Universidade Federal do Rio Grande do Sul

Instituto de Matemática

GRANDES DESVIOS NO CONTEXTO DE VARIÁVEIS ALEATÓRIAS INDEPENDENTES E IDENTICAMENTE DISTRIBUÍDAS

MARCO ANTÔNIO GIACOMELLI

Dissertação submetida ao Curso de Pós-Graduação em Matemática como requisito parcial para a obtenção do grau de mestre.

Orientação: Profa. Dra. Sara Ianda Corrêa Carmona (UFRGS - IM)

Banca examinadora: Prof. Dr. Nelson Ithiro Tanaka (USP - IME)
Profa. Dra. Sílvia Regina Costa Lopes (UFRGS - IM)
Profa. Dra. Sara Ianda Corrêa Carmona (UFRGS - IM)

Abril de 1995

Porto Alegre

A meus pais

Issuile e Elenice,

pelo empenho e determinação em minha formação.

AGRADECIMENTOS

Muitas são as pessoas a quem devo agradecer. Particularmente menciono meus familiares, cujo apoio e incentivo foram fundamentais, os professores do Curso de Pós-Graduação, em especial a professora Sara Carmona pela orientação deste trabalho, e aos colegas do curso.

RESUMO

Esta dissertação é o resultado de um estudo sobre a Teoria de Grandes Desvios, no qual deu-se ênfase ao contexto de variáveis aleatórias independentes e identicamente distribuídas (v.a's i.i.d.). Com a finalidade de discutir alguns dos principais resultados sobre Grandes Desvios, inicialmente apresentamos as definições e propriedades básicas. Numa etapa seguinte apresentamos o teorema de Cramér-Chernoff para v.a's i.i.d. a valores em $\mathbf{R}$. A seguir enunciamos um princípio de Grandes Desvios para quaisquer abertos e fechados de $\mathbf{R}$. Numa etapa posterior estendemos o teorema de Cramér-Chernoff para v.a's i.i.d. a valores em $\mathbf{R}^d$. No Capítulo final apresentamos, de maneira sintética, outras extensões do teorema de Cramér-Chernoff, tais como: o teorema de Sanov para v.a's i.i.d. e o princípio de Grandes Desvios para Cadeias de Markov finitas. Além disso, apresentamos algumas aplicações de Grandes Desvios em Estatística Matemática.

ABSTRACT

This thesis is the result of a study about the Large Deviations Theory, in which the context of independent and identically distributed (i.i.d.) random variables was emphasized. Aiming to discuss some of the main results about Large Deviations, first we present the definitions and the basic properties. In a second moment, we present Cramér-Chernoff's theorem for i.i.d random variables in $\mathbf{R}$. After, we enunciate a Large Deviations principle for any open and closed sets in $\mathbf{R}$. In another stage, we extend Cramér-Chernoff's theorem for i.i.d. random variables in $\mathbf{R}^d$. In last chapter, we present, briefly, other extensions of Cramér-Chernoff's theorem, such as Sanov's theorem for i.i.d. random variables and the Large Deviations principle for finite state Markov Chains. Besides, we present some applications of Large Deviations in Mathematical Statisitics.

ÍNDICE

0 - INTRODUÇÃO

O objetivo da teoria de Grandes Desvios consiste em estimar probabilidades de eventos raros. A denominação "eventos raros" refere-se a determinados eventos cuja probabilidade tende a zero.

Para compreendermos melhor a qualificação "Grandes Desvios" consideremos a situação mais simples possível, ou seja, assuma que $\{X_n\}_{n\geq 1}$ é uma seqüência de v.a's i.i.d. a valores em $\mathbf{R}$, com lei de probabilidade μ e esperança m. Seja $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ e μ_n a lei de probabilidade de $\bar{X}_n$, isto é, $\mu_n(A) = P(\bar{X}_n \in A)$ para $A \in \mathcal{B}(\mathbf{R})$, sendo $\mathcal{B}(\mathbf{R})$ a $\sigma-$álgebra de Borel em $\mathbf{R}$. Gostaríamos de estudar o comportamento assintótico de $P(\bar{X}_n \in A)$, com $m \notin \bar{A}$ (onde $\bar{A}$ é o fecho de A), quando $n \to \infty$. Pela Lei Fraca dos Grandes Números, sabe-se que $\lim_{n\to\infty} P(\bar{X}_n \in A) = 0$. Isto quer dizer que para n muito grande $[\bar{X}_n \in A]$ é um evento muito raro, ou seja, sua ocorrência seria "atípica" ou um "Grande Desvio" em relação à Lei dos Grandes Números.

