Vinicio Rodrigues Moreira

Uma Introdução a Grandes Desvios e Finanças Quantitativas

Vitória

2022

Vinicio Rodrigues Moreira

Uma Introdução a Grandes Desvios e Finanças Quantitativas

Dissertação de mestrado apresentada ao PPG-MAT como parte dos requisitos exigidos para a obtenção do título de Mestre em Matemática

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Orientador: Prof. Dr. Fábio Júlio da Silva Valentim
Coorientador: Prof. Dr. Yuri Fahham Saporito

Vitória
2022

Ficha catalográfica disponibilizada pelo Sistema Integrado de
Bibliotecas - SIBI/UFES e elaborada pelo autor

Sobrenome do autor, Nome do autor, 1989-
S677t Título da obra : Subtítulo da obra / Nome do autor Sobrenome do autor. - 2021.
Total de folhas f. : il.

Orientador: Nome do orientador Sobrenome do orientador.
Coorientador: Nome do coorientador Sobrenome do coorientador.
Dissertação (Mestrado em Matemática) - Universidade Federal do Espírito Santo, Centro de Ciências Exatas.

1. Análise matemática. I. Sobrenome do orientador, Nome do orientador. II. Sobrenome do coorientador, Nome do coorientador. III. Universidade Federal do Espírito Santo. Centro de Ciências Exatas. IV. Título.

CDU: 51

Nome do Autor

Título da Dissertação

Dissertação de mestrado apresentada ao PPG-MAT como parte dos requisitos exigidos para a obtenção do título de Mestre em Matemática

Trabalho aprovado. Vitória, 3 de agosto de 2021:

Prof. Dr. Nome do Orientador
Universidade Federal do Espírito Santo
Orientador

Prof. Dr. Convidado 1
Universidade Federal do Espírito Santo
Membro Interno

Prof. Dr. Convidado 2
Instituição
Membro Externo

Vitória
2021

Editável

Vinicio Rodrigues Moreira

Uma Introdução a Grandes Desvios e Finanças Quantitativas

Dissertação de mestrado apresentada ao PPG-MAT como parte dos requisitos exigidos para a obtenção do título de Mestre em Matemática

Trabalho aprovado. Vitória, 29 de novembro de 2022:

Prof. Dr. Fábio Júlio da Silva Valentim
Universidade Federal do Espírito Santo
Orientador

Prof. Dr. Yuri Fahham Saporito
Fundação Getúlio Vargas
Coorientador

Diogo Manuel Fernandes Bessam
Universidade Federal do Espírito Santo
Membro Interno

Glauco Valle
Universidade Federal do Rio de Janeiro
Membro Externo

Vitória

2022

Texto referente a dedicatória.

Agradecimentos

Texto referente aos agradecimentos.

“Texto referente a epigrafe.”
(Indicar fonte do texto)

Resumo

Neste trabalho iremos introduzir ao leitor A Teoria de Grandes Desvios. Desenvolveremos resultados que nos permitirão estimar a velocidade do decaimento da probabilidade de um evento, assim como uma aplicação em finanças quantitativas. Os principais teoremas desse trabalho são o Teorema de Sanov, que nos permite estudar a velocidade de decaimento de probabilidades em um conjunto finito, o Teorema de Crámer, que nos permite estudar a velocidade de decaimento de probabilidades em $\mathbb{R}$, e um Teorema na parte de Finanças Quantitativas, que nos introduz ao estudo de Perdas em Grandes Portfólios de Crédito de Risco.

Palavras-chave: Teorema de Crámer, Teoria dos Grandes Desvios, Teorema de Sanov, Finanças Quantitativas.

Abstract

In this dissertation we will introduce the reader to The Theory of Large Deviations. We will develop results that will allow us to estimate the decay rate of the probability of an event, as well as an application in quantitative finance. The main results of this work are Sanov's Theorem, which allows us to study the decay speed of probabilities in a finite set, Crámer's Theorem, which allows us to study the decay speed of probabilities in $\mathbb{R}$, and a Theorem in the Sanov's Theorem, which introduces us to the study of Losses in Large Risk Credit Portfolios.

Keywords: Crámer's Theorem, Theory of Large Deviations, Sanov's Theorem, Quantitative Finance

Sumário

Introdução

A Teoria Grandes Desvios estuda o decaimento exponencial de probabilidades de determinados eventos, que chamaremos de eventos raros, em certos sistemas aleatórios. Surgiu como teoria geral durante as décadas de 70 e 80, a partir dos trabalhos independentes de Donssker e Varadhan ((DONSKER; VARADHAN, 1975a) (DONSKER; VARADHAN, 1975b), (DONSKER; VARADHAN, 1976), (DONSKER; VARADHAN, 1983)), e de Freidlin e Wentzell ((FREIDLIN; WENTZELL, 1984)). Contudo resultados sobre Grandes Desvios existiam antes deste período, como por exemplo o trabalho feito por Boltzmann em mecânica estatística no ano de 1877 (ver (ELLIS, 1999)), mas não existia uma teoria unificada que lidava com esses resultados. A teoria de Grandes Desvios atua em diversas áreas do conhecimento, como por exemplo na já mencionada área de mecânica estatística no trabalho realizado por Boltzmann, também pode ser encontrada área de sistemas de comunicações, onde o evento raro pode ser entendido como um colapso do sistema, conseguimos encontrar também aplicações na área de finanças, como o apreçamento de opções e no caso em que estudaremos mais ao final deste trabalho a Perda em Grandes Portfólios de Crédito de Risco.

O capítulo 1 deste trabalho é baseado nos capitulo 1 e 2 do livro Large Deviations and Techinques and Aplications (ver (DEMBO; ZEITOUNE, 1998)), no qual iremos dar ao leitor uma noção do que é a teoria dos Grandes Desvios, e introduziremos conceitos, como por exemplo o Principio dos Grades Desvios, que nos fala que o decaimento da probabilidade de um evento é exponencial, e também calcularemos taxas deste decaimento, uma no caso discreto com o teorema de Sanov, e outra no caso contínuo com o Teorema de Cramér.

O capítulo 2 deste trabalho é baseado no artigo Large Deviations in Multifactor Portfolio Credit Risk (ver (GLASSERMAN; KANG; SHAHABUDDIN, 2007)), onde ao leitor será mostrado uma modelagem básica de um problema de grande importância para instituições financeiras, que é o tema de Grandes Perdas em Portfólios de Crédito de Risco. Esta modelagem sera divida em duas partes, a primeira é quando os devedores que estão neste portfólio são independes entre-se e a segunda é quando eles tem um grau de dependência.

No apêndice deste trabalho o leitor encontrará duas seções com definições e enunciados de teoremas das teorias de medida e probabilidade, como por exemplo as definições de medida e σ-álgebra na teoria de medida e definição de variável aleatória e o Teorema da Convergência Dominada em teoria da probabilidade. Também encontrara uma seção dedicada a Esperança Condicional e Probabilidade Condicional, que tem por objetivo deixar claro ao leitor a definição de Esperança Condicional usada neste trabalho e mostrar que existe uma "implicação"da definição utilizada com outra forma de definir Esperança Condicional. Recomenda-se que o leitor tenha um conhecimento prévio de topologia e análise na reta.

1 Teoria dos Grandes Desvios

Nesta seção temos o objetivo de introduzir ao leitor a Teoria dos Grandes Desvios. Uma pergunta óbvia a ser feita: "Por que Grande Desvios?". O próprio nome nos oferecerá uma pista daquilo que poderemos ver mais a frente. Quando a palavra "Desvio"vem em algum contexto, geralmente é por que algo fora do comum, esperado, etc..., pode acontecer. Então quando lemos "Grande Desvio"podemos esperar que algo muito fora do padrão pode acontecer. Para trazermos essa ideia de Desvio para nosso contexto matemático, consideremos a função

$$1_x : \mathbb{R} \longrightarrow \{0, 1\},$$

dada por $1_x(y) = 0$, se $y \neq x$ e $1_x(y) = 1$, caso $y = x$. Tomando $\epsilon = \frac{1}{t}$, com $t \to +\infty$, e pegando uma família de probabilidades em $\mathbb{R}$, $\{\mu_\epsilon\}_{\epsilon \in \mathbb{R}^+}$, tal que $\epsilon \to 0$ implica que $\mu_\epsilon \to 1_x$, ou seja dado $A \subset \mathbb{R}$ mensurável, tal que $p \notin A$, então $\mu_\epsilon(A) \to 0$. Notemos que o evento $\{x\}$, vai se tornando algo esperado a medida que t tende a $+\infty$ e os eventos A no qual x não acontece irão se tornar eventos inesperados, desvios de comportamento.

Para estudar esse comportamento de desvio, iremos considerar $\mathcal{X}$ um espaço topológico, $\mathcal{B}$ uma sigma álgebra de $\mathcal{X}$ e uma família de medidas de probabilidades $\{\mu_\epsilon\}_{\epsilon>0}$ em $(\mathcal{X}, \mathcal{B})$. Para isso iremos introduzir ao leitor O Principio dos Grandes Desvios, no qual nosso objetivo é encontrar uma função L, que chamaremos de taxa, tal que dado $A \in \mathcal{B}$, vale

$$\exp\left(-\epsilon \inf_{x \in A^o} L(x)\right) \leq \mu_\epsilon(A) \leq \exp\left(-\epsilon \inf_{x \in \overline{A}} L(x)\right)$$

para ϵ suficiente próximo de zero, onde A^o é o interior de A e $\overline{A}$ é o fecho de A.

A partir disso iremos primeiramente tomar $\mathcal{X}$ como sendo um conjunto discreto e para encontrar nossa função L, olharemos para as probabilidades que serão induzidas por uma sequências finitas de elementos de $\mathcal{X}$, onde essas probabilidades representarão as frequências com que cada elemento aparece na sequência. Com isso veremos que essas probabilidades tem a características de serem densas no espaço de todas a probabilidades possíveis de $\mathcal{X}$. Usando então essa propriedade de densidade e por meio de limites e desigualdades iremos demonstrar nosso primeiro grande resultado conhecido como Teorema de Sanov.

Para expandir a ideia que vimos no Teorema de Sanov, iremos fazer o Teorema de Cramér, este que atua em variáveis aleatórias que tomam valores em $\mathbb{R}$. Assim tomaremos $\mathcal{X} = \mathbb{R}$ e iremos considerar a $\sigma-$álgebra de Lebesgue como sendo a $\sigma-$álgebra de $\mathbb{R}$.

Veremos que nos dois resultados anteriores, estamos considerando sequências de variáveis aleatórias independentes e identicamente distribuídas, para que o leitor deste trabalho não seja induzido a pensar que essa hipótese é necessária, iremos enunciar o Teorema de Gartner-Ellis. Para mais detalhes deste resultado, indicamos que o leitor veja (DEMBO; ZEITOUNE, 1998).

1.1 O Principio dos Grandes Desvios

Para podermos estudas os dois principais resultados de seção é necessário introduzir alguns conceitos que nos serão de grande ajuda. Sejam $\mathcal{X}$ um espaço topológico e $\mathcal{B}$ uma σ-álgebra de $\mathcal{X}$.

Definição 1.1.1. *Uma função $I : \mathcal{X} \longrightarrow [0, \infty]$ é chamada de **função taxa** se é semi-contínua inferiormente, isto é, para todo $\alpha \in [0, \infty)$ o conjunto **nível** $\Psi_I(\alpha) = \{x \in \mathcal{X} : I(x) \leq \alpha\}$ é fechado. Dizemos que I é uma **função taxa bem comportada** se o conjunto nível $\Psi_I(\alpha)$ é compacto para todo $\alpha \in [0, \infty)$. O **Domínio Efetivo** de I, denotado por $\mathcal{D}_I$, é o conjunto dos pontos de $\mathcal{X}$ que tem taxa finita, ou seja $\mathcal{D}_I = \{x \in \mathcal{X} : I(x) < \infty\}$.*

Exemplo 1.1.2. *Considere*

$$f(x) = \begin{cases} x^2 + 1 & , \ x \neq 0 \\ 0 & , \ x = 0. \end{cases} \tag{1.1}$$

Observemos que $\Psi_f(\alpha)$ é um conjunto compacto para todo $\alpha \in [0, +\infty)$ e portanto f é semicontínua inferiormente. Como f é não negativa, segue que f é uma função taxa bem comportada.

Observação: Se $g : \mathcal{X} \longrightarrow \mathbb{R}$, é continua, então segue diretamente da definição que g é uma função taxa. Porem notemos que a volta não é verdadeira, basta tomar (1.1).

A partir de agora iremos considerar o ínfimo de uma função no conjunto vazio como sendo $+\infty$. A seguir daremos a definição do Princípio dos Grandes Desvios.

Definição 1.1.3. *Dizemos que uma família $\{\mu_\epsilon\}_{\epsilon>0}$ de distribuições de probabilidade satisfaz o **Princípio dos Grandes Desvios** com a função taxa I, se para todo $\Gamma \in \mathcal{B}$*

$$-\inf_{x\in\Gamma^o} I(x) \leq \liminf_{\epsilon\to 0} \epsilon \ln \mu_\epsilon(\Gamma) \leq \limsup_{\epsilon\to 0} \epsilon \ln \mu_\epsilon(\Gamma) \leq -\inf_{x\in\overline{\Gamma}} I(x). \tag{1.2}$$

O que basicamente dissemos na definição 1.1.3, é que conseguimos controlar a taxa de decaimento da probabilidade de qualquer conjunto em $\mathcal{B}$ por meio de duas exponenciais. Denotando $\mathcal{B}_\mathcal{X}$, como sendo a sigma-álgebra de Borel no espaço topológico $\mathcal{X}$, temos o seguinte resultado.

Lema 1.1.4. *Se $\mathcal{B}_\mathcal{X} \subset \mathcal{B}$ e I é uma função taxa, então são equivalentes as seguintes sentenças:*

1. *Definição 1.1.3.*

2. *i) Para todo* $\alpha < \infty$ *e todo conjunto mensurável* $\Gamma \in \mathcal{B}$*, com* $\overline{\Gamma} \subset \Psi_I(\alpha)^c$*, onde* $\Psi_I(\alpha)^c$ *é complementar do conjunto* $\Psi_I(\alpha)$*, vale*

$$\limsup_{\epsilon \to 0} \epsilon \ln \mu_\epsilon(\Gamma) \leq -\alpha. \tag{1.3}$$

   ii) Para todo $x \in \mathcal{D}_I$ *e qualquer conjunto mensurável* $\Gamma \in \mathcal{B}$*, com* $x \in \Gamma^o$*, vale*

$$-I(x) \leq \liminf_{\epsilon \to 0} \epsilon \ln \mu_\epsilon(\Gamma). \tag{1.4}$$

3. *i) Para todo conjunto fechado* $F \subseteq \mathcal{X}$*, vale*

$$\limsup_{\epsilon \to 0} \epsilon \ln \mu_\epsilon(\Gamma) \leq -\inf_{x \in F} I(x). \tag{1.5}$$

   ii) Para todo conjunto aberto $G \subseteq \mathcal{X}$*, vale*

$$-\inf_{x \in G} I(x) \leq \liminf_{\epsilon \to 0} \epsilon \ln \mu_\epsilon(\Gamma). \tag{1.6}$$

Demonstração:(1 $\Rightarrow$ 2) Primeiro mostraremos a validade de 2.i). Por hipótese temos que para todo $\Gamma \in \mathcal{B}$, vale

$$\limsup_{\epsilon \to 0} \epsilon \ln \mu_\epsilon(\Gamma) \leq -\inf_{x \in \overline{\Gamma}} I(x). \tag{1.7}$$

Dado $\alpha < \infty$, suponhamos que $\Gamma \in \mathcal{B}$ e $\overline{\Gamma} \subset \Psi_I(\alpha)^c$, logo para todo $x \in \overline{\Gamma}$ vale que $\alpha < I(x)$, então $-I(x) < -\alpha$. Desse modo temos que

$$-\inf_{x \in \overline{\Gamma}} I(x) < -\alpha. \tag{1.8}$$

Aplicando (1.7) em (1.8), verificamos a validade de 2.i).

Agora mostraremos a validade de 2.ii). Por hipótese temos que

$$-\inf_{x \in \Gamma^o} I(x) \leq \liminf_{\epsilon \to 0} \epsilon \ln \mu_\epsilon(\Gamma). \tag{1.9}$$

Dado $x \in \mathcal{D}_I$, seja $\Gamma \in \mathcal{B}$, tal que $x \in \Gamma^o$. Então $I(x) < \infty$ e segue da definição de ínfimo que $\inf_{y \in \Gamma^o} I(y) \leq I(x)$. Aplicando (1.9), concluímos que

$$-I(x) \leq -\inf_{y \in \Gamma^o} I(y) \leq \liminf_{\epsilon \to 0} \epsilon \ln \mu_\epsilon(\Gamma)$$

com isso concluímos a validade de 2.ii). Portanto mostramos a validade de $1 \Rightarrow 2$.
(2 $\Rightarrow$ 3) Primeiro mostraremos a validade de 3.i). Com efeito, dado $F \subseteq \mathcal{X}$ fechado. Se $I(x) = \infty$, para todo $x \in F$, então dado $\alpha < \infty$, tem-se $F \subseteq \Psi_I(\alpha)^c$. Como estamos assumindo que $\mathcal{B}_\mathcal{X} \subseteq \mathcal{B}$, segue que $F \in \mathcal{B}$ e portanto de (1.3) vale que

$$\limsup_{\epsilon \to 0} \epsilon \ln \mu_\epsilon(F) \leq -\alpha.$$

Tomando $\alpha \to \infty$, temos a conclusão da demonstração de $3.i)$ para o caso em que $I(x) = +\infty$, para todo $x \in F$. Se existir $I(x) < +\infty$, tal que $x \in F$, definamos $\alpha_\lambda := \inf_{x\in F}(I(x) - \lambda) < +\infty$, com $\lambda > 0$. Temos então por (1.3) que

$$\limsup_{\epsilon\to 0} \epsilon \ln \mu_\epsilon(F) \leq -\alpha_\lambda$$

e como $\lambda > 0$ é arbitrário, segue-se que tomando $\lambda \to 0$ temos a conclusão da demonstração de $3.i)$ para o caso em que $I(x) < +\infty$, para algum $x \in F$. Assim concluímos a demonstração de $3.i)$.

Agora mostraremos a validade de $3.ii)$. Com efeito, seja $G \subseteq \mathcal{X}$ aberto, como $G \in \mathcal{B}_\mathcal{X}$, segue que $G \in \mathcal{B}$, e portanto é mensurável. Primeiramente notemos que se $I(x) = +\infty$, para todo $x \in G$, a desigualdade $3.ii)$ é trivial. Se $G \cap D_I \neq \emptyset$,então por (1.4), temos que dado $x \in G \cap D_I$ vale

$$-I(x) \leq \liminf_{\epsilon\to 0} \epsilon \ln \mu_\epsilon(G),$$

logo tem-se

$$-\inf_{x\in G} I(x) = \sup_{x\in G} -I(x) \leq \liminf_{\epsilon\to 0} \epsilon \ln \mu_\epsilon(G)$$

e com isso concluímos a validade de $3.ii)$ para o caso em que $G \cap D_I \neq \emptyset$. Com isso fechamos a demonstração de $2 \Rightarrow 3$.

$(3 \Rightarrow 1)$ Basta notarmos que dado Γ mensurável, vale que Γ^o é aberto e $\overline{\Gamma}$ é fechado, e com isso segue trivialmente a validade de 1. Com isso fechamos a demonstração de $3 \Rightarrow 1$. □

1.2 Teorema de Sanov para Alfabetos finitos

Chamaremos de **alfabeto finito** ou simplesmente **alfabeto**, o conjunto finito $\sum = \{a_1, ..., a_R\}$. Usaremos a notação $M_1(\sum)$ para representar o conjunto de todas as medidas de probabilidade possíveis no alfabeto $\sum$. Observemos que trivialmente podemos identificar $M_1(\sum)$ com o conjunto R-dimensional de vetores reais, tais que suas entradas são não-negativas e a soma dessas entradas é 1. Denotaremos por $\sum_\mu = \{a_i \in \sum; \mu(a_i) > 0\}$, o suporte de $\mu \in M_1(\sum)$. Nesta subseção iremos assumir que nossas variáveis aleatórias assumem valores em $\sum$, e com isso em mente nosso objetivo será encontrar uma função taxa para os subconjuntos de $\sum$. Denotaremos por $\mathbb{P}$ a medida de probabilidade na σ-álgebra de Ω, onde Ω é o espaço que nossas variáveis aleatórias estão atuando.

Definição 1.2.1. *Um* ***modelo***[1] L_n^y *de uma sequência finita* $y = (y_1, ..., y_n) \in \sum^n$ *é uma medida de probabilidade empírica induzida da seguinte forma:* $L_n^y = (L_n^y(a_1), ..., L_n^y(a_R))$*, é um elemento de* $M_1(\sum)$ *e*

$$L_n^y(a_i) = \frac{1}{n}\sum_{j=1}^{n} 1_{a_i}(y_j), \ \ se\ i = 1, ..., R. \tag{1.10}$$

[1] O termo Modelo utilizado em 1.2.1, é escrito em inglês como Type na referência (DEMBO; ZEITOUNE, 1998)

Notemos que na definição acima o termo $L_n^y(a_i)$ representa a média que a_i aparece na sequência finita $y \in \sum^n$.

Exemplo 1.2.2. *Consideremos* $\sum = \{1, 2, 3\}$. *Tomando* $y = (1, 1, 3, 3) \in \sum^4$, *segue que*

$$L_4^y = \left(\frac{1}{2}, 0, \frac{1}{2}\right).$$

Consideraremos a partir de agora $Y_1, ..., Y_n$ variáveis aleatórias que são independentes e identicamente distribuídas (i.i.d.), com distribuição $\mu \in M_1(\sum)$, ou seja dado $A \subseteq \sum$, então $\mathbb{P}[Y_1 \in A] = \mu(A)$. Denotaremos por $\mathcal{L}_n$ o conjunto de todos os possíveis modelos de sequências de tamanho n, ou seja, o conjunto $\mathcal{L}_n = \{\nu \in M_1(\sum); \nu = L_n^y$ para algum $y \in \sum^n\}$. Definindo $Y = (Y_1, ..., Y_n)$ um vetor aleatório, usaremos da notação L_n^Y, para denotar a medida empírica induzida pelo vetor aleatório Y.

Definição 1.2.3. *Dadas duas medidas de probabilidade* ν, $\nu' \in M_1(\sum)$. *Chamamos de* ***distância variacional***

$$d_V(\nu, \nu') := \sup_{A \subseteq \sum} |\nu(A) - \nu'(A)|. \tag{1.11}$$

Lema 1.2.4. *Sejam* ν, ν' *duas medidas de probabilidade em* $M_1(\sum)$. *Então vale a seguinte igualdade*

$$d_V(\nu, \nu') = \frac{1}{2}\sum_{i=1}^{R} |\nu(a_i) - \nu'(a_i)|. \tag{1.12}$$

Demonstração: Dados duas medidas de probabilidade ν, ν' em $M_1(\sum)$, definamos os seguintes conjuntos:

1. $A_+ := \{x \in \sum; \nu(x) - \nu'(x) \geq 0\}$
2. $A_- := \{x \in \sum; \nu'(x) - \nu(x) > 0\}$.

