nu}{d\mu} \leqslant C < +\infty$. Porém, essa limitação nem sempre acontece.

Vamos agora supor que existe c tal que $0 < c \leqslant \frac{d\nu}{d\mu}$. Definindo $h_n = \min \left\{ \frac{d\nu}{d\mu}, n \right\}$, temos que $h_n \uparrow \frac{d\nu}{d\mu} > 0$. Pelo Teorema da convergência monótona (Teorema A.14), obtemos

$$\int \log h_n d\nu \longrightarrow \int \log \left(\frac{d\nu}{d\mu} \right) d\nu = \int \log \left(\frac{d\nu}{d\mu} \right) \frac{d\nu}{d\mu} d\mu = H(\nu|\mu).$$

Reescrevendo, temos

$$\lim_{n \to \infty} \left\{ \langle \log h_n, \nu \rangle - \log \left\langle e^{\log h_n}, \mu \right\rangle \right\} + \lim_{n \to \infty} \log \langle e^{\log h_n}, \mu \rangle = H(\nu|\mu).$$

Utilizando novamente o Teorema da convergência monótona, concluímos que $\langle h_n, \mu \rangle \to 0$ e, portanto, a segunda parcela da soma acima vale zero. Logo, segue da igualdade acima que

$$H(\nu|\mu) = \lim_{n \to \infty} \left\{ \langle \log h_n, \nu \rangle - \log \left\langle e^{\log h_n}, \mu \right\rangle \right\} \leqslant \sup_{V \in B_b(E)} \left\{ \langle V, \nu \rangle - \log \left\langle e^V, \mu \right\rangle \right\} = I_\mu(\nu),$$

pois $\log h_n \in B_b(E)$.

Até aqui, concluímos que $H(\nu|\mu) \leqslant I_\mu(\nu)$, quando $\frac{d\nu}{d\mu}$ é estritamente positiva. Agora vamos provar no caso geral. Para isso, definimos $\nu_\varepsilon = \varepsilon\mu + (1-\varepsilon)\nu$. Como $\frac{d\nu_\varepsilon}{d\mu} \geqslant \varepsilon > 0$, pelo caso anterior, temos que

$$H(\nu_\varepsilon|\mu) \leqslant I_\mu(\nu_\varepsilon) = I_\mu(\varepsilon\mu + (1-\varepsilon)\nu) = \sup_V \left\{\langle V, \varepsilon\mu + (1-\varepsilon)\nu\rangle - \log\langle e^V, \mu\rangle\right\}.$$

Podemos reescrever a desigualdade acima como

$$H(\nu_\varepsilon|\mu) \leqslant \sup_V \left\{\varepsilon\langle V, \mu\rangle + (1-\varepsilon)\langle V, \nu\rangle - \varepsilon\log\langle e^V, \mu\rangle - (1-\varepsilon)\log\langle e^V, \mu\rangle\right\},$$

ou ainda,

$$H(\nu_\varepsilon|\mu) \leqslant \varepsilon \sup_V \left\{\langle V, \mu\rangle - \log\langle e^V, \mu\rangle\right\} + (1-\varepsilon) \sup_V \left\{\langle V, \nu\rangle - \log\langle e^V, \mu\rangle\right\} = \varepsilon I_\mu(\mu) + (1-\varepsilon) I_\mu(\nu).$$

Repare que $I_\mu(\mu) = 0$, pois, para qualquer V, $\langle V, \mu\rangle - \log\langle e^V, \mu\rangle \leqslant \langle V, \mu\rangle - \log e^{\langle V, \mu\rangle} = 0$, pela desigualdade de Jensen (Teorema A.10). Logo,

$$H(\nu_\varepsilon|\mu) \leqslant (1-\varepsilon) I_\mu(\nu) \leqslant I_\mu(\nu),$$

pois $I_\mu(\nu) \geqslant 0$. Portanto,

$$I_\mu(\nu) \geqslant H(\nu_\varepsilon|\mu) = \int \frac{d\nu_\varepsilon}{d\mu} \log \frac{d\nu_\varepsilon}{d\mu} d\mu.$$

Para concluirmos a prova do item *(iii)* do Teorema de Sanov, falta apenas mostrar que $H(\nu_\varepsilon|\mu) \to H(\nu|\mu)$, quando $\varepsilon \to 0$. Para isso, lembre que a função $\nu \mapsto H(\nu|\mu)$ é convexa (pelo Lema 4.2) e note que

$$H(\mu, \mu) = \int \frac{d\mu}{d\mu} \log\left(\frac{d\mu}{d\mu}\right) d\mu = 0.$$

Portanto,

$$H(\nu_\varepsilon|\mu) = H(\varepsilon\mu + (1-\varepsilon)\nu|\mu) \leqslant \varepsilon \underbrace{H(\mu|\mu)}_{0} + (1-\varepsilon) H(\nu|\mu) \leqslant (1-\varepsilon) H(\nu|\mu) \leqslant H(\nu|\mu).$$

Como a desigualdade acima vale para qualquer $\varepsilon > 0$, concluímos que

$$\varlimsup_{\varepsilon \to 0} H(\nu_\varepsilon|\mu) \leqslant H(\nu|\mu). \tag{4.2}$$

Por outro lado, como $\nu_\varepsilon = \varepsilon\mu + (1-\varepsilon)\nu$,

$$H(\nu_\varepsilon|\mu) = \int \frac{d\nu_\varepsilon}{d\mu} \log\left(\frac{d\nu_\varepsilon}{d\mu}\right) d\mu = \varepsilon \int \log\left(\frac{d\nu_\varepsilon}{d\mu}\right) d\mu + (1-\varepsilon) \int \log\left(\frac{d\nu_\varepsilon}{d\mu}\right) d\mu. \tag{4.3}$$

Considerando que

$$\log \frac{d\nu_\varepsilon}{d\mu} = \log\left(\varepsilon + (1-\varepsilon)\frac{d\nu}{d\mu}\right) \geqslant \max\left\{\log\varepsilon, \log\left[(1-\varepsilon)\frac{d\nu}{d\mu}\right]\right\}$$

segue da igualdade (4.3) que

$$H(\nu_\varepsilon|\mu) \geqslant \varepsilon \int \log \varepsilon d\mu + (1-\varepsilon) \int \left[\log(1-\varepsilon) + \log \frac{d\nu}{d\mu}\right] d\nu.$$

Como a desigualdade acima vale para qualquer $\varepsilon > 0$, concluímos que

$$\varliminf_{\varepsilon \to 0} H(\nu_\varepsilon|\mu) \geqslant \int \log \left(\frac{d\nu}{d\mu}\right) d\nu = \int \log \left(\frac{d\nu}{d\mu}\right) \frac{d\nu}{d\mu} d\mu = H(\nu|\mu). \tag{4.4}$$

Assim, segue de (4.2) e (4.4) que $H(\nu_\varepsilon|\mu) \to H(\nu|\mu)$, quando $\varepsilon \to 0$.

Com isso, mostramos que $H(\nu|\mu) \leqslant I_\mu(\nu)$, o que, juntamente com a desigualdade oposta, que já tinhamos provado anteriormente, conclui a demonstração do item *(iii)* do Teorema de Sanov.

4.2 Teorema de Gärtner-Ellis

Decidimos acrescentar esta seção, pois o Teorema de Gärtner-Ellis nos fornece uma outra maneira de provar o Teorema 3.1. Como veremos, para seguir por este caminho, a análise convexa é uma ferramenta fundamental. Por este motivo, optamos por não utilizá-lo e demonstrar o Teorema 3.1 conforme feito no Capítulo 3, pois tal caminho é mais probabilista e não precisa de resultados tão sofisticados.

Esta seção será dividida em duas partes. Inicialmente, trataremos do Teorema de Gärtner-Ellis em si, percorrendo todo o caminho necessário para a demonstração do mesmo. Em seguida, vamos dar uma ideia de uma aplicação deste teorema. Mostraremos como poderia ser provado o princípio de grandes desvios para Cadeias de Markov utilizando este teorema.

4.2.1 O Teorema

Vamos considerar $\{W_n\}$ elementos aleatórios tomando valores em $\mathbb{R}^m$. Sejam μ_n a distribuição de W_n em $\mathbb{R}^m$ e $\varphi_n : \mathbb{R}^m \to \mathbb{R}$ a função geradora de momentos, isto é,

$$\varphi_n(\zeta) = \mathbb{E}[e^{\langle \zeta, W_n \rangle}],$$

onde $\langle \cdot, \cdot \rangle$ denota um produto interno em $\mathbb{R}^m$.

Hipótese 4.5. *Suponhamos que existe uma sequência $a_n \uparrow +\infty$ tal que, para cada $\zeta \in \mathbb{R}^m$, existe*

$$\varphi(\zeta) = \lim_{n \to \infty} \frac{1}{a_n} \log \varphi_n(a_n \zeta),$$

com $\varphi(\zeta) \in (-\infty, +\infty]$.

Notação. *Vamos denotar $\mathcal{D}_\varphi = \{\zeta \in \mathbb{R}^m : \varphi(\zeta) < +\infty\}$ o conjunto dos pontos onde o limite anterior é finito.*

Hipótese 4.6. *Suponhamos que* $0 \in int(\mathcal{D}_\varphi)$.

Notação. *Denotaremos a transformada de Fenchel-Legendre de* φ *por*

$$\varphi^*(x) = \sup_{\zeta \in \mathbb{R}^m} (\langle \zeta, x \rangle - \varphi(\zeta)).$$

Definição 4.7. *Um ponto* $\zeta \in \mathbb{R}^m$ *é chamado de subdiferencial de* φ^* *em* $y \in \mathbb{R}^m$ *se, para todo* $z \in \mathbb{R}^m$ *com* $z \neq y$,

$$\varphi^*(z) > \varphi^*(y) + \langle \zeta, z - y \rangle.$$

Neste caso, definimos

$$\mathcal{E}(\varphi, \varphi^*) = \{y \in \mathbb{R}^m : \textit{existe } \zeta \in int(\mathcal{D}_\varphi) \textit{ subdiferencial de } \varphi^* \textit{ em } y\}.$$

Algumas das propriedades básicas de φ e φ^* são dadas no lema abaixo:

Lema 4.8. *Sob as Hipóteses 4.5 e 4.6, temos:*

(i) $\log \varphi_n$ *é convexa;*

(ii) $\varphi : \mathbb{R}^m \to (-\infty, +\infty]$ *é convexa;*

(iii) $\varphi^* : \mathbb{R}^m \to [0, +\infty]$ *é convexa e semicontínua inferiormente.*

Demonstração. Sejam $\zeta_1, \zeta_2 \in \mathbb{R}^m$ e $\beta \in [0,1]$. Então,

$$\varphi_n(\beta\zeta_1 + (1-\beta)\zeta_2) = \mathbb{E}[e^{\langle \beta\zeta_1 + (1-\beta)\zeta_2, W_n \rangle}].$$

Lembrando que o produto interno é linear nas duas entradas, obtemos

$$\varphi_n(\beta\zeta_1 + (1-\beta)\zeta_2) = \mathbb{E}[e^{\beta\langle \zeta_1, W_n \rangle + (1-\beta)\langle \zeta_2, W_n \rangle}].$$

Pela propriedade de exponencial, podemos escrever

$$\varphi_n(\beta\zeta_1 + (1-\beta)\zeta_2) = \mathbb{E}[(e^{\langle \zeta_1, W_n \rangle})^\beta (e^{\langle \zeta_2, W_n \rangle})^{(1-\beta)}].$$

Como a exponencial é positiva, aplicando a desigualdade de Hölder (Teorema A.13) para $\frac{1}{\beta}$ segue que

$$\varphi_n(\beta\zeta_1 + (1-\beta)\zeta_2) \leqslant \mathbb{E}[e^{\langle \zeta_1, W_n \rangle}]^\beta \mathbb{E}[e^{\langle \zeta_2, W_n \rangle}]^{(1-\beta)}.$$

Logo,

$$\log \varphi_n(\beta\zeta_1 + (1-\beta)\zeta_2) \leqslant \log \mathbb{E}[e^{\langle \zeta_1, W_n \rangle}]^\beta \mathbb{E}[e^{\langle \zeta_2, W_n \rangle}]^{(1-\beta)}.$$

Por propriedades de logaritmos,

$$\begin{aligned}\log \varphi_n(\beta\zeta_1 + (1-\beta)\zeta_2) &\leqslant \beta \log \mathbb{E}[e^{\langle \zeta_1, W_n \rangle}] + (1-\beta) \log \mathbb{E}[e^{\langle \zeta_2, W_n \rangle}] \\ &= \beta \log \varphi_n(\zeta_1) + (1-\beta) \log \varphi_n(\zeta_2).\end{aligned}$$

Assim, concluímos que $\log \varphi_n$ é convexa, o que finaliza o item *(i)*. Disto segue que

$$\lim_{n\to\infty} \frac{1}{a_n} \log \varphi_n(a_n(\beta\zeta_1 + (1-\beta)\zeta_2)) \leqslant \beta \lim_{n\to\infty} \frac{1}{a_n} \log \varphi_n(a_n\zeta_1) + (1-\beta) \lim_{n\to\infty} \frac{1}{a_n} \log \varphi_n(a_n\zeta_2).$$

Portanto,

$$\varphi(\beta\zeta_1 + (1-\beta)\zeta_2) \leqslant \beta\varphi(\zeta_1) + (1-\beta)\varphi(\zeta_2),$$

concluíndo o item *(ii)*.