Uma questão a qual o leitor poderia indagar é o porquê da restrição $m \notin \bar{A}$. Para responder a essa questão suponha $A = (m, +\infty)$ e que μ_n é absolutamente contínua em relação à medida de Lebesgue, ou seja, existe uma função f_n não negativa, mensurável e finita tal que $\mu_n(A) = \int_A f_n(x)dx$. Então, $\int_A f_n(x)dx = \int_{\bar{A}} f_n(x)dx$, e portanto, $\mu_n(A) = \mu_n(\bar{A})$, $\forall n \geq 1$. Ora, como $m \in \bar{A}$ então pela Lei dos Grandes Números $\lim_{n\to\infty} \mu_n(\bar{A}) = 1$, levando-nos a concluir que $[\bar{X}_n \in A]$ não é uma situação de Grande Desvio em relação à esperança m. Logo, a restrição $m \notin \bar{A}$ é para garantir que $[\bar{X}_n \in A]$ seja um evento do tipo Grande Desvio.

Atualmente os teoremas-limite na teoria das probabilidades dividem-se em três grandes grupos: teoremas do tipo Lei dos Grandes Números, Teorema Central do Limite e Grandes Desvios. No contexto de v.a's i.i.d., analisaremos a seguir quais informações podemos obter sobre o comportamento assintótico de $P(\bar{X}_n \in A)$, $A \in \mathcal{B}(\mathbf{R})$, em cada um destes grupos.

Para $m \notin \bar{A}$, $A \in \mathcal{B}(\mathbf{R})$, a Lei dos Grandes Números afirma que $\lim_{n\to\infty} P(\bar{X}_n \in A) = 0$. Já o princípio de Grandes Desvios para a média $\bar{X}_n$ de v.a's i.i.d. diz que $\underline{\lim}_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in A) \geq -\inf_{y\in A^\circ} \lambda(y)$, sendo A° o interior do conjunto A, e $\overline{\lim}_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in A) \leq -\inf_{y\in \bar{A}} \lambda(y)$, onde λ é a transformada de Cramér, a qual será definida rigorosamente no Capítulo 1. No caso de A ser uma semireta, por exemplo $A = (-\infty, x]$, $x < m$, veremos no Capítulo 1 que

$$\lim_{n\to\infty} \frac{\ln P(\bar{X}_n \leq x)}{-n\lambda(x)} = 1. \tag{0.1}$$

Quando (0.1) é válido, diz-se que existe uma "equivalência logarítmica" entre $P(\bar{X}_n \leq x)$ e $e^{-n\lambda(x)}$. De (0.1) vemos que $\ln P(\bar{X}_n \leq x)$ e $-n\lambda(x)$ convergem para $-\infty$ na mesma velocidade, logo, pode-se concluir que $\lambda(x)$ é a taxa de decaimento exponencial de $P(\bar{X}_n \leq x)$ para zero quando $n \to \infty$.

Se $m \in A$ então $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in A) = 0$, pela Lei Fraca dos Grandes Números, e pelo princípio de Grandes Desvios $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in A) = -\inf_{x\in A} \lambda(x) = -\lambda(m) = 0$, como será provado no Capítulo 1. É claro que não estamos interessados na situação $m \in A$ pois, como vimos anteriormente, neste caso $[\bar{X}_n \in A]$ não seria um evento do tipo Grande Desvio. A intenção aqui é apenas mostrar que nesta situação ambos teoremas-limite fornecem o mesmo resultado.

Agora iremos comparar o Teorema Central do Limite e o princípio de Grandes Desvios para a média amostral. Para isto, considere $A = [m + \frac{\delta}{\sqrt{n}}, +\infty)$, $\delta > 0$, e $\{X_n\}_{n\geq 1}$ uma seqüência de v.a's i.i.d. de esperança m e variância σ^2. O evento $[\bar{X}_n \in A]$ é denominado de um "Pequeno Desvio" em relação à esperança m. Podemos observar que $[m + \frac{\delta}{\sqrt{n}}, +\infty)$ tende a $[m, +\infty)$ quando $n \to \infty$, ao contrário de uma situação de Grande Desvio, onde o conjunto em questão seria $[m + \delta, +\infty)$. Sendo $E(\bar{X}_n) = m$ e $Var(\bar{X}_n) = \frac{\sigma^2}{n}$,