Observemos então que dado $B \subseteq \sum$, tem-se

$$|\nu(B) - \nu'(B)| = \left|\sum_{x \in B} \nu(x) - \nu'(x)\right| = \left|\sum_{x \in B_+} \nu(x) - \nu'(x) - \sum_{x \in B_-} \nu'(x) - \nu(x)\right|,$$

onde $B_+ := \{x \in B; \nu(x) - \nu'(x) \geq 0\}$ e $B_- := \{x \in B; \nu'(x) - \nu(x) > 0\}$. Vemos que vale a desigualdade

$$|\nu(B) - \nu'(B)| \leq \sup\{\nu(A_+) - \nu'(A_+), \nu'(A_-) - \nu(A_-)\},$$

logo

$$d_V(\nu, \nu') = \sup\{\nu(A_+) - \nu'(A_+), \nu'(A_-) - \nu(A_-)\}.$$

Notemos agora que

$$\nu'(A_-) - \nu(A_-) = \sum_{x \in A_-} \nu'(x) - \nu(x) = 1 - \sum_{x \in A_+} \nu'(x) - 1 + \sum_{x \in A_+} \nu(x) = \sum_{x \in A_+} \nu(x) - \nu'(x) = \nu(A_+) - \nu'(A_+).$$

Portanto concluímos que

$$d_V(\nu, \nu') = \nu'(A_-) - \nu(A_-) = \frac{1}{2} \sum_{i=1}^{R} |\nu(a_i) - \nu'(a_i)|.$$

□

Abaixo enunciaremos um resultado que nos mostrará uma cota superior para a cardinalidade de $\mathcal{L}_n$.

Lema 1.2.5. *Seja $\mathcal{L}_n$ o conjunto de todos os possíveis modelos de sequências de tamanho n, temos então que:*
a) Vale a seguinte desigualdade

$$|\mathcal{L}_n| \leq (n+1)^R. \tag{1.13}$$

b) Dado um vetor probabilidade $\nu \in M_1(\sum)$, vale que

$$d_V(\nu, \mathcal{L}_n) := \inf_{\nu' \in \mathcal{L}_n} d_V(\nu, \nu') \leq \frac{R}{2n}. \tag{1.14}$$

Demonstração: a) Notemos que dado $\nu \in \mathcal{L}_n$, tem-se $\nu = (b_1, ..., b_R)$, com $b_j \in \{0, \frac{1}{n}, ..., \frac{n}{n}\}$, para todo $j \in \{1, ..., R\}$. Logo para cada entrada de ν temos $n+1$ possibilidades. Como temos R entradas, segue que $|\mathcal{L}_n| \leq (n+1)^R$.

b) Seja $\nu \in M_1(\sum)$ arbitrário, mostraremos primeiramente que existe ν' em $\mathcal{L}_n$, tal que $|\nu(a_i) - \nu'(a_i)| < \frac{1}{n}$, para todo $i = 1, ..., R$. Com efeito, definamos p_i como sendo a probabilidade de a_i. Sabemos que existe $k_1 \in \{0, ..., n\}$, tal que

$$\frac{k_1}{n} \leq p_1 < \frac{k_1 + 1}{n}.$$

Agora suponhamos que $1 < i \leq R$ e que dado $1 \leq j \leq i-1$, existem $k_1, ..., k_j \in \{1, ..., n\}$, tais que

$$\frac{k_j}{n} \leq p_j + \sum_{l=0}^{j-1} (p_l - \frac{k_l}{n}) < \frac{k_j + 1}{n},$$

com $p_0 = k_0 = 0$. Notemos então que

$$0 \leq \sum_{l=0}^{j} (p_l - \frac{k_l}{n}) < \frac{1}{n},$$

para todo $1 \leq j \leq i-1$. Deste modo temos que o termo $p_i + \sum_{l=0}^{i-1} \left(p_l - \frac{k_l}{n}\right)$ esta no intervalo $\left[0, 1 + \frac{1}{n}\right)$. Assim existe $k_i \in \{0, ..., n\}$, tal que

$$\frac{k_i}{n} \leq p_i + \sum_{l=0}^{i-1} \left(p_l - \frac{k_l}{n}\right) < \frac{k_i + 1}{n}.$$

Portanto pelo segundo método de indução vemos que existem $k_1, ..., k_R$, tais que

$$\frac{k_i}{n} \leq p_i + \sum_{l=0}^{i-1}\left(p_l - \frac{k_l}{n}\right) < \frac{k_i+1}{n}, \tag{1.15}$$

para todo $i \in \{1, ..., R\}$. De (1.24), conseguimos as seguintes cadeias de desigualdades

$$-\frac{1}{n} = -\frac{k_i+1}{n} + \frac{k_i}{n} < -p_i - \sum_{l=0}^{i-1}\left(p_l - \frac{k_l}{n}\right) + \frac{k_i}{n} \leq -p_i + \frac{k_i}{n} \tag{1.16}$$

e

$$p_i - \frac{k_i}{n} \leq p_i + \sum_{l=0}^{i-1}\left(p_l - \frac{k_l}{n}\right) - \frac{k_i}{n} < \frac{k_i+1}{n} - \frac{k_i}{n} < \frac{1}{n}. \tag{1.17}$$

Por (1.16) e (1.17) vale

$$\left|p_i - \frac{k_i}{n}\right| < \frac{1}{n},$$

para todo $i \in \{1, ..., R\}$. Novamente de (1.24) temos

$$0 \leq \sum_{l=0}^{R}\left(p_l - \frac{k_l}{n}\right) < \frac{1}{n}.$$

Como

$$\sum_{l=0}^{R} p_l = 1 \text{ e } \sum_{l=0}^{R}\frac{k_l}{n} = \frac{t}{n},$$

com $t \in \mathbb{N}$, segue que

$$\sum_{l=0}^{R}\frac{k_l}{n} = 1.$$

Tomando $\nu' := (\frac{k_1}{n}, ..., \frac{k_R}{n})$, vemos que $\nu' \in \mathcal{L}_n$, e vale que $|\nu(a_i) - \nu'(a_i)| < \frac{1}{n}$. Segue então de (1.12) que

$$d_V(\nu, \mathcal{L}_n) \leq d_V(\nu, \nu') = \tfrac{1}{2}\textstyle\sum_{i=1}^{R} |\nu(a_i) - \nu'(a_i)| \leq \frac{R}{2n}.$$

□

Definição 1.2.6. *O conjunto* ***classe do modelo*** *da medida de probabilidade $\nu \in \mathcal{L}_n$, é definido como*

$$T_n(\nu) := \{y \in \sum^n; L_n^y = \nu\}.$$

Observemos que o conjunto classe do modelo de uma medida $\nu \in \mathcal{L}_n$ é o conjunto de todos os vetores em $\sum^n$ que apresentam a mesma frequência para cada $a_i \in \sum$, ou seja são todas as permutações possíveis.

Exemplo 1.2.7. *Consideremos $\sum$ o mesmo do exemplo 1.2.2 e tome $\nu := \left(\frac{1}{2}, 0, \frac{1}{2}\right)$. Temos então que*

$$T_4(\nu) = \{(1,1,3,3), (1,3,1,3), (1,3,3,1), (3,1,3,1), (3,1,1,3), (3,3,1,1)\}.$$

Observação: A partir de agora convencionaremos que $0 \ln 0 := 0$ e $0 \ln(0/0) := 0$.

Definição 1.2.8. *a) A **entropia** de um vetor probabilidade $\nu \in M_1(\sum)$ é*

$$H(\nu) := -\sum_{i=1}^{R} \nu(a_i) \ln \nu(a_i). \tag{1.18}$$

*b) A **entropia relativa** de um vetor probabilidade ν com respeito a outro vetor de probabilidade μ é*

$$H(\nu|\mu) := \sum_{i=1}^{R} \nu(a_i) \ln \frac{\nu(a_i)}{\mu(a_i)}. \tag{1.19}$$

Exemplo 1.2.9. *Consideremos os vetores de probabilidade $\nu = \left(\frac{1}{2}, 0, \frac{1}{2}\right)$ e $\mu = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. Deste modo temos que a entropia de ν é*

$$H(\nu) = -\left(\frac{1}{2} \ln \frac{1}{2} + 0 \ln 0 + \frac{1}{2} \ln \frac{1}{2}\right) = \ln 2.$$

A entropia de ν com respeito a μ é

$$H(\nu|\mu) = \left(\frac{1}{2} \ln \frac{\frac{1}{2}}{\frac{1}{3}} + 0 \ln \frac{0}{\frac{1}{3}} + \frac{1}{2} \ln \frac{\frac{1}{2}}{\frac{1}{3}}\right) = \ln 3 - \ln 2.$$

Consideremos agora a seguinte função

$$H(\cdot|\mu) : M_1\left(\sum\right) \longrightarrow \mathbb{R} \cup \{+\infty\}$$

definida por

$$H(\nu|\mu) := \sum_{i=1}^{R} \nu(a_i) \ln \frac{\nu(a_i)}{\mu(a_i)}.$$

Lema 1.2.10. *Sejam $\mu \in M_1(\sum)$ e $H(\cdot|\mu) : M_1(\sum) \longrightarrow \mathbb{R} \cup \{+\infty\}$. Então:*
a) $H(\cdot|\mu)$ esta bem definida.
b) A função $H(\cdot|\mu)$ é não negativa.
c) O conjunto $\{\nu \in M_1(\sum); \sum_\nu \subseteq \sum_\mu\}$ é compacto e $H(\cdot|\mu)$ é finita e continua em $\{\nu \in M_1(\sum); \sum_\nu \subseteq \sum_\mu\}$.
d) A função $H(*|\mu) = \infty$ fora do conjunto $\{\nu \in M_1(\sum); \sum_\nu \subseteq \sum_\mu\}$ e $H(*|\mu)$ é um função taxa bem comportada.

Demonstração: a) Seja $\nu \in M_1(\sum)$. Se $\sum_\nu \subset \sum_\mu$, então $H(\nu|\mu)$ é uma somatória de termos finitos, logo $H(\nu|\mu) \in \mathbb{R}$. Se $\sum_\nu \not\subset \sum_\mu$, definimos o conjunto $A := \{a_i \in \sum; \nu(a_i) \neq 0, \mu(a_i) = 0\}$, pelo fato de $\sum_\nu \not\subset \sum_\mu$ temos que $A \neq \emptyset$. Assim vemos que vale a seguinte igualdade

$$H(\nu|\mu) = \sum_{a_i \in A} \nu(a_i) \ln \frac{\nu(a_i)}{0} + \sum_{a_i \notin A} \nu(a_i) \ln \frac{\nu(a_i)}{\mu(a_i)} = +\infty.$$

Dados $\nu, \nu' \in M_1(\sum)$, tais que $\nu = \nu'$. Deste modo temos que

$$H(\nu|\mu) = \sum_{i=1}^{R} \nu(a_i) \ln \frac{\nu(a_i)}{\mu(a_i)} = H(\nu|\mu) = \sum_{i=1}^{R} \nu'(a_i) \ln \frac{\nu'(a_i)}{\mu(a_i)} = H(\nu'|\mu).$$

Portanto concluímos que $H(*|\mu)$ esta bem definida.
b) Primeiramente notemos que tomando $f(x) = x\ln(x)$ para todo $x > 0$ e $f(0) = 0$, temos que $f'(x) = 1 + \ln x$, para $x \neq 0$. Como a derivada de f é crescente em $(0, +\infty)$, segue que f é uma função convexa em $(0, +\infty)$. Agora pela regra de L'Hôpital tem-se

$$\lim_{x\to 0} x \ln\left(\frac{x}{b}\right) = 0,$$

para $b > 0$, logo f é continua e disso podemos concluir que f é convexa em $[0, +\infty)$.

Notemos que se $\sum_\nu \not\subseteq \sum_\mu$, segue de a) que $H(\nu|\mu) = \infty$. Agora se $\sum_\nu \subseteq \sum_\mu$, então pela convenção que fizemos segue que

$$H(\nu|\mu) = \sum_{l=1}^{n} \nu(a_{i_l}) \ln \frac{\nu(a_{i_l})}{\mu(a_{i_l})},$$

onde $\sum_\mu = \{a_{i_1}, ..., a_{i_n}\}$. Pela convexidade de f, vale que

$$0 = f\left(\sum_{l=1}^{n} \mu(a_{i_l}) \frac{\nu(a_{i_l})}{\mu(a_{i_l})}\right) \leq \sum_{l=1}^{n} \mu(a_{i_l}) f\left(\frac{\nu(a_{i_l})}{\mu(a_{i_l})}\right) = \sum_{l=1}^{n} \nu(a_{i_l}) \ln \frac{\nu(a_{i_l})}{\mu(a_{i_l})} = H(\nu|\mu).$$

Portanto $H(*|\mu)$ é não negativa.
c) Primeiramente mostraremos que $J := \{\nu \in M_1(\sum); \sum_\nu \subseteq \sum_\mu\}$ é compacto. Mostraremos primeiramente que J é limitado. Com efeito, fixemos $\nu_0 \in M_1(\sum)$ arbitrário, deste modo observamos que dado $\mu \in M_1(\sum)$ tem-se

$$d_V(\nu_0, \mu) \leq \frac{R}{2},$$

onde a desigualdade acima segue diretamente de (1.12). Então vemos que $M_1(\sum) \subset B_{\frac{R}{2}} = \left\{\nu \in M_1(\sum); d_V(\nu_0, \nu) \leq \frac{R}{2}\right\}$, em particular J é limitado.

Mostraremos agora que J é fechado. Com efeito, seja $\nu_j \in J$, uma sequência convergente e ν o limite dessa sequência. Consideremos $\{a_{i_1}, ..., a_{i_k}\}$ o conjunto de todos os elementos que tem probabilidade 0 com relação a μ. Temos então por (1.12) que

$$\nu(a_{i_s}) \leq \sum_{i=1}^{R} |\nu(a_i) - \nu_j(a_i)| = 2d_V(\nu, \nu_j) \longrightarrow 0, \text{ quando } j \to +\infty$$

para $s = 1, ..., k$, logo $\nu(a_{i_s}) = 0$ e disso concluímos que $\sum_\nu \subseteq \sum_\mu$. Portanto J é fechado e com isso concluímos que o conjunto é compacto.

Mostraremos agora que $H(*|\mu)$ é continua em $\{\nu \in M_1(\sum); \sum_\nu \subseteq \sum_\mu\}$. Com efeito, seja $\sum_\mu = \{a_{i_1}, ..., a_{i_t}\}$. Dado $\nu \in M_1(\sum)$, tal que $\sum_\nu \subseteq \sum_\mu$. Se $\nu(a_{i_j}) > 0$, com $j \in \{1, ..., t\}$, então para todo $\epsilon > 0$ dado, existe δ_j, tal que $|x - \nu(a_{i_j})| < \delta_j$ implica que

$$\left| x \ln\left(\frac{x}{\mu(a_{i_j})}\right) - \nu(a_{i_j}) \ln\left(\frac{\nu(a_{i_j})}{\mu(a_{i_j})}\right)\right| < \frac{\epsilon}{R}.$$

Como $\lim_{x\to 0} x\ln(\frac{x}{b}) = 0$, para $b > 0$, segue que a desigualdade acima vale também para o caso em que $\nu(a_{i_j}) = 0$. Tomando $\delta = \min_{1\leq j\leq t}\left\{\frac{\delta_j}{2}\right\}$, temos que se $\nu' \in J$ e $d_V(\nu', \nu) < \delta$, então

$$|\nu'(a_{i_j}) - \nu(a_{i_j})| \leq \sum_{i=1}^{R} |\nu(a_i) - \nu'(a_i)| < 2\delta \leq \delta_j,$$

para todo $j = 1, ..., t$. Temos então que

$$\left|\sum_{i=1}^{R}\left(\nu'(a_i)\ln\frac{\nu'(a_i)}{\mu(a_i)} - \nu'(a_i)\ln\frac{\nu(a_i)}{\mu(a_i)}\right)\right| < \epsilon.$$

Portanto $H(*|\mu)$ é continua em $\{\nu \in M_1(\sum); \sum_\nu \subseteq \sum_\mu\}$.

A finitude de $H(\cdot|\mu)$ em J é imediata.

d) Trivialmente vê-se que $H(\cdot|\mu) = \infty$ fora do conjunto J. E $H(*|\mu)$ ser uma função taxa bem comportada segue dos fatos de $H(*|\mu)$ ser continua em J, finita somente em J e J ser limitado, pois assim o conjunto de nível $\Psi_I(\alpha)$ é compacto. Assim pela definição 1.1.1, segue que $H(\cdot|\mu)$ é uma função taxa bem comportada. □

Sejam $\{Y_j\}_{j\in\mathbb{N}}$ uma sequência infinita de variáveis aleatórias independente e identicamente distribuídas (i.i.d.) de acordo com $\mu \in M_1(\sum)$. Usaremos a notação $\mathbb{P}_\mu$ para representar probabilidade que sequência infinita $\{Y_j\}_{j\in\mathbb{N}}$.

Lema 1.2.11. *Sejam $y \in T_n(\nu)$, tal que $\nu \in \mathcal{L}_n$, e $\{Y_j\}_{j\in\mathbb{N}}$ uma sequência infinita de variáveis aleatórias independente e identicamente distribuídas de acordo com $\mu \in M_1(\sum)$. Então*

$$\mathbb{P}_\mu((Y_1, ..., Y_n) = y) = e^{-n[H(\nu)+H(\nu|\mu)]} \tag{1.20}$$

Demonstração: Denotando $\mathbb{P}$ como sendo a probabilidade do espaço onde as Y_j atuam, temos

$$\mathbb{P}_\mu((Y_1, ..., Y_n) = y) = \mathbb{P}((Y_1, ..., Y_n) = y).$$

Pelo fato das Y_i serem i.i.d., segue que

$$\mathbb{P}_\mu((Y_1, ..., Y_n) = y) = \mathbb{P}_\mu(Y_1 = y_1, ..., Y_n = y_n) = \prod_{j=1}^{n} \mathbb{P}(Y_j = y_j) = \prod_{j=1}^{n} \mu(y_j). \tag{1.21}$$

Se $\sum_\nu \not\subset \sum_\mu$, então existe $a_i \in \sum$ tal que $\nu(a_i) \neq 0$ e $\mu(a_i) = 0$. Como $y \in T_n(\nu)$ podemos concluir que a_i aparece em alguma entrada do vetor y. Logo existe $j_0 \in \{1, ..., n\}$, tal que $y_{j_0} = a_i$. Por esse fato e de (1.21), tem-se

$$\mathbb{P}_\mu((Y_1, ..., Y_n) = y) = \mu(y_1)...\mu(a_i)...\mu(y_n) = 0.$$

Pelo Lema 1.2.10, item d), temos que $H(\nu|\mu) = \infty$. Assim vale

$$e^{-n[H(\nu)+H(\nu|\mu)]} = 0,$$

com isso verificamos a validade da igualdade para o caso em que $\sum_\nu \not\subset \sum_\mu$. Suponhamos agora $\sum_\nu \subseteq \sum_\mu$, segue de 1.2.10, item c), que $H(\nu|\mu) < \infty$. Agora observemos que como $y \in T_n(\nu)$ e $\nu \in \mathcal{L}_n$, segue que cada elemento a_i de $\sum$, aparece $n\nu(a_i)$ vezes na sequência finita y e $1 = 0^0$, pois $0 \ln 0 = 0$. Segue de de (1.21) que

$$\mathbb{P}_\mu((Y_1, ..., Y_n) = y) = \prod_{i=1}^{R} \mu(a_i)^{n\nu(a_i)}. \tag{1.22}$$

Agora observemos que

$$\begin{gathered}\textstyle\sum_{i=1}^R \nu(a_i) \ln \mu(a_i) = \sum_{i=1}^R ((\nu(a_i) \ln \mu(a_i) + \nu(a_i) \ln \nu(a_i) - \nu(a_i) \ln \nu(a_i)) = \\ \textstyle\sum_{i=1}^R \nu(a_i) \ln \nu(a_i) + \sum_{i=1}^R (\nu(a_i) \ln \mu(a_i) - \nu(a_i) \ln \nu(a_i)) = -[H(\nu) + H(\nu|\mu)],\end{gathered}$$

logo temos

$$\mathbb{P}_\mu((Y_1, ..., Y_n) = y) = \exp\left\{n \textstyle\sum_{i=1}^R \nu(a_i) \ln \mu(a_i)\right\} = e^{-n[H(\nu)+H(\nu|\mu)]}.$$

Com esta ultima cadeia igualdades finalizamos a demonstração do resultado. □

Lema 1.2.12. *Para todo $\nu \in \mathcal{L}_n$, tem-se*

$$(n+1)^{-R} e^{nH(\nu)} \leq |T_n(\nu)| \leq e^{nH(\nu)} \tag{1.23}$$

Demonstração: Notemos primeiramente que $|T_n(\nu)| = n!/((n\nu(a_1))!...(n\nu(a_R))!)$. Consideremos agora a sequência $\{\widehat{\nu}_i\}_{i=1}^{+\infty}$, onde $\widehat{\nu}_i = \nu$ para todo $i \in \mathbb{N}$, então por resultado (ver apêndice B, Teorema B.2.1) existe uma sequência $\{Y_i\}_{i=1}^{+\infty}$ que é i.i.d., tal que Y_1 tem distribuição ν.

Dado $y \in T_n(\nu)$, temos pelo Lema 1.2.11 que

$$\mathbb{P}_\nu((Y_1, ..., Y_n) = y) = e^{-n[H(\nu)+H(\nu|\nu)]}.$$

Observando que $T_n(\nu) = \dot{\bigcup}_{y \in T_n(\nu)} \{y\}$, temos

$$\mathbb{P}_\nu((Y_1, ..., Y_n) \in T_n(\nu)) = \sum_{y \in T_n(\nu)} \mathbb{P}_\nu((Y_1, ..., Y_n) = y). \tag{1.24}$$

logo
$$\mathbb{P}_\nu((Y_1,...,Y_n)\in T_n(\nu)) = |T_n(\nu)|e^{-n[H(\nu)+H(\nu|\nu)]} = |T_n(\nu)|e^{-nH(\nu)}, \tag{1.25}$$
onde $|X|$ é cardinalidade do conjunto X, e disso podemos concluir que
$$|T_n(\nu)| \leq e^{nH(\nu)}.$$

Agora iremos demonstrar o limite inferior. Com efeito, suponhamos $\nu' \in \mathcal{L}_n$, tal que $\sum_{\nu'} \subseteq \sum_\nu$. Notemos que dado $\mu \in \mathcal{L}_n$, temos $\{L_n^Y = \mu\} = \{(Y_1,...,Y_n) \in T_n(\mu)\}$, logo
$$\frac{\mathbb{P}_\nu(L_n^Y = \nu)}{\mathbb{P}_\nu(L_n^Y = \nu')} = \frac{\mathbb{P}_\nu((Y_1,...,Y_n)\in T_n(\nu))}{\mathbb{P}_\nu((Y_1,...,Y_n)\in T_n(\nu'))},$$
observemos que a igualdade da equação (1.20) independe do y escolhido, logo fixando $y_0 \in T_n(\nu)$ e $y_0' \in T_n(\nu')$, segue diretamente de (1.24) que
$$\frac{\mathbb{P}_\nu(L_n^Y = \nu)}{\mathbb{P}_\nu(L_n^Y = \nu')} = \frac{|T(\nu)|\mathbb{P}_\nu((Y_1,...,Y_n) = y_0)}{|T(\nu')|P_\nu((Y_1,...,Y_n) = y_0')}.$$
Segue da (1.22) que
$$\frac{\mathbb{P}_\nu(L_n^Y = \nu)}{\mathbb{P}_\nu(L_n^Y = \nu')} = \frac{|T(\nu)|\prod_{i=1}^R \nu(a_i)^{n\nu(a_i)}}{|T(\nu')|\prod_{i=1}^R \nu(a_i)^{n\nu'(a_i)}}.$$
Como $|T(\nu)| = n!/((n\nu(a_1))!...\ (n\nu(a_R))!)$, segue que
$$\frac{\mathbb{P}_\nu(L_n^Y = \nu)}{\mathbb{P}_\nu(L_n^Y = \nu')} = \prod_{i=1}^R \frac{(n\nu'(a_i))!}{(n\nu(a_i))!}\nu(a_i)^{n(\nu(a_i)-\nu'(a_i))}.$$

Mostraremos agora que dados $m, l \in \mathbb{N}$, vale que $\frac{m!}{l!} \geq l^{(m-l)}$. Com efeito, se $m > l$, tem-se
$$\frac{m!}{l!} = \frac{m...(l+1)l!}{l!} = m...(l+1) \geq l^{(m-l)}.$$
Se $m = l$, é imediato a desigualdade. Se $l > m$, temos que
$$\frac{m!}{l!} = \frac{m!}{l...(m+1)m!} = \frac{1}{l...(m+1)} \geq \frac{1}{l^{l-m}} = l^{m-l}$$
Com isso concluímos a afirmação.