Para a demonstração do item *(iii)*, sejam $x_1, x_2 \in \mathbb{R}^m$ e $\beta \in [0,1]$. Então,

$$\varphi^*(\beta x_1 + (1-\beta)x_2) = \sup_{\zeta\in\mathbb{R}^m} (\langle \zeta, \beta x_1 + (1-\beta)x_2 \rangle - \varphi(\zeta)).$$

Usando a linearidade do produto interno e escrevendo $\varphi(\zeta) = \beta\varphi(\zeta) + (1-\beta)\varphi(\zeta)$, concluímos que

$$\varphi^*(\beta x_1 + (1-\beta)x_2) = \sup_{\zeta\in\mathbb{R}^m} (\beta(\langle \zeta, x_1\rangle - \varphi(\zeta)) + (1-\beta)(\langle \zeta, x_2\rangle - \varphi(\zeta))).$$

Por propriedades do supremo, obtemos

$$\varphi^*(\beta x_1 + (1-\beta)x_2) \leqslant \beta \sup_{\zeta\in\mathbb{R}^m} (\langle \zeta, x_1\rangle - \varphi(\zeta)) + (1-\beta) \sup_{\zeta\in\mathbb{R}^m} (\langle \zeta, x_2\rangle - \varphi(\zeta))).$$

Segue da desigualdade acima que φ^* é convexa. O fato de ser semicontínua inferiormente segue diretamente do Lema A.2. Além disso, φ^* de fato é positiva, pois

$$\varphi^*(x) = \sup_{\zeta\in\mathbb{R}^m} (\langle \zeta, x\rangle - \varphi(\zeta)) \geqslant \langle 0, x\rangle - \varphi(0) = 0.$$

□

Para o Teorema de Gärtner-Ellis, precisaremos da noção de interior relativo, que definimos a seguir.

Definição 4.9. *Seja $C \subseteq \mathbb{R}^m$ um conjunto convexo e não vazio. O interior relativo de C, denotado por $ri(C)$, é definido por*

$$ri(C) = \{x \in C : \textit{para qualquer } y \in C, x - \varepsilon(y-x) \in C, \textit{para algum } \varepsilon > 0\}.$$

Repare que $int(C) \subseteq ri(C)$, pois dado $x \in int(C)$, existe $\delta > 0$ tal que $B(x,\delta) \subseteq int(C)$, já que o interior é um conjunto aberto. Então, para cada $y \in C$, exise algum ε suficientemente pequeno de modo que $x - \varepsilon(y-x) \in B(x,\delta)$ e, portanto, $x \in ri(C)$. Já a inclusão oposta não é necessariamente válida, isto é, o interior relativo pode ser diferente do interior usual, pois tomando $C = \{x\}$, temos $ri(C) = \{x\}$ e $int(C) = \emptyset$.

Propriedades. *Algumas das propriedades básicas do interior relativo são:*

(i) $ri(C) \neq \emptyset$*;*

(ii) *Se* $z \in ri(C)$ *e* $y \in C$*, então* $\varepsilon z + (1-\varepsilon)y \in ri(C)$*, para qualquer* $\varepsilon \in (0,1]$*;*

(iii) *Se* $g : \mathbb{R}^m \to (-\infty, +\infty]$ *é uma função convexa, então* g *é contínua em* $ri(\mathcal{D}_g)$*, onde* $\mathcal{D}_g = \{x : g(x) < +\infty\}$*.*

Na demonstração do teorema principal deste capítulo, precisaremos de um último lema, que enunciaremos a seguir.

Definição 4.10. *O vetor gradiente (ou simplesmente gradiente) de* $\varphi \in \mathbb{R}^m$ *é definido por*

$$\nabla\varphi = \left(\frac{\partial\varphi}{\partial x_1}, \ldots, \frac{\partial\varphi}{\partial x_m}\right).$$

Lema 4.11. *Além das Hipóteses 4.5 e 4.6, suponhamos que*

(a) φ *é semicontínua em* $\mathbb{R}^m$ *e diferenciável em* $int(\mathcal{D}_\varphi)$*;*

(b) $\mathcal{D}_\varphi = \mathbb{R}^d$ *ou* $|\nabla\varphi(\zeta)| \to +\infty$*, quando* $\zeta \to \partial\mathcal{D}_\varphi$ *com* $\zeta \in int(\mathcal{D}_\varphi)$*.*

Então, $ri(\mathcal{D}_{\varphi^*}) \subseteq \mathcal{E}(\varphi, \varphi^*)$*.*

Este lema não será aqui demonstrado, pois sua demonstração precisa de elementos da teoria de análise convexa que não condizem com o tema deste trabalho. Tal demonstração pode ser encontrada em Rockafellar (1997, p. 257). Demonstraremos apenas o próximo teorema, pois, apesar do seu enunciado ser um pouco diferente da definição de LDP que usamos durante este texto, trata-se de um LDP e sua demonstração é permeada das ideias que apareceram nos Capítulos 1 e 3.

Teorema 4.12 (Teorema de Gärtner-Ellis)**.** *Sob as Hipóteses 4.5 e 4.6 temos:*

(i) $\displaystyle\overline{\lim_{n\to\infty}} \frac{1}{a_n} \log \mu_n(F) \leqslant -\inf_{x\in F} \varphi^*(x)$*, para qualquer* $F \subseteq \mathbb{R}^m$ *fechado;*

(ii) $\displaystyle\underline{\lim_{n\to\infty}} \frac{1}{a_n} \log \mu_n(A) \geqslant -\inf_{x\in A\cap\mathcal{E}} \varphi^*(x)$*, para todo* $A \subseteq \mathbb{R}^m$ *aberto, onde* $\mathcal{E} = \mathcal{E}(\varphi, \varphi^*)$*.*

Ainda, se assumirmos (a) e (b) do Lema 4.11, podemos substituir $A \cap \mathcal{E}$ *por* A *no item (ii) acima. Neste caso, obtemos que a sequência* $\{\mu_n\}$ *satisfaz o LDP (veja Definição A.32) com função taxa* φ^**.*

Demonstração. Fixados $F \subseteq \mathbb{R}^m$ e $\zeta \in \mathbb{R}^m$, aplicamos a desigualdade de Chebyshev (Teorema A.12) com $\psi(y) = e^{\langle\zeta,y\rangle}$ e obtemos

$$\inf_{y\in F}(e^{\langle\zeta,y\rangle})\mathbb{P}[W_n \in F] \leqslant \mathbb{E}[e^{\langle\zeta,W_n\rangle}].$$

Como μ_n é a distribuição de W_n e a exponencial é crescente, concluímos que

$$e^{\inf_{y\in F}\langle\zeta,y\rangle}\mu_n(F) \leqslant \mathbb{E}[e^{\langle\zeta,W_n\rangle}].$$

Pela definição de φ_n, podemos reescrever a desigualdade acima como

$$\mu_n(F) \leqslant e^{-\inf\limits_{y\in F}\langle\zeta,y\rangle}\varphi_n(\zeta).$$

Tomando $\frac{1}{a_n}\log$ na expressão acima, temos, para qualquer $\zeta \in \mathbb{R}^m$,

$$\frac{1}{a_n}\log\mu_n(F) \leqslant -\inf_{y\in F}\left\langle \frac{\zeta}{a_n}, y\right\rangle + \frac{1}{a_n}\log\varphi_n(\zeta).$$

Como a desigualdade acima vale para qualquer $\zeta \in \mathbb{R}^m$, podemos fazer a mudança de variável $\zeta \mapsto a_n\zeta$ e obter

$$\frac{1}{a_n}\log\mu_n(F) \leqslant -\inf_{y\in F}\langle\zeta, y\rangle + \frac{1}{a_n}\log\varphi_n(a_n\zeta).$$

Minimizando em ζ a desigualdade acima, obtemos

$$\frac{1}{a_n}\log\mu_n(F) \leqslant -\sup_{\zeta\in\mathbb{R}^m}\left(\inf_{y\in F}\langle\zeta, y\rangle - \frac{1}{a_n}\log\varphi_n(a_n\zeta)\right). \tag{4.5}$$

Seja K um compacto. Repare que, se $\inf\limits_{x\in K}\varphi^*(x) = 0$, então vale este limite:

$$\overline{\lim_{n\to\infty}}\frac{1}{a_n}\log\mu_n(K) \leqslant 0 = -\inf_{x\in K}\varphi^*(x),$$

pois $a_n \uparrow +\infty$ e $\mu_n(K) \in [0,1]$. Agora, vamos supor que $\inf\limits_{x\in K}\varphi^*(x) > 0$. Então, existe $a \in \mathbb{R}$ tal que $0 < a < \inf\limits_{x\in K}\varphi^*(x)$. Fixamos $y \in K$. Assim, $0 < a < \varphi^*(y)$. Pela forma como φ^* foi definido, existe ζ_y tal que $0 < a < \langle\zeta_y, y\rangle - \varphi(\zeta_y)$. Portanto, para qualquer $y \in K$, existem $\zeta_y \in \mathbb{R}^m$ e $\delta > 0$ tais que

$$\langle\zeta_y, y\rangle - \varphi(\zeta_y) = a + \delta.$$

Disto segue que $\varphi(\zeta_y) < +\infty$. Além disso, observando a forma como φ foi definida, repare que podemos tomar $n_y \in \mathbb{N}$ tal que, para todo $n \geqslant n_y$,

$$\langle\zeta_y, y\rangle - \frac{1}{a_n}\log\varphi_n(a_n\zeta_y) > a + \frac{\delta}{2}.$$

Seja $\delta_y > 0$ e $B(y,\delta_y)$ a bola centrada em y de raio δ_y. Assim, tomando δ_y suficientemente pequeno ($\delta_y|\zeta_y| < \frac{\delta}{2}$), segue de (4.5) e da desigualdade acima que, para todo $n \geqslant n_y$,

$$\frac{1}{a_n}\log\mu_n(B(y,\delta_y)) \leqslant -a.$$

Logo,

$$\overline{\lim_{n\to\infty}}\frac{1}{a_n}\log\mu_n(B(y,\delta_y)) \leqslant -a.$$

Como K é compacto, podemos cobrí-lo com um numero finito de bolas $B(y_i, \delta_{y_i})$. Vamos supor que utilizamos N bolas para cobrir K. Neste caso,

$$\mu_n(K) \leqslant \sum_{i=1}^{N}\mu_n(B(y_i,\delta_{y_i}))$$

e, portanto, aplicando o Lema A.3,

$$\varlimsup_{n\to\infty}\frac{1}{a_n}\log\mu_n(K)\leqslant\varlimsup_{n\to\infty}\frac{1}{a_n}\log\sum_{i=1}^{N}\mu_n(B(y_i,\delta_{y_i}))=\max_{1\leqslant i\leqslant N}\varlimsup_{n\to\infty}\frac{1}{a_n}\log\mu_n(B(y_i,\delta_{y_i}))\leqslant -a.$$

Lembre que tomamos a qualquer de modo que $0<a<\inf_{y\in K}\varphi^*(y)$. Portanto, fazendo $a\to\inf_{y\in K}\varphi^*(y)$, concluímos que

$$\varlimsup_{n\to\infty}\frac{1}{a_n}\log\mu_n(K)\leqslant -\inf_{y\in K}\varphi^*(y).$$

Por fim, basta estender a desigualdade acima para qualquer F fechado. Para isso, basta mostrar que W_n é exponencialmente rígida (veja Definição A.18) e a extensão será consequência da Proposição A.19.

Seja $a>0$. Além disso, pela Hipótese 4.6, existe $\varepsilon>0$ tal que $B(0,\varepsilon)\subseteq\mathcal{D}_\varphi$. Portanto, basta tomar r com $0<r<\varepsilon$ e teremos que $-re_i, re_i\in\mathcal{D}_\varphi$, para todo $i\in\{1,\dots,m\}$, onde $\{e_1,\dots,e_m\}$ denota a base canônica. Suponhamos que $K_a=[-k,k]^m$, onde k é uma constante que depende de a e r. Vamos estudar a condição que precisamos utilizando este compacto que definimos acima e, com base no resultado obtido, encontraremos o valor de k.