$$P(\bar{X}_n \in A) = P\left(\bar{X}_n \geq m + \frac{\delta}{\sqrt{n}}\right) = P\left(\sqrt{n}\left(\frac{\bar{X}_n - m}{\sigma}\right) \geq \frac{\delta}{\sigma}\right),$$

e pelo Teorema Central do Limite $\sqrt{n}\left(\frac{\bar{X}_n - m}{\sigma}\right)$ converge em distribuição para uma normal padrão. Logo,

$$\lim_{n\to\infty} P(\bar{X}_n \in A) = 1 - \Phi\left(\frac{\delta}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{\frac{\delta}{\sigma}}^{+\infty} e^{-z^2/2} dz > 0,$$

onde Φ é a função de distribuição de uma normal padrão.

Como veremos no Capítulo 1, o princípio de Grandes Desvios para a média amostral de uma seqüência de v.a's i.i.d. a valores em $\mathbf{R}$ dependerá da transformada de Cramér. Isto significa que será preciso conhecer a lei de probabilidade da seqüência de v.a's. Exemplificando, assuma $\{X_n\}_{n\geq 1}$ como sendo uma seqüência de v.a's com distribuição de probabilidade normal de esperança $m = 0$ e variância $\sigma^2 = 1$. Considere o evento $[|\bar{X}_n| > \delta]$, $\delta > 0$. Logo,

$$P(|\bar{X}_n| > \delta) = 2 \int_{\delta\sqrt{n}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

Agora observe que

$$\int_{\delta\sqrt{n}}^{+\infty} e^{-z^2/2} dz = \int_{\delta\sqrt{n}}^{+\infty} \frac{1}{z} z e^{-z^2/2} dz = -\left.\frac{e^{-z^2/2}}{z}\right|_{\delta\sqrt{n}}^{+\infty} - \int_{\delta\sqrt{n}}^{+\infty} \frac{1}{z^2} e^{-z^2/2} dz$$

$$= \frac{1}{\delta\sqrt{n}} e^{-n\delta^2/2} - \int_{\delta\sqrt{n}}^{+\infty} \frac{1}{z^2} e^{-z^2/2} dz \leq \frac{1}{\delta\sqrt{n}} e^{-n\delta^2/2}, \quad \forall n \geq 1. \tag{0.2}$$

Disso vem que $P(|\bar{X}_n| > \delta) \leq \frac{2}{\delta\sqrt{2\pi n}} e^{-n\delta^2/2}$, $\forall\ n \geq 1$. De (0.2)

$$\int_{\delta\sqrt{n}}^{+\infty} e^{-z^2/2} dz + \int_{\delta\sqrt{n}}^{+\infty} \frac{1}{z^2} e^{-z^2/2} dz = \frac{1}{\delta\sqrt{n}} e^{-n\delta^2/2},$$

e pelo fato de

$$\int_{\delta\sqrt{n}}^{+\infty} e^{-z^2/2} dz \geq \int_{\delta\sqrt{n}}^{+\infty} \frac{1}{z^2} e^{-z^2/2} dz, \quad \text{para } \delta\sqrt{n} \geq 1,$$

tem-se

$$2\int_{\delta\sqrt{n}}^{+\infty} e^{-z^2/2} dz \geq \frac{1}{\delta\sqrt{n}} e^{-n\delta^2/2}, \quad \text{se } \delta\sqrt{n} \geq 1.$$

Portanto,

$$\frac{1}{\delta\sqrt{2\pi n}} e^{-n\delta^2/2} \leq P(|\bar{X}_n| > \delta) \leq \frac{2}{\delta\sqrt{2\pi n}} e^{-n\delta^2/2}, \delta\sqrt{n} \geq 1, \tag{0.3}$$

e segue de (0.3) que $\lim_{n\to\infty} \frac{1}{n} \ln P(|\bar{X}_n| > \delta) = -\frac{\delta^2}{2}$. Esta última igualdade nos dá a transformada de Cramér (ver Definição 1.1.2) para uma v.a. com distribuição de probabilidade normal de esperança zero e variância 1, isto é, $\lambda(\delta) = \frac{\delta^2}{2}$. No Exemplo 1.1.3 do Capítulo 1 mostraremos como obter λ quando X é v.a. com distribuição normal de esperança m e variância σ^2 quaisquer.