Da afirmação acima segue que
$$\frac{\mathbb{P}_\nu(L_n^Y = \nu)}{\mathbb{P}_\nu(L_n^Y = \nu')} \geq \prod_{i=1}^R (n\nu(a_i))^{n(\nu'(a_i)-\nu(a_i))}\nu(a_i)^{n(\nu(a_i)-\nu'(a_i))}$$
$$= \prod_{i=1}^R n^{n(\nu(a_i)-\nu'(a_i))} = n^{n\sum_{i=1}^R \nu(a_i)-\nu'(a_i)} = 1.$$
Assim
$$\mathbb{P}_\nu(L_n^Y = \nu) \geq \mathbb{P}_\nu(L_n^Y = \nu'),$$

quando $\sum_{\nu'} \subseteq \sum_{\nu}$. Pela demonstração de 1.2.11, segue que $\sum_{\nu'} \not\subseteq \sum_{\nu}$ implica em $\mathbb{P}_{\nu}(L_n^Y = \nu') = 0$. Portanto podemos concluir que

$$\mathbb{P}_{\nu}(L_n^Y = \nu) \geq \mathbb{P}_{\nu}(L_n^Y = \nu'),$$

para todo $\nu' \in \mathcal{L}_n$.

Observando que $\sum^n = \dot{\bigcup}_{\mu \in \mathcal{L}_n} T_n(\mu)$, para $\mu \in \mathcal{L}_n$ arbitrário, temos

$$1 = \mathbb{P}_{\nu}\left(\sum^n\right) = \sum_{\nu' \in \mathcal{L}_n} \mathbb{P}_{\nu}((Y_1, ..., Y_n) \in T_n(\nu')) = \sum_{\nu' \in \mathcal{L}_n} \mathbb{P}_{\nu}(L_n^Y = \nu') \leq |\mathcal{L}_n| \mathbb{P}_{\nu}(L_n^Y = \nu),$$

segue então de (1.13), que

$$1 \leq (n+1)^R \mathbb{P}_{\nu}(L_n^Y = \nu) = (n+1)^R |T(\nu)| e^{-nH(\nu)},$$

e disso conseguimos que

$$(n+1)^{-R} e^{nH(\nu)} \leq |T_n(\nu)|.$$

Com esta última desigualdade concluímos a demonstração do lema. □

Observação: Em particular de (1.25), temos que se (Y_n) é uma sequência de variáveis aleatórias i.i.d, com distribuição $\mu \in M_1(\sum)$, então dado $\nu \in \mathcal{L}_n$, vale

$$\mathbb{P}_{\mu}((Y_1, ..., Y_n) \in T_n(\nu)) = |T_n(\nu)| e^{-n[H(\nu) + H(\nu|\mu)]}. \tag{1.26}$$

Lema 1.2.13. *Para todo $\nu \in \mathcal{L}_n$ e $\mu \in M_1(\sum)$, vale que*

$$(n+1)^{-R} e^{-nH(\nu|\mu)} \leq \mathbb{P}_{\mu}(L_n^Y = \nu) \leq e^{-nH(\nu|\mu)} \tag{1.27}$$

Demonstração: Pelo Lema 1.2.12 segue que dado $\tilde{\nu} \in \mathcal{L}_v$, vale

$$(n+1)^{-R} e^{nH(\nu)} \leq |T_n(\nu)| \leq e^{nH(\nu)}.$$

Por (1.26), tem-se

$$(n+1)^{-R} e^{nH(\nu)} \leq \mathbb{P}_{\mu}((Y_1, ..., Y_n) \in T_n(\nu)) e^{n[H(\nu) + H(\nu|\mu)]} \leq e^{nH(\nu)}.$$

Como $\mathbb{P}_{\mu}((Y_1, ..., Y_n) \in T_n(\nu)) = \mathbb{P}_{\mu}(L_n^Y = \nu)$, segue a demonstração do lema. □

Partiremos agora para o nosso principal resultado desta subseção, o Teorema de Sanov, no qual encontraremos a função taxa para alfabetos finitos.

Teorema 1.2.14 (Teorema de Sanov). *Dado $\{Y_j\}_{j \in \mathbb{N}}$ uma sequência infinita de variáveis i.i.d. distribuídas de acordo com $\mu \in M_1(\sum)$. Então para todo conjunto $\Gamma \subseteq M_1(\sum)$, vale que:*

1. *O limite inferior*

$$-\inf_{\nu\in\Gamma^o} H(\nu|\mu) \leq \liminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}_\mu(L_n^Y \in \Gamma). \tag{1.28}$$

2. *O limite superior*

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}_\mu(L_n^Y \in \Gamma) \leq -\inf_{\nu\in\Gamma} H(\nu|\mu). \tag{1.29}$$

Demonstração: Pelo Lema 1.2.13, segue que dado $\widetilde{\nu} \in \mathcal{L}_n$, vale

$$(n+1)^{-R} e^{-nH(\nu|\mu)} \leq \mathbb{P}_\mu(L_n^Y = \nu) \leq e^{-nH(\nu|\mu)}.$$

Observemos também que $L_n^Y \in \Gamma$ se e somente se existe $\nu \in \Gamma$, tal que $L_n^Y = \nu$. Logo $\nu \in \mathcal{L}_n \cap \Gamma$.

Como $\mathcal{L}_n \cap \Gamma$ é finito, segue que

$$\mathbb{P}_\mu(L_n^Y \in \Gamma) = \sum_{\nu\in\Gamma\cap\mathcal{L}_n} \mathbb{P}_\mu(L_n^Y = \nu) \leq \sum_{\nu\in\Gamma\cap\mathcal{L}_n} e^{-nH(\nu|\mu)} \leq |\Gamma\cap\mathcal{L}_n| e^{-n\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)},$$

então por (1.13) temos

$$\mathbb{P}_\mu(L_n^Y \in \Gamma) \leq (n+1)^R e^{-n\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)}. \tag{1.30}$$

Segue também do fato de $\mathcal{L}_n \cap \Gamma$ ser finito que

$$\mathbb{P}_\mu(L_n^Y \in \Gamma) = \sum_{\nu\in\Gamma\cap\mathcal{L}_n} \mathbb{P}_\mu(L_n^Y = \nu) \geq \sum_{\nu\in\Gamma\cap\mathcal{L}_n} (n+1)^{-R} e^{-nH(\nu|\mu)} \geq (n+1)^{-R} e^{-n\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)},$$

e como existe $\nu_0 \in \mathcal{L}_n \cap \Gamma$, tal que $H(\nu_0|\mu) = \inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)$, temos

$$\mathbb{P}_\mu(L_n^Y \in \Gamma) \geq (n+1)^{-R} e^{-n\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)}. \tag{1.31}$$

Por (1.30) e (1.31) temos que

$$\frac{\ln(n+1)^{-R}}{n} - \inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu) \leq \frac{1}{n} \ln \mathbb{P}_\mu(L_n^Y \in \Gamma) \leq \frac{\ln(n+1)^R}{n} - \inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu).$$

Observemos que $\left|R\frac{\ln(n+1)}{n}\right| \to 0$, quando $n \to \infty$. Como a desigualdade acima vale para n arbitrário, temos

$$\limsup_{n\to\infty} \left\{-\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)\right\} \leq \limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}_\mu(L_n^Y \in \Gamma) \leq \limsup_{n\to\infty} \left\{-\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)\right\}.$$

Disto podemos concluir que

$$-\liminf_{n\to\infty} \left\{\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)\right\} = \limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}_\mu(L_n^Y \in \Gamma) \tag{1.32}$$

e de maneira análoga concluímos que

$$-\limsup_{n\to\infty} \left\{\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)\right\} = \liminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}_\mu(L_n^Y \in \Gamma). \tag{1.33}$$

Como $\nu \in \Gamma \cap \mathcal{L}_n \subseteq \Gamma$, segue que $\inf_{\nu\in\Gamma} H(\nu|\mu) \leq \inf_{n\leq m}\{\inf_{\nu\in\Gamma\cap\mathcal{L}_m} H(\nu|\mu)\}$, para todo $n \in \mathbb{N}$. Temos então por (1.32) que

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}_\mu(L_n^Y \in \Gamma) \leq -\inf_{\nu\in\Gamma} H(\nu|\mu)$$

e com isso terminamos a demonstração do limite superior.

Agora iremos demonstra a validade de (1.28). Com efeito, se $\Gamma^o = \emptyset$, então o resultado é imediato. Se $\Gamma^o \neq \emptyset$, então dado $\nu \in \Gamma^o$, temos que $\sum_\nu \nsubseteq \sum_\mu$ ou $\sum_\nu \subseteq \sum_\mu$. Se $\sum_\nu \nsubseteq \sum_\mu$ temos pelo Lema 1.2.10, item d), que $H(\nu|\mu) = \infty$ e disto temos

$$-\limsup_{n\to\infty}\left\{\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)\right\} \geq -H(\nu|\mu). \tag{1.34}$$

Agora se ν é tal que $\sum_\nu \subseteq \sum_\mu$, mostraremos que existe uma sequência $(\nu_n)_{n\in\mathbb{N}}$ em Γ, tal que $\nu_n \to \nu$, e para todo $n \in \mathbb{N}$ valem que $\nu_n \in \Gamma \cap \mathcal{L}_n$, $\sum_{\nu_n} \subseteq \sum_\mu$. Com efeito, sejam $p_1, ..., p_R$, os elementos de ν, ou seja $\nu = (p_1, ..., p_R)$. Escrevendo $\sum_\nu = \{a_{i_1}, ..., a_{i_j}\}$, vemos que tomando $\bar{\nu} := (p_{i_1}, ..., p_{i_j})$, vale que $\bar{\nu} \in M_1\left(\sum_\nu\right)$, segue então da demonstração do lema 1.2.5 b), que para cada $n \in \mathbb{N}$, existe um $\bar{\nu}_n \in \mathcal{L}_n\left(\sum_\nu\right) = \{\nu \in M_1(\sum_\nu); \nu = L_n^y \text{ para algum } y \in \left(\sum_\nu\right)^n\}$, tal que $d_V(\bar{\nu}, \bar{\nu_n}) \leq \frac{R}{2n}$. Escrevendo $\bar{\nu}_n = (a_{n,i_1}, ..., a_{n,i_j})$, definiremos a medida $\nu_n = (b_{n,1}, ..., b_{n,R})$, como sendo $b_{n,l} = a_{n,i_k}$, se $l = i_k$, com $k = 1, ..., j$ e 0 caso contrario. Segue então de (1.12) que $d_{V'}(\bar{\nu}, \bar{\nu}_n) = d_V(\nu, \nu_n)$, onde

$$d_{V'}(\bar{\mu}, \bar{\mu}') := \sup_{A\subseteq\sum_\nu} |\bar{\mu}(A) - \bar{\mu}'(A)|$$

e em particular $\nu_n \to \nu$ e trivialmente vemos que $\nu_n \in \mathcal{L}_n$. Como $\nu \in \Gamma^o$, então para n suficientemente grande $\nu_n \in \Gamma \cap \mathcal{L}_n$. Observemos que pela construção dos ν_n, vale que $\sum_{\nu_n} \subseteq \sum_\nu$ para todo $n \in \mathbb{N}$, em particular $\sum_{\nu_n} \subseteq \sum_\mu$ para todo $n \in \mathbb{N}$. Disso segue que

$$\begin{gathered}-\limsup_{n\to\infty}\{\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)\} = \liminf_{n\to\infty}\left\{\sup_{\nu\in\Gamma\cap\mathcal{L}_n} -H(\nu|\mu)\right\} \geq \\ \liminf_{n\to\infty} -H(\nu_n|\mu) = \lim_{n\to\infty} -H(\nu_n|\mu) = -H(\nu|\mu),\end{gathered}$$

onde a ultima igualdade segue diretamente do Lema 1.2.10, item c). Temos então por (1.34) e pela desigualdade acima que

$$-\limsup_{n\to\infty}\left\{\inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu)\right\} \geq \sup_{\nu\in\Gamma^o} -H(\nu|\mu).$$

Segue então de (1.33) que

$$\liminf_{n\to\infty} \tfrac{1}{n} \ln \mathbb{P}_\mu(L_n^Y \in \Gamma) \geq \sup_{\nu\in\Gamma^o} -H(\nu|\mu) = -\inf_{\nu\in\Gamma^o} H(\nu|\mu).$$

E com isso concluímos a demonstração do limite inferior. $\square$

Exemplo 1.2.15. *Considere $|\sum| = 2$. Temos então que se $\bar{\mu} \in M_1(\sum)$, podemos escrever $\bar{\mu} = (x, 1-x)$, $x \in [0,1]$. Tome $\mu = \left(\frac{1}{2}, \frac{1}{2}\right)$, temos então que dado $\nu = (x, 1-x)$, a entropia relativa de ν com relação a μ é*

$$H(\nu|\mu) = -x \ln 2x - (1-x)\ln(2-2x). \tag{1.35}$$

Corolário 1.2.16. *Se Γ é um conjunto aberto vale que*

$$\lim_{n\to\infty} \frac{1}{n} \ln P_\mu(L_n^Y \in \Gamma) = -\liminf_{n\to\infty} \left\{ \inf_{\nu\in\Gamma\cap\mathcal{L}_n} H(\nu|\mu) \right\} = -\inf_{\nu\in\Gamma} H(\nu|\mu). \tag{1.36}$$

Demonstração: Basta observar que pelo teorema de Sanov vale seguinte cadeia de desigualdades

$$-\inf_{\nu\in\Gamma^o} H(\nu|\mu) \le \liminf_{n\to\infty} \frac{1}{n} \ln P_\mu(L_n^Y \in \Gamma) \le \limsup_{n\to\infty} \frac{1}{n} \ln P_\mu(L_n^Y \in \Gamma) \le -\inf_{\nu\in\Gamma} H(\nu|\mu)$$

e pelas igualdades (1.32) e (1.33), concluímos o resultado □

Exemplo 1.2.17. *Consideremos $\sum$ e μ os mesmos do exemplo 1.2.15. Tomando $\Gamma = \left\{\nu \in M_1(\sum); \nu(x, 1-x), x \in \left(\frac{1}{3}, \frac{1}{2}\right)\right\}$. Dado $\nu \in \Gamma$, tal que $\nu = (x, 1-x)$, tomemos $\delta := \min\left\{x - \frac{1}{3}, \frac{1}{2} - x\right\}$. Notemos que se $\bar{\nu} = (\bar{x}, 1-\bar{x})$, com $\bar{x} \in \left(\frac{1}{3}, \frac{1}{2}\right)$, segue então de (1.12) que*

$$d_V(\nu, \bar{\nu}) < \delta$$

e vemos facilmente que $\bar{\nu} \in \Gamma$. Logo Γ é aberto. E por meio de uma derivação simples vemos que $f(x) = -x\ln 2x - (1-x)\ln(2-2x)$, é crescente em $\left(\frac{1}{3}, \frac{1}{2}\right)$ e portanto pelo corolário acima segue que

$$\lim_{n\to\infty} \frac{1}{n} \ln P_\mu(L_n^Y \in \Gamma) = \frac{1}{3} \ln \frac{2}{3} + \frac{2}{3} \ln \frac{4}{3}.$$

1.3 Teorema de Cramér para $\mathbb{R}$

Nesta subseção iremos considerar a $\sigma-$álgebra de Lebesgue em $\mathbb{R}$. Sejam $X_1, ..., X_n, ...$, variáveis aleatórias i.i.d que assumem valores em $\mathbb{R}$, onde X_1 tem distribuição de acordo com a probabilidade $\mu \in M_1(\mathbb{R})$. Definimos $\widehat{S}_n := \frac{1}{n}\sum_{j=1}^n X_j$. Chamaremos de **logaritmo da função geradora de momento** associado a μ a função

$$\Lambda(\lambda) := \ln M(\lambda) := \ln \mathbb{E}[e^{\lambda X_1}]. \tag{1.37}$$

Observemos que $e^{\lambda X_1}$ é uma variável aleatória não negativa, logo vale que $e^{\lambda X_1} = 0$ q.t.p., somente quando $X_1 = -\infty$ q.t.p. e disto concluímos que $\mathbb{E}[e^{\lambda X_1}] > 0$, portanto $\ln \mathbb{E}[e^{\lambda X_1}] > -\infty$ para todo $\lambda \in \mathbb{R}$. Lembrando que para cada $n \in \mathbb{N}$ podemos induzir uma medida de probabilidade μ_n em $\mathbb{R}$, por meio da variável aleatória $\widehat{S}_n$, da seguinte forma

$$\mu_n = \mathbb{P}\left[\widehat{S}_n \in A\right],$$

onde $\mathbb{P}$ é a probabilidade que age na $\sigma-$álgebra do espaço amostral onde X_i atua. Denotaremos a esperança de X_1, quando ela existir, por $\overline{x} := \mathbb{E}[X_1]$.

Definição 1.3.1. *A* ***Transformação de Fenchel-Legendre*** *de* $\Lambda(\lambda)$ *é definida como*

$$\Lambda^*(x) := \sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda(\lambda)\} \tag{1.38}$$

Iremos agora calcular duas transformações de Fenchel-Legendre.

Exemplo 1.3.2. *Consideremos* X_1 *com distribuição Poisson*(θ)*. Temos primeiramente que* $\Lambda_{X_1}(\lambda) = \theta(e^\lambda - 1)$*, logo* $\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \theta(e^\lambda - 1)\}$*. Tomando* $f(\lambda) = \lambda x - \theta(e^\lambda - 1)$*, temos* $f'(\lambda) = x - \theta e^\lambda$*. Observemos que se* $x < 0$*, então* f *é decrescente e portanto*

$$\sup_{\lambda\in\mathbb{R}} = \lim_{\lambda\to-\infty} f(\lambda) = +\infty.$$

Se $x = 0$*, então*

$$\sup_{\lambda\in\mathbb{R}} f(\lambda) = \lim_{\lambda\to-\infty} f(\lambda) = \theta.$$

Agora, se $\lambda > 0$*, vemos que*

$$f'(\lambda)\begin{cases} < 0, & se\ \lambda > \ln\frac{x}{\theta}; \\ = 0, & se\ \lambda = \ln\frac{x}{\theta}; \\ > 0, & se\ \lambda < \ln\frac{x}{\theta}. \end{cases}$$

Logo f *assume seu supremo em* $\ln\frac{x}{\theta} = \lambda$*, assim*

$$\sup_{\lambda\in\mathbb{R}} f(\lambda) = x\ln\frac{x}{\theta} - \theta\frac{x}{\theta} + \theta.$$

Portanto

$$\Lambda^*(x)\begin{cases} = +\infty, & se\ x < 0; \\ = x\ln\frac{x}{\theta} - \theta\frac{x}{\theta} + \theta, & se\ x \geq 0. \end{cases} \tag{1.39}$$

Exemplo 1.3.3. *Consideremos agora* X_1 *com distribuição Bernoulli(p). Temos primeiramente que* $\Lambda(\lambda) = \ln(1 - p + pe^\lambda)$*. Tomando agora* $f(\lambda) = \lambda x - \ln(1 - p + pe^\lambda)$*. Temos que*

$$f'(\lambda) = x - \frac{pe^\lambda}{1 - p + pe^\lambda}.$$

Se $x < 0$*, então como o termo* $\frac{pe^\lambda}{1-p+pe^\lambda} \geq 0$*, segue que* f' *é não-positiva, e portanto* f *é decrescente, logo*

$$\sup_{\lambda\in\mathbb{R}} f(\lambda) = \lim_{\lambda\to-\infty} f(\lambda).$$

Como $\ln(1 - p + pe^\lambda) \to \ln(1 - p)$*, quando* $\lambda \to -\infty$*, concluímos que*

$$\sup_{\lambda\in\mathbb{R}} f(\lambda) = +\infty.$$

Agora se $x = 0$*, então* $f'(\lambda) = -\frac{pe^{\lambda}}{1-p+pe^{\lambda}}$ *e por argumentação análoga vemos que*

$$\sup_{\lambda\in\mathbb{R}} f(\lambda) = \lim_{\lambda\to-\infty} f(\lambda) = -\ln(1-p).$$

Se $x > 1$*, então notemos que* $f' \geq 0$*, logo*

$$\sup_{\lambda\in\mathbb{R}} f(\lambda) = \lim_{\lambda\to+\infty} f(\lambda) = \lim_{\lambda\to\infty} \ln \frac{e^{\lambda x}}{1-p+pe^{\lambda}} = +\infty.$$

Se $x \in (0,1)$*, então segue que*

$$f'(\lambda) \begin{cases} < 0, & se\ \lambda > \ln \frac{x-xp}{p-xp}; \\ = 0, & se\ \lambda = \ln \frac{x-xp}{p-xp}; \\ > 0, & se\ \lambda < \ln \frac{x-xp}{p-xp}. \end{cases}$$

Logo vemos que

$$\sup_{\lambda\in\mathbb{R}} f(\lambda) = f\left(\ln \frac{x-xp}{p-xp}\right) = x\ln\frac{x-xp}{p-xp} - \ln\left(1-p+p\frac{x-xp}{p-xp}\right) =$$

$$x\ln\left(\frac{x}{p}\right)\left(\frac{1-p}{1-x}\right) - \ln\left(\frac{1-p}{1-x}\right) = x\ln\left(\frac{x}{p}\right) + (1-x)\ln\left(\frac{1-x}{1-p}\right)$$

Agora se $x = 1$*, então* $f' \geq 0$*. Logo*

$$\sup\nolimits_{\lambda\in\mathbb{R}} f(\lambda) = \lim\nolimits_{\lambda\to\infty} \ln \tfrac{e^{\lambda x}}{1-p+pe^{\lambda}} = \ln \tfrac{1}{p}$$

Portanto concluímos que

$$\Lambda^*(x) \begin{cases} = x\ln\left(\frac{x}{p}\right) + (1-x)\ln\left(\frac{1-x}{1-p}\right), & se\ x \in [0,1] \\ = +\infty, & C.C. \end{cases} \tag{1.40}$$

Faremos agora um resultado que nos auxiliara na demonstração do proximo lema.