Denotaremos por W_n^i a i-ésima coordenada de W_n, isto é, $W_n^i=\langle e_i,W_n\rangle$. Então,

$$\mathbb{P}[W_n\notin K_a]=\mathbb{P}[W_n\in\mathbb{R}^m\backslash[-k,k]^m]=\mathbb{P}[W_n^i>k \text{ ou } W_n^i<-k,\ \text{para algum } i\in\{1,\dots,m\}].$$

Como $\mathbb{P}$ é uma probabilidade, concluímos que

$$\mathbb{P}[W_n\notin K_a]\leqslant\sum_{i=1}^{m}\Big(\mathbb{P}[W_n^i>k]+\mathbb{P}[W_n^i<-k]\Big)=\sum_{i=1}^{m}\Big(\mathbb{P}[W_n^i>k]+\mathbb{P}[-W_n^i>k]\Big).$$

Aplicando a desigualdade de Chebyshev (Teorema A.12) com $\psi(t)=e^{ra_nt}$, obtemos

$$\mathbb{P}[W_n\notin K_a]\leqslant\sum_{i=1}^{m}\left(\frac{\mathbb{E}[e^{ra_nW_n^i}]}{e^{ra_nk}}+\frac{\mathbb{E}[e^{-ra_nW_n^i}]}{e^{ra_nk}}\right).$$

Pela forma como W_n^i foi definido, usando a linearidade do produto interno e considerando que e^{ra_nk} é constante, segue da desigualdade acima que

$$\begin{aligned}\mathbb{P}[W_n\notin K_a]&\leqslant e^{-ra_nk}\sum_{i=1}^{m}(\mathbb{E}[e^{\langle ra_ne_i,W_n\rangle}]+\mathbb{E}[e^{\langle -ra_ne_i,W_n\rangle}])\\&=e^{-ra_nk}\sum_{i=1}^{m}(\varphi_n(ra_ne_i)+\varphi_n(-ra_ne_i)).\end{aligned}$$

Portanto,

$$\varlimsup_{n\to\infty}\frac{1}{a_n}\log\mathbb{P}[W_n\notin K_a]=-rk+\varlimsup_{n\to\infty}\frac{1}{a_n}\log\sum_{i=1}^{m}(\varphi_n(ra_ne_i)+\varphi_n(-ra_ne_i)).$$

Aplicando o Teorema A.3, temos

$$\overline{\lim_{n\to\infty}} \frac{1}{a_n} \log \mathbb{P}[W_n \notin K_a] \leqslant -rk + \max_{1\leqslant i\leqslant m} \left\{ \overline{\lim_{n\to\infty}} \frac{1}{a_n} \log(\varphi_n(ra_ne_i) + \varphi_n(-ra_ne_i)) \right\}.$$

Novamente, pelo Teorema A.3,

$$\overline{\lim_{n\to\infty}} \frac{1}{a_n} \log \mathbb{P}[W_n \notin K_a] \leqslant -rk + \max_{1\leqslant i\leqslant m} \max \left\{ \overline{\lim_{n\to\infty}} \frac{1}{a_n} \log \varphi_n(ra_ne_i), \overline{\lim_{n\to\infty}} \frac{1}{a_n} \log \varphi_n(-ra_ne_i) \right\}.$$

Observando a Hipótese 4.5, temos que

$$\overline{\lim_{n\to\infty}} \frac{1}{a_n} \log \mathbb{P}[W_n \notin K_a] \leqslant -rk + \max_{1\leqslant i\leqslant m} \max \left\{ \varphi(re_i), \varphi(-re_i) \right\}.$$

Repare que, para provarmos a rigidez exponencial, precisamos tomar

$$k = \frac{a + \max\limits_{1\leqslant i\leqslant m} \max\{\varphi(re_i), \varphi(-re_i)\}}{r},$$

pois, com este k definido acima, para $K_a = [-k, k]^m$, temos que

$$\overline{\lim_{n\to\infty}} \frac{1}{a_n} \log \mathbb{P}[W_n \notin K_a] \leqslant -a,$$

provando que W_n é exponencialmente rígida e, como consequência da Proposição A.19, concluímos a demonstração do item *(i)*. Vale ressaltar que este k está bem definido, pois tomamos r de modo que $re_i, -re_i \in \mathcal{D}_\varphi$.

Para a demonstração do item *(ii)*, vamos fixar $x \in \mathcal{E}$ e $\delta > 0$. Portanto, podemos escolher $\zeta \in int(D_\varphi)$ subdiferencial de φ^* em x. Como $\zeta \in int(D_\varphi)$, pela forma como φ foi definida, concluímos que existe $n_x \in \mathbb{N}$ tal que $a_n\zeta \in \mathcal{D}_{\varphi_n}$, para qualquer $n \geqslant n_x$. Assim, vamos definir novas probabilidades ν_n tais que, para cada $y \in \mathbb{R}^m$,

$$\frac{d\nu_n}{d\mu_n}(y) = \frac{e^{a_n\langle\zeta,y\rangle}}{\varphi_n(a_n\zeta)}. \tag{4.6}$$

Observe que ν_n satisfaz as Hipóteses 4.5 e 4.6. Seja δ' tal que $0 < \delta' < \delta$. Vamos denotar por $B(x, \delta)$ a bola com centro em x e raio δ. Então,

$$\frac{1}{a_n} \log \mu_n(B(x,\delta)) \geqslant \frac{1}{a_n} \log \mu_n(B(x,\delta')).$$

Pela forma como definimos ν_n, podemos reescrever a desigualdade acima como

$$\frac{1}{a_n} \log \mu_n(B(x,\delta)) \geqslant \frac{1}{a_n} \log \int_{B(x,\delta')} \frac{\varphi_n(a_n\zeta)}{e^{a_n\langle\zeta,y\rangle}} d\nu_n(y),$$

ou ainda,

$$\frac{1}{a_n} \log \mu_n(B(x,\delta)) \geqslant \frac{1}{a_n} \log \varphi_n(a_n\zeta) + \frac{1}{a_n} \log \int_{B(x,\delta')} e^{-a_n\langle\zeta,y\rangle} d\nu_n(y).$$

Como $|y - x| \leqslant \delta'$, para qualquer $y \in B(x, \delta')$, temos que $|\langle\zeta, y - x\rangle| \leqslant \delta'|\zeta|$, pois é um produto interno. Portanto, para qualquer $y \in B(x, \delta')$,

$$\langle\zeta, x\rangle - \delta'|\zeta| \leqslant \langle\zeta, y\rangle \leqslant \langle\zeta, x\rangle + \delta'|\zeta|.$$

Logo,

$$\frac{1}{a_n}\log\mu_n(B(x,\delta)) \geqslant \frac{1}{a_n}\log\varphi_n(a_n\zeta) - \langle\zeta, x\rangle - \delta'|\zeta| + \frac{1}{a_n}\log\nu_n(B(x,\delta')).$$

Tomando o limite inferior quando $n \to \infty$ na desigualdade acima e observando a definição de φ, segue que

$$\varliminf_{n\to\infty}\frac{1}{a_n}\log\mu_n(B(x,\delta)) \geqslant \varphi(\zeta) - \langle\zeta, x\rangle - \delta'|\zeta| + \varliminf_{n\to\infty}\frac{1}{a_n}\log\nu_n(B(x,\delta')).$$

Fazendo $\delta' \downarrow 0$,

$$\varliminf_{n\to\infty}\frac{1}{a_n}\log\mu_n(B(x,\delta)) \geqslant \varphi(\zeta) - \langle\zeta, x\rangle + \lim_{\delta'\downarrow 0}\varliminf_{n\to\infty}\frac{1}{a_n}\log\nu_n(B(x,\delta')).$$

Repare que, pela definição de φ^*, temos $\langle\zeta, x\rangle - \varphi(\zeta) \leqslant \varphi^*(x)$. Portanto,

$$\varliminf_{n\to\infty}\frac{1}{a_n}\log\mu_n(B(x,\delta)) \geqslant -\varphi^*(x) + \lim_{\delta'\downarrow 0}\varliminf_{n\to\infty}\frac{1}{a_n}\log\nu_n(B(x,\delta')). \tag{4.7}$$

O próximo passo será provar que o último limite na desigualdade acima é zero. De modo análogo ao que fizemos para μ_n no início desta seção, vamos denotar por ψ_n a função geradora de momentos de ν_n e, para cada $\tau \in \mathbb{R}^m$, definir ψ por

$$\psi(\tau) = \lim_{n\to\infty}\frac{1}{a_n}\log\psi_n(a_n\tau) = \lim_{n\to\infty}\frac{1}{a_n}\log\int e^{\langle a_n\tau, y\rangle}d\nu_n(y).$$

Pela forma como ν_n foi definida em (4.6), segue que

$$\psi(\tau) = \lim_{n\to\infty}\frac{1}{a_n}\log\int e^{\langle a_n\tau, y\rangle}\frac{e^{\langle a_n\zeta, y\rangle}}{\varphi_n(a_n\zeta)}d\mu_n(y).$$

Repare que o termo $\varphi_n(a_n\zeta)$ é uma constante dentro da integral. Usando linearidade da integral e as propriedades de limites, exponenciais, logaritmos e do produto interno, podemos reescrever a equação acima como

$$\psi(\tau) = \lim_{n\to\infty}\frac{1}{a_n}\log\int e^{\langle a_n(\tau+\zeta), y\rangle}d\mu_n(y) - \lim_{n\to\infty}\frac{1}{a_n}\log\varphi_n(a_n\zeta).$$

Portanto,

$$\psi(\tau) = \lim_{n\to\infty}\frac{1}{a_n}\log\varphi_n(a_n(\tau+\zeta)) - \lim_{n\to\infty}\frac{1}{a_n}\log\varphi_n(a_n\zeta) = \varphi(\tau+\zeta) - \varphi(\zeta).$$

Denotando por ψ^* a transformada de Fenchel-Legendre de ψ, temos

$$\psi^*(x) = \sup_{\tau\in\mathbb{R}^m}(\langle\tau, x\rangle - \psi(\tau)) = \sup_{\tau\in\mathbb{R}^m}(\langle\tau+\zeta, x\rangle - \langle\zeta, x\rangle - \varphi(\zeta+\tau) + \varphi(\zeta)).$$

Pela forma como φ^* foi definida, concluímos que

$$\psi^*(x) = \varphi^*(x) - \langle\zeta, x\rangle - \varphi(\zeta).$$

Aplicando a parte *(i)* deste teorema para as medidas ν_n, temos que, para qualquer $F \subseteq \mathbb{R}^m$ fechado,

$$\overline{\lim_{n\to\infty}} \frac{1}{a_n} \log \nu_n(F) \leqslant -\inf_{z\in F} \psi^*(z).$$

Em particular, vale para o fechado $\mathbb{R}^m \setminus B(x,\delta')$, isto é,

$$\overline{\lim_{n\to\infty}} \frac{1}{a_n} \log \nu_n(\mathbb{R}^m \setminus B(x,\delta')) \leqslant -\inf_{z\in\mathbb{R}^m\setminus B(x,\delta')} \psi^*(z) = -\psi^*(y),$$

para algum $y \in \mathbb{R}^m \setminus B(x,\delta')$. Observe que, pela definião de φ^*, segue que $\varphi(\zeta) \geqslant \langle\zeta, x\rangle - \varphi^*(x)$, para qualquer $\zeta \in \mathbb{R}^m$. Além disso, lembre que $x \in \mathcal{E}$ e ζ é um subdiferencial de φ^* em x. Portanto,

$$\psi^*(y) = \varphi^*(y) - \langle\zeta, y\rangle + \varphi(\zeta) \geqslant (\varphi^*(y) - \langle\zeta, y\rangle) - (\varphi^*(x) - \langle\zeta, x\rangle) > 0.$$

Logo, para qualquer $\delta' > 0$,

$$\overline{\lim_{n\to\infty}} \frac{1}{a_n} \log \nu_n(\mathbb{R}^m \setminus B(x,\delta')) < 0. \tag{4.8}$$

Como ν_n é uma probabilidade, temos que $\nu_n(\mathbb{R}^m) = 1$. Queremos mostrar que $\lim_{n\to\infty} \nu_n(B(x,\delta')) = 1$. Suponhamos, por absurdo, que isto não vale, ou seja, existe $\varepsilon > 0$ tal que $\nu_n(B(x,\delta')) \leqslant 1 - \varepsilon$, para qualquer $n \in \mathbb{N}$. Neste caso,

$$\overline{\lim_{n\to\infty}} \frac{1}{a_n} \log \nu_n(\mathbb{R}^m \setminus B(x,\delta')) \geqslant \overline{\lim_{n\to\infty}} \frac{1}{a_n} \log \varepsilon = 0,$$

contradizendo (4.8). Logo,

$$\lim_{n\to\infty} \nu_n(B(x,\delta')) = 1.$$

Portanto, segue de (4.7) que, para qualquer $x \in \mathcal{E}$,

$$\underline{\lim_{n\to\infty}} \frac{1}{a_n} \log \mu_n(B(x,\delta)) \geqslant -\varphi^*(x). \tag{4.9}$$

Com a desigualdade acima, podemos agora concluir a prova do item *(ii)*. Para isso, seja $A \subseteq \mathbb{R}^m$ aberto e $x \in A \cap \mathcal{E}$. Como A é aberto, existe $\varepsilon_0(x)$ tal que $B(x,\varepsilon_0) \subseteq A$. Disto segue que $B(x,\varepsilon) \subseteq A$, para qualquer $\varepsilon \leqslant \varepsilon_0(x)$. Logo,

$$\mu_n(B(x,\varepsilon)) \leqslant \mu_n(A).$$

Juntando isso com a equação (4.9), concluímos que, para qualquer $x \in A \cap \mathcal{E}$,

$$\underline{\lim_{n\to\infty}} \frac{1}{a_n} \log \mu_n(A) \geqslant \underline{\lim_{n\to\infty}} \frac{1}{a_n} \log \mu_n(B(x,\delta)) \geqslant -\varphi^*(x).$$

Maximizando a desigualdade anterior em x, obtemos

$$\underline{\lim_{n\to\infty}} \frac{1}{a_n} \log \mu_n(A) \geqslant -\inf_{x\in A\cap\mathcal{E}} \varphi^*(x),$$

concluindo a prova do item *(ii)*.