Os primeiros resultados sobre Grandes Desvios são devidos a Cramér [8], o qual provou que para uma seqüência $\{X_n\}_{n\geq 1}$ de v.a's i.i.d., $P(\bar{X}_n \leq x)$ tem uma "equivalência exata" em relação a $e^{-n\lambda(x)}$, $x < m$, isto é, $\lim_{n\to\infty} \frac{P(\bar{X}_n \leq x)}{e^{-n\lambda(x)}} = 1$. As hipóteses por ele consideradas foram: X_1 com lei de probabilidade absolutamente contínua em relação à medida de Lebesgue e $E\left(e^{tX_1}\right) < +\infty$, com $t \in (-\delta, \delta)$ para algum $\delta > 0$. Se pensarmos em termos de limite, a equivalência exata diz que para $\forall \epsilon > 0$, $\exists N > 0$ tal que se $\forall n \geq N$ então

$$(1-\epsilon)e^{-n\lambda(x)} < P(\bar{X}_n \leq x) < (1+\epsilon)e^{-n\lambda(x)}. \tag{0.4}$$

Através de (0.4) também vemos que λ determina a taxa de convergência exponencial de $P(\bar{X}_n \leq x)$ para zero quando $n \to \infty$.

Neste trabalho inicial, Cramér estava interessado em estudar o comportamento assintótico de $\frac{1-F_n(x)}{1-\Phi(x)}$, onde $F_n(x) = P(\bar{X}_n \leq x)$, quando $x = x_n$ tende a $+\infty$ com n e $x_n = o(\sqrt{n})$. Posteriormente, Chernoff [7], eliminando a hipótese de v.a's absolutamente contínuas considerada por Cramér, provou que

$\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \leq x) = -\lambda(x)$, se $x \leq m$, e $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \geq x) = -\lambda(x)$, se $x \geq m$, permitindo que os resultados de Cramér fossem estendidos para um universo maior de v.a's. Chernoff ainda abordou questões sobre Estatística Matemática, como a eficiência assintótica de testes de hipóteses. Um dos exemplos trata do caso em que $\{X_n\}_{n\geq 1}$ é uma seqüência de v.a's i.i.d. com distribuição de probabilidade de Bernoulli, onde deseja-se estimar a probabilidade dos erros tipo I e II no teste da razão de verossimilhança simples. No Exemplo 1.2.1 do Capítulo 1 analisaremos com maiores detalhes este exemplo. Assim, no universo de v.a's i.i.d. a valores em **R**, os resultados de Cramér e Chernoff passam então a constituir o que hoje chama-se de "contexto clássico de Grandes Desvios".

Posteriormente aos trabalhos de Cramér e Chernoff houve generalizações basicamente em duas direções: quanto ao espaço onde as v.a's estão definidas ($\mathbf{R}^d$, espaços de medidas, espaços infinito-dimensionais e espaços mais gerais) e quanto à existência de algum tipo de dependência estocástica entre as v.a's. Para detalhar essas generalizações, considere $\{X_n\}_{n\geq 1}$ uma seqüência de vetores aleatórios i.i.d. a valores em $\mathbf{R}^d$, $d \geq 1$. Assuma $f : \mathbf{R}^d \longrightarrow \mathbf{R}^d$ uma função mensurável e limitada. Analogamente ao caso da média amostral em **R**, tem-se interesse em estimar o comportamento assintótico de

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} f_1(X_i)\right| > \delta, \cdots\cdots\cdots, \left|\frac{1}{n}\sum_{i=1}^{n} f_d(X_i)\right| > \delta\right) \tag{0.5}$$

quando $n \to \infty$, sendo $f = (f_1, \cdots, f_d)$. Naturalmente, a média em **R** é um caso particular de (0.5), com $f(x) = x$ e $d = 1$. Generalizações como em (0.5) foram tratadas por Lanford [20]. Donsker-Varadhan [11] fizeram generalizações para espaços gerais, tais como: Banach, Polonês, Hausdorff e espaços infinito-dimensionais, e consideraram variáveis aleatórias com dependência fraca, ou seja, processos de Markov. Também no contexto de vetores aleatórios assumindo valores em espaços gerais (Hausdorff, Banach) Bahadur & Zabel [3] fizeram importantes contribuições.