Afirmação 1.3.4. *Seja* $\bar{x} = \mathbb{E}[X_1]$*, então:*
1)Se $\bar{x} \in \mathbb{R}$*, então*

$$\Lambda(\lambda) \geq \lambda\overline{x}, \tag{1.41}$$

para todo $\lambda \in \mathbb{R}$*.*
2) Se $\bar{x} = -\infty$*, então* $\Lambda(\lambda) = +\infty$*, para todo* $\lambda \in (-\infty, 0)$*.*
3) Se $\bar{x} = +\infty$*, então* $\Lambda(\lambda) = +\infty$*, para todo* $\lambda \in (0, +\infty)$*.*

Demonstração: Com efeito, suponhamos primeiramente que $\bar{x} \in \mathbb{R}$. Se $\lambda \in D_\Lambda$ e observando que $-\ln$ é uma função convexa, segue da Desigualdade de Jensen (ver apêndice B, Teorema B.2.4) que

$$-\Lambda(\lambda) = -\ln\mathbb{E}[e^{\lambda X_1}] \leq \mathbb{E}[-\ln e^{\lambda X_1}] = -\lambda\overline{x}.$$

Caso $\lambda \notin D_\Lambda$, a desigualdade acima é imediata. Disso concluímos que

$$\Lambda(\lambda) \geq \lambda\overline{x},$$

para todo $\lambda \in \mathbb{R}$, quando $\bar{x} \in \mathbb{R}$. Com isso terminamos 1).

Se $\bar{x} = -\infty$, segue que

$$\int_{-\infty}^{0} x d\mu = -\infty,$$

pois $\bar{x} = \int_0^{+\infty} x d\mu + \int_{-\infty}^{0} x d\mu$. Logo vemos que dado $\lambda < 0$, tem-se

$$1_{X_1 \in (-\infty,0]} \lambda X_1 \leq e^{\lambda X_1},$$

logo

$$\lambda \int_{-\infty}^{0} x d\mu \leq M(\lambda).$$

Assim $M(\lambda) = +\infty$ e disso concluímos que $\Lambda(\lambda) = \infty$, para todo $\lambda \in (-\infty, 0)$, quando $\overline{x} = -\infty$. Com isso concluímos 2).

O processo para a demonstração de 3) é análogo ao de 2). □

Para facilitar a demonstração dos resultados a seguir iremos usar a seguinte notação: dado uma função $f : \mathbb{R} \longrightarrow (-\infty, +\infty]$, denotaremos como $D_f := \{\lambda \in \mathbb{R} : f(\lambda) < +\infty\}$ o conjunto de todos pontos onde f assume um valor real. Iremos agora fazer dois resultados que nos auxiliarão na demonstração do Teorema de Cramér.

Lema 1.3.5. *Sejam Λ e Λ^* o logaritmo da função geradora de momento e a transformação de Fenchel-Legendre associadas a variável aleatória X_1. Temos então:*
a) Λ é uma função convexa e Λ^* é uma função taxa convexa.
b) Se $D_\Lambda = \{0\}$, então $\Lambda^* \equiv 0$. Se $\Lambda(\lambda) < \infty$ para algum $\lambda > 0$, então $\overline{x} < \infty$ (podendo ser $-\infty$) e para todo $x \geq \overline{x}$ vale que

$$\Lambda^*(x) := \sup_{\lambda \geq 0}\{\lambda x - \Lambda(\lambda)\}, \tag{1.42}$$

mais ainda, para $x \geq \overline{x}$, Λ^* é não-decrescente. Similarmente, se $\Lambda(\lambda) < \infty$ para algum $\lambda < 0$, então $\overline{x} > -\infty$ (podendo ser $+\infty$) e para todo $x \leq \overline{x}$ vale que

$$\Lambda^*(x) := \sup_{\lambda \leq 0}\{\lambda x - \Lambda(\lambda)\}, \tag{1.43}$$

mais ainda, para $x \leq \overline{x}$, Λ^* é não-crescente. Quando $\overline{x}$ é finito, $\Lambda^*(\overline{x}) = 0$. E sempre vale que

$$\inf_{x \in \mathbb{R}} \Lambda^*(x) = 0. \tag{1.44}$$

c) $\Lambda(*)$ é diferenciável em D_Λ^o com

$$\Lambda'(\eta) = \frac{1}{M(\eta)}\mathbb{E}[X_1 e^{\eta X_1}] \tag{1.45}$$

e

$$\Lambda'(\eta) = y \Longrightarrow \Lambda^*(y) = \eta y - \Lambda(\eta) \tag{1.46}$$

Demonstração: a) Mostraremos primeiramente que Λ é convexa D_Λ. Com efeito, sejam λ_1 e λ_2 pertencentes a D_Λ arbitrários. Provaremos primeiro que D_Λ é um conjunto convexo. Então seja $\theta \in [0,1]$, segue diretamente da Desigualdade de Holder (ver apêndice A, teorema B.1.1) que

$$\Lambda(\theta\lambda_1 + (1-\theta)\lambda_2) = \ln \mathbb{E}\left[\left(e^{\lambda_1 X_1}\right)^{\theta}\left(e^{\lambda_2 X_1}\right)^{1-\theta}\right] \leq \ln\left(\mathbb{E}\left[\left(e^{\lambda_1 X_1}\right)\right]\right)^{\theta}\left(\mathbb{E}\left[\left(e^{\lambda_2 X_1}\right)\right]\right)^{1-\theta} =$$

$$\ln\left(\mathbb{E}\left[\left(e^{\lambda_1 X_1}\right)\right]\right)^{\theta} + \ln\left(\mathbb{E}\left[\left(e^{\lambda_2 X_1}\right)\right]\right)^{1-\theta} = \theta\Lambda\left(\lambda_1\right) + (1-\theta)\,\Lambda\left(\lambda_2\right).$$

Logo

$$\Lambda(\theta\lambda_1 + (1-\theta)\lambda_2) \leq \theta\Lambda\left(\lambda_1\right) + (1-\theta)\,\Lambda\left(\lambda_2\right) < +\infty$$

e com isso mostramos que D_Λ é um conjunto convexo e em particular mostramos a convexidade de Λ em D_Λ.

Mostraremos agora a convexidade de Λ^* em D_{Λ^*}. Com efeito, seja x_1 e x_2 pertencentes a D_{Λ^*}, dado $\theta \in [0,1]$ temos que

$$\theta\Lambda^*(x_1) + (1-\theta)\Lambda^*(x_2) = \sup_{\lambda\in\mathbb{R}}\{\lambda\theta x_1 - \theta\Lambda(\lambda)\} + \sup_{\lambda\in\mathbb{R}}\{\lambda(1-\theta)x_2 - (1-\theta)\Lambda(\lambda)\}$$

$$\geq \sup_{\lambda\in\mathbb{R}}\{\lambda\theta x_1 - \theta\Lambda(\lambda) + \lambda(1-\theta)x_2 - (1-\theta)\Lambda(\lambda)\} = \sup_{\lambda\in\mathbb{R}}\{\lambda(\theta x_1 + (1-\theta)x_2) - \Lambda(\lambda)\} =$$

$$\Lambda^*(\theta x_1 + (1-\theta)x_2)$$

Portanto D_{Λ^*} é um conjunto convexo e Λ^* é convexa em D_{Λ^*}.

Mostraremos agora que Λ^* é uma função taxa. Com efeito notemos primeiramente que $\Lambda^*(x) \geq 0x - \Lambda(0) = 0$, logo Λ^* é não negativa. Iremos partir agora para a demonstração da semi-continuidade inferior, para isso consideremos $\alpha \in [0,+\infty)$ e o conjunto $\Psi_{\Lambda^*}(\alpha)$. Então seja x_n uma sequência em $\Psi_{\Lambda^*}(\alpha)$, tal que $x_n \to x$. Dado $\lambda \in \mathbb{R}$, vale que

$$\lambda x - \Lambda(\lambda) = \lim_{n\to\infty} \lambda x_n - \Lambda(\lambda) \leq \alpha,$$

logo $\Lambda^*(x) \leq \alpha$. Como λ é arbitrário, segue que 1.1.1 que Λ^* é uma função taxa.
b) Notemos que se $D_\Lambda = \{0\}$, então temos que $\lambda x - \Lambda(\lambda) = -\infty$, para todo $\lambda \neq 0$, logo

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda(\lambda)\} = 0x - \Lambda(0) = 0$$

e portanto $\Lambda^* \equiv 0$.

Agora suponhamos que $\Lambda(\lambda_0) < +\infty$, para algum $\lambda_0 > 0$. Primeiro mostraremos que $\overline{x} < +\infty$, para isto notemos que por (1.37), tem-se $M(\lambda_0) < +\infty$ e também temos que

$$1_{X_1\in[0,+\infty)}\lambda_0 X_1 \leq e^{\lambda_0 X_1},$$

assim

$$\int_0^{+\infty} x d\mu \leq \frac{M(\lambda_0)}{\lambda_0} < +\infty,$$

disto concluímos que $\overline{x}$ existe (podendo ser $-\infty$), pois

$$\overline{x} \leq \int_0^{\infty} x d\mu \leq \frac{M(\lambda_0)}{\lambda_0} < +\infty.$$

Mostraremos agora que $\Lambda^*(x) = \sup_{\lambda \geq 0}\{\lambda x - \Lambda(\lambda)\}$, para todo $x \geq \bar{x}$. Com efeito, notemos que se $\overline{x} = -\infty$, segue de 1.3.4, item 2), que $\Lambda(\lambda) = \infty$, para todo $\lambda \in (-\infty, 0)$, logo $\Lambda^*(x) := \sup_{\lambda \geq 0}\{\lambda x - \Lambda(\lambda)\}$. Agora se $\overline{x} > -\infty$, temos por 1.3.4, item 1), que $\lambda \bar{x} - \Lambda(\lambda) \leq 0$, para todo $\lambda \in \mathbb{R}$. Notemos agora que dados $x \geq \overline{x}$ e $\lambda < 0$, temos

$$\lambda x - \Lambda(\lambda) \leq \lambda \overline{x} - \Lambda(\lambda) \leq 0 = \Lambda^*(\overline{x}),$$

e como $\Lambda^*(x) \geq 0$ para todo $x \in \mathbb{R}$, pois Λ^* é uma função taxa, segue que para $x \geq \overline{x}$, tem-se

$$\Lambda^*(x) := \sup_{\lambda \geq 0}\{\lambda x - \Lambda(\lambda)\}.$$

Notemos agora que dados $\overline{x} \leq x_1 < x_2$ e $\lambda \geq 0$, vale que $\lambda x_1 - \Lambda(\lambda) \leq \lambda x_2 - \Lambda(\lambda)$, e disto concluímos que $\Lambda^*(x_1) \leq \Lambda^*(x_2)$, e portanto Λ^* é crescente para $x \geq \overline{x}$.

O caso em que $\Lambda(\lambda_0) < +\infty$ para algum $\lambda_0 < 0$ é feito de maneira analoga ao caso anterior.

Mostraremos agora que $\Lambda^*(\overline{x}) = 0$, quando $\bar{x} = 0$. Com efeito, segue de 1.3.4, item 1), que $\lambda \overline{x} - \Lambda(\lambda) \leq 0$, para todo $\lambda \in \mathbb{R}$. Disso concluímos que $\Lambda^*(\overline{x}) = 0$.

Agora nos resta mostrar que sempre vale

$$\inf_{x \in \mathbb{R}} \Lambda^*(x) = 0.$$

Com efeito, se $D_\Lambda = \{0\}$, então $\Lambda^* \equiv 0$, e portanto $\inf_{x \in \mathbb{R}} \Lambda^*(x) = 0$. Se $D_\Lambda \neq \{0\}$, existe $\lambda_0 \neq 0$, tal que $\Lambda(\lambda_0) < +\infty$. Se $\lambda_0 > 0$, temos $\overline{x} < +\infty$. Consideremos primeiramente que $\overline{x} \in \mathbb{R}$, então é imediato que $\inf_{x \in \mathbb{R}} \Lambda^*(x) = 0$, pois $\Lambda^*(\overline{x}) = 0$. Agora se $\overline{x} = -\infty$, então dado $\lambda \geq 0$, temos que se $\mathbb{E}[e^{\lambda X_1}] < \infty$, vale

$$\mu([x, \infty)) = \mathbb{P}[X_1 \in [x, +\infty)] = \mathbb{P}\left[e^{\lambda X_1} \in \left[e^{\lambda x}, +\infty\right)\right] \leq \mathbb{E}\left[e^{\lambda X_1 - \lambda x}\right],$$

onde a ultima desigualdade segue da Desigualdade de Chebycheff (ver apêndice B, Teorema B.2.5), aplicado na função valor absoluto. Se $\mathbb{E}[e^{\lambda X_1}] = +\infty$ vemos facilmente que $\mu([x, +\infty)) < \mathbb{E}[e^{\lambda X_1 - \lambda x}]$. Logo temos que

$$\ln \mu([x, +\infty)) \leq \inf_{\lambda \in [0, +\infty)} \ln \mathbb{E}[e^{\lambda X_1 - \lambda x}] = -\sup_{\lambda \geq 0}\{\lambda x - \Lambda(\lambda)\} = -\Lambda^*(x)$$

e portanto

$$0 \leq \lim_{x \to -\infty} \Lambda^*(x) \leq \lim_{x \to -\infty} -\ln \mu([x, +\infty)) = 0.$$

Se $\lambda_0 < 0$, então $\overline{x} > -\infty$. Suponhamos primeiramente que $\overline{x} \in \mathbb{R}$, então segue de imediato que $\inf_{x\in\mathbb{R}} \Lambda^*(x) = 0$. Se $\overline{x} = +\infty$, dado $\lambda \leq 0$, vamos supor que $\mathbb{E}[e^{\lambda X_1}] < \infty$. Considerando $Y = -X_1$ e $\alpha = -\lambda$, então pela Desigualdade de Chebycheff vale

$$\mu((-\infty, x]) = \mathbb{P}[Y \in [-x, +\infty)] = \mathbb{P}\left[e^{\alpha Y} \in \left[e^{\alpha(-x)}, +\infty\right)\right] \leq \mathbb{E}\left[e^{\alpha Y - \alpha(-x)}\right] = \mathbb{E}\left[e^{\lambda X_1 - \lambda x}\right].$$

De maneira imediata vemos que se $\mathbb{E}[e^{\lambda X_1}] = \infty$, vale que $\mu([x,\infty)) < \mathbb{E}[e^{\lambda X_1 - \lambda x}]$ e portanto temos que

$$\ln \mu((-\infty, 0]) \leq \inf_{\lambda \in (-\infty, 0]} \ln \mathbb{E}[e^{\lambda X_1 - \lambda x}] = -\sup_{\lambda \leq 0}\{\lambda x - \Lambda(\lambda)\} = -\Lambda^*(x)$$

assim vale

$$\lim_{x\to\infty} \Lambda^*(x) \leq \lim_{x\to\infty} -\ln \mu((-\infty, 0])) = 0.$$

Com isso concluímos a demonstração do item b).

c) Seja $\eta \in D_\Lambda^o$. Notemos primeiramente que $\Lambda(\lambda) = foM(\lambda)$, onde $f(w) = \ln w$. Com isso vemos que para demonstrarmos a diferenciabilidade de Λ em η, basta que mostremos a diferenciabilidade de M em η. Com efeito, dado $\delta > 0$, afirmamos que para todo $\epsilon \in (-\delta, 0) \cup (0, \delta)$ tem-se

$$\left|\frac{e^{(\eta+\epsilon)x} - e^{\eta x}}{\epsilon}\right| \leq e^{\eta x}\frac{e^{\delta|x|} - 1}{\delta}. \tag{1.47}$$

Observemos que para a desigualdade (1.47) acontecer, basta provarmos que vale

$$\left|\frac{e^{\epsilon x} - 1}{\epsilon}\right| \leq \frac{e^{\delta|x|} - 1}{\delta},$$

para todo $\epsilon \in (-\delta, 0) \cup (0, \delta)$. Suponhamos $x \geq 0$. Notemos que se $\epsilon \in (-\delta, 0) \cup (0, \delta)$, então

$$\left|\frac{e^{\epsilon x} - 1}{\epsilon}\right| = \frac{e^{\epsilon x} - 1}{\epsilon}.$$

Tomando $g(y) = e^{yx}$ e assumindo que $\epsilon \in (-\delta, 0)$, segue do Teorema do Valor Intermediário de Lagrange que existem $c \in (-\epsilon, 0)$ e $d \in (0, \delta)$ tais que

$$g'(c) = \frac{e^{\epsilon x} - 1}{\epsilon} \text{ e } g'(d) = \frac{e^{\delta x} - 1}{\delta}.$$

Como $x \geq 0$, segue que $g'(c) \leq g'(d)$, logo

$$\left|\frac{e^{\epsilon x} - 1}{\epsilon}\right| \leq \frac{e^{\delta|x|} - 1}{\delta},$$

quando $\epsilon \in (-\delta, 0)$. Se $\epsilon \in (0, \delta)$, então $\epsilon = t\delta$, para algum $t \in (0, 1)$. Pela convexidade de g, segue que

$$e^{\epsilon x} = g(\epsilon) = g(t\delta) \leq tg(\delta) + (1-t)g(0) = \frac{\epsilon}{\delta}\left(e^{\delta} - 1\right) + 1,$$

logo
$$\left|\frac{e^{\epsilon x}-1}{\epsilon}\right| = \frac{e^{\epsilon x}-1}{\epsilon} \leq \frac{e^{\delta|x|}-1}{\delta}.$$
Suponhamos agora que $x < 0$, então
$$\left|\frac{e^{\epsilon x}-1}{\epsilon}\right| = \frac{1-e^{\epsilon x}}{\epsilon},$$
para todo $\epsilon \in (-\delta, 0) \cup (0, \delta)$. Escrevendo $\widetilde{x} = -x$ e $\widetilde{\epsilon} = -\epsilon$, vemos que
$$\left|\frac{e^{\epsilon x}-1}{\epsilon}\right| = \frac{e^{\widetilde{\epsilon}\,\widetilde{x}}-1}{\widetilde{\epsilon}},$$
assim, pelo que provamos para o caso em que $x \geq 0$, segue que
$$\left|\frac{e^{\epsilon x}-1}{\epsilon}\right| \leq \frac{e^{\delta y}-1}{\delta} = \frac{e^{\delta|x|}-1}{\delta}.$$
Com isso concluímos a demonstração da afirmação.

Mostraremos agora que vale a desigualdade
$$\mathbb{E}\left[e^{\eta X_1+\delta|X_1|}\right] < +\infty.$$
para $\delta > 0$ suficientemente pequeno, tal que $(\eta-\delta, \eta+\delta) \in D_\Lambda$. Vale que
$$+\infty > \mathbb{E}\left[e^{(\delta+\eta)X}\right] \geq \int_{X\geq 0} e^{(\delta+\eta)X_+} d\mathbb{P}$$
e de maneira análoga vemos que
$$+\infty > \mathbb{E}\left[e^{(-\delta+\eta)X}\right] \geq \int_{X<0} e^{-(-\delta+\eta)X_-} d\mathbb{P}.$$
Como
$$\mathbb{E}\left[e^{\eta X+\delta|X|}\right] = \int_{X\geq 0} e^{(\delta+\eta)X_+} d\mathbb{P} + \int_{X<0} e^{-(-\delta+\eta)X_-} d\mathbb{P},$$
segue que
$$\mathbb{E}\left[e^{\eta X+\delta|X|}\right] < +\infty,$$
para $\delta > 0$ suficientemente pequeno. Em particular vemos que
$$\mathbb{E}\left[\frac{e^{\eta X+\delta|X|}-e^{\eta X}}{\delta}\right] < \infty, \tag{1.48}$$
para δ suficientemente pequeno.

Segue do Teorema de Convergência Dominada (ver apêndice B, Teorema B.2.6) e das desigualdades (1.47) e (1.48) que
$$\lim_{\epsilon\to 0} \mathbb{E}\left[\frac{e^{(\eta+\epsilon)X}-e^{\eta X}}{\epsilon}\right] = \mathbb{E}\left[\lim_{\epsilon\to 0} \frac{e^{(\eta+\epsilon)X}-e^{\eta X}}{\epsilon}\right] = \mathbb{E}[Xe^{\eta X}].$$

Portanto M é diferenciável em η e

$$\Lambda'(\eta) = \frac{1}{M(\eta)}\mathbb{E}[Xe^{\eta X}].$$

Mostraremos agora a validade de (1.46). Com efeito, consideremos então a função $h(\lambda) = \lambda y - \Lambda(\lambda)$, onde $y = \Lambda'(\eta)$. Notemos que h é côncava em D_Λ, pois é a soma de duas funções côncavas. Como $y = \Lambda'(\eta)$, vale $h'(\eta) = 0$. Dado $\delta > 0$ tal que $(\eta-\delta, \eta+\delta) \subseteq D_\Lambda$, vale que $h(\lambda) \leq h(\delta)$, isso segue dos fatos de $h'(\eta) = 0$ e h ser côncava em D_Λ. Se $\lambda \notin D_\Lambda$, é imediato que $h(\lambda) < h(\eta)$. Agora se $\lambda \in D_\Lambda$, suponhamos por absurdo que $h(\lambda) > h(\eta)$, então para $t \in (0,1)$ suficientemente pequeno vale que $\lambda_t := t\lambda + (1-t)\eta \in (\eta - \delta, \eta + \delta)$ e assim vemos que

$$h(\lambda_t) \geq th(\lambda) + (1-t)h(\eta) > h(\eta),$$

onde a primeira desigualdade segue do fato de h ser concava em D_Λ. Essa desigualdade gera um absurdo, esse absurdo acontece por supormos que $h(\lambda) > h(\eta)$. Com isso concluímos que $h(\eta) = \sup_{\lambda\in\mathbb{R}} h(\lambda)$, e portanto

$$\Lambda^*(y) = \eta y - \Lambda(\eta).$$

Com isso concluímos a demonstração de c). $\square$

Lema 1.3.6. *Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias i.i.d., com distribuição $\mu \in M_1(\mathbb{R})$ arbitraria. Dado $\delta > 0$, vale que*

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mu_n((-\delta, +\delta)) \geq \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) = -\Lambda^*(0), \tag{1.49}$$

onde Λ é o logaritmo da função geradora de momento e Λ^* a Transformação de Fenchel-Legendre associadas a variável aleatória X_1.