Vamos seguir para a parte final deste teorema. Para isso, vamos provar que, para qualquer $A \subseteq \mathbb{R}^m$ aberto,

$$\inf_{x \in A \cap ri(\mathcal{D}_{\varphi^*})} \varphi^*(x) \leqslant \inf_{x \in A} \varphi^*(x).$$

Suponhamos que $A \cap \mathcal{D}_{\varphi^*} \neq \emptyset$, pois, caso contrário, a afirmação acima é trivial. Neste caso, seja $y \in A \cap \mathcal{D}_{\varphi^*}$. Além disso, como o interior relativo é não vazio, podemos tomar $z \in ri(\mathcal{D}_{\varphi^*})$. Repare que, como A é aberto, existe δ_0 tal que $B(y, \delta_0) \subseteq A$. Assim, para β suficientemente pequeno,

$$\beta z + (1 - \beta) y \in B(y, \delta_0) \subseteq A.$$

Por outro lado, como $z \in ri(\mathcal{D}_{\varphi^*})$ e $y \in \mathcal{D}_{\varphi^*}$, pela Propriedade *(ii)* do interior relativo, temos que, para qualquer $\beta \in (0, 1]$,

$$\beta z + (1 - \beta) y \in ri(\mathcal{D}_{\varphi^*}).$$

Juntando as duas conclusões obtidas anteriormente, segue que, para $\beta > 0$ suficientemente pequeno,

$$\beta z + (1 - \beta) y \in A \cap ri(\mathcal{D}_{\varphi^*}).$$

Fazendo $\beta \downarrow 0$ e usando a convexidade de φ^*, concluímos que

$$\inf_{x \in A \cap ri(\mathcal{D}_{\varphi^*})} \varphi^*(x) \leqslant \lim_{\beta \downarrow 0} \varphi^*(\beta z + (1 - \beta) y) \leqslant \varphi^*(y).$$

Como a desigualdade acima vale para qualquer $y \in A \cap \mathcal{D}_{\varphi^*}$, minimizando ela em y, obtemos

$$\inf_{x \in A \cap ri(\mathcal{D}_{\varphi^*})} \varphi^*(x) \leqslant \inf_{y \in A \cap \mathcal{D}_{\varphi^*}} \varphi^*(y).$$

Lembre que $\mathcal{D}_{\varphi^*} = \{x \in \mathbb{R}^m : \varphi^*(x) < +\infty\}$. Portanto, $\varphi^*(x) = +\infty$, para qualquer $x \in A \cap \mathcal{D}^{\complement}_{\varphi^*}$. Logo,

$$\inf_{x \in A \cap ri(\mathcal{D}_{\varphi^*})} \varphi^*(x) \leqslant \inf_{y \in A \cap \mathcal{D}_{\varphi^*}} \varphi^*(y) = \inf_{y \in A} \varphi^*(y),$$

provando o que queríamos. Para completar a demonstração, repare que, como $ri(\mathcal{D}_{\varphi^*}) \subseteq \mathcal{E}$ pelo Lema 4.11, obtemos

$$\inf_{x \in A \cap \mathcal{E}} \varphi^*(x) \leqslant \inf_{x \in A \cap ri(\mathcal{D}_{\varphi^*})} \varphi^*(x) \leqslant \inf_{y \in A} \varphi^*(y).$$

Pelo que provamos no item *(ii)* deste teorema e pela desigualdade acima, segue que

$$\varliminf_{n \to \infty} \frac{1}{a_n} \log \mu_n(A) \geqslant - \inf_{x \in A \cap \mathcal{E}} \varphi^*(x) \geqslant - \inf_{x \in A} \varphi^*(x),$$

provando que a sequência $\{W_n\}$ satisfaz o LDP com função taxa φ^*. $\square$

4.2.2 LDP para Cadeias de Markov como consequência do Gärtner-Ellis

Agora que concluímos a prova do Teorema de Gärtner-Ellis, vamos dar uma ideia, sem mostrar todos os detalhes, de como o Teorema 3.1 poderia ser provado como consequência deste teorema. Para isso, vamos considerar o caso $m = 1$ e utilizar o produto interno dado por

$$\langle V, W \rangle = \sum_{x \in E} V(x)W(x),$$

onde E é finito e $V, W \in \mathbb{R}^E$. Seja q uma probabilidade sobre E. Repare que, utilizando este produto interno, temos

$$\varphi^*(q) = \sup_V \left(\langle V, q \rangle - \varphi(V)\right) = \sup_V \left(\sum_x V(x)q(x) - \varphi(V)\right).$$

Comparando a função acima com a função I definida em (3.1), perceba que $I = \varphi^*$ se, e somente se, $\varphi(V) = \log \lambda_p(V)$. Essa igualdade é o que provaremos agora.

Lembre que, no Capítulo 3, definimos a medida empírica $L_n(\omega, \cdot) = \frac{1}{n} \sum\limits_{i=1}^{n} \delta_{X_i(\omega)}(\cdot)$. Por simplicidade, omitimos ω da notação. Assim, podemos escrever

$$L_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[X_i = y]}.$$

Então, considerando $W_n = L_n$,

$$\varphi_n(nV) = \mathbb{E}_\mu[e^{\langle nV, L_n \rangle}] = \mathbb{E}_\mu\left[\exp\left(\sum_{y \in E} nV(y)L_n(y)\right)\right] = \mathbb{E}_\mu\left[\exp\left(\sum_{y \in E} \sum_{i=1}^{n} V(y)\mathbb{1}_{[X_i = y]}\right)\right].$$

Observe que

$$\sum_{y \in E} \sum_{i=1}^{n} V(y)\mathbb{1}_{[X_i = y]} = \sum_{i=1}^{n} \sum_{y \in E} V(X_i)\mathbb{1}_{[X_i = y]} = \sum_{i=1}^{n} V(X_i) \underbrace{\sum_{y \in E} \mathbb{1}_{[X_i = y]}}_{1} = \sum_{i=1}^{n} V(X_i)$$

e, portanto,

$$\varphi_n(nV) = \mathbb{E}_\mu\left[\exp\left(\sum_{i=1}^{n} V(X_i)\right)\right].$$

Seja P^V conforme definida no Teorema 3.1. Então, segue do Lema 3.2 que

$$\varphi_n(nV) = \sum_{x_0 \in E} \mu(x_0) \sum_{y \in E} (P^V)^n(x_0, y),$$

onde $(P^V)^n(x_0, y)$ denota a respectiva entrada da matriz $(P^V)^n$. Aplicando o Corolário 2.25 com $A = P^V$, $B = E$, $C = E$, $\nu = \mu$ e $\mu \equiv 1$, concluímos que, quando $n \to \infty$,

$$\frac{1}{n} \log \sum_{x_0 \in E} \mu(x_0) \sum_{y \in E} (P^V)^n(x_0, y) \longrightarrow \log \lambda_p(V).$$

Logo,

$$\frac{1}{n} \log \varphi_n(nV) \longrightarrow \log \lambda_p(V).$$

Se $a_n = n$, pela definição de φ, segue do resultado acima que

$$\varphi(V) = \log \lambda_p(V).$$

Neste caso, $I = \varphi^*$. Como consequência do Teorema de Gärtner-Ellis, se, para $a_n = n$, forem satisfeitas as Hipóteses 4.5 e 4.6, além das suposições *(a)* e *(b)* do Lema 4.11, concluímos que a sequência $\{X_n\}$ satisfaz o LDP com função taxa I, ou seja, vale o Teorema 3.1.

Repare que, para obtermos essa conclusão, precisamos verificar todas essas hipóteses. Porém, esta não é uma tarefa simples, pois esbarra em alguns pontos cuja demonstração não é trivial. Um desses pontos delicados, por exemplo, é a necessidade de diferenciabilidade da função que associa cada V ao respectivo autovalor maximal $\lambda_p(V)$, fato que, segundo Olivieri e Vares (2005), é consequência do Teorema da Função Implícita.

Como podemos ver, para demonstrar o Teorema 3.1 utilizando o Teorema de Gärtner-Ellis, precisaríamos de alguns resultados mais sofisticados. Por este motivo, optamos pela demonstração utilizada no Capítulo 3, já que ela é mais intuitiva e construtiva.

A APÊNDICE

A.1 Análise

Nesta seção, consideraremos que o domínio das funções é o espaço polonês D.

Definição A.1. *Uma função $f : D \to [-\infty, +\infty]$ é dita **semicontínua superiormente** em $a \in D$ se, para qualquer sequência $\{x_n\}$ com $x_n \to a$, vale que*

$$\overline{\lim_{x_n \to a}} f(x_n) \leqslant f(a).$$

*A função f é dita **semicontínua inferiormente** em a quando*

$$\underline{\lim_{x_n \to a}} f(x_n) \geqslant f(a).$$

*Uma função é dita **semicontínua superiormente (inferiormente)** quando o for em todo seu domínio.*

Lema A.2. *Seja $\{f_\lambda\}_{\lambda \in \Lambda}$ uma sequência de funções contínuas tomando valores em $[-\infty, +\infty]$. Então, $f = \sup_{\lambda \in \Lambda} f_\lambda$ é semicontínua inferiormente e $g = \inf_{\lambda \in \Lambda} f_\lambda$ é semicontínua superiormente.*

Demonstração. Sejam $a \in D$ e $\{x_n\}$ uma sequência em D tal que $x_n \to a$. Como todas as f_λ são contínuas, temos que $f_\lambda(x_n) \to f_\lambda(a)$, para qualquer $\lambda \in \Lambda$. Assim,

$$f(x_n) = \sup_{\lambda \in \Lambda} f_\lambda(x_n) \geqslant f_\lambda(x_n),$$

para qualquer $\lambda \in \Lambda$. Portanto,

$$\underline{\lim_{x_n \to a}} f(x_n) \geqslant \underline{\lim_{x_n \to a}} f_\lambda(x_n) = f_\lambda(a),$$

para qualquer $\lambda \in \Lambda$. Então, maximizando em λ a desiqualdade acima, obtemos

$$\underline{\lim_{x_n \to a}} f(x_n) \geqslant \sup_{\lambda \in \Lambda} f_\lambda(a) = f(a).$$

Logo, f é semicontínua inferiormente.

Observe que $-g = \sup_{\lambda \in \Lambda}(-f_\lambda)$. Como as $-f_\lambda$ são contínuas, pelo que provamos antes, segue que $-g$ é semicontínua inferiormente. Logo, g é semicontínua superiormente. $\square$

Lema A.3. *Dadas as sequências de números reais $a_n, b_n \geqslant 0$ e $c_n \uparrow +\infty$,*

$$\overline{\lim_{n \to \infty}} \frac{1}{c_n} \log(a_n + b_n) = \max\left\{ \overline{\lim_{n \to \infty}} \frac{1}{c_n} \log a_n, \overline{\lim_{n \to \infty}} \frac{1}{c_n} \log b_n \right\}.$$

Demonstração. Como a função logarítmica é crescente e as sequências $a_n, b_n \geqslant 0$, é fácil ver que vale a desigualdade

$$\varlimsup_{n\to\infty} \frac{1}{c_n}\log(a_n+b_n) \geqslant \max\left\{\varlimsup_{n\to\infty} \frac{1}{c_n}\log a_n, \varlimsup_{n\to\infty} \frac{1}{c_n}\log b_n\right\}. \tag{A.1}$$

Para a desigualdade oposta, note que

$$\varlimsup_{n\to\infty} \frac{1}{c_n}\log(a_n+b_n) \leqslant \varlimsup_{n\to\infty} \frac{1}{c_n}\log(2\max\{a_n,b_n\}) = \varlimsup_{n\to\infty} \frac{1}{c_n}(\log 2 + \log\max\{a_n,b_n\}).$$

Como $c_n \uparrow +\infty$, temos que $\varlimsup\limits_{n\to\infty} \frac{1}{c_n}\log 2 = 0$. Então,

$$\varlimsup_{n\to\infty} \frac{1}{c_n}\log(a_n+b_n) \leqslant \varlimsup_{n\to\infty} \frac{1}{c_n}(\log 2 + \log\max\{a_n,b_n\}) = \varlimsup_{n\to\infty} \frac{1}{c_n}\log\max\{a_n,b_n\}.$$

Para cada n temos que $\log\max\{a_n,b_n\} = \max\{\log a_n, \log b_n\}$. Além disso, note que $\varlimsup \max\{\alpha_n,\beta_n\} = \max\{\varlimsup \alpha_n, \varlimsup \beta_n\}$. Portanto,

$$\varlimsup_{n\to\infty} \frac{1}{c_n}\log(a_n+b_n) \leqslant \max\left\{\varlimsup_{n\to\infty} \frac{1}{c_n}\log a_n, \varlimsup_{n\to\infty} \frac{1}{c_n}\log b_n\right\}.$$

Portanto, vale a igualdade (A.1). □

A.2 Probabilidade Básica

Definição A.4. *Uma coleção $\mathcal{X}$ de subconjuntos de um conjunto X é dita uma σ-álgebra se satisfaz as seguintes condições:*
(i) $\emptyset$ e X pertencem a $\mathcal{X}$.
(ii) Se A pertence a $\mathcal{X}$, então o seu complementar $A^{\complement} = X \setminus A$ pertence a $\mathcal{X}$.
(iii) Se $\{A_n\}$ é uma sequência de conjuntos em $\mathcal{X}$, então a união $\bigcup\limits_{n=1}^{\infty} A_n$ pertence a $\mathcal{X}$.