Outra importante generalização do Teorema de Cramér-Chernoff é o princípio de Grandes Desvios para a medida empírica de uma seqüência de v.a's i.i.d., devido a Sanov [26]. No contexto de Cadeias de Markov finitas, Gärtner [14] estabeleceu um princípio de Grandes Desvios para a medida empírica. No Capítulo 3 discutiremos estas duas generalizações.

Generalizações do Teorema de Cramér-Chernoff para famílias de processos Markovianos tem sido feitas por vários pesquisadores, tais como Gärtner [15], Mackean [24], Freidlin & Wentzell [13] e outros. No contexto de famílias de processos de difusão, caracterizados pela introdução de um pequeno parâmetro, a teoria de Wentzell-Freidlin, sintetizada na referência [13], é de fundamental importância.

Entre as aplicações de Grandes Desvios, a Mecânica Estatística tem se destacado nos últimos anos, contribuindo com o surgimento de novas técnicas. Para maiores detalhes citamos Ellis [12], Holley & Stroock [16], Varadhan [29], Lanford [20] e Cassandro et alii [6].

Na Estatística Matemática encontramos aplicações em testes de hipóteses e estimação de parâmetros. Uma generalização do contexto tratado por Chernoff [7], em testes de hipóteses, pode ser encontrada em Dembo & Zeitouni [9]. Já em Bucklew [5] encontramos aplicações em estimação de parâmetros, discernimento entre duas populações e na teoria da informação.

Nesta monografia ficaremos restritos a alguns dos principais resultados sobre Grandes Desvios no contexto de v.a's i.i.d. a valores em $\mathbf{R}$ e $\mathbf{R}^d$, $d > 1$, assim como uma breve exposição no contexto de Cadeias de Markov. Na seção 1.1 do Capítulo 1 abordaremos as principais propriedades da transformada de Cramér e na seção 1.2 daremos ênfase ao teorema de Cramér-Chernoff para a média amostral. Na seção 1.3 desse mesmo capítulo enunciaremos alguns conceitos gerais em Grandes Desvios, além de estabelecer um princípio de Grandes Desvios para quaisquer conjuntos abertos e fechados de $\mathbf{R}$. No capítulo 2 iremos estender os resultados do Capítulo 1 para $\mathbf{R}^d$, além de fazer uma comparação com o teorema de Cramér-Chernoff em $\mathbf{R}$. No Capítulo 3 apresentaremos, de forma suscinta, algumas extensões do teorema de Cramér-Chernoff, como o teorema de Sanov para medidas empíricas de v.a's i.i.d., e também algumas aplicações.

Como principais referências bibliográficas foram utilizadas as obras de Azencott [2], Bucklew [5] Dembo & Zeitouni [9], Stroock [28] e Vares [30].

CAPÍTULO 1

GRANDES DESVIOS PARA A MÉDIA AMOSTRAL DE VARIÁVEIS ALEATÓRIAS I.I.D. EM R

Neste capítulo discutiremos a formulação de um princípio de Grandes Desvios para a média amostral $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ de v.a's $X_1, X_2, \cdots, X_n$ i.i.d. a valores em $\mathbf{R}$. Na seção 1.1 daremos definições básicas e as principais propriedades da transformada de Cramér, a qual permitirá enunciar na seção 1.2 um importante teorema sobre Grandes Desvios envolvendo conjuntos do tipo $(-\infty, x]$ ou $[x, +\infty)$. Na seção 1.3 daremos algumas definições gerais em Grandes Desvios, além de estabelecer um princípio de Grandes Desvios para quaisquer conjuntos abertos e fechados de $\mathbf{R}$.

1.1 - A transformada de Cramér e suas propriedades

Definição 1.1.1. *Seja μ uma medida em $\mathbf{R}$. A transformada de Laplace $\hat{\mu} : \mathbf{R} \longrightarrow (0, +\infty]$ para a medida μ é definida por*

$$\hat{\mu}(t) \equiv \int_{-\infty}^{+\infty} e^{tx} d\mu(x), \quad t \in \mathbf{R}.$$

Se μ é uma medida de probabilidade então a correspondente transformada de Laplace é denominada de função geradora de momentos da v.a. cuja distribuição de probabilidade é μ. Note que $\hat{\mu}$ pode assumir $+\infty$ para todo $t \neq 0$ e $\hat{\mu}(0) = 1$.