Demonstração: Dado $\delta > 0$. Suponhamos primeiramente que $\mu((-\infty, 0)) > 0$, $\mu((0, +\infty)) > 0$ e que μ tem suporte em um conjunto limitado de $\mathbb{R}$. Como μ tem suporte em um conjunto limitado, segue que existem $A < 0 < B$ tais que $\mu([A, B]) = 1$. Como $\mu([A, B]) = 1$, vale que $A \leq X_1 \leq B$ q.t.p.. Então dado $\lambda > 0$, temos $\ln \mathbb{E}[e^{\lambda A}] \leq \Lambda(\lambda) \leq \ln \mathbb{E}[e^{\lambda B}]$, logo $\lambda A \leq \Lambda(\lambda) \leq \lambda B$. Tomando agora $\lambda < 0$, vale por inequações análogas que $\lambda A \geq \Lambda(\lambda) \geq \lambda B$. Portanto $D_\Lambda = \mathbb{R}$ e assim pelo lema 1.3.5, item c), concluímos que Λ é continua e diferenciável em $\mathbb{R}$.

Mostraremos agora que $|\lambda| \to +\infty$ implica em $\Lambda(\lambda) \to +\infty$. Com efeito, observemos que existe $\alpha > 0$ tal que $\mu((\alpha, +\infty)) > 0$, onde esse α pode ser encontrado a partir da continuidade da medida de probabilidade μ. Logo

$$\Lambda(\lambda) = \ln \int e^{\lambda x} d\mu \geq \ln \int_{(\alpha,+\infty)} e^{\lambda x} d\mu \geq \ln \int_{(\alpha,+\infty)} e^{\lambda\alpha} d\mu = \lambda\alpha + \ln \mu((\alpha, +\infty))$$

Disto vemos facilmente que quando $\lambda \to +\infty$, tem-se $\Lambda(\lambda) \to +\infty$. De maneira análoga conseguimos que $\Lambda(\lambda) \to +\infty$, quando $\lambda \to -\infty$.

Como $\Lambda(\lambda) \to +\infty$ quando $|\lambda| \to +\infty$, segue da continuidade de Λ que existe $\eta \in \mathbb{R}$ tal que $\Lambda(\eta) = \inf_{\lambda \in \mathbb{R}} \Lambda(\lambda)$ e como Λ é diferenciável em $\mathbb{R}$, temos $\Lambda'(\eta) = 0$. Definamos então a seguinte medida

$$\widetilde{\mu}(A) = \int_A e^{\eta x - \Lambda(\eta)} d\mu \tag{1.50}$$

e como vale que

$$\widetilde{\mu}(\mathbb{R}) = \int_{\mathbb{R}} e^{\eta x - \Lambda(\eta)} d\mu = \frac{1}{M(\eta)} \int_{\mathbb{R}} e^{\eta x} d\mu = 1,$$

segue que $\widetilde{\mu}$ é uma medida de probabilidade. Consideremos agora a sequência $\{\nu_n\}$ em $M_1(\mathbb{R})$, tal que $\nu_n = \widetilde{\mu}$, para todo $j \in \mathbb{N}$. Por resultado (ver apêndice B, Teorema B.2.1) existe uma sequência de variáveis aleatórias $\{Y_n\}_{n \in \mathbb{N}}$ independentes, tal que Y_n tem distribuição ν_n. Denotaremos por $\widetilde{\mu}_n$ a distribuição induzida por $\widehat{S}_{Y,n} = \frac{1}{n}\sum_{j=1}^n Y_j$.

Agora consideremos $(\Omega, \mathcal{F}, \mathbb{P})$ o espaço de probabilidade no qual as nossas variáveis aleatórias X_i estão atuando. Sabemos que elas induzem uma medida μ^n em $(\mathbb{R}^n, \mathcal{L}^n)$, por meio do vetor aleatório $(X_1, ..., X_n)$, da seguinte forma

$$\mu^n(B) = \mathbb{P}\left[(X_1, ..., X_n) \in B\right],$$

onde $\mathcal{L}^n$ é a σ-álgebra gerada pelos retângulos $L_1 \times ... \times L_n$, tal que $L_i \in \mathcal{L}$, com $i = 1, ..., n$, e $\mathcal{L}$ de Lebesgue para $\mathbb{R}$. Temos então que tomando $A_i \in \mathcal{L}$, com $i = 1, ..., n$, vale que

$$\mu^n(A_1 \times ... \times A_n) = \mathbb{P}[X_1 \in A_1, ..., X_n \in A_n] = \prod_{i-1}^{n} \mu(A_i),$$

logo μ^n é a medida produto de $\theta_1, ..., \theta_n$, onde $\theta_i = \mu$. Observemos também que $(\mathbb{R}^n, \mathcal{L}^n, \mu^n)$ induz uma distribuição, em $(\mathbb{R}, \mathcal{L})$, que denotaremos por μ_n', por meio da variável aleatória $\frac{1}{n}\sum_{j=1}^n x_j$, com $(x_1, ..., x_n) \in \mathbb{R}^n$. Assim

$$\mu_n(A) = \mathbb{P}\left[\widehat{S}_n \in A\right] = \mathbb{P}\left[(X_1, ..., X_n)^{-1}\left(\widehat{S}_n \in A\right)\right] = \mu^n\left((x_1, ..., x_n); \frac{1}{n}\sum_{j=1}^{n} x_j \in A\right) = \mu_n'(A),$$

com isso concluímos que $\mu_n = \mu_n'$, em particular vemos que μ^n induz μ_n. Notemos que podemos fazer essa mesma construção para a sequência $\{Y_j\}$. Temos então por resultado (ver apêndice B, Teorema B.2.2) que

$$\mu_n((-\epsilon, +\epsilon)) = \int_{x \in (-n\epsilon, +n\epsilon)} d\mu_n = \int_{\sum_{i=1}^n x_i \in (-n\epsilon, +n\epsilon)} d\mu^n$$

e segue do Teorema de Fubini (ver apêndice A, Teorema B.1.2) que vale a igualdade

$$\mu_n((-\epsilon, +\epsilon)) = \int_{\mathbb{R}} ... \int_{\mathbb{R}} 1_{\left[\sum_{i=1}^n x_i \in (-n\epsilon, +n\epsilon)\right]} \mu(dx_1)...\mu(dx_n).$$

Então de (1.50) e pela derivada de Radon-Nikodin, segue que

$$\mu_n((-\epsilon, +\epsilon)) = \int_{\mathbb{R}} ... \int_{\mathbb{R}} 1_{\left[\sum_{i=1}^n x_i \in (-n\epsilon, +n\epsilon)\right]} e^{\sum_{i=1}^n (-\eta x_i + \Lambda(\eta))} \widetilde{\mu}(dx_1)...\widetilde{\mu}(dx_n) =$$

$$\int_{\mathbb{R}^n} 1_{[\sum_{i=1}^n x_i \in (-n\epsilon, +n\epsilon)]} e^{\sum_{i=1}^n (-\eta x_i + \Lambda(\eta))} d\widetilde{\mu}^n \geq e^{n\Lambda(\eta) - |\eta| n\epsilon} \int_{\mathbb{R}^n} 1_{[\sum_{i=1}^n x_i \in (-n\epsilon, +n\epsilon)]} d\widetilde{\mu}^n,$$

onde a segunda igualdade vale pelo Teorema de Fubini (ver apêndice A, Teorema B.1.2). Assim

$$\mu_n((-\epsilon, +\epsilon)) \geq e^{n\Lambda(\eta) - |\eta| n\epsilon} \widetilde{\mu}_n((-\epsilon, +\epsilon)). \tag{1.51}$$

Observemos agora que aplicando o lema 1.3.5, item c), temos

$$\mathbb{E}_{\widetilde{\mu}}[Y_1] = \int_{\mathbb{R}} y d\widetilde{\mu} = \frac{1}{M(\eta)} \int_{\mathbb{R}} y e^{\eta y} d\mu = \Lambda'(\eta) = 0,$$

temos então pela Lei dos Grandes Números que

$$\lim_{n \to +\infty} \widetilde{\mu}_n((-\epsilon, +\epsilon)) = 1.$$

Então tomando $0 < \epsilon < \delta$ e considerando a desigualdade (1.51), tem-se

$$\liminf_{n \to +\infty} \frac{1}{n} \ln \mu_n((-\epsilon, +\epsilon)) \geq \liminf_{n \to +\infty} \left(\frac{1}{n} \ln \widetilde{\mu}_n((-\epsilon, +\epsilon)) + \Lambda(\eta) - |\eta|\epsilon \right) = \Lambda(\eta) - |\eta|\epsilon$$

e fazendo $\epsilon \to 0$, vale que

$$\liminf_{n \to +\infty} \frac{1}{n} \ln \mu_n((-\epsilon, +\epsilon)) \geq \Lambda(\eta) \geq \inf_{\lambda \in \mathbb{R}} \Lambda(\lambda) = -\Lambda^*(0).$$

Suponhamos agora que μ tem suporte ilimitado e $\mu((-\infty, 0)) > 0$ e $\mu((0, +\infty)) > 0$. Existe $M > 0$, tal que $\mu([-M, 0)) > 0$ e $\mu((0, M]) > 0$. Assim definiremos a seguinte função

$$\Lambda_M(\lambda) := \ln \int_{-M}^{M} e^{\lambda x} d\mu.$$

Denotemos por ν a distribuição induzida pela probabilidade condicional de X_1 dado $\{|X_1| \leq M\}$ e denotemos por ν_n a distribuição induzida pela probabilidade condicional de $\widehat{S}_n$ dado $\{|X_i| \leq M, i = 1, ..., n\}$. Temos então que dado $\delta > 0$, vale

$$\mu_n((-\delta, +\delta)) \geq \frac{\mathbb{P}[\{\widehat{S}_n \in (-\delta, +\delta)\} \cap \{\cap_{i=1}^n |X_i| \leq M\}]}{\mathbb{P}[\{\cap_{i=1}^n |X_i| \leq M\}]} \mathbb{P}[\{\cap_{i=1}^n |X_i| \leq M\}] =$$
$$\nu_n((-\delta, +\delta))(\mu([-M, +M]))^n.$$

Por resultado (ver apêndice B, Teorema B.2.1), conseguimos uma sequência $\{T_j\}$ de variáveis aleatórias i.i.d., onde cada T_j tem distribuição ν e atuam em um espaço de probabilidade $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$. Novamente conseguimos induzir em $\mathbb{R}$ uma distribuição $\widetilde{\nu}_n$ por meio da variável aleatória $h(T_1, ..., T_n) = \frac{1}{n} \sum_{j=1}^n T_j$, tal que

$$\widetilde{\nu}_n(A) = \widetilde{\mathbb{P}}\left[\tfrac{1}{n} \textstyle\sum_{j=1}^n T_j \in A\right] = \widetilde{\mathbb{P}}[(T_1, ..., T_n) \in h^{-1}(A)] = \widetilde{\nu}^n(h^{-1}(A)),$$

onde $\widetilde{\nu}^n$ é uma medida em $\mathbb{R}^n$, induzida pelo vetor aleatório $(T_1, ..., T_n)$. Agora notemos que conseguimos uma distribuição ν^n em $\mathbb{R}^n$, induzida pela probabilidade condicional e pelo vetor aleatório $(X_1, ..., X_n)$, tal que

$$\nu_n(A) = \mathbb{P}[(X_1, ..., X_n) \in h^{-1}(A) | |X_i| \leq M, i = 1, ..., n] = \nu^n(h^{-1}(A))$$

Então dado $B_1 \times ... \times B_n \subseteq \mathbb{R}^n$, com $B_i \in \mathcal{L}$ para $i = 1, ..., n$, tem-se

$$\widetilde{\nu}^n(B_1 \times \ldots \times B_n) = \widetilde{\mathbb{P}}[T_i \in B_i, i = 1, \ldots, n] = \prod_{i=1}^n \nu(B_i) = \mathbb{P}[X_i \in B_i, i = 1, \ldots, n||X_i| \leq M, i = 1, \ldots, n] = \nu^n(B_1 \times \ldots \times B_n).$$

Como $\widetilde{\nu}^n$ e ν^n são medidas de probabilidade, segue por resultado(ver apêndice B, Teorema B.2.3) que $\nu^n = \widetilde{\nu}^n$, portanto $\nu_n = \widetilde{\nu}_n$. Como ν tem suporte limitado e $\nu(-\infty, 0) > 0$ e $\nu(0, +\infty) > 0$ podemos aplicar a demonstração que fizemos no caso anterior, no qual supomos que μ tinha suporte limitado, em ν.

Observemos agora que

$$\Lambda_\nu(\lambda) = \ln \int_{\mathbb{R}} e^{\lambda x} d\nu = \ln \int_{-M}^{M} \frac{e^{\lambda x}}{\mu([-M, M])} d\mu = \Lambda_M(\lambda) - \ln \mu([-M, M]), \tag{1.52}$$

então tomando $0 < \epsilon < \delta$, vale que

$$\liminf_{n \to \infty} \frac{1}{n} \ln \mu_n((-\delta, \delta)) \geq \liminf_{n \to \infty} \left(\frac{1}{n} \ln \nu_n((-\epsilon, \epsilon)) + \ln \mu([-M, M]) \right) \geq \Lambda_M(\eta_\nu) - \epsilon |\eta_\nu|,$$

onde $\Lambda_\nu(\eta_\nu) = \inf_{\lambda \in \mathbb{R}} \Lambda_\nu(\lambda)$. Fazendo $\epsilon \to 0$, vemos que

$$\liminf_{n \to \infty} \frac{1}{n} \ln \mu_n((-\delta, \delta)) \geq \Lambda_M(\eta_\nu) = \inf_{\lambda \in \mathbb{R}} \Lambda_M(\lambda), \tag{1.53}$$

com a ultima igualdade acima seguindo diretamente de (1.52). Tomando $I_M := -\inf_{\lambda \in \mathbb{R}} \Lambda_M(\lambda)$ e $I^* := \limsup_{M \to \infty} I_M$. Por (1.53), vemos que se M_0 é tal que $\mu([-M_0, 0)) > 0$ e $\mu((0, M_0]) > 0$, vale que

$$\liminf_{n \to \infty} \frac{1}{n} \ln \mu_n((-\delta, \delta)) \geq \inf_{N \geq M_0} -I_N.$$

Logo

$$\liminf_{n \to \infty} \frac{1}{n} \ln \mu_n((-\delta, \delta)) \geq \liminf_{M \to \infty} -I_M = -I^* \tag{1.54}$$

Notemos agora que pela forma que definimos Λ_M, vemos que se fixarmos $\lambda \in \mathbb{R}$, vale que $\Lambda_M(\lambda)$ é crescente em M, logo $-I_M$ também é crescente em M. Observemos também que dado $M > 0$, vale que $-I_M \leq \Lambda_M(0) \leq \Lambda(0) = 0$, então de (1.54) concluímos que $-I^* \leq 0$. Agora notemos que para M suficientemente grande vale que $\mu((0, M]) > 0$ e $\mu([-M, 0)) > 0$, logo dado $\lambda \geq 0$ vemos que

$$-\infty < \ln \mu((0, M]) \leq \Lambda_M(\lambda)$$

e de maneira análoga vemos que

$$-\infty < \ln \mu([-M, 0)) \leq \Lambda_M(\lambda)$$

para $\lambda < 0$. Logo

$$-I_M \geq \min\{\ln \mu((0, M]), \ln \mu([-M, 0))\}.$$

Pelo fato de $-I_M$ ser crescente segue que $-I^* > -\infty$.

Mostraremos agora que $A_M := \{\lambda \in \mathbb{R}; \Lambda_M(\lambda) \leq I^*\} \neq \emptyset$. Com efeito, notemos que para M suficientemente grande, tal que $\mu([-M_0, 0)) > 0$ e $\mu((0, M_0]) > 0$, existe η_M, tal que $-I_M = \Lambda_M(\eta_M)$, pois observe que $n_M = n_\nu$, onde esta ν é a medida induzida pela probabilidade condicional que definimos acima, a qual também depende de M. Observemos agora que

$$-I^* = \liminf_{M\to\infty} -I_M = \lim_{M\to\infty} -I_M,$$

pois $-I_M$ é crescente em M e segue também desse fato que $-I_M \leq -I^*$ e portanto para M suficientemente grande temos $A_M \neq \emptyset$. E como $\Lambda_{M_1}(\lambda) \leq \Lambda_{M_2}(\lambda)$, para $M_1 < M_2$, segue que para todo $M > 0$, vale que $A_M \neq \emptyset$.

Pela argumentação que fizemos acima segue facilmente que $A_{M_2} \subseteq A_{M_1}$, se $M_1 < M_2$. Como Λ_ν é diferenciável em $\mathbb{R}$, segue que Λ_M é continua e como $\Lambda_\nu(\lambda) \to \infty$ quando $|\lambda| \to \infty$, vemos facilmente que A_M é compacto para M, tal que $\mu([-M_0, 0)) > 0$ e $\mu((0, M_0]) > 0$. Como $\{A_M\}$ é uma sequência de compactos não vazios, segue que existe $\lambda_0 \in \bigcap_{M>0} A_M$. Então tomando $X_M(x) = 1_{[-M,M]}e^{\lambda_0 x}$, segue do Teorema da Convergência Monótona (ver apêndice B, Teorema B.2.8) que $\Lambda(\lambda_0) = \lim_{M\to\infty} \Lambda_M(\lambda_0) \leq -I^*$. Portanto

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mu_n((-\delta, \delta)) \geq -I^* \geq \Lambda(\lambda_0).$$

Com isto concluímos a demonstração para o caso em que μ tem suporte ilimitado e $\mu((-\infty, 0)) > 0$ e $\mu((0, +\infty)) > 0$. Suponhamos agora que $\mu((-\infty, 0)) = 0$. Assim

$$\Lambda(\lambda) = \ln \int_{-\infty}^{+\infty} e^{\lambda x} d\mu = \ln \int_0^{+\infty} e^{\lambda x} d\mu.$$

Como $1_{[0,+\infty)}e^{\lambda x} \leq 1$, para $\lambda \leq 0$ e $1_{[0,+\infty)}e^{\lambda' x} \geq 1$, para $\lambda' \geq 0$, temos pelo Teorema da Convergência Dominada (ver apêndice B, Teorema B.2.6) que

$$\inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) = \lim_{\lambda\to-\infty} \Lambda(\lambda) = \ln \mu(\{0\}).$$

De maneira análoga vemos que $\inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) = \ln \mu(\{0\})$, caso $\mu((0, +\infty)) = 0$. Vemos então que

$$\mu_n((-\delta, +\delta)) \geq \mu_n(\{0\}) \geq (\mu(\{0\}))^n.$$

Com esta desigualdade concluímos a demonstração do lema. □

Seguiremos agora para o teorema de Crámer.

Teorema 1.3.7 (Teorema de Cramér). *Sejam $X_i \in \mathbb{R}$ uma sequência de variáveis aleatórias i.i.d., com X_1 tendo distribuição $\mu \in M_1(\mathbb{R})$. Então a família $\{\mu_n\}$ satisfaz Principio dos Grandes Desvios. Mais precisamente:*
a) Para qualquer conjunto fechado $F \subseteq \mathbb{R}$,

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mu_n(F) \leq -\inf_{x\in F} \Lambda^*(x). \tag{1.55}$$

b)Para qualquer conjunto aberto $G \subseteq \mathbb{R}$,

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mu_n(G) \geq -\inf_{x\in G} \Lambda^*(x). \tag{1.56}$$

Demonstração:a) Seja $F \subseteq \mathbb{R}$ um conjunto fechado. Se $F = \emptyset$, então o resultado é imediato. Suponhamos $F \neq \emptyset$. Se $I_F := \inf_{x\in F} \Lambda^*(x) = 0$, como $\mu_n(F) \leq 1$, temos $\ln \mu_n(F) \leq 0$ e portanto

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mu_n(F) \leq 0 = -\inf_{x\in F} \Lambda^*(x).$$

Se $I_F > 0$, existe $x \in F$, tal que $\Lambda^*(x) > 0$, logo existe $\lambda_0 \in \mathbb{R}$, tal que $\lambda_0 x - \Lambda(\lambda_0) > 0$, então $\lambda_0 x > \Lambda(\lambda_0)$, assim $\Lambda(\lambda_0) < \infty$. Então pelo Lema 1.3.5, item b), vale que existe $\overline{x}$, podendo ser $-\infty$ ou $+\infty$. Suponhamos agora que $\overline{x} < \infty$. Temos para todo $x \in \mathbb{R}$ e para todo $\lambda \geq 0$ que

$$1_{[\widehat{S}_n - x \geq 0]} \leq e^{n\lambda\left(\widehat{S}_n - x\right)}.$$

Portanto

$$\mu_n([x,\infty)) = \int_{\mathbb{R}} 1_{[(\sum_{i=1}^n x_n)-x\geq 0]} d\mu_n = \mathbb{E}\left[1_{\widehat{S}_n - x\geq 0}\right] \leq \mathbb{E}\left[e^{n\lambda(\widehat{S}_n - x)}\right]$$
$$= e^{-n\lambda x}\mathbb{E}\left[e^{n\lambda \widehat{S}_n}\right] = e^{-n\lambda x}\mathbb{E}\left[\prod_{i=1}^n e^{\lambda X_i}\right].$$

Como $e^{\lambda x}$ é continua e X_i são i.i.d., segue que as variáveis aleatórias $e^{\lambda X_i}$, são i.i.d.. Assim

$$\mu_n([x,\infty)) \leq e^{-n\lambda x} \prod_{i=1}^n \mathbb{E}\left[e^{\lambda X_i}\right] = e^{-n\lambda x} \prod_{i=1}^n e^{\Lambda(\lambda)} = e^{-n[\lambda x - \Lambda(\lambda)]}.$$

Então pelo Lema 1.3.5, item b), vale para todo $x > \overline{x}$ que

$$\mu_n([x,\infty)) \leq e^{-n\Lambda^*(x)}. \tag{1.57}$$

De maneira análoga vemos que se $\overline{x} > -\infty$, temos para todo $x < \overline{x}$ que

$$\mu_n((-\infty, x]) \leq e^{-n\Lambda^*(x)}.$$

Se $\overline{x} \in \mathbb{R}$, temos pelo lema 1.3.5 b) que $\Lambda^*(\overline{x}) = 0$ e como estamos assumindo que $I_F > 0$, segue que $\overline{x} \notin F$. Definimos então o seguinte conjunto $(x_-, x_+) := \bigcup_{(a,b)\subseteq F^c}(a,b)$, tal que $\overline{x} \in (a,b)$. Evidentemente $x_- < x_+$ e pelos menos um dos x_-, x_+, tem que ser real, pois $F \neq \emptyset$. Se x_- é real, então $x_- \in F$, pois se $x_- \in F^c$, existe $\delta > 0$ tal que $(x_- - \delta, x_- + \delta) \subseteq F^c$, logo $(x_-, x_+) \subsetneq (x_- - \delta, x_+)$, o que não pode acontecer pela forma que definimos (x_-, x_+). Portanto $\Delta^*(x_-) \geq I_F$. Notemos que argumentação análoga vale para o caso em que x_+ é real. Logo se x_+ é real, então $x_+ \in F$ e $\Lambda^*(x_+) \geq I_F$. Observemos então que vale $F \subseteq (-\infty, x_-]\dot{\cup}[x_+, +\infty)$, logo

$$\mu_n(F) \leq \mu_n((-\infty, x_-]) + \mu_n([x_+, +\infty)) \leq e^{-n\Lambda^*(x_-)} + e^{-n\Lambda^*(x_+)} \leq 2e^{-nI_F},$$

então

$$\sup_{m\geq n} \frac{1}{m} \ln \mu_m(F) \leq \sup_{m\geq n} \left(\frac{\ln 2}{m} - I_F \right)$$

e portanto

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mu_n(F) \leq -I_F.$$

Se $\overline{x} = -\infty$, então pelo Lema 1.3.5, item b), vale que Λ^* é não-decrescente e também pelo Lema 1.3.5, item b), vale que $\lim_{x\to-\infty} \Lambda^*(x) = 0$. Como estamos assumindo que $I_F > 0$, temos que $x_+ := \inf\{x : x \in F\}$ é real. Como F é fechado segue que $x_+ \in F$, e consequentemente $\Lambda^*(x_+) \geq I_F$. Observando que $F \subseteq [x_+, +\infty)$, teremos (1.57) que

$$\mu_n(F) \leq \mu_n([x_+, +\infty)) \leq +e^{-n\Lambda^*(x_+)} \leq 2e^{-nI_F},$$

assim seguindo os passos finais da demonstração para o caso em que $\bar{x}$ é finito, conseguimos o limite superior (1.55). Se $\overline{x} = +\infty$, veremos por argumentação análoga que $x_- := \sup\{x : x \in F\}$ é real e $x_- \in F$, e como $F \subseteq (-\infty, x_-]$. A partir disto conseguimos mostra que o limite superior é valido de maneira análoga ao caso em que $\overline{x} = -\infty$.