A dupla $(X,\mathcal{X})$, formada por um conjunto X e uma σ-álgebra $\mathcal{X}$ é chamada de espaço mensurável.

Definição A.5. *Uma função $f : X \to Y$, onde $(X,\mathcal{X})$ e $(Y,\mathcal{Y})$ são espaços mensuráveis, é dita mensurável se para qualquer $E \in \mathcal{Y}$, temos que $f^{-1}(E) \in \mathcal{X}$.*

Definição A.6. *Uma função $\mathcal{Z} : X \to Y$ é dita variável aleatória se ela for uma função mensurável.*

Para alguns autores o termo variável aleatória é usado só quando $Y = \mathbb{R}$, para os outros casos é usado o termo elemento aleatório. Aqui como na maioria dos casos $Y \subseteq \mathbb{R}$ usaremos sempre a nomenclatura variável aleatória.

Notação. *Vamos denotar por $\overline{\mathbb{R}}$ a coleção $\mathbb{R}\cup\{-\infty,+\infty\}$, chamada de sistema de números reais estendido.*

Definição A.7. *Uma função* $\mu : \mathcal{X} \to \overline{\mathbb{R}}$ *é chamada de medida se satizfaz as seguintes condições:*
(i) $\mu(\emptyset) = 0$.
(ii) $\mu(B) \geqslant 0$ *para todo* $B \in \mathcal{X}$.
(iii) μ *é* σ*-aditiva, ou seja, se* $\{B_n\}$ *é uma sequência disjunta de conjuntos de* $\mathcal{X}$*, então*

$$\mu\Big(\bigcup_{n=1}^{\infty} B_n\Big) = \sum_{n=1}^{\infty} \mu(B_n).$$

Se, além disto, uma medida satisfizer a condição $\mu(X) = 1$*, então ela é chamada de probabilidade.*

Uma terna ordenada $(X, \mathcal{X}, \mu)$ formada por um conjunto X, uma σ-álgebra $\mathcal{X}$ e uma medida μ chama-se espaço de medida. Já a terna ordenada $(X, \mathcal{X}, \mathbb{P})$ formada por um conjunto X, uma σ-álgebra $\mathcal{X}$ e uma probabilidade $\mathbb{P}$ chama-se espaço de probabilidade.

Definição A.8. *Dizemos que a sequência de variáveis aletórias* $\{\mathcal{Z}_n\}$ *(definidas no espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbb{P})$ *e tomando valores no espaço de estados* $(E, \mathcal{B}(E))$*, onde* $\mathcal{B}(E)$ *é a* σ*-álgebra de Borel sobre E) é independente e identicamente distribuida (i.i.d.) se*

(i) $\{\mathcal{Z}_n\}$ *for independente:*

Para todo $k \in \mathbb{N}$, $n_1 < \cdots < n_k$ *e* $A_1, \ldots, A_k \in \mathcal{B}(E)$ *vale*

$$\mathbb{P}[\mathcal{Z}_{n_1} \in A_1, \ldots, \mathcal{Z}_{n_k} \in A_k] = \prod_{i=1}^{k} \mathbb{P}[\mathcal{Z}_{n_i} \in A_i];$$

(ii) $\{\mathcal{Z}_n\}$ *for identicamente distribuida:*

Para todo $n \in \mathbb{N}$ *e* $A \in \mathcal{B}(E)$,

$$\mathbb{P}[\mathcal{Z}_n \in A] = \mathbb{P}[\mathcal{Z}_1 \in A].$$

Proposição A.9 (Lei da Probabilidade Total)**.** *Seja* $\{A_1, \ldots, A_n\}$ *uma partição do conjunto* Ω*, ou seja, uma coleção de conjuntos não-vazios dois a dois disjuntos tal que* $\bigcup_{i=1}^{n} A_i = \Omega$*. Suponhamos que* $\mathbb{P}(A_i) > 0$*, para qualquer* $i \in \{1, \ldots, n\}$*. Então, dado* $B \subseteq \Omega$,

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

Demonstração. Como $\{A_1, \ldots, A_n\}$ é uma partição de Ω, podemos escrever

$$B = B \cap \Omega = B \cap (A_1 \uplus \cdots \uplus A_n) = (B \cap A_1) \uplus \cdots \uplus (B \cap A_n).$$

Disto segue que

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i) = \sum_{i=1}^{n} \frac{\mathbb{P}(B \cap A_i)\mathbb{P}(A_i)}{\mathbb{P}(A_i)},$$

ou seja,

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

$\square$

Notação. *Denotamos por* $\mathbb{E}$ *a esperança (integral) com respeito à probabilidade* $\mathbb{P}$. *Assim,* $\mathbb{P}[A] = \mathbb{E}[\mathbb{1}_A]$, *onde*

$$\mathbb{1}_A(x) = \begin{cases} 1, & se\ x \in A \\ 0, & se\ x \notin A \end{cases}.$$

Teorema A.10 (Desigualdade de Jensen)**.** [1] *Seja* $\psi : \mathbb{R} \to \mathbb{R}$ *uma função convexa, isto é, para quaisquer* $\lambda \in [0, 1]$ *e* $x, y \in \mathbb{R}$, *vale* $\lambda\psi(x) + (1 - \lambda)\psi(y) \geqslant \psi(\lambda x + (1 - \lambda)y)$. *Se* $f : D \to \mathbb{R}$ *satisfaz* $\mathbb{E}[f] < +\infty$ *e* $\mathbb{E}[\psi(f)] < +\infty$, *então*

$$\psi\left(\mathbb{E}[f]\right) \leqslant \mathbb{E}[\psi(f)].$$

Demonstração. Seja $c = \mathbb{E}[f]$. Como ψ é convexa, existe uma função $\ell(x) = ax + b$ tal que $\ell(c) = \psi(c)$ e $\psi(x) \geqslant \ell(x)$, para qualquer $x \in \mathbb{R}$. Note que esta função de fato existe, já que, para qualquer $a \in [D_-\psi(c), D_+\psi(c)]$, a função $\ell(x) = a(x - c) + \psi(c)$ satisfaz essas hipóteses, onde D_- e D_+ denotam as derivadas laterais pela esquerda e pela direita, respectivamente. Então,

$$\mathbb{E}[\psi(f)] \geqslant \mathbb{E}[\ell(f)] \geqslant \mathbb{E}[(af + b)].$$

Portanto,

$$\mathbb{E}[\psi(f)] \geqslant a\mathbb{E}[f] + b = ac + b = \ell(c) = \psi(c) = \psi\left(\mathbb{E}[f]\right).$$

$\square$

Corolário A.11. *Se* $\psi : \mathbb{R} \to \mathbb{R}$ *é uma função côncava e* $f : D \to \mathbb{R}$ *satisfaz* $\mathbb{E}[f] < +\infty$ *e* $\mathbb{E}[\psi(f)] < +\infty$, *então*

$$\psi(\mathbb{E}[f]) \geqslant \mathbb{E}[\psi(f)].$$

Demonstração. Seja ψ uma função côncava. Então, $-\psi$ é convexa. Aplicando a Desigualdade de Jensen, concluímos que

$$-\psi(\mathbb{E}[f]) \leqslant \mathbb{E}[-\psi(f)].$$

Logo,

$$\psi(\mathbb{E}[f]) \geqslant -\mathbb{E}[-\psi(f)] = \mathbb{E}[\psi(f)].$$

$\square$

[1] Adaptado de Durrett (2013, p.21)

Teorema A.12 (Desigualdade de Chebyshev)**.** [2] *Sejam $\psi : \mathbb{R} \to [0, +\infty)$, $A \in \mathcal{B}(\mathbb{R})$ e $i_A = \inf_{y \in A} \psi(y)$. Então,*

$$i_A \mathbb{P}[X \in A] \leqslant \mathbb{E}[\psi(X)\mathbb{1}_{[X \in A]}] \leqslant \mathbb{E}[\psi(X)].$$

Demonstração. O modo como i_A foi definido e o fato de $\psi \geqslant 0$ implicam que

$$i_A \mathbb{1}_{[X \in A]} \leqslant \psi(X)\mathbb{1}_{[X \in A]} \leqslant \psi(X).$$

Portanto,

$$\mathbb{E}\left[i_A \mathbb{1}_{[X \in A]}\right] \leqslant \mathbb{E}\left[\psi(X)\mathbb{1}_{[X \in A]}\right] \leqslant \mathbb{E}\left[\psi(X)\right],$$

provando o resultado desejado, já que $\mathbb{E}\left[\mathbb{1}_{[X \in A]}\right] = \mathbb{P}[X \in A]$. □

Teorema A.13 (Desigualdade de Hölder)**.** [3] *Sejam $p, q \in (1, +\infty)$ com $\frac{1}{p} + \frac{1}{q} = 1$. Se $f : D \to \mathbb{R}$ e $g : D \to \mathbb{R}$ são funções tais que $\mathbb{E}[|f|^p] < +\infty$ e $\mathbb{E}[|g|^q] < +\infty$, então*

$$\mathbb{E}[|fg|] \leqslant \mathbb{E}[|f|^p]^{\frac{1}{p}} + \mathbb{E}[|g|^q]^{\frac{1}{q}}.$$

Demonstração. Se $|f| = 0$ ou $|g| = 0$, a desigualdade vale trivialmente. Então, suponhamos que $|f| > 0$ e $|g| > 0$. Fixado $y \geqslant 0$, definimos $\psi : [0, +\infty) \to \mathbb{R}$ por

$$\psi(x) = \frac{x^p}{p} + \frac{y^q}{q} - xy.$$

Assim,

$$\psi'(x) = x^{p-1} - y,$$
$$\psi''(x) = (p-1)x^{p-2}.$$

Como $p > 1$, temos que $\psi''(x) \geqslant 0$. Então, ψ tem um mínimo em $x_0 = y^{\frac{1}{p-1}}$. Como

$$\psi(x_0) = \frac{y^{\frac{p}{p-1}}}{p} + \frac{y^q}{q} - y^{\frac{1}{p-1}}y = y^q\left(\frac{1}{p} + \frac{1}{q}\right) - y^q = 0,$$

temos que $\psi(x) \geqslant 0$, para qualquer $x \geqslant 0$. Logo,

$$xy \leqslant \frac{x^p}{p} + \frac{y^q}{q}.$$

Sejam $x = \dfrac{|f|}{\mathbb{E}[|f|^p]^{\frac{1}{p}}}$ e $y = \dfrac{|g|}{\mathbb{E}[|g|^q]^{\frac{1}{q}}}$. Assim,

$$\frac{|fg|}{\mathbb{E}[|f|^p]^{\frac{1}{p}}\mathbb{E}[|g|^q]^{\frac{1}{q}}} \leqslant \frac{|f|^p}{p\mathbb{E}[|f|^p]} + \frac{|g|^q}{q\mathbb{E}[|g|^q]}.$$

Calculando a esperança de ambos os lados da desigualdade acima e usando a lidearidade da esperança, obtemos

$$\frac{\mathbb{E}[|fg|]}{\mathbb{E}[|f|^p]^{\frac{1}{p}}\mathbb{E}[|g|^q]^{\frac{1}{q}}} \leqslant \frac{1}{p} + \frac{1}{q} = 1,$$

provando a Desigualdade de Hölder. □

[2] Retirado de Durrett (2013, p.25)
[3] Adaptado de Durrett (2013, p.21)

Teorema A.14 (Teorema da convergência monótona). *Seja μ uma medida sobre D. Se f e $\{f_n\}$ são funções definidas em D e tomando valores em $\mathbb{R}$ tais que $f_n \geqslant 0$, para todo $n \in \mathbb{N}$ e $f_n \uparrow f$, então*

$$\int f_n d\mu \uparrow \int f d\mu.$$

A demonstração deste teorema pode ser encontrada em Durrett (2013, p.23).