Definição 1.1.2. *A transformada de Cramér λ_μ: $\mathbf{R} \longrightarrow [0, +\infty]$ para a medida de probabilidade μ é definida por*

$$\lambda_\mu(x) \equiv \sup_{t \in \mathbf{R}} \{tx - \ln \hat{\mu}(t)\}, \quad x \in \mathbf{R}.$$

Observe que $\lambda_\mu \geq 0$, pois $tx - \ln \hat{\mu}(t) = 0$ quando $t = 0$ seja qual for x. Por outro lado, λ_μ pode ou não assumir $+\infty$. Posteriormente veremos os casos onde $\lambda_\mu(x) = +\infty$ para x pertencente a um subconjunto de $\mathbf{R}$. Na seção 1.3 mostraremos que λ_μ é uma "função taxa" e além disso a seqüência de medidas $\{\mu_n\}_{n \geq 1}$, onde $\mu_n(A) = P(\bar{X}_n \in A)$ para $A \subseteq \mathbf{R}$, satisfaz um princípio de Grandes Desvios através de λ_μ. A notação λ_μ para a transformada de Cramér será substituída por λ quando não houver dúvidas quanto à medida de probabilidade em questão.

Uma interpretação geométrica de λ é dada na figura 1.1.1. Quando existe $\tau \in \mathbf{R}$ tal que $\lambda(x) = x\tau - \ln \hat{\mu}(\tau)$ então da equação $\frac{d}{dt}(tx - \ln \hat{\mu}(t)) = 0$ segue que $\frac{\hat{\mu}'(\tau)}{\hat{\mu}(\tau)} = x$.

O próximo objetivo será estudar as propriedades da função λ, as quais serão úteis nesta e em seções posteriores.

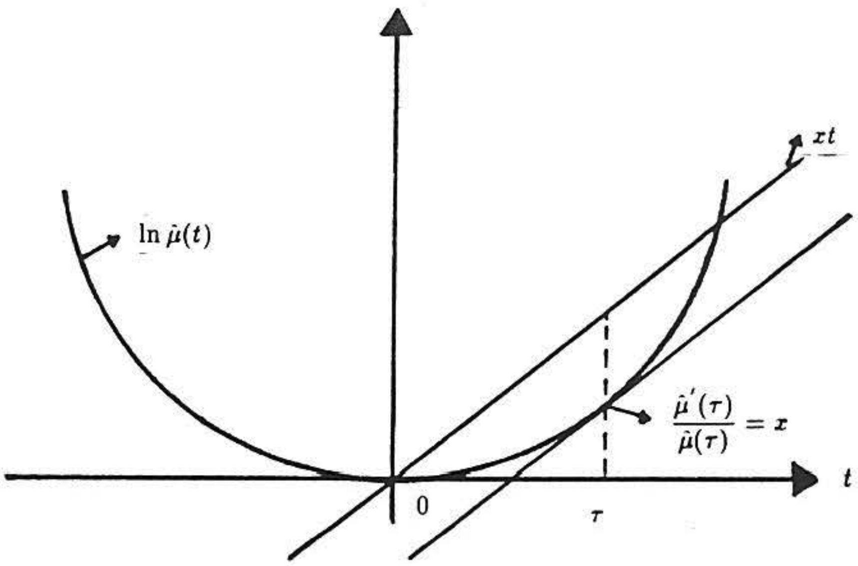

figura 1.1.1 - Interpretação geométrica de λ

Proposição 1.1.1. *Seja X uma v.a. com distribuição de probabilidade μ e respectiva transformada de Laplace $\hat{\mu}(t)$. Então:*

$$-\lambda(x) = \ln \inf_{t \in \mathbf{R}} \{e^{-tx}\hat{\mu}(t)\}, \quad x \in \mathbf{R}. \tag{1.1.1}$$

Prova:

$$-\lambda(x) = -\sup_{t \in \mathbf{R}}\{tx - \ln \hat{\mu}(t)\} = \inf_{t \in \mathbf{R}}\{-tx + \ln \hat{\mu}(t)\} = \inf_{t \in \mathbf{R}} \ln \left(e^{-tx}\hat{\mu}(t)\right).$$