b) Seja $G \subseteq \mathbb{R}$, um conjunto aberto. Dado $x \in G$, existe $\delta > 0$, tal que $(x - \delta, x + \delta) \subseteq G$. Se $0 = 0$, então pelo Lema 1.3.6, segue que

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mu_n((-\delta, \delta)) \geq \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) = -\Lambda^*(0).$$

Se $x \neq 0$, definamos a sequência de variáveis aleatórias $Y_n = X_n - x$. Assim vemos que o logaritmo da função geradora de momento e a transformada de Fenchel-Legendre relacionadas a Y são $\Lambda_{Y_1} = \Lambda(\lambda) - \lambda x$ e $\Lambda^*(y) = \Lambda^*_{Y_1}(y + x)$.. Vemos que Y_1 tem distribuição $\nu(A) = \mu(A - x)$. Em particular vale que Y_n é uma sequência i.i.d.. Tomando $\widehat{S}_{n,Y} := \sum_{i=1}^n Y_i$, e considerando ν_n a medida induzida por $\widehat{S}_{n,Y}$, vemos que vale a seguinte igualdade

$$\nu_n(B) = \mathbb{P}\left[\widehat{S}_{n,Y} \in B\right] = \mathbb{P}\left[\widehat{S}_n \in B + x\right] = \mu_n(B + x).$$

Aplicando o Lema 1.3.6 agora a sequência Y_n, tem-se

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mu_n((x-\delta, x+\delta)) = \liminf_{n\to\infty} \frac{1}{n} \ln \nu_n((-\delta, +\delta)) \geq \inf_{\lambda\in\mathbb{R}} \Lambda_{Y_1}(\lambda) = -\Lambda^*_{Y_1}(0) = -\Lambda^*(x).$$

Portanto vemos que

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mu_n(G) \geq \sup_{x\in G} -\Lambda^*(x) = -\inf_{x\in G} \Lambda^*(x).$$

Com isso finalizamos a demonstração de b). $\square$

Vemos então que se $X_i \in \mathbb{R}$ é uma sequência de variáveis aleatórias i.i.d., com X_1 tendo distribuição Bernolli(p), vale pelo exemplo 1.3.3 e pelo teorema de Crámer que dado $F \in \mathbb{R}$ fechado, tem-se então que

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mu_n(F) \leq - \inf_{x\in F\cap[0,1]} x \ln\left(\frac{x}{p}\right) + (1-x)\ln\left(\frac{1-x}{1-p}\right),$$

e de maneira análoga, vale que dado $G \in \mathbb{R}$ aberto, tem-se

$$- \inf_{x\in G\cap[0,1]} x \ln\left(\frac{x}{p}\right) + (1-x)\ln\left(\frac{1-x}{1-p}\right) \leq \liminf_{n\to\infty} \frac{1}{n} \ln \mu_n(G).$$

1.4 Teorema de Gärtner–Ellis

Com o objetivo de mostrar ao leitor desse trabalho que o resultado que vimos no Teorema de Cramér não esta restrito a hipótese das variáveis aleatórias serem i.i.d.. Enunciaremos um resultado (ver (DEMBO; ZEITOUNE, 1998), capitulo 2.3), que mostrara a possibilidade de estender o Teorema de Crámer para um sequência de variáveis aleatórias $\{X_i\}$ que não são necessariamente uma sequência i.i.d., no sentido de conseguirmos um cota inferior e uma superior para o conjunto que estamos calculando.

Consideremos primeiramente uma sequência de variáveis aleatórias $Z_n \in \mathbb{R}$, tal que Z_n tem distribuição μ_n e logaritmo da função geradora de momento

$$\Lambda_n(\lambda) := \ln \mathbb{E}[e^{\lambda Z_n}].$$

Daremos agora algumas definições, as quais iremos usar no enunciado.

Definição 1.4.1. *Seja $y \in R$ é um ponto* ***exposto*** *de Λ^*, se existe $\lambda \in \mathbb{R}$ e para todo $x \neq y$ tal que*

$$\lambda x - \Lambda^*(y) > \lambda x - \Lambda^*(x), \tag{1.58}$$

o y que satisfaz (1.58) é chamado de um ***hiperplano expositor***

Definição 1.4.2. *Uma função convexa $\Lambda : \mathbb{R} \longrightarrow (-\infty, +\infty]$ é* ***essencialmente suave*** *se:*

1. *D^o_Λ é não-vazio;*
2. *$\Lambda(*)$ é diferenciável em D^o_Λ;*
3. *$\lim_{n\to+\infty} \Lambda'(\lambda_n) = +\infty$, sempre que λ_n é uma sequência de pontos em D^o_Λ convergindo para um ponto de fronteira de D^o_Λ.*

Enunciaremos agora o resultado que estende o teorema de Crámer.

Teorema 1.4.3 (Teorema de Gärtner–Ellis). *Seja $Z_n \in \mathbb{R}$ uma sequência de v.a.'s, tais que Z_n tem distribuição μ_n. Suponhamos então que para cada $\lambda \in \mathbb{R}$ o logaritmo da função geradora de momento definido como o limite*

$$\widetilde{\Lambda}(\lambda) := \lim_{n\to+\infty} \frac{1}{n}\Lambda_n(\lambda)$$

existe como um número da reta real estendida, e suponhamos também que a origem pertence a D^o_Λ. Tomando $\widetilde{\Lambda}^*$ como sendo a transformação de Fenchel-Legendre de $\widetilde{\Lambda}$, valem
a) Para qualquer conjunto fechado $F \subseteq \mathbb{R}$,

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mu_n(F) \leq -\inf_{x\in F} \Lambda^*(x). \tag{1.59}$$

b)Para qualquer conjunto aberto $G \subseteq \mathbb{R}$,

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mu_n(G) \geq -\inf_{x\in G\cap\mathcal{F}} \Lambda^*(x), \tag{1.60}$$

onde $\mathcal{F}$ é o conjunto dos pontos expostos Λ^*, os quais o hiperplano expositor pertence a D^o_Λ.
c)Se Λ é essencialmente suave, semicontínua inferiormente, então o P.D.G. vale para a função taxa bem comportada Λ^*.

Observemos que na hipótese assumida neste teorema, se tomarmos Z_n como sendo a média empírica de uma sequência de vetores aleatórios $X_i \in \mathbb{R}$ i.i.d., ou seja

$$Z_n := \frac{1}{n}\sum_{j=1}^{n} X_j$$

para todo $n \in \mathbb{N}$, então vale que

$$\frac{1}{n}\Lambda_n(n\lambda) = \frac{1}{n}\ln\mathbb{E}\left[\exp\left\{n\lambda\frac{1}{n}\sum_{i=1}^{n}X_i\right\}\right] = \frac{1}{n}\ln\left(\mathbb{E}\left[e^{\lambda X_1}\right]\right)^n = \ln\mathbb{E}\left[e^{\lambda X_1}\right].$$

A partir disso, segue teorema de Gärtner-Ellis que quando consideramos uma sequência i.i.d. de variáveis aleatórias, conseguimos encontrar uma cota inferior e outra superior.

2 Perda em Grandes Portfólios de Crédito de Risco

Para começar esta seção, devemos primeiramente nos fazer a seguinte pergunta: "O que são Portfólios de Credito de Risco?". Podemos responder isso da seguinte forma: É uma coleção de atividades financeiras de uma instuição financeira que podem gerar perdas financeiras para o mesmo. No caso em que estudaremos nesse capitulo, essas atividades financeiras serão os empréstimos realizados a milhares ou milhões de entidades.

Estaremos interessados em estudar eventos que acontecem raramente, mas que geram perdas financeiras significativas ao banco. São os eventos em que devedores, entidades que pediram empréstimo, não conseguem pagar o que devem até data pré-fixada no contrato, dando assim um calote no banco. Determinaremos a distribuição de probabilidade dessas perdas, por meio do uso da teoria de grandes desvios.

2.1 Modelagem

Começaremos agora a modelagem do nosso problema, que será estudado neste capítulo. Para isso usaremos as seguintes notações:

1. n:= Numero de devedores expostos no portfólio.
2. Y_k:= Indicador de não pagamento, ou seja $Y_k = 1$, se não houve pagamento divida e $Y_k = 0$, caso o pagamento da divida for efetuado.
3. p_k:= probabilidade do k-ésimo devedor não efetuar o pagamento, isto é, $p_k = \mathbb{P}[Y_k = 1]$
4. c_k:= perda gerada pelo não pagamento da divida do k-ésimo devedor.
5. $L_n = \sum_{j=1}^{n} c_j Y_j$:=Perda total gerada gerada pela falta de pagamento.

Nós estamos interessados em estimar a probabilidade da calda $\mathbb{P}[L_n > l_n]$, em um regime de grandes perdas l_n e em consequência disso calculamos também os raros eventos de grandes perdas que acontecem por causa de um grande número de devedores e múltiplas faltas de pagamento dos mesmos.

Por simplicidade, consideraremos um portfólio homogêneo onde todos os p_k são iguais a p e $c_k = 1$, para $k = 1, ..., n$. A dependência entre os devedores será modelada por meio da dependência entre os Y_k. Essa dependência será introduzida através da seguinte maneira: cada indicador de não pagamento será definida como

$$Y_k = 1_{\{X_k > x_k\}}, k = 1, ..., n \tag{2.1}$$

onde cada X_k tem distribuição normal padrão e x_k foi escolhido de tal forma que $x_k = \Phi^{-1}(1 - p_k) = -\Phi^{-1}(p_k)$, onde Φ distribuição acumulada da normal padrão. Denotaremos também por $\varphi = \Phi'$ a função densidade de probabilidade da normal padrão. As correlações entre as X_k, as quais determinaram a dependência entre as Y_k, serão especificadas através de

$$X_k = \rho Z + \sqrt{1 - \rho^2}\epsilon_k, k = 1, ..., n, \tag{2.2}$$

onde Z tem a distribuição normal $\mathcal{N}(0, 1)$ e $(\epsilon_k)_{k\in\{1,..,n\}}$ são i.i.d., com distribuição $\mathcal{N}(0, 1)$. Assumiremos que Z e ϵ_k são independentes uma das outras para $k = 1, ..., n$. Chamamos de Z o **fator de risco sistemático** (uma parte dos risco que esta além do controle do banco, exemplos: catástrofes naturais,corrupção politica,...) e ϵ_k é o risco associado ao k-ésimo devedor. A constante $\rho \in [0, 1)$ é o grau de dependência sobre a variável aleatória Z. Observemos que X_k, definida dessa forma se mantém um v.a. com distribuição normal padrão, pois

$$\mathbb{E}[X_k] = \rho\mathbb{E}[Z] + \sqrt{1 - \rho^2}\mathbb{E}[\epsilon_k] = 0$$
$$\sigma(X_k) = \mathbb{E}[X_k^2] = \mathbb{E}[(\rho Z)^2] + \mathbb{E}[2(\rho - \rho^3)Z\epsilon_k] + \mathbb{E}[(1 - \rho^2)(\epsilon_k)^2] = 1$$

e como a soma de duas normais é uma normal, segue que que X_k tem distribuição $\mathcal{N}(0, 1)$.

As próximas subseções estarão focadas em estudar a independência entre os devedores ($\rho = 0$) e o caso de dependência entre os devedores ($\rho \in (0, 1)$).

2.2 Devedores Independentes

Suponhamos que $\rho = 0$, notemos que pela equação (2.2) vale $X_k = \epsilon_k$, portanto as X_k são i.i.d. e disso concluímos que as Y_k serão i.i.d., com distribuição binomial com parâmetro p. Tomando l_n tal que $\frac{l_n}{n} \to q \in (p, 1)$, então S.P.G. podemos considerar $l_n = nq$. Definindo $F := \{x; x > q\}$, temos por 1.3.5, item b), que para $x \geq p$, vale que Λ^* é não-decrescente, disto vemos que $\inf_{\lambda\in F} \Lambda^*(\lambda) = \Lambda^*(q)$ e de maneira similar vemos que $\inf_{\lambda\in F^o} \Lambda^*(\lambda) = \Lambda^*(q)$. Por (1.40) e pelo Teorema de Crámer, temos então que

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left[\frac{L_n}{n} > q\right] = \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left[\frac{L_n}{n} \in F\right] = q \ln\left(\frac{q}{p}\right) + (1 - q) \ln\left(\frac{1 - q}{1 - p}\right).$$

2.3 Devedores Dependentes

Recomendamos ao leitor deste trabalho, a prévia leitura o Apêndice C, para que possa ter um entendimento prévio do por que podemos fazer as contas a seguir, a partir da definição C.1.1.

Para começarmos, seja $\mathcal{B}$ a σ-álgebra de Borel na reta. Para dados $k \in \mathbb{N}$ e $B \in \mathcal{B}$, consideremos a seguinte probabilidade condicional

$$\mathbb{P}[\epsilon_k \in B|Z], \tag{2.3}$$

assim vemos que conseguimos uma família $\{\varphi_{k,B}\}_{B\in\mathcal{B}}$ (ver Teorema C.2.1) de funções Borel mensuráveis em $\mathbb{R}$, onde $\mathcal{B}$ representa a σ-álgebra de Borel da reta, tal que

$$\mathbb{P}[\epsilon_k \in B|Z] = \varphi_{k,B}(Z).$$

Segue de C.5 que esta família satisfaz a definição C.3.1, para ϵ_k dado Z. Pelo Teorema C.4.2 e da definição C.3.2, existe $B_0 \in \mathcal{B}$, tal que $\mathbb{P}[Z \in B_0] = 1$ e

$$\mathbb{P}[\epsilon_k \in B|Z = z] = \lim_{\Delta_z \to 0} \mathbb{P}[\epsilon_k \in B|Z \in I_z] \text{ para todo } z \in B_0, \tag{2.4}$$

onde I_z é um intervalo contendo z, Δ_z é o tamanho do intervalo I_z e $\mathbb{P}[\epsilon_k \in B|Z = z] = \varphi_{k,B}(z)$. Como $\mathbb{P}[Z \in I_z] > 0$, então para dado $z \in \mathbb{R}$ vale

$$\lim_{\Delta_z \to 0} \mathbb{P}[\epsilon_k \in B|Z \in I_z] = \mathbb{P}[\epsilon_k \in B].$$

Assim vemos que

$$\varphi_{k,B}(Z) = \mathbb{P}[\epsilon_k \in B] \text{ q.t.p.}.$$

Então sem perda de generalidade tomaremos

$$\varphi_{k,B}(z) = \mathbb{P}[\epsilon_k \in B] \text{ para todo } z \in \mathbb{R},$$

em particular temos

$$\mathbb{P}[\epsilon_k \in B|Z = z] = \mathbb{P}[\epsilon_k \in B] \text{ para todo } z \in \mathbb{R}.$$

Com isto feito, consideremos $\rho > 0$. Assim dado $k \in \mathbb{N}$, tomemos p_k dada pelo Teorema C.2.1, com relação variável aleatória $1_{\{Y_k=1\}}$ dada Z, temos novamente pela definição C.3.1 que

$$p_k(z) := \mathbb{P}[Y_k = 1|Z = z] = \mathbb{P}[X_k > x_k|Z = z].$$

Segue do Teorema C.4.1 que

$$p_k(z) = \mathbb{P}\left[\rho Z + \sqrt{1-\rho^2}\epsilon_k > -\Phi^1(p)|Z = z\right] = \mathbb{P}\left[\rho z + \sqrt{1-\rho^2}\epsilon_k > -\Phi^1(p)\right].$$

Portanto concluímos que

$$p_k(z) = \Phi\left(\frac{\rho z + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right).$$

Como a igualdade acima independe de $k \in \mathbb{N}$, definimos então a função

$$p(z) = \Phi\left(\frac{\rho z + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right) \tag{2.5}$$

Para que os eventos $\{L_n \geq l_n\}$ se tornem raros e também não impossíveis a medida que n tende ao infinito, vamos assumir que $\frac{l_n}{n}$ ira se aproximar de 1 por baixo. Definimos então em particular que

$$l_n = nq_n, \text{ com } q_n \nearrow 1, \text{ quando } n \to \infty \text{ e } q_n < 1. \tag{2.6}$$

Vamos assumir que q_n converge para 1 tal que

$$1 - q_n \simeq n^{-a}, \text{ com } a < 1, \tag{2.7}$$

onde o simbolo $f(x) \simeq g(x)$ significa que

$$\lim_{x \to +\infty} \frac{f(x)}{g(x)} = 1.$$

Faremos agora um resultado de grandes desvios para regime de limite de grandes perdas.

Teorema 2.3.1. *Consideremos um modelo de fator único de um portfólio de risco de credito, modelado por 2.2, e com o regime limite l_n como em 2.6-2.7. Temos então que*

$$\lim \frac{1}{\ln n} \ln \mathbb{P}[L_n \geq nq_n] = -a\frac{1-\rho^2}{\rho} \tag{2.8}$$

Demonstração: Primeiramente mostraremos o limite inferior

$$\liminf_{n \to \infty} \frac{1}{\ln n} \ln \mathbb{P}[L_n \geq nq_n] \geq -a\frac{1-\rho^2}{\rho}. \tag{2.9}$$

Notemos primeiramente que como p é uma bijeção crescente de $\mathbb{R}$ em $(0,1)$, segue que para cada z_n, existe um único q_n, tal que $p(z_n) = q_n$. Mostraremos agora que para todo $n \in \mathbb{N}$, vale que

$$\mathbb{P}[L_n \geq nq_n] \geq \mathbb{P}[L_n \geq nq_n | Z = z_n]\mathbb{P}[Z \geq z_n]. \tag{2.10}$$

Com efeito, notemos que fixado $z \in \mathbb{R}$, a função dada por $\mathbb{P}[L_n(z) \in B | Z = x] := \mathbb{P}[L_n(z) \in B]$, onde

$$L_n(z) = \sum_{k=1}^{n} Y_k(z),$$

com $Y_k(z) = 1_{\{\rho z + \sqrt{1-\rho^2}\epsilon_k > -\Phi^{-1}(p)\}}$, satisfaz a definição C.3.1. Logo vale que $\mathbb{P}[L_n(z) \in B]$[1] é uma distribuição condicional (regular) para $L_n(z)$ dada Z. Segue então do Teorema C.4.1 que

$$\mathbb{P}[L_n(z_n) \geq nq_n | Z = z_n]\mathbb{P}[Z \geq z_n] = \mathbb{P}[L_n(z_n) \geq nq_n]\mathbb{P}[Z \geq z_n] = \mathbb{P}[L_n(z_n) \geq nq_n, Z \geq z_n].$$

[1] Aqui estamos considerando apenas a definição C.3.1, pois é mais conveniente assim, entretanto podemos entender $\mathbb{P}[L_n(z) \in B]$ como sendo a família de funções Borel mensuráveis dadas pelo Teorema C.2.1, para $L_n(z)$ dado Z, onde a construção desta família é feita de maneira análoga construção que fizemos para ϵ_k dado Z.

Observemos que pelo fato de $L_n(s) \leq L_n(r)$, se $s < r$, segue que $\{L_n(z_n) \geq nq_n, Z \geq z_n\} \subseteq \{L_n \geq nq_n\}$, com isso concluímos a validade de (2.10).

Notemos agora que quando fixarmos $Z = z_n$, ás variáveis aleatórias Y_k, ficarão dependendo só das ϵ_k, assim vemos que elas são independentes e observando que

$$\mathbb{P}[Y(z_n)_k = 1] = \mathbb{P}\left[\rho z + \sqrt{1-\rho^2}\epsilon_k > -\Phi^1(p)\right] = p(z_n),$$

concluímos que elas são identicamente distribuídas. Portanto $L_n(z_n)$ é uma binomial de parâmetros n, q_n e como $q_n \to 1$, segue de (GREENBERG; MOHRI, 2013) e do Teorema C.4.1 que

$$\mathbb{P}[L_n \geq nq_n | Z = z_n] = \mathbb{P}[L_n(z_n) \geq nq_n | Z = z_n] = \mathbb{P}[L_n(z_n) \geq nq_n] \geq \frac{1}{4}, \tag{2.11}$$

para n suficientemente grande.