Teorema A.15 (Teorema da convergência dominada). *Seja μ uma medida sobre D. Se f, g e $\{f_n\}$ são funções definidas em D e tomando valores em $\mathbb{R}$ tais que $\mu(f_n \nrightarrow f) = 0$ (isto é, f_n não converge a f somente em um conjunto de medida μ nula), $|f_n| \leqslant g$, para todo $n \in \mathbb{N}$ e g é integrável, então*

$$\int f_n d\mu \to \int f d\mu.$$

A demonstração deste teorema pode ser encontrada em Durrett (2013, p.23).

Teorema A.16 (Teorema da extensão de Kolmogorov). *Sejam E^n, $E^{\mathbb{N}}$, $\mathcal{E}^n$ e $\mathcal{E}^{\mathbb{N}}$ definidos conforme a Definição 2.5. Suponhamos que existam medidas de probabilidade μ_n em $(E^n, \mathcal{E}^n)$ satisfazendo*

$$\mu_{n+1}(A_0 \times \cdots \times A_{n-1} \times E) = \mu_n(A_0 \times \cdots \times A_{n-1}). \tag{A.2}$$

Então, existe uma única medida de probabilidade $\mathbb{P}$ em $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$ com

$$\mathbb{P}(w : w_i \in A_i, 0 \leqslant i \leqslant n-1) = \mu_n(A_0 \times \cdots \times A_{n-1}).$$

A demonstração deste teorema pode ser encontrada em Durrett (2013, p.366-368).

Corolário A.17. *Seja μ uma probabilidade sobre um espaço polonês E. Então, existe uma sequência de variáveis aleatórias i.i.d. com distribuição μ.*

Demonstração. Vamos denotar $\Omega = E^{\mathbb{N}}$ e $\mathcal{F} = \mathcal{E}^{\mathbb{N}}$. Definimos as variáveis aleatórias $Y_i : \Omega \to E$ como sendo as projeções em cada uma das entradas, isto é, $Y_i(\omega) = \omega_i$. Em cada $(E^n, \mathcal{E}^n)$ definimos uma medida μ_n por

$$\mu_n(A_0 \times \cdots \times A_{n-1}) = \prod_{i=0}^{n-1} \mu(A_i).$$

Repare que as medidas μ_n definidas acima satisfazem a hipótese (A.2) do Teorema da extensão de Kolmogorov. Portanto, existe uma probabilidade $\mathbb{P}$ em $(\Omega, \mathcal{F})$ tal que

$$\mathbb{P}(w : w_i \in A_i, 0 \leqslant i \leqslant n-1) = \prod_{i=0}^{n-1} \mu(A_i).$$

Pela forma como definimos Y_i, podemos reescrever

$$\mathbb{P}[Y_0 \in A_0, \ldots, Y_{n-1} \in A_{n-1}] = \prod_{i=0}^{n-1} \mu(A_i). \tag{A.3}$$

Fixados $n \in \mathbb{N}$ e $A \subseteq E$, considerando $A_0 = \cdots = A_{n-2} = E$ e $A_{n-1} = A$ na igualdade acima, obtemos $\mathbb{P}[Y_{n-1} \in A] = \mu(A)$, ou seja, a sequência $\{Y_n\}$ é identicamente distribuida com distribuição μ.

Para mostrar a independência, basta voltar em (A.3) e usar que cada variável aleatória Y_i possui distribuição μ para provar que

$$\mathbb{P}[Y_0 \in A_0, \ldots, Y_{n-1} \in A_{n-1}] = \prod_{i=0}^{n-1} \mu(A_i) = \prod_{i=0}^{n-1} \mathbb{P}[Y_i \in A_i].$$

Portanto, $\{Y_i\}$ é i.i.d. $\square$

Definição A.18. *Dizemos que uma sequência de variáveis aleatórias $\{X_n\}$ é exponencialmente rígida se, para cada $a > 0$, existe um conjunto compacto K_a tal que*

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}[X_n \notin K_a] \leqslant -a.$$

Proposição A.19. *Seja $\{X_n\}$ uma sequência de variáveis aleatórias exponencialmente rígida. Se*

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[X_n \in K\right] \leqslant -\inf_{x\in K} f(x),$$

para qualquer K compacto, então $\{X_n\}$ satisfaz

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[X_n \in F\right] \leqslant -\inf_{x\in F} f(x),$$

para qualquer F fechado.

Demonstração. Seja F um conjunto fechado e $a > 0$. Então, como $\mathbb{P}$ é uma probabilidade,

$$\mathbb{P}\left[X_n \in F\right] \leqslant \mathbb{P}\left[X_n \in F \cap K_a\right] + \mathbb{P}\left[X_n \in K_a^{\complement}\right].$$

Logo,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[X_n \in F\right] \leqslant \max\left\{\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[X_n \in F \cap K_a\right], \overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[X_n \in K_a^{\complement}\right]\right\}.$$

Como F é fechado e K_a é compacto, temos que $F \cap K_a$ é compacto. Então, aplicando a hipótese para $K = F \cap K_a$ e utilizando o fato de $\{X_n\}$ ser exponencialmente rígida, concluimos que

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[X_n \in F\right] \leqslant \max\left\{-\inf_{x\in F\cap K_a} f(x), -a\right\} \leqslant \max\left\{-\inf_{x\in F} f(x), -a\right\}.$$

Essa desigualdade vale para qualquer $a > 0$. Fazendo $a \to \infty$, obtemos

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[X_n \in F\right] \leqslant -\inf_{x\in F} f(x).$$

$\square$

Lema A.20. *Sejam S um espaço polonês, $K \subseteq S$ um conjunto compacto, Λ um conjunto de parâmetros e $\{J_\lambda : S \to \mathbb{R}; \lambda \in \Lambda\}$ uma família de funções semicontínuas superiormente. Assim,*

$$\inf_{O_1,\dots,O_n} \max_{1 \leqslant j \leqslant n} \inf_\lambda \sup_{y \in O_j} J_\lambda(y) \leqslant \sup_{x \in K} \inf_\lambda J_\lambda(x),$$

onde o primeiro ínfimo do lado esquerdo da desigualdade é tomado sobre todas coberturas abertas $\{O_1, \dots, O_n\}$ de K.

Demonstração. Sejam $R = \sup_{x \in K} \inf_\lambda J_\lambda(x)$ e $\varepsilon > 0$. Para cada $x \in K$, $\inf_\lambda J_\lambda(x) \leqslant R$ e, portanto, existe λ_x tal que $J_{\lambda_x}(x) \leqslant R + \varepsilon$. Por hipótese, cada J_λ é semicontínua superiormente. Logo, existe uma vizinhança O_x de x tal que

$$\sup_{y \in O_x} J_{\lambda_x}(y) \leqslant R + 2\varepsilon.$$

A família $\{O_x, x \in K\}$ é uma cobertura aberta para K. Como K é compacto, podemos extrair uma subcobertura finita $O_{x_1}, \dots, O_{x_n}$. A desigualdade anterior é satisfeita para cada um desses conjuntos, ou seja,

$$\sup_{y \in O_{x_j}} J_{\lambda_{x_j}}(y) \leqslant R + 2\varepsilon, \text{ para } 1 \leqslant j \leqslant n.$$

Como o lado direito dessa desigualdade é uma constante, concluímos que

$$\inf_{O_1,\dots,O_n} \max_{1 \leqslant j \leqslant n} \inf_\lambda \sup_{y \in O_{x_j}} J_\lambda(y) \leqslant R + 2\varepsilon,$$

onde o primeiro ínfimo do lado esquerdo da desigualdade é tomado sobre todas coberturas abertas $\{O_1, \dots, O_n\}$ de K.

Por fim, basta fazer $\varepsilon \to 0$ para concluir a demonstração. ☐

O próximo resultado é muito importante no estudo de grandes desvios, pois ele permite trocar a ordem do supremo e do ínfimo apenas com a hipótese de semicontinuidade superior de J_λ. Muitas vezes será necessário fazer uso desse resultado para provar a cota superior do LDP.

Lema A.21. *Sejam S um espaço polonês, Λ um conjunto de parâmetros, $\{J_\lambda : S \to \mathbb{R}; \lambda \in \Lambda\}$ uma família de funções semicontínuas superiormente e $\{X_n\}$ uma sequência de variáveis aleatórias. Supondo que, para cada $O \subseteq S$ aberto,*

$$\overline{\lim_{n \to \infty}} \frac{1}{n} \log \mathbb{P}[X_n \in O] \leqslant \inf_\lambda \sup_{x \in O} J_\lambda(x). \tag{A.4}$$

Então, para cada $K \subseteq X$ compacto,

$$\overline{\lim_{n \to \infty}} \frac{1}{n} \log \mathbb{P}[X_n \in K] \leqslant \sup_{x \in K} \inf_\lambda J_\lambda(x).$$

Demonstração. Seja $O_1, \dots, O_n$ uma cobertura finita de K. Então,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[X_n \in K\right] \leqslant \overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[X_n \in \bigcup_{j=1}^{n} O_j\right] \leqslant \overline{\lim_{n\to\infty}} \frac{1}{n} \log\left(\sum_{j=1}^{n} \mathbb{P}\left[X_n \in O_j\right]\right).$$

Utilizando o Lema A.3, concluímos que

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[X_n \in K\right] \leqslant \max_{1\leqslant j\leqslant n} \overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[X_n \in O_j\right].$$

Pela suposição (A.4), temos

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathbb{P}\left[X_n \in K\right] \leqslant \max_{1\leqslant j\leqslant n} \inf_{\lambda} \sup_{x\in O_j} J_\lambda(x).$$

Para concluir a demonstração, basta minimizar a desigualdade acima sobre todas as coberturas finitas de K e utilizar o Lema A.20. □

Proposição A.22. *Seja $\{X_n\}$ uma sequência de variáveis aleatórias i.i.d. integráveis tomando valores em $\mathbb{R}$ tais que $\varphi(\lambda) = \mathbb{E}[e^{\lambda X_1}] < +\infty$ para todo $\lambda \in \mathbb{R}$. Então, X_i possui todos os momentos finitos, para qualquer $i \in \mathbb{N}$ e φ é infinitamente derivável, isto é, $\varphi \in C^\infty(\mathbb{R})$.*

Demonstração. Como a sequência é i.i.d, basta mostrar que X_1 possui todos os momentos finitos. Para isso, note que

$$\begin{aligned}\mathbb{E}[|X_1|^k] &= \mathbb{E}[|X_1|^k \mathbb{1}_{[X_1\geqslant 0]}] + \mathbb{E}[|X_1|^k \mathbb{1}_{[X_1<0]}] \\ &\leqslant \mathbb{E}\left[\exp\left((k!)^{\frac{1}{k}} X_1\right) \mathbb{1}_{[X_1\geqslant 0]}\right] + \mathbb{E}\left[\exp\left(-(k!)^{\frac{1}{k}} X_1\right) \mathbb{1}_{[X_1<0]}\right] \\ &\leqslant \varphi\left((k!)^{\frac{1}{k}}\right) + \varphi\left(-(k!)^{\frac{1}{k}}\right) < +\infty,\end{aligned}$$

para qualquer $k \in \mathbb{N}$.

Vamos agora mostrar que $\varphi \in C^\infty(\mathbb{R})$. Observe que, para qualquer $\lambda \in \mathbb{R}$,

$$\frac{\varphi(\lambda + h) - \varphi(\lambda)}{h} = \mathbb{E}\left[e^{\lambda X_1}\left(\frac{e^{hX_1} - 1}{h}\right)\right].$$

Repare que, para provarmos que φ é derivável, precisamos mostrar que a expressão acima converge quando $h \to 0$. Como $\frac{e^{hX_1}-1}{h} \to X_1$, quando $h \to 0$, é suficiente mostrar que $e^{\lambda X_1}\left(\frac{e^{hX_1} - 1}{h}\right)$ é integrável, pois neste caso, o resultado desejado será consequência direta do Teorema da convergência dominada (Teorema A.15).

Expandindo e^{hX_1} em Série de Taylor, podemos escrever

$$\frac{e^{hX_1} - 1}{h} = X_1 + \frac{hX_1^2}{2} + \frac{h^2 X_1^3}{3!} + \cdots$$

e, portanto,

$$\mathbb{E}\left[\left|e^{\lambda X_1}\left(\frac{e^{hX_1} - 1}{h}\right)\right|\right] \leqslant \mathbb{E}[e^{2\lambda X_1}]^{\frac{1}{2}} + \mathbb{E}[|X_1|^2]^{\frac{1}{2}} < +\infty,$$

pois $\varphi(2\lambda) < +\infty$ e X_1 possui todos os momentos finitos. Portanto, pelo Teorema da convergência dominada,

$$\lim_{h\to 0} \frac{\varphi(\lambda+h)-\varphi(\lambda)}{h} = \mathbb{E}[e^{\lambda X_1} X_1].$$

Repetindo o mesmo argumento diversas vezes, concluímos que $\varphi \in C^\infty(\mathbb{R})$ e

$$\varphi^{(n)}(\lambda) = \mathbb{E}[e^{\lambda X_1} X_1^2].$$

□

Proposição A.23. *Seja $\{X_n\}$ uma sequência de variáveis aleatórias i.i.d. integráveis tomando valores em $\mathbb{R}$. Suponhamos que $\varphi(\lambda) = \mathbb{E}[e^{\lambda X_1}] < +\infty$ para todo $\lambda \in \mathbb{R}$. Então, as funções $\varphi(\lambda)$, $\log \varphi(\lambda)$ e I (definida em (1.6)) são convexas.*

Demonstração. Como $\varphi(\lambda) < +\infty$ para todo $\lambda \in \mathbb{R}$, pela Proposição A.22, temos que $\varphi \in C^\infty(\mathbb{R})$. Assim,

$$\begin{aligned}\varphi'(\lambda) &= \mathbb{E}[X_1 e^{\lambda X_1}],\\ \varphi''(\lambda) &= \mathbb{E}[X_1^2 e^{\lambda X_1}].\end{aligned}$$

Repare que $\varphi''(\lambda) \geqslant 0$. Logo, φ é convexa. Seja $\psi(\lambda) = \log \varphi(\lambda)$. Então, para $\theta \in [0,1]$,

$$\psi(\theta t + (1-\theta)s) = \log \mathbb{E}[e^{\theta t X_1} e^{(1-\theta)sX_1}].$$

Aplicando a Desigualdade de Hölder (Teorema A.13) com $p = \frac{1}{\theta}$, $q = \frac{1}{1-\theta}$, $f = e^{\theta t X_1}$ e $g = e^{(1-\theta)sX_1}$, concluímos que

$$\begin{aligned}\psi(\theta t + (1-\theta)s) &\leqslant \log \left(\mathbb{E}\left[\left(e^{\theta t X_1}\right)^{\frac{1}{\theta}} \right]^{\theta} \mathbb{E}\left[\left(e^{(1-\theta)sX_1}\right)^{\frac{1}{1-\theta}} \right]^{1-\theta} \right)\\ &= \log \mathbb{E}[e^{tX_1}]^{\theta} + \log \mathbb{E}[e^{sX_1}]^{1-\theta}\\ &= \theta \log \mathbb{E}[e^{tX_1}] + (1-\theta) \log \mathbb{E}[e^{sX_1}]\\ &= \theta\psi(t) + (1-\theta)\psi(s).\end{aligned}$$

Logo, $\log \varphi(\lambda)$ é convexa.

Falta mostrar que I é convexa. Para isso, seja $a \in [0,1]$. Então,

$$\begin{aligned}I(ax + (1-a)y) &= \sup_{\lambda\in\mathbb{R}} [\lambda(ax + (1-a)y) - \log \varphi(\lambda)]\\ &= \sup_{\lambda\in\mathbb{R}} [\lambda a x + (1-a)\lambda y - a \log \varphi(\lambda) - (1-a) \log \varphi(\lambda)]\\ &\leqslant \sup_{\lambda\in\mathbb{R}} [\lambda a x - a \log \varphi(\lambda)] + \sup_{\lambda\in\mathbb{R}} [(1-a)\lambda y - (1-a) \log \varphi(\lambda)]\\ &= a \sup_{\lambda\in\mathbb{R}} [\lambda x - \log \varphi(\lambda)] + (1-a) \sup_{\lambda\in\mathbb{R}} [\lambda y - \log \varphi(\lambda)]\\ &= aI(x) + (1-a)I(y).\end{aligned}$$

Logo, I é convexa. □

Lema A.24. *Seja φ conforme definida anteriormente. Suponhamos que o conjunto $\Phi^* = \{\lambda \in \mathbb{R} : \varphi(\lambda) < +\infty\}$ é aberto. Então, φ é contínua em Φ^*.*

Demonstração. Tomamos $\lambda_0 \in \Phi^*$. Para provarmos que φ é contínua em λ_0, precisamos verificar que $\varphi(\lambda_0 + h) - \varphi(\lambda_0) \to 0$, quando $h \to 0$. Note que

$$\varphi(\lambda_0 + h) - \varphi(\lambda_0) = \mathbb{E}[e^{(\lambda_0+h)X_1} - e^{\lambda_0 X_1}].$$

Repare que $e^{(\lambda_0+h)X_1} - e^{\lambda_0 X_1} \to 0$, quando $h \to 0$. Ainda, como $\lambda_0 \in \Phi^*$ e Φ^* é aberto, temos que, para h suficientemente pequeno, $\varphi(\lambda_0 + h) = \mathbb{E}[e^{(\lambda_0+h)X_1}] < +\infty$. Portanto, pelo Teorema da convergência dominada (Teorema A.15), $\varphi(\lambda_0 + h) - \varphi(\lambda_0) \to 0$, quando $h \to 0$. Logo, φ é contínua em todo $\lambda_0 \in \Phi^*$. □

A.3 Convergência de Medidas

Definição A.25. *Sejam $\{\mathbb{P}_n\}$ e $\mathbb{P}$ probabilidades em $(E, \mathcal{B}(E))$. Dizemos que $\mathbb{P}_n$ converge fracamente para $\mathbb{P}$, e denotamos $\mathbb{P}_n \Rightarrow \mathbb{P}$, se, para toda f contínua e limitada,*

$$\lim_{n\to\infty} \int f d\mathbb{P}_n = \int f d\mathbb{P}.$$

Ainda, se $\{\mathbb{P}_n\}$ e $\mathbb{P}$ forem as distribuições das variáveis aleatórias $\{X_n\}$ e X, respectivamente, dizemos que X_n converge em distribuição para X, e denotamos $X_n \xrightarrow{d} X$, se $\mathbb{P}_n \Rightarrow \mathbb{P}$.

Definição A.26. *Um conjunto de probabilidades Π sobre um espaço métrico $(E, \mathcal{B}(E))$ é dito rígido se, para qualquer $\varepsilon > 0$, existe um compacto $K \subseteq E$ tal que $\mathbb{P}(K) > 1 - \varepsilon$, para qualquer $\mathbb{P} \in \Pi$.*

Teorema A.27 (Teorema de Prohorov)**.** *Seja Π um conjunto de probabilidades sobre $(E, \mathcal{B}(E))$. Se Π é rígido, então qualquer sequência em Π possui subsequência fracamente convergente. Ainda, se E for polonês, vale a recíproca.*

A demonstração deste teorema pode ser encontrada em Baldasso (2013, p.5-12). Outra versão também pode ser encontrada em Billingsley (1995, p.57-65).

Seja E um conjunto finito, que assumiremos, sem perda de generalidade, $E = \{1, \ldots, d\}$. Denotando $\mathcal{M} = \mathcal{M}(E)$ como o conjunto das probabilidades sobre E, podemos mostrar que a convergência fraca em $\mathcal{M}$ é equivalente à convergência no simplex do $\mathbb{R}^d$, que definimos em (3.2) e denotamos por Θ_d. Para isto, vamos metrizar a convergência fraca, isto é, definiremos uma métrica d_P sobre $\mathcal{M}$ tal que $\mu_n \Rightarrow \mu$ se, e somente se, $d_P(\mu_n, \mu) \to 0$ e depois mostrar que esta métrica d_P é equivalente à métrica no simplex.

Iniciamos apresentando a definição de métrica de Prohorov no caso geral, em que E é um espaço polonês (espaço métrico completo e separável). Denotamos por d a métrica em E e definimos a distância de um ponto $x \in E$ até um conjunto fechado $F \subseteq E$ por $d(x, F) = \inf_{y \in F} d(x, y)$.

Definição A.28. *A métrica d_P, chamada métrica de Prohorov, é definida por*

$$d_P(\mu, \nu) = \inf\{\delta > 0 : \forall F \subseteq E \textit{ fechado}, \mu(F) \leqslant \nu(F^\delta) + \delta \textit{ e } \nu(F) \leqslant \mu(F^\delta) + \delta\},$$

onde $F^\delta = \{x \in E : d(x, F) < \delta\}$.

É importante salientar que a métrica de Prohorov definida acima é uma metrização da convergência fraca mesmo sem assumirmos que E é finito. O estudo no caso geral pode ser encontrado em Oliveira (s.d., p. 18). Aqui, faremos o caso em que E é finito.

Teorema A.29. *Se E é finito, então*

$$d_P(\mu, \nu) = \inf\{\delta > 0 : \forall F \subseteq E, |\mu(F) - \nu(F)| \leqslant \delta\}.$$

Demonstração. Repare que, como E é finito, qualquer subconjunto de E é fechado. Portanto, podemos reescrever a métrica de Prohorov como

$$d_P(\mu, \nu) = \inf\{\delta > 0 : \forall F \subseteq E, \mu(F) \leqslant \nu(F^\delta) + \delta \text{ e } \nu(F) \leqslant \mu(F^\delta) + \delta\}.$$

Note que, como μ e ν são probabilidades, as duas desigualdades acima são válidas quando $\delta \geqslant 1$. Portanto, $d_P(\mu, \nu) \leqslant 1$, ou seja, precisamos apenas analisar quando $\delta < 1$. Neste caso, como supomos $E = \{1, \ldots, d\}$ e $F \subseteq E$, temos que $d(x, F) < \delta$ se, e somente se, $x \in F$. Logo, $F^\delta = F$. Assim,

$$d_P(\mu, \nu) = \inf\{\delta > 0 : \forall F \subseteq E, \mu(F) \leqslant \nu(F) + \delta \text{ e } \nu(F) \leqslant \mu(F) + \delta\}.$$

Então,

$$d_P(\mu, \nu) = \inf\{\delta > 0 : \forall F \subseteq E, |\mu(F) - \nu(F)| \leqslant \delta\}.$$

□

Repare que o teorema anterior nos permite escrever a métrica de Prohorov de uma maneira muito mais simples. A hipótese de E ser finito e a consequente possibilidade de escrever a métrica desta forma nos permite provar as equivalências abaixo.

Teorema A.30. *Se $\{\mu_n\}$ e μ são probabilidades sobre um conjunto E finito, então as seguintes afirmações são equivalentes:*

(i) $\mu_n \Rightarrow \mu$;

(ii) Para todo $F \subseteq E$, $|\mu_n(F) - \mu(F)| \to 0$, *quando* $n \to \infty$;

(iii) $d_P(\mu_n, \mu) \to 0$, *quando* $n \to \infty$.

Demonstração. Vamos começar provando que as duas primeiras afirmações são equivalentes. Suponhamos que $\mu_n \Rightarrow \mu$. Por definição de convergência fraca, temos que, para qualquer f contínua e limitada,

$$\int f d\mu_n \longrightarrow \int f d\mu,$$

quando $n \to \infty$. Portanto, se $n \to \infty$, então

$$\left| \int f d\mu_n - \int f d\mu \right| \longrightarrow 0.$$

Como supomos $E = \{1, \dots, d\}$, temos que, para qualquer $\mathbb{P}$ probabilidade sobre E, $\int f d\mathbb{P} = \sum_{i=1}^{d} f(i)\mathbb{P}(i)$. Assim, para toda f contínua e limitada, vale que

$$\left| \sum_{i=1}^{d} f(i)(\mu_n(i) - \mu(i)) \right| \longrightarrow 0.$$

Fixado $F \subseteq E$, seja $g = \mathbb{1}_F$. Tomamos $f : \mathbb{R} \to \mathbb{R}$ uma função contínua e limitada que seja uma extensão de g, isto é, f coincide com g nos elementos do conjunto E e pode assumir quaisquer outros valores fora desse conjunto. Note que, para esta f em particular,

$$\left| \sum_{i=1}^{d} f(i)(\mu_n(i) - \mu(i)) \right| = \left| \sum_{i \in F} (\mu_n(i) - \mu(i)) \right| = |\mu_n(F) - \mu(F)|.$$

Portanto, para todo $F \subseteq E$, $|\mu_n(F) - \mu(F)| \to 0$.

Para a recíproca, suponhamos que vale *(ii)*. No caso particular $F = \{i\}$, temos que, para qualquer $i \in E$, $|\mu_n(i) - \mu(i)| \to 0$, quando $n \to \infty$. Portanto,

$$\sum_{i=1}^{d} |\mu_n(i) - \mu(i)| \longrightarrow 0.$$

Seja f uma função contínua e limitada. Assim,

$$\left| \int f d\mu_n - \int f d\mu \right| = \left| \sum_{i=1}^{d} f(i)(\mu_n(i) - \mu(i)) \right| \leqslant \sum_{i=1}^{d} |f(i)| \, |\mu_n(i) - \mu(i)| \longrightarrow 0,$$

pois f é limitada e o somatório converge a 0 pelo que vimos anteriormente.