Como q_n tende a 1 quando n tende ao $+\infty$, vemos que z_n tende a $+\infty$ quando n tende a $+\infty$. Pelas hipóteses que temos e usando o fato que $1 - \Phi(x) \simeq \frac{\varphi(x)}{x}$ quando $x \to +\infty$. Segue que

$$n^{-a} \simeq 1 - q_n = 1 - p(z_n) = 1 - \Phi\left(\frac{\rho z_n + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right) \simeq \frac{\sqrt{1-\rho^2}}{\rho z_n + \Phi(p)} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\rho z_n + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right)^2\right\},$$

logo

$$n^{-a} \simeq \frac{\sqrt{1-\rho^2}}{\rho z_n + \Phi(p)} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\rho z_n + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right)^2\right\}.$$

Portanto temos que dado $\epsilon \in (0, 1)$, vale que

$$1 - \epsilon < n^a \frac{\sqrt{1-\rho^2}}{\rho z_n + \Phi(p)} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\rho z_n + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right)^2\right\} < 1 + \epsilon,$$

para n suficientemente grande. Logo existe $R > 0$, tal que

$$\left| a \ln n - \frac{1}{2}\frac{\rho^2 z_n^2}{1-\rho^2} - \frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2} - \ln\left(z_n + \frac{\Phi^{-1}(p)}{\rho}\right)\right| < R,$$

então tem-se

$$\frac{2(1-\rho^2)}{\rho^2}\left(-\frac{\ln\left(z_n + \frac{\Phi^{-1}(p)}{\rho}\right)}{\ln n} - \frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n} - \frac{R}{\ln n} + a\right) < \frac{z_n^2}{\ln n} < \frac{2(1-\rho^2)}{\rho^2}\left(-\frac{\ln\left(z_n + \frac{\Phi^{-1}(p)}{\rho}\right)}{\ln n} - \frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n} + \frac{R}{\ln n} + a\right).$$

Mostraremos agora que os termos

$$\frac{\ln\left(z_n + \frac{\Phi^{-1}(p)}{\rho}\right)}{\ln n} \text{ e } \frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n}$$

tendem a zero quando $n \to +\infty$. Com efeito, notemos que se $p < \frac{1}{2}$, então $\Phi^{-1}(p) < 0$, logo

$$\frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n} < 0 \text{ e } \frac{\ln\left(z_n + \frac{\Phi^{-1}(p)}{\rho}\right)}{\ln n} > 0$$

para n suficientemente grande. Logo

$$\frac{z_n^2}{\ln n} < \frac{2(1-\rho^2)}{\rho^2}\left(-2\frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n} + \frac{R}{\ln n} + a\right).$$

Como $\frac{R}{\ln n} \to 0$, quando $n \to +\infty$, vale que existe $C > 0$ tal que

$$\frac{z_n^2}{\ln n} < \frac{2(1-\rho^2)}{\rho^2}\left(-2\frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n} + C + a\right). \tag{2.12}$$

Notemos que

$$\lim_{n\to+\infty} -2\frac{\rho z_n \Phi^{-1}(p)}{(\ln n)(1-\rho^2)}\frac{\ln n}{z_n^2} = 0, \tag{2.13}$$

assim vale que se $\frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n}$ não converge para zero, existe uma subsequência z_{n_k} tal que

$$\frac{z_{n_k}^2}{\ln n_k} \longrightarrow +\infty.$$

Segue de 2.13 e de 2.12, que para k suficientemente grande vale

$$\frac{z_{n_k}^2}{\ln n_k} < \frac{2(1-\rho^2)}{\rho^2}\left(-2\frac{\frac{\rho z_{n_k} \Phi^{-1}(p)}{1-\rho^2}}{\ln n_k} + C + a\right) \leq \frac{z_{n_k}^2}{\ln n_k},$$

o que seria um absurdo. Esse absurdo acontece se assumirmos que $\frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n}$ não converge para zero. Logo concluímos que $\frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n}$ converge para zero. Segue de L'Hospital que

$$\lim_{z\to+\infty} \frac{ln\left(z + \frac{\Phi^{-1}(p)}{\rho}\right)}{\frac{z\Phi^{-1}(p)}{1-\rho^2}} = 0$$

e pelo fato de

$$\frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n} \longrightarrow 0, \text{ quando } n \to +\infty$$

segue que

$$\frac{\ln\left(z_n + \frac{\Phi^{-1}(p)}{\rho}\right)}{\ln n} \longrightarrow 0.$$

Se $p \geq \frac{1}{2}$, então $\Phi^1(p) \geq 0$, logo existe $0 < S < +\infty$ tal que

$$\frac{z_n^2}{\ln n} < \frac{2(1-\rho^2)}{\rho^2}\left(-\frac{\ln\left(z_n + \frac{\Phi^{-1}(p)}{\rho}\right)}{\ln n} - \frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n} + \frac{R}{\ln n} + a\right) < S,$$

para n suficientemente grande. Então por argumentação análoga, vemos que se $\frac{\ln\left(z_n+\frac{\Phi^{-1}(p)}{\rho}\right)}{\ln n}$ e $\frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n}$ não tendem a zero quando $n \to +\infty$, teríamos que existe uma subsequência n_k, tal que

$$\frac{z_{n_k}^2}{\ln n_k} \longrightarrow +\infty,$$

o que seria um absurdo, pois assim teríamos que $S = +\infty$. Portanto concluímos que $\frac{\ln\left(z_n+\frac{\Phi^{-1}(p)}{\rho}\right)}{\ln n}$ e $\frac{\frac{\rho z_n \Phi^{-1}(p)}{1-\rho^2}}{\ln n}$ tendem a zero quando $n \to +\infty$. A partir deste fato demonstrado, vemos que

$$\lim_{n\to\infty} \frac{z_n^2}{\ln n} = 2a\frac{(1-\rho^2)}{\rho^2}.$$

Observando então que

$$\mathbb{P}[Z \geq z_n] \geq \mathbb{P}[z_n + 1 \geq Z \geq z_n] = \int_{z_n}^{z_n+1} \varphi(x)dx \geq \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(z_n+1)^2\},$$

segue então 2.10 e 2.11 que

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{\ln n} \ln \mathbb{P}[L_n \geq nq_n] &\geq \liminf_{n\to\infty} \left(\frac{\ln \mathbb{P}[Z \geq z_n]}{\ln n} + \frac{\ln \frac{1}{4}}{\ln n}\right) \\ &\geq \liminf_{n\to+\infty} \left(\frac{\ln \frac{1}{\sqrt{2\pi}}}{\ln n} - \frac{1}{2}\frac{(z_n+1)^2}{\ln n}\right) = -a\frac{1-\rho^2}{\rho^2}.\end{aligned}$$

Com isso demonstramos (2.9). Vamos iniciar agora a demonstração do limite superior

$$\limsup_{n\to\infty} \frac{1}{\ln n} \ln \mathbb{P}[Z \geq z_n] \leq -a\frac{1-\rho^2}{\rho^2}. \tag{2.14}$$

Com efeito, começaremos introduzindo a seguinte função

$$\Gamma(\theta, z) := \ln \mathbb{E}[e^{\theta Y_k}|Z=z] = \ln \mathbb{E}[1 + Y_k\left(e^{\theta}-1\right)|Z=z] = \ln(1-p(z)+e^{\theta}p(z)). \tag{2.15}$$

Observemos agora que para todo $\theta > 0$ vale que

$$1_{[L_n - nq_n \geq 0]} \leq e^{\theta(L_n - nq_n)},$$

logo temos para $\theta > 0$ que

$$\mathbb{E}\left[1_{[L_n-nq_n\geq 0]}|Z\right] \leq \mathbb{E}[e^{\theta([L_n-nq_n)}|Z] \Longrightarrow \mathbb{P}[L_n - nq_n \geq 0|Z] \leq \mathbb{E}[e^{\theta([L_n-nq_n)}|Z].$$

Agora observemos que pelo Teorema C.4.1, segue a seguinte cadeia de igualdades

$$\mathbb{E}\left[e^{\theta L_n}|Z\right] = \sum_{i=1}^{n} e^{\theta i}\mathbb{E}\left[1_{L_n=i}|Z\right] = \sum_{i=1}^{n} e^{\theta i}\mathbb{P}\left[L_n = i|Z\right] = \sum_{i=1}^{n} e^{\theta i}\mathbb{P}\left[L_n(Z) = i\right]^{2} =$$

[2] Aqui estamos considerando como sendo a família de funções Borel mensuráveis dadas pelo Teorema C.2.1.

$$\sum_{i=1}^{n} e^{\theta i} p(Z)^i (1-p(Z))^{n-i} = (e^\theta p(Z) + (1-p(Z)))^n = \left(\mathbb{E}\left[e^{\theta Y_k}|Z\right]\right)^n = e^{n\Gamma(\theta,Z)}.$$

Temos então para $\theta > 0$ que

$$\mathbb{P}[L_n - nq_n \geq 0|Z] \leq e^{-n\theta q_n}\mathbb{E}\left[\prod_{i=1}^{n} e^{\theta Y_i}|Z\right] = e^{-n(\theta q_n - \Gamma(\theta,Z))}. \tag{2.16}$$

Definiremos agora a função

$$\Gamma^*(q,z) := \sup_{\theta\geq 0}[\theta q - \Gamma(\theta,z)], \tag{2.17}$$

logo

$$\mathbb{P}[L_n - nq_n \geq 0|Z] \leq e^{-n\Gamma^*(q_n,Z)}.$$

Consideremos agora a função

$$g(\theta) := \theta q - \Gamma(\theta,z), \tag{2.18}$$

assim

$$g'(\theta) = \frac{q - qp(z) + e^\theta qp(z) - e^\theta p(z)}{1 - p(z) + e^\theta p(z)}.$$

Disto conseguimos que

$$g'(\theta) \begin{cases} < 0, \text{ se } \theta > \ln\frac{q}{p(z)} + \ln\frac{1-p(z)}{1-q} \\ = 0, \text{ se } \theta = \ln\frac{q}{p(z)} + \ln\frac{1-p(z)}{1-q} \\ > 0, \text{ se } \theta < \ln\frac{q}{p(z)} + \ln\frac{1-p(z)}{1-q}. \end{cases} \tag{2.19}$$

Notemos agora que se $q \leq p(z)$, então

$$\ln\frac{q}{p(z)} + \ln\frac{1-p(z)}{1-q} \leq 0,$$

segue de (2.19) que $\Gamma^*(q,z) = 0$. Agora se $q > p(z)$, então

$$\ln\frac{q}{p(z)} + \ln\frac{1-p(z)}{1-q} > 0,$$

então novamente de (2.19), segue que $\Gamma^*(q,z) = g\left(\ln\frac{q}{p(z)} + \ln\frac{1-p(z)}{1-q}\right)$. Portanto vale

$$\Gamma^*(q,z) \begin{cases} = 0 & , \text{ se } q \leq p(z) \\ = q\ln\frac{q}{p(z)} + (1-q)\ln\frac{1-q}{1-p(z)} & , \text{ se } q > p(z). \end{cases}$$

Tomando agora

$$F_n(z) = -n\Gamma^*(q_n,z), \tag{2.20}$$

temos por (2.16) que

$$\mathbb{E}[\mathbb{P}[L_n - nq_n \geq 0|Z]] \leq \mathbb{E}[e^{-n\Gamma^*(q_n,Z)}] \Longrightarrow \mathbb{P}[L_n \geq nq_n] \leq \mathbb{E}[e^{F_n(Z)}].$$

Notemos agora que como $\rho > 0$, p é uma função crescente de z, então dados $z_2 > z_1$, vale para $\theta \geq 0$ que

$$1 - p(z_2) + e^{\theta}p(z_2) - (1 - p(z_1) + e^{\theta}p(z_1)) \geq 0 \Longrightarrow \Gamma(\theta, z_2) - \Gamma(\theta, z_1) \geq 0,$$

disso concluímos que Γ é uma função não-decrescente em z, para $\theta \geq 0$. Assim vemos que Γ^* é uma função não-crescente em z, para todo $q \in [0, 1]$, e disto podemos concluir que $F_n(z)$ é uma função não-decrescente não positiva que atinge seu valor máximo em 0.

Mostraremos agora que F_n é côncava. Lembrando que $p(z_n) = q_n$, vemos facilmente que F_n é diferenciável nos intervalos $(-\infty, z_n)$ e $(z_n, +\infty)$, com

$$F_n'(z) = n\varphi\left(\frac{\rho z + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right)\left(\frac{q_n}{p(z)} - \frac{1-q_n}{1-p(z)}\right)\frac{\rho}{\sqrt{1-\rho^2}}, \tag{2.21}$$

se $z \in (-\infty, z_n)$ e

$$F_n'(z) = 0,$$

se $z \in (z_n, +\infty)$. Observando que

$$\lim_{z \to z_n^-} \frac{F_n(z) - F_n(z_n)}{z - z_n} = \lim_{z \to z_n^+} F_n'(z) = 0 = \lim_{z \to z_n^+} \frac{F_n(z) - F_n(z_n)}{z - z_n},$$

podemos então concluir que F_n é diferenciável em z_n, em particular vemos facilmente que ela é C^1. Como $F_n'(z)$ é não-crescente em $(-\infty, z_n)$, pois p é uma função crescente em $(-\infty, z_n)$, e como em $[z_n, \infty)$ vale que F_n' é constante, segue da continuidade de F_n' que ela é não crescente e portanto F_n é uma função côncava.

Consideremos então a medida $\mathbb{P}_\mu$, dada por

$$\frac{d\mathbb{P}_\mu}{d\mathbb{P}} = \exp(\mu Z - \frac{1}{2}\mu^2),$$

notemos que por argumentação análoga a utilizada no Teorema de Cramer que $\mathbb{P}_\mu$ é uma medida de probabilidade. A esperança de $\mathbb{P}_\mu$ sera denotada por

$$\mathbb{E}_\mu[Y] := \int_\Omega Y d\mathbb{P}_\mu.$$

Temos então

$$\mathbb{E}\left[e^{F_n(Z)}\right] = \mathbb{E}_\mu\left[e^{F_n(Z) - \mu Z + \frac{1}{2}\mu^2}\right].$$

Pelo fato de F_n ser concava, segue que

$$F_n(Z) \leq F_n(\mu) + F_n'(\mu)(Z - \mu),$$

assim

$$\mathbb{E}\left[e^{F_n(Z)}\right] \leq \mathbb{E}_\mu\left[e^{F_n(\mu) + (F_n'(\mu) - \mu)Z - \mu F_n' + \frac{1}{2}\mu^2}\right].$$

Observemos agora que para $z \to -\infty$, vale que

$$\left(\frac{q_n}{p(z)} - \frac{1-q_n}{1-p(z)}\right) \to +\infty,$$

segue então de (2.21) que $F_n'(z) - z > 0$, para z suficientemente pequeno e agora observando que para $z \geq z_n$ vale $F_n'(z) = 0$, segue que $F_n'(z) - z < 0$, para $z \geq z_n$. Portanto temos por continuidade que existe μ_n, tal que $F_n'(\mu_n) = \mu_n$. Então vale que

$$\mathbb{P}[L_n \geq nq_n] \leq \mathbb{E}[e^{F_n(Z)}] \leq e^{F_n(\mu_n) - \frac{1}{2}\mu_n^2}.$$

Mostraremos agora que $\frac{\mu_n}{z_n} \to 1$, quando $n \to +\infty$. Para isso mostraremos que dado $0 < \epsilon < 2$, vale para n suficientemente grande que $z_n(1-\epsilon) \leq \mu_n \leq z_n$. Notemos que é suficiente mostrar que

$$F_n'(z_n(1-\epsilon)) - z_n(1-\epsilon) \geq F_n'(\mu_n) - \mu_n \geq F_n'(z_n) - z_n,$$

pois $F_n'(z) - z$ é a derivada de uma função côncava diferenciável e portanto não-crescente. Observemos que como $F_n'(z_n) = 0$ e $z_n \to +\infty$, quando $n \to +\infty$, temos

$$F_n'(\mu_n) - \mu_n = 0 \geq F_n'(z_n) - z_n$$

para n suficientemente grande. Mostraremos agora que $F_n'(z_n(1-\epsilon)) - z_n(1-\epsilon) \geq 0 = F_n'(\mu_n) - \mu_n$. Com efeito, consideremos primeiramente o limite

$$\lim_{n\to\infty}\left(\frac{\rho z_n(1-\epsilon)+\phi^{-1}(p)}{\rho z_n+\phi^{-1}(p)}\right)\exp\left(-\frac{1}{2}\left(\frac{\rho z_n+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)^2+\frac{1}{2}\left(\frac{\rho z_n(1-\epsilon)+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)^2\right)$$

e observemos que termo

$$\left(\frac{\rho z_n(1-\epsilon)+\phi^{-1}(p)}{\rho z_n+\phi^{-1}(p)}\right) \to 1-\epsilon,$$

quando $n \to +\infty$, e o termo

$$(\rho z_n \epsilon)^2 - 2(\rho z_n)^2\epsilon - 2\rho z_n \epsilon \phi^{-1}(p) \to -\infty,$$

quando $n \to +\infty$. Portanto podemos concluir que o limite acima é igual a 0. Lembrando agora que $1 - \Phi(x) \simeq \frac{\varphi(x)}{x}$ quando $x \to \infty$ e observando que

$$\frac{\varphi\left(\frac{\rho z_n+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)}{\left(\frac{\rho z_n+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)}\frac{\left(\frac{\rho z_n(1-\epsilon)+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)}{\varphi\left(\frac{\rho z_n(1-\epsilon)+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)} =$$

$$\left(\frac{\rho z_n(1-\epsilon)+\phi^{-1}(p)}{\rho z_n+\phi^{-1}(p)}\right)\exp\left(-\frac{1}{2}\left(\frac{\rho z_n+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)^2+\frac{1}{2}\left(\frac{\rho z_n(1-\epsilon)+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)^2\right),$$

tem-se que

$$\lim_{n\to+\infty} \frac{1-p(z_n)}{1-p(z_n(1-\epsilon))} =$$

$$\lim_{n\to+\infty 0} \frac{1-p(z_n)}{1-p(z_n(1-\epsilon))} \left(\frac{\varphi\left(\frac{\rho z_n+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)}{\left(\frac{\rho z_n+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)} \frac{\left(\frac{\rho z_n(1-\epsilon)+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)}{\varphi\left(\frac{\rho z_n(1-\epsilon)+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)} \right)^{-1} \frac{\varphi\left(\frac{\rho z_n+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)}{\left(\frac{\rho z_n+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)} \frac{\left(\frac{\rho z_n(1-\epsilon)+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)}{\varphi\left(\frac{\rho z_n(1-\epsilon)+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right)}$$

$$= 1 * 0 = 0.$$

E também temos que

$$\lim_{n\to+\infty} \frac{p(z_n)}{p(z_n(1-\epsilon))} = \lim_{n\to+\infty} \frac{-1+1-p(z_n)}{-1+1-p(z_n(1-\epsilon))} = 1,$$

lembrando que $p(z_n) = q_n$. Notemos então que valem:

1. $\frac{n}{\sqrt{\ln n}} \to +\infty$,
2. $\varphi\left(\frac{\rho z_n(1-\epsilon)+\phi^{-1}(p))}{\sqrt{1-\rho^2}}\right) \to +\infty$,
3. $\left(\frac{q_n}{p(z_n(1-\epsilon))} - \frac{1-q_n}{1-p(z_n(1-\epsilon)}\right) > \frac{1}{2}$,

quando $n \to +\infty$. Segue então de (2.21) que para todo $M > 0$, vale que existe $n_M \in \mathbb{N}$, tal que $n > n_M$ implica que

$$M\sqrt{\ln n} < F_n'(z_n(1-\epsilon)).$$

Lembrando que

$$\frac{z_n^2}{\ln n} \to 2a\frac{(1-\rho^2)}{\rho^2},$$

quando $n \to +\infty$, segue que $z_n(1-\epsilon) = O(\sqrt{\ln n})$, ou seja existe $N_0 > 0$ tal que

$$z_n(1-\epsilon) < N_0\sqrt{\ln n},$$

para n suficientemente grande. Logo temos para n suficientemente grande que

$$F_n'(z_n(1-\epsilon)) - z_n(1-\epsilon) \geq F_n'(\mu_n) - \mu_n.$$

Portanto concluímos que $\frac{\mu_n}{z_n} \to 1$.

Como F_n é não positiva, segue que

$$\limsup_{n\to+\infty} \frac{1}{\ln n} \ln \mathbb{P}[L_n \geq nq_n] \leq \limsup_{n\to+\infty} \frac{1}{\ln n} \ln e^{F_n - \frac{1}{2}\mu_n^2} \leq \limsup_{n\to+\infty} \frac{1}{\ln n}(-\frac{1}{2}\mu_n^2) =$$

$$-\frac{1}{2}\lim_{n\to+\infty}\frac{z_n^2}{\ln n} = -a\frac{(1-\rho^2)}{\rho^2}.$$

Com esta ultima desigualdade concluímos a nossa demonstração. □

Caso o leitor tenha interesse em continuar seus estudos nessa área, indicamos ao mesmo que veja o artigo Large Deviations of Multifactor Portfolio Credit Risk(ver (GLASSERMAN; KANG; SHAHABUDDIN, 2007)), no qual as notas de aula Some Applications and Methods of Large Deviations in Finance and Insurance foram baseadas. Indicaremos também o artigo Sharp asymptotics for large portfolio losses under extreme risks(ver (TANG; TANG; YANG, 2019)), que é um artigo mais recente e que foi feito com base na preocupação com os eventos da crise econômica que aconteceu entre os anos de 2008 e 2009.

Referências

BARTLE, R. *The Elements of Integration*. Wiley, 1966. (Wiley Classics Library). ISBN 9780471054573. Disponível em: <https://books.google.com.br/books?id=UrLvAAAAMAAJ>. Citado 2 vezes nas páginas 59 e 62.

CHUNG, K. *A Course in Probability Theory*. Elsevier Science, 2001. ISBN 9780121741518. Disponível em: <https://books.google.com.br/books?id=zu80w7wd40UC>. Citado 4 vezes nas páginas 61, 62, 64 e 67.

DEMBO, A.; ZEITOUNE, O. *Large Deviations and Techinques and Aplications*. [S.l.]: ACADEMIC PRESS, INC, 1998. v. 2. Citado 4 vezes nas páginas 11, 14, 16 e 43.

DONSKER, M.; VARADHAN, S. Asymptotic evaluation of certain markov process expectations for large time—iii. *Communications on Pure and Applied Mathematics*, Wiley-Liss Inc., v. 29, n. 4, p. 389–461, jul. 1976. ISSN 0010-3640. Copyright: Copyright 2016 Elsevier B.V., All rights reserved. Citado na página 11.

DONSKER, M.; VARADHAN, S. Asymptotic evaluation of certain markov process expectations for large time. iv. *Communications on Pure and Applied Mathematics*, Wiley-Liss Inc., v. 36, n. 2, p. 183–212, mar. 1983. ISSN 0010-3640. Citado na página 11.

DONSKER, M. D.; VARADHAN, S. R. S. Asymptotic evaluation of certain markov process expectations for large time, i. *Communications on Pure and Applied Mathematics*, v. 28, n. 1, p. 1–47, 1975. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160280102>. Citado na página 11.

DONSKER, M. D.; VARADHAN, S. R. S. Asymptotic evaluation of certain markov process expectations for large time, ii. *Communications on Pure and Applied Mathematics*, v. 28, n. 2, p. 279–301, 1975. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160280206>. Citado na página 11.

ELLIS, R. S. The theory of large deviations: from boltzmann's 1877 calculation to equilibrium macrostates in 2d turbulence. *Physica D: Nonlinear Phenomena*, v. 133, p. 106–136, 1999. Citado na página 11.

FREIDLIN, M. I.; WENTZELL, A. D. Random perturbations. In: ______. *Random Perturbations of Dynamical Systems*. New York, NY: Springer US, 1984. p. 15–43. ISBN 978-1-4684-0176-9. Disponível em: <https://doi.org/10.1007/978-1-4684-0176-9_2>. Citado na página 11.

GLASSERMAN, P.; KANG, W.; SHAHABUDDIN, P. Large deviations in multifactor portfolio credit risk. *Mathematical Finance*, v. 17, n. 3, p. 345–379, 2007. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9965.2006.00307.x>. Citado 2 vezes nas páginas 11 e 56.

GREENBERG, S.; MOHRI, M. Tight lower bound on the probability of a binomial exceeding its expectation. *CoRR*, abs/1306.1433, 2013. Disponível em: <http://arxiv.org/abs/1306.1433>. Citado na página 49.

JAMES, B. R. *Probabilidade:Um Curso de Nível Intermediário*. [S.l.]: Grafica Editora Hamburg Ltda., 1981. Citado na página 64.

TANG, Q.; TANG, Z.; YANG, Y. Sharp asymptotics for large portfolio losses under extreme risks. *European Journal of Operational Research*, v. 276, n. 2, p. 710–722, 2019. Disponível em: <https://ideas.repec.org/a/eee/ejores/v276y2019i2p710-722.html>. Citado na página 56.

APÊNDICE A – Definições

Esta seção conterá as definições de objetos que foram utilizados neste trabalho.

A.1 Definições da Teoria da Medida

As definições desta seção foram retiradas do livro (BARTLE, 1966).

Definição A.1.1 (σ-Álgebra)**.** *Seja X um conjunto. Dizemos que uma família $\mathcal{X}$ de subconjuntos de X é uma σ-**Álgebra** se:*

1. *$\emptyset$ e X pertencem a $\mathcal{X}$.*
2. *Se $A \in \mathcal{X}$, então o complementar de A pertence a $\mathcal{X}$, ou seja $A^c \in \mathcal{X}$.*
3. *Se $A_n \in \mathcal{X}$ é uma sequência de conjuntos então $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{X}$.*

Definição A.1.2. *O par ordenado $(X, \mathcal{X})$ consistindo de um conjunto X e uma σ-Álgebra $\mathcal{X}$ de X, é chamado de **Espaço Mensurável**. Se $A \in \mathcal{X}$, chamamos A de conjunto $\mathcal{X}$-mensurável, quando a $\sigma-$Álgebra é fixada chamaremos A de conjunto mensurável.*

Definição A.1.3 (Medida)**.** *Uma **medida** μ é uma função real estendida da σ-Álgebra $\mathcal{X}$ em $[0, +\infty]$, tal que*

1. *$\mu(\emptyset) = 0$.*
2. *$\mu(E) \geq 0$ para todo $E \in \mathcal{X}$.*
3. *$E_n \in \mathcal{X}$ uma sequência de conjuntos disjuntos quaisquer, então*

$$\mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \sum_{n=1}^{+\infty} E_n.$$

Definição A.1.4. *Um **espaço de medida** é uma tripla $(X, \mathcal{X}, \mu)$ consistindo de um conjunto X, uma σ-Álgebra $\mathcal{X}$ de X, e uma medida μ em $\mathcal{X}$.*

Definição A.1.5. *Seja B o conjunto de todos os intervalos abertos (a, b) de $\mathcal{R}$. chamaremos de σ-**Álgebra de Borel** a σ-Álgebra gerada por B.*

Definição A.1.6 (Função Mensurável)**.** *Seja f uma função de X em $\mathbb{R}$ é chamada de $\mathcal{X}$-**mensurável** se para todo numero real a o conjunto*

$$\{x \in X; f(x) > a\}$$

*pertence a $\mathcal{X}$. Quando $\mathcal{X}$ estiver fixado, chamaremos f de **mensurável**.*

Definição A.1.7 (Função Simples)**.** *Uma função f de X em $\mathbb{R}$ é dita ser simples quando pode ser escrita da seguinte forma*

$$f = \sum_{j=1}^{l} a_j 1_{E_j},$$

onde $E_j \in \mathcal{X}$.

Definição A.1.8 (Função Integrável)**.** *Seja f uma função do espaço de medida $(X, \mathcal{X}, \mu)$ em $\mathbb{R}$, então:*

1. *Se $f = \sum_{j=1}^{l} a_j 1_{E_j}$, com $a_j \geq 0$, chamamos de **integral** de f o número real*

$$\int f d\mu = \sum_{j=1}^{l} a_j \mu(E_j).$$

2. *Se f é uma função não-negativa, chamamos de **integral** de f o numero real ou estendido*

$$\int f d\mu = \sup_{\varphi \in H} \int \varphi d\mu,$$

onde H é o conjunto de todas as funções simples não-negativas de $(X, \mathcal{X}, \mu)$ em $\mathbb{R}$.

3. *Se f é uma função arbitraria, então $f = f^+ - f^-$, onde $f^+ = \max\{0, f\}$ e $f^- = \max\{0, -f\}$. Dizemos então que f é **integrável** quando*

$$\textstyle\int f^+ d\mu \in \mathbb{R} \text{ e } \int f^- d\mu \in \mathbb{R},$$

*e definimos a **integral** de f como*

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

Observação: Na definição acima no caso 2, a integral pode assumir o valor $+\infty$.

Definição A.1.9. *Chamaremos de espaço L^p, o conjunto*

$$L^p(X, \mathcal{X}, \mu) := \{f : X \longrightarrow \mathbb{R};\ f \text{ é mensurável e } ||f||_p < +\infty\},$$

onde

$$||f||_p := \left(\int |f|^p d\mu\right)^{\frac{1}{p}}$$

Definição A.1.10 (Medida σ-Finita)**.** *Dizemos que μ é uma medida σ**-finita** quando existe uma sequência $E_n \in \mathcal{X}$ tal que $X = \cup_{n\in\mathbb{N}}$ e $\mu(E_n) < +\infty$*

A.2 Definições de Teoria da Probabilidade

As definições desta seção foram retiradas do livro (CHUNG, 2001). Então sejam Ω um conjunto e $\mathcal{F}$ uma σ-Álgebra de Ω.

Definição A.2.1 (Medida de Probabilidade)**.** *Seja* $\mathbb{P}$ *um medida no espaço mensurável* $(\Omega, \mathcal{F})$. *Dizemos que* $\mathbb{P}$ *é uma* ***Medida de Probabilidade*** *quando* $\mathbb{P}(\Omega) = 1$.

Definição A.2.2 (Variável Aleatória)**.** *Uma* ***Variável Aleatória Real*** *é uma função* f *com domínio em* Ω *e contra-domínio* $\mathbb{R}$*, tal que para cada* $B \in \mathcal{B}$*, onde* $\mathcal{B}$ *é a* σ*-álgebra de Borel, tem-se*

$$\{\omega \in \Omega; f(\omega) \in B\} \in \mathcal{F}.$$

Definição A.2.3 (Esperança)**.** *A* ***esperança*** *de uma variável aleatória* Y*, quando existir, sera dada por*

$$\mathbb{E}[Y] = \int Y d\mathbb{P}$$

Definição A.2.4. *Seja* $\mathcal{A}$ *uma coleção não vazia de conjuntos de Ω. Dizemos que* $\mathcal{A}$ *é um* ***corpo****, se satisfaz:*
i) $E \in \mathcal{A} \Longrightarrow E^c \in \mathcal{A}$.
ii)$E_1 \in \mathcal{A}, E_2 \in \mathcal{A} \Longrightarrow E_1 \cup E_2 \in \mathcal{A}$

Definição A.2.5. *Sejam* X *e* Y *duas variáveis aleatórias definidas no mesmo espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbb{P})$. *Uma função* $\mathbb{P}\left[X \in B | Y = y\right]$, *definida para* B *boreliano e* $y \in \mathbb{R}$ *sera chamada de* ***distribuição condicional*** *para* X *dada* Y*, se:*
(i) para todo $y \in \mathbb{R}$ *fixo,* $\mathbb{P}\left[X \in B | Y = y\right]$ *define uma probabilidade em* $\mathcal{B}$*, a* $\sigma-$*álgebra de Borel em* $\mathbb{R}$.
(ii) para todo $B \in \mathrm{B}$ *fixo,* $\mathbb{P}\left[X \in B | Y = y\right]$ *é uma função mensurável de* y *e para todo* $(x, y) \in \mathbb{R}^2$

$$\int_{-\infty}^{y} \mathbb{P}\left[X \leq x | Y = t\right] d\mu = \mathbb{P}\left[X \leq x,\, Y \leq y\right].$$

onde μ *é a probabilidade induzida por* $\mathbb{P}[X \leq x | Y = t]$.

APÊNDICE B – Teoremas

Esta seção conterá os resultados que foram utilizados neste trabalho, das áreas da teoria de Probabilidade e Teoria da Medida.

B.1 Teoremas de Teoria da Medida

Para mais detalhes e demonstrações ver (BARTLE, 1966).

Teorema B.1.1 (Desigualdade de Holder)**.** *Suponhamos que* $1 < p < +\infty$ *e* $p^{-1} + q^{-1} = 1$. *Se* f *e* g *são função mensuráveis em* X*, então*

$$||fg||_1 \leq ||f||_p ||g||_q.$$

Teorema B.1.2 (Teorema de Fubini)**.** *Sejam* $(X_n, \mathcal{X}_n, \mu_n), ..., (X_n, \mathcal{X}_n, \mu_n)$ *espaços de medida* σ*-finitos e* $\mathcal{Z}$ *a* σ*-álgebra gerada pelos conjuntos* $A_1 \times ... \times A_n$*, com* $A_i \in \mathcal{X}_i$*. Se* π *é a medida produto de* $\mu_1, ..., \mu_n$ *em* $\mathcal{Z}$ *e* f *uma função* $\mathcal{Z}$*-mensurável e integrável com relação a* π *então*

$$\int f d\pi = \int ... \int f d\mu_1 ... d\mu_n.$$

B.2 Teoremas de Teoria da Probabilidade

Para mais detalhes e demonstrações ver (CHUNG, 2001).

Teorema B.2.1. *Seja* $\{\mu_n\}$ *uma sequência finita ou infinita de medidas em* $(\mathbb{R}, \mathcal{B})$*. Então existe um espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbb{P})$ *e uma sequência de variáveis aleatórias independentes* $\{X_n\}$ *definidas neste espaço, tal que* X_n *tem distribuição* μ_n*, para cada* j*.*

Teorema B.2.2. *Seja* X *uma variável aleatória no espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbb{P})$*, tal que* X *induz o espaço de probabilidade* $(\mathbb{R}, \mathcal{B}, \mu)$*. Se* f *uma função Borel mensurável, então*

$$\int_\Omega f(X) d\mathbb{P} = \int_\mathbb{R} f(x) d\mu.$$

Teorema B.2.3. *Sejam* ν *e* μ *duas medidas de probabilidade definidas em uma mesma* $\sigma-$*Álgebra* $\mathcal{F}$*, que é gerada pelo corpo* $\mathcal{F}_0$*. Se* $\mu(E) = \nu(E)$*, para todo* $E \in \mathcal{F}_0$*, então* $\mu = \nu$ *em* $\mathcal{F}$*.*

Teorema B.2.4 (Desigualdade de Jensen)**.** *Se φ é uma função convexa $\mathbb{R}$, e X e $\varphi(X)$ são variáveis aleatórias integráveis, então*

$$\varphi\left(\mathbb{E}[X]\right) \leq \mathbb{E}\left[\varphi(\mathbb{X})\right].$$

Teorema B.2.5 (Desigualdade de Chebshev)**.** *Seja φ uma função estritamente positiva e crescente em $(0, +\infty)$, $\varphi(u) = \varphi(-u)$, e X uma variável aleatória tal que $\mathbb{E}[\varphi(X)] < +\infty$, então vale para cada $u > 0$ que*

$$\mathbb{P}[|X| \geq u] \leq \frac{\mathbb{E}[\varphi(X)]}{\varphi(u)}$$

Teorema B.2.6 (Teorema Convergência Dominada)**.** *Se $\lim_{n\to+\infty} X_n = X$ q.t.p. ou em medida em um conjunto E e para todo $n \in \mathbb{N}$ vale que $|X_n| \leq Y$ q.t.p.em E e com $\int Y < +\infty$, então*

$$\lim_{n\to+\infty} \int_E X_n d\mathbb{P} = \int_E X d\mathbb{P} \int_E \lim_{n\to+\infty} X_n d\mathbb{P}.$$

Teorema B.2.7 (Teorema da Convergência Limitada)**.** *Se $\lim_{n\to+\infty} X_n = X$ q.t.p. ou em medida em um conjunto E e existe $M > 0$ constante, tal que para todo $n \in \mathbb{N}$ vale que $|X_n| \leq M$ q.t.p. em E, então*

$$\lim_{n\to+\infty} \int_E X_n d\mathbb{P} = \int_E X d\mathbb{P} \int_E \lim_{n\to+\infty} X_n d\mathbb{P}.$$

Teorema B.2.8 (Teorema da Convergência Monótona)**.** *Se $X_n \geq 0$ e $X_n \uparrow X$ q.t.p. em um conjunto E, então*

$$\lim_{n\to+\infty} \int_E X_n d\mathbb{P} = \int_E X d\mathbb{P} \int_E \lim_{n\to+\infty} X_n d\mathbb{P}.$$

Observemos que o teorema da convergência monótona permite que o valor da integral seja $+\infty$

Teorema B.2.9 (Integração Termo a Termo)**.** *Sejam X_n uma sequência de variáveis aleatórias. Se*

$$\sum_n \int_\Delta |X_n| d\mathbb{P} < +\infty,$$

então $\sum_n |X_n| < +\infty$ q.t.p. em Δ, assim $\sum_n X_n$ converge q.t.p. em Δ e

$$\sum_n \int_\Delta X_n d\mathbb{P} = \int_\Delta \sum_n X_n d\mathbb{P}.$$

APÊNDICE C – Esperança/Probabilidade Condicional

Esta seção é dedicada ao estudo da Probabilidade Condicional, que mesmo fazendo parte da área de Probabilidade exerce um papel importante na seção 3 deste trabalho, mais especificamente na subseção 2.3. Observando que as nossas definições e resultados da Teoria de Probabilidade, estão todos em A Course in Probability Theory (ver (CHUNG, 2001)), deste modo também usaremos a definição de Probabilidade Condicional do mesmo. Entretanto resultados uteis ao nosso estudo podem ser encontrados em Probabilidade:Um Curso de Nível Intermediário (ver (JAMES, 1981)), assim mostraremos que a definição de Probabilidade Condicional do primeiro se relaciona com a definição do segundo, por meio de função Borel Mensúravel em $\mathbb{R}$, para que desta forma mostremos que podemos utilizar os resultados que nos ajudaram.

C.1 Definições do livro: A Course in Probability Theory

Definição C.1.1. *Dado uma variável aleatória integrável Y e uma sub-σ-álgebra $\mathcal{G}$ da σ-álgebra $\mathcal{F}$, A* ***Esperança Condicional*** $\mathbb{E}[Y|\mathcal{G}]$ ***de*** Y ***com relação*** $\mathcal{G}$ *é definida como a classe de variáveis aleatórias em Ω tais que:*
(a) são $\mathcal{G}$ mensuráveis e
(b) tem a mesma integral e Y em todos os conjuntos de $\mathcal{G}$, ou seja dado $\Delta \in \mathcal{G}$ e $X \in \mathbb{E}[Y|\mathcal{G}]$ vale

$$\int_\Delta Y d\mathbb{P} = \int_\Delta X d\mathbb{P}.$$

Quando $\mathcal{G}$ é gerado por uma variável aleatória W denotaremos a Esperança Condicional por $\mathbb{E}[Y|W]$.

Notemos que como todas as funções de $\mathbb{E}[Y|\mathcal{G}]$ são iguais em todos os conjuntos de $\mathcal{G}$, podemos concluir que elas são iguais q.t.p, assim usaremos $\mathbb{E}[Y|\mathcal{G}]$ para representar uma variável aleatória que pertence a classe que satisfaz as propriedades (a) e (b).

Definição C.1.2. *Dado um conjunto $\Delta \in \mathcal{F}$. Definiremos como* ***Probabilidade Condicional de*** Δ ***com relação a*** $\mathcal{G}$ *a variável aleatória*

$$\mathbb{P}(\Delta|\mathcal{G}) := \mathbb{E}[1_\Delta|\mathcal{G}].$$

Definição C.1.3. *Considerando $\{\mathcal{F}_a;\, a \in A\}$ uma família de sub-σ-álgebras de $\mathcal{F}$. Dizemos que esta família é* ***Condicionalmente Independente com Relação a $\mathcal{G}$****, quando para qualquer coleção finita de conjuntos Λ_1, ..., Λ_n, tais que $\Lambda_j \in \mathcal{F}_{a_j}$, com os a_j's sendo indices distintos de A, temos*

$$\mathbb{P}\left[\bigcap_{j=1}^{n} \Lambda_j | \mathcal{G}\right] = \prod_{j=1}^{n} \mathbb{P}\left[\Lambda_j | \mathcal{G}\right].$$

Em particular quando as $\mathcal{F}_a$'s são geradas por variáveis aleatórias Y_a's, dizemos que as Y_a's ***são Condicionalmente Independente com Relação a $\mathcal{G}$****.*

C.2 Teorema do livro: A Course in Probability Theory

Teorema C.2.1. *Existe φ uma função Borel Mensurável em $\mathbb{R}$ tal que podemos escrever $\mathbb{E}\left[Y|X\right] = \varphi(X)$. Mais ainda, se nós definirmos a medida λ no borelianos de $\mathbb{R}$, por*

$$\lambda(B) = \int_{X^{-1}(B)} Y d\mathbb{P}, \text{ para todo } B \in \mathcal{B}$$

e tomando $\mu(A) = \mathbb{P}[Y \in A]$, com $A \in \mathcal{B}$, então φ é uma versão da derivada de Radon-Nikodym $\frac{d\lambda}{d\mu}$, onde $\mathcal{B}$ representa a σ-álgebra de Borel.

C.3 Definição do livro: Probabilidade: Um Curso de Nível Intermediário

Definição C.3.1. *Sejam X e Y variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$. Uma função $\mathbb{P}[X \in B|Y = y]$, definida para B boreliano e $y \in \mathbb{R}$, sera chamada uma* ***Distribuição Condicional (regular) para X dada Y*** *se:*
a) para todo $y \in \mathbb{R}$ fixo, $\mathbb{P}[X \in B|Y = y]$ define uma probabilidade em $\mathcal{B}$, a $\sigma-$álgebra de Borel em $\mathbb{R}$ e
b) para todo $B \in \mathcal{B}$ fixo $\mathbb{P}[X \in B|Y = y]$ é uma função mensurável de y e para todo $(x, y) \in \mathbb{R}^2$,

$$\int_{-\infty}^{y} \mathbb{P}[X \leq x|Y = t] dF_Y(t) = \mathbb{P}[X \leq x, Y \leq y],$$

onde F_Y é distribuição de Y e a integral acima representa a integral de Riemann-Stieltjes.

Definição C.3.2. *Seja $y \in \mathbb{R}$ e $B \in \mathcal{B}$. A* ***Probabilidade Condicional*** *de $X \in B$ dado que $Y = y$, é definida por*

$$\mathbb{P}[X \in B|Y = y] = \lim_{\Delta_y \to 0} \mathbb{P}[X \in B|Y \in I_y],$$

onde I_y representa um intervalo contendo y e Δ_y representa o comprimento do intervalo I_y.

C.4 Teoremas do livro: Um Curso de Nível Intermediário

Teorema C.4.1 (Principio da substituição para distribuição condicional). *Sejam K e W variáveis aleatórias, $\varphi(k,w)$ uma função mensurável. Se a distribuição condicional de K dado W é*

$$\mathbb{P}[K \in B|W = w],\ B \in \mathcal{B},\ w \in \mathbb{R},$$

então a distribuição condicional para $\varphi(K,W)$ dado W é

$$\mathbb{P}[\varphi(K,W) \in B|W = w] = \mathbb{P}[K \in \{k;\ \varphi(k,w) \in B\}|W = w],\ B \in \mathcal{B},\ w \in \mathbb{R}.$$

Teorema C.4.2. *Para cada $B \in \mathcal{B}$ fixo o limite na definição C.3.2 existe quase certamente, ou seja $\mathbb{P}\left[Y \in \{y \in \mathbb{R};\ o\ limite\ existe\ em\ y\}\right] = 1$. Além disso, para cada B fixo o limite é igual a $\mathbb{P}[X \in B|Y = y]$, como definida em C.3.1, quase certamente.*

C.5 Relação entre as Definições

Observemos primeiramente que a definição C.3.1, nos da uma família funções Borel mensuráveis que satisfazem as propriedades a) e b), que são indexadas por $B \in \mathcal{B}$. Sendo assim nesta subseção veremos que a família de funções Borel mensuráveis $\{\varphi_B\}_{B\in\mathcal{B}}$, onde φ_B é escolhida de tal forma que (ver Teorema C.2.1)

$$\varphi_B(Y) = \mathbb{P}[X \in B|Y] = \mathbb{E}\left[1_{\{X\in B\}}|Y\right],$$

satisfaz a definição C.3.1. Por um abuso de notação, usaremos

$$\varphi_B(y) := \mathbb{P}[X \in B|Y = y],$$

para representar a variável aleatória φ_B aplicada em $y \in \mathbb{R}$.

Mostraremos primeiramente o item a) da definição C.3.1. Com efeito, notemos que se $B = \Omega$, então dado $\Delta \in \mathcal{F}(X)$, onde $\mathcal{F}(X)$ é σ-álgebra gerada por X, temos

$$\int_\Delta 1_{\{X\in\Omega\}} d\mathbb{P} = \int_\Delta 1 d\mathbb{P},$$

assim podemos escrever $\mathbb{P}[X \in \Omega|Y = y] = 1$, para todo $y \in \mathbb{R}$. Agora notemos que como $1_{\{X\in B\}} \geq 0$, para B boreliano arbitrário, segue que podemos assumir que $\mathbb{P}[X \in B|Y] \geq 0$, pois a menos de um conjunto de medida nula em Ω, vale que $\mathbb{P}[X \in B|Y]$ é maior que zero, assim vale que $\mathbb{P}[X \in B|Y = y] \geq 0$, para todo B boreliano e para todo y. Agora nos resta mostrar a propriedade aditiva, assim consideremos B_n uma sequência de borelianos dois a dois disjuntos, então mostraremos que

$$\mathbb{P}\left[X \in \dot{\bigcup_n} B_n|Y = y\right] = \sum_n \mathbb{P}\left[X \in B_n|Y = y\right]. \tag{C.1}$$

Notemos que como

$$\sum_n \int_\Delta \mathbb{P}\left[X \in B_n | Y\right] d\mathbb{P} = \mathbb{P}\left[\Delta \bigcap \left\{\bigcup_n B_n\right\}\right],$$

para dado $\Delta \in \mathcal{F}(X)$. Segue então do Teorema B.2.9 que $\sum_n \mathbb{P}\left[X \in B_n | Y\right]$ converge q.t.p. em Ω, pois $\Delta \in \mathcal{F}(X)$ é arbitrário, e

$$\sum_n \int_\Delta \mathbb{P}\left[X \in B_n | Y\right] d\mathbb{P} = \int_\Delta \sum_n \mathbb{P}\left[X \in B_n | Y\right] d\mathbb{P}.$$

Logo

$$\mathbb{P}\left[X \in \dot{\bigcup_n} B_n | Y\right] = \sum_n \mathbb{P}\left[X \in B_n | Y\right] \text{ q.t.p.}$$

e portanto podemos concluir que

$$\mathbb{P}\left[X \in \dot{\bigcup}_n B_n | Y = y\right] = \sum_n \mathbb{P}\left[X \in B_n | Y = y\right],$$

para todo y. Portanto vemos que essa coleção φ_B satisfaz o item a) da definição C.3.1.

Mostraremos agora que a coleção satisfaz o item b) da definição C.3.1. Com efeito observemos que a primeira parte de b) segue diretamente Teorema C.2.1. Agora mostraremos a segunda parte, temos que

$$\int_{-\infty}^{y} \mathbb{P}[X \leq x | Y = t] dF_Y(t) = \int_{Y \leq y} \mathbb{P}[X \leq x | Y] d\mathbb{P} = \mathbb{P}[X \leq x, Y \leq y],$$

onde a primeira igualdade acima segue diretamente de (CHUNG, 2001), pagina 315. Com isso mostramos que $\{\varphi_B\}_{B \in \mathcal{B}}$ satisfaz a definição C.3.1, assim podemos utilizar os resultados da seção C.3.