Concluiremos a demonstração provando que as afirmações *(ii)* e *(iii)* são equivalentes. Suponhamos que vale *(ii)*. Fixado $\delta_0 > 0$. Então, pela definição de limite, para cada $F \subseteq E$, existe um N_F tal que $|\mu_n(F) - \mu(F)| < \delta_0$, para todo $n \geqslant N_F$. Seja $N_0 = \max N_F$. Assim, para qualquer $F \subseteq E$, temos que $|\mu_n(F) - \mu(F)| < \delta_0$, para todo $n \geqslant N_0$. Logo,

$$\delta_0 \in \{\delta > 0 : \forall F \subseteq E, |\mu_n(F) - \mu(F)| \leqslant \delta\}.$$

Então, para todo $n \geqslant N_0$,

$$d_P(\mu_n, \mu) = \inf\{\delta > 0 : \forall F \subseteq E, |\mu_n(F) - \mu(F)| \leqslant \delta\} \leqslant \delta_0.$$

Portanto, $d_P(\mu_n, \mu) \to 0$, quando $n \to \infty$.

Para a recíroca desta equivalência, assumimos *(iii)* e fixamos $F \subseteq E$. Suponhamos por absurdo que $|\mu_n(F) - \mu(F)| \nrightarrow 0$, quando $n \to \infty$, isto é, existe $\varepsilon > 0$ tal que $|\mu_n(F) - \mu(F)| > \varepsilon$, a partir de um certo n. Portanto,

$$d_P(\mu_n, \mu) = \inf\{\delta > 0 : \forall F \subseteq E, |\mu_n(F) - \mu(F)| \leqslant \delta\} \geqslant \varepsilon.$$

Logo, $d_P(\mu_n, \mu) \nrightarrow 0$, o que é um absurdo, pois contraria a afirmação *(iii)*. Então, $|\mu_n(F) - \mu(F)| \to 0$, quando $n \to \infty$. □

Pelo Teorema acima, temos que d_P é a metrização da convergência fraca de medidas. O próximo passo é mostrar que esta métrica é equivalente à métrica no simplex.

Teorema A.31. *Sejam μ e ν probabilidades sobre $E = \{1, \ldots, d\}$. Com as notações inseridas anteriormente, vale que*

$$\max_{1 \leqslant i \leqslant d} |\mu(i) - \nu(i)| \leqslant d_P(\mu, \nu) \leqslant d \max_{1 \leqslant i \leqslant d} |\mu(i) - \nu(i)|.$$

Demonstração. Lembre que, pelo Teorema A.29,

$$d_P(\mu, \nu) = \inf\{\delta > 0 : \forall F \subseteq E, |\mu(F) - \nu(F)| \leqslant \delta\}.$$

Se $|\mu(i) - \nu(i)| = 0$, para todo $i \in E$, o resultado é imediato. Vamos então supor que isto não acontece. Repare que,

$$\{\delta > 0 : \forall F \subseteq E, |\mu(F) - \nu(F)| \leqslant \delta\} \subseteq \{\delta > 0 : |\mu(i) - \nu(i)| \leqslant \delta, \text{ para todo } i \in E\},$$

pois no lado direito é pedido que satisfaça a exigência apenas nos subconjuntos de E com exatamente um elemento. Disto segue que

$$\inf\{\delta > 0 : |\mu(i) - \nu(i)| \leqslant \delta, \text{ para todo } i \in E\} \leqslant \inf\{\delta > 0 : \forall F \subseteq E, |\mu(F) - \nu(F)| \leqslant \delta\}.$$

Portanto,

$$\inf\{\delta > 0 : |\mu(i) - \nu(i)| \leqslant \delta, \text{ para todo } i \in E\} \leqslant d_P(\mu, \nu). \tag{A.5}$$

Suponhamos, por absurdo, que

$$\inf\{\delta > 0 : |\mu(i) - \nu(i)| \leqslant \delta, \text{ para todo } i \in E\} < \max_{1 \leqslant i \leqslant d} |\mu(i) - \nu(i)|.$$

Neste caso, existe $\delta_0 > 0$ tal que $\delta_0 < \max_{1 \leqslant i \leqslant d} |\mu(i) - \nu(i)|$ e $|\mu(i) - \nu(i)| \leqslant \delta_0$, para todo $i \in E$. Seja $i_0 \in E$ tal que $\max_{1 \leqslant i \leqslant d} |\mu(i) - \nu(i)| = |\mu(i_0) - \nu(i_0)|$. Então,

$$\max_{1 \leqslant i \leqslant d} |\mu(i) - \nu(i)| = |\mu(i_0) - \nu(i_0)| \leqslant \delta_0 < \max_{1 \leqslant i \leqslant d} |\mu(i) - \nu(i)|,$$

chegando a um absurdo. Assim,

$$\inf\{\delta > 0 : |\mu(i) - \nu(i)| \leqslant \delta, \text{ para todo } i \in E\} \geqslant \max_{1 \leqslant i \leqslant d} |\mu(i) - \nu(i)|. \tag{A.6}$$

Por outro lado, para qualquer $j \in E$,

$$|\mu(j) - \nu(j)| \leqslant \max_{1\leqslant i\leqslant d} |\mu(i) - \nu(i)|.$$

Como supomos inicialmente que $|\mu(i) - \nu(i)| \neq 0$, para algum $i \in E$, concluímos que

$$\max_{1\leqslant i\leqslant d} |\mu(i) - \nu(i)| \in \{\delta > 0 : |\mu(i) - \nu(i)| \leqslant \delta, \text{ para todo } i \in E\}.$$

Logo,

$$\inf\{\delta > 0 : |\mu(i) - \nu(i)| \leqslant \delta, \text{ para todo } i \in E\} \leqslant \max_{1\leqslant i\leqslant d} |\mu(i) - \nu(i)|. \tag{A.7}$$

Segue de (A.6) e (A.7) que

$$\inf\{\delta > 0 : |\mu(i) - \nu(i)| \leqslant \delta, \text{ para todo } i \in E\} = \max_{1\leqslant i\leqslant d} |\mu(i) - \nu(i)|. \tag{A.8}$$

Juntando (A.5) e (A.8), concluímos que

$$\max_{1\leqslant i\leqslant d} |\mu(i) - \nu(i)| \leqslant d_P(\mu, \nu).$$

Para verificar a outra desigualdade, note que

$$|\mu(F) - \nu(F)| \leqslant \sum_{i\in F} |\mu(i) - \nu(i)|.$$

Portanto,

$$\left\{\delta > 0 : |\mu(i) - \nu(i)| \leqslant \frac{\delta}{d}, \text{ para todo } i \in E\right\} \subseteq \{\delta > 0 : \forall F \subseteq E, |\mu(F) - \nu(F)| \leqslant \delta\}.$$

Logo,

$$\inf\{\delta > 0 : \forall F \subseteq E, |\mu(F) - \nu(F) \leqslant \delta\} \leqslant \inf\left\{\delta > 0 : |\mu(i) - \nu(i)| \leqslant \frac{\delta}{d}, \text{ para todo } i \in E\right\}.$$

Seja $\alpha > 0$ tal que $\delta = d\alpha$. Pela forma como d_P e α foram definidos, segue da desigualdade acima que

$$d_P(\mu, \nu) \leqslant \inf\{d\alpha : |\mu(i) - \nu(i)| \leqslant \alpha, \text{ para todo } i \in E\}.$$

Pela igualdade (A.8), segue da desigualdade acima que

$$d_P(\mu, \nu) \leqslant d \inf\{\alpha : |\mu(i) - \nu(i)| \leqslant \alpha, \text{ para todo } i \in E\} = d \max_{1\leqslant i\leqslant d} |\mu(i) - \nu(i)|.$$

Portanto,

$$\max_{1\leqslant i\leqslant d} |\mu(i) - \nu(i)| \leqslant d_P(\mu, \nu) \leqslant d \max_{1\leqslant i\leqslant d} |\mu(i) - \nu(i)|.$$

$\square$

Como $\mathcal{M}$ é homeomorfo ao simplex Θ_d através do homeomorfismo $\Phi : \mathcal{M} \to \Theta_d$ tal que $\Phi(\mu) = (\mu(1), \ldots, \mu(d))$, podemos identificar $\mathcal{M}$ com Θ_d. Como consequência do teorema anterior, temos que a convergência fraca de medidas, que é metrizada por d_P, é equivalente a convergência em Θ_d, que pode ser metrizada por d tal que

$$d(x, y) = \max_{1 \leqslant i \leqslant d} |x_i - y_i|,$$

para quaisquer $x, y \in \mathbb{R}^d$.

A definição de LDP apresentada na introdução não está na forma mais geral, pois podemos ter o enunciado de LDP posto somente em termos de sequências de medidas $\{\mu_n\}$ sobre um espaço polonês R.

Definição A.32. *Sejam R um espaço polonês e $\mathcal{B}(R)$ a sua σ-álgebra de Borel. Dizemos que a sequência de probabilidades $\{\mu_n\}$ em $(R, \mathcal{B}(R))$ satisfaz o LDP (princípio de grandes desvios) se existe uma função $I : R \to [0, +\infty]$ tal que*

(i) Cota superior: Para cada $F \in \mathcal{B}(R)$ fechado,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log \mu_n(F) \leqslant - \inf_{x \in F} I(x);$$

(ii) Cota inferior: Para cada $O \in \mathcal{B}(R)$ um aberto,

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mu_n(O) \geqslant - \inf_{x \in O} I(x).$$

Repare que, se μ_n é a distribuição de $\mathcal{Z}_n$, a definição acima implica a definição de LDP apresentada na introdução. O ponto importante é que, para esta definição apresentada aqui, nem precisamos ter variáveis aleatórias. Observamos que esta definição mais geral não foi apresentada na parte inicial deste texto (Capítulos 1 e 3), pois ela tornaria os enunciados dos resultados que provamos mais carregados na notação.

REFERÊNCIAS

BALDASSO, R. **Comportamento hidrodinâmico para o processo de exclusão com taxa lenta no bordo**. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, 2013.

BILLINGSLEY, P. **Probability and measure**. New York: John Wiley & Sons, 1995.

DONSKER, M.; VARADHAN, S. Asymptotic evaluation of certain Markov process expectations for large time I. **Communication on Pure and Applied Mathematics**, Wiley Online Library, v. 28, n. 1, p. 1–47, 1975a.

DONSKER, M.; VARADHAN, S. Asymptotic evaluation of certain Markov process expectations for large time II. **Communication on Pure and Applied Mathematics**, Wiley Online Library, v. 28, n. 2, p. 279–301, 1975b.

DONSKER, M.; VARADHAN, S. Asymptotic evaluation of certain Wiener integrals for large time. In: ARTHURS, A. M. (Ed.). **Functional Integration and Its Applications**. Proceedings of the International Conference held at Cumberland Lodge, Windsor Great Park, London. Oxford: Clarendon Press, 1975c. p. 15–33.

DONSKER, M.; VARADHAN, S. Asymptotic evaluation of certain Markov process expectations for large time III. **Communication on Pure and applied Mathematics**, Wiley Online Library, v. 29, n. 4, p. 389–461, 1976.

DONSKER, M.; VARADHAN, S. Large deviations from a hydrodynamic scaling limit. **Communication on Pure and Applied Mathematics**, Wiley Online Library, v. 42, n. 3, p. 243–270, 1989.

DURRETT, R. **Probability: theory and examples**. Cambridge: Cambridge University Press, 2013.

HOLLANDER, F. D. **Large deviations**. Providence: American Mathematical Soc., 2008.

HORN, R. A.; JOHNSON, C. R. **Matrix analysis**. Cambridge: Cambridge University Press, 2012.

KIPNIS, C.; LANDIM, C. **Scaling limits of interacting particle systems**. New York: Springer Science & Business Media, 2013.

LEVIN, D. A.; PERES, Y.; WILMER, E. L. **Markov chains and mixing times**. Providence: American Mathematical Soc., 2009.

MÜLLER, G. H. **Cadeias de Markov e sua aplicação no algoritmo de buscas do Google**. 2015. 45 f. Trabalho de Conclusão de Curso (Licenciatura em Matemática) - Curso de Matemática, Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, 2015.

NORRIS, J. R. **Markov chains**. Cambridge: Cambridge university press, 1998.

OLIVEIRA, R. I. M. F. de. **Notas de aula sobre probabilidade em espaços poloneses**. s.d. Disponível em: <http://w3.impa.br/~rimfo/notas_polones.pdf>. Acesso em: 11 jul. 2016.

OLIVIERI, E.; VARES, M. E. **Large deviations and metastability**. Cambridge: Cambridge University Press, 2005.

ROCKAFELLAR, R. T. **Convex Analysis**. Princeton: Princeton University Press, 1997.

ROUSSEAU, C. et al. **Matemática e Atualidade Volume 1**. Rio de Janeiro: SBM, 2015.

STROOCK, D. W. **An introduction to the theory of large deviations**. New York: Springer Science & Business Media, 1984.

VARADHAN, S. **Large deviations and applications**. Philadelphia: SIAM, 1984.

VARADHAN, S. Special invited paper. Large deviations. **The annals of probability**, Institute of Mathematical Statistics, v. 36, n. 2, p. 397–419, 2008.