Vamos definir $g(t) \equiv e^{-tx}\hat{\mu}(t)$ e $I \equiv \inf_{t \in \mathbf{R}} g(t)$. Então $\ln g(t) \geq \ln I$, $\forall t \in \mathbf{R}$. Logo, $\inf_{t \in \mathbf{R}} \ln g(t) \geq \ln I = \ln \inf_{t \in \mathbf{R}} g(t)$. Agora suponha $\inf_{t \in \mathbf{R}} \ln g(t) > \ln I$. Disso decorre que $\exists \epsilon > 0$ tal que $\inf_{t \in \mathbf{R}} \ln g(t) \geq \ln I + \epsilon$, e daí, $\ln g(t) \geq \ln I + \epsilon$, $\forall t \in \mathbf{R}$, o qual implica em $g(t) \geq e^{\ln I + \epsilon} > e^{\ln I} = I$, $\forall t \in \mathbf{R}$, ou seja, I não é mais o ínfimo de $g(t)$, levando-nos a uma contradição. Assim,

$$\inf_{t\in\mathbf{R}} \ln\left(e^{-tx}\hat{\mu}(t)\right) = \ln \inf_{t\in\mathbf{R}} \left\{e^{-tx}\hat{\mu}(t)\right\}.$$

□

Proposição 1.1.2. *Seja λ a transformada de Cramér para a medida de probabilidade μ. Então:*

(i) λ é semicontínua inferiormente;

(ii) λ é convexa;

(iii) $\lambda(m) = 0$, onde $m = \int_{-\infty}^{+\infty} y\, d\mu(y)$ é finita;

(iv) $\lambda(x)$ é não decrescente para $x \geq m$ e não crescente para $x \leq m$;

(v) Se λ for duas vezes diferenciável em $x = m$, m finita, então $\lambda'(m) = 0$ e $\lambda''(m) > 0$.

Prova: A prova que segue é baseada em Stroock [28].

(i) Uma definição equivalente para uma função semicontínua inferiormente (s.c.i.) é $\lambda(x) \leq \underline{\lim}_{n\to\infty} \lambda(x_n)$ para qualquer seqüência $\{x_n\}_{n\geq 1}$ em $\mathbf{R}$ tal que $x_n \to x$, com $x \in \mathbf{R}$ (ver Proposição A.2.1 do Apêndice 2). Assim, tome uma seqüência qualquer $\{x_n\}_{n\geq 1}$ tal que $x_n \to x$. Da definição de λ temos que para todo $t \in \mathbf{R}$, $\lambda(x_n) \geq tx_n - \ln\hat{\mu}(t)$. Logo,

$$\underline{\lim}_{n\to\infty} \lambda(x_n) \geq \sup_{t\in\mathbf{R}}\{tx - \ln\hat{\mu}(t)\} = \lambda(x).$$

(ii) Sejam $x, y \in \mathbf{R}$ e $0 \leq \alpha \leq 1$. Para qualquer $t \in \mathbf{R}$, tem-se

$$\alpha\lambda(x) + (1-\alpha)\lambda(y) \geq \alpha(tx - \ln\hat{\mu}(t)) + (1-\alpha)(ty - \ln\hat{\mu}(t)) = t(\alpha x + (1-\alpha)y) - \ln\hat{\mu}(t),$$

e portanto, $\alpha\lambda(x) + (1-\alpha)\lambda(y) \geq \lambda(\alpha x + (1-\alpha)y)$.

(iii) Pela desigualdade de Jensen,

$$\ln \int_{-\infty}^{+\infty} e^{ty} d\mu(y) \geq t \int_{-\infty}^{+\infty} y d\mu(y) = t\, m,\ \forall t \in \mathbf{R}. \tag{1.1.2}$$

Assim, $tm - \ln\hat{\mu}(t) \leq 0$ para $\forall t \in \mathbf{R}$ e disto conclui-se que $\lambda(m) = 0$, já que $\lambda(x) \geq 0$, $\forall x \in \mathbf{R}$.

(iv) Tome $m < x_1 < x_2$ e $0 < \alpha < 1$ tal que $x_1 = (1-\alpha)m + \alpha x_2$. Por (ii) e (iii)

$$\lambda(x_1) = \lambda((1-\alpha)m + \alpha x_2) \leq (1-\alpha)\lambda(m) + \alpha\lambda(x_2) = \alpha\lambda(x_2) \leq \lambda(x_2).$$

Agora tome $x_1 < x_2 < m$ e $0 < \alpha < 1$ tal que $x_2 = (1-\alpha)m + \alpha x_1$. Novamente por (ii) e (iii)

$$\lambda(x_2) = \lambda((1-\alpha)m + \alpha x_1) \leq (1-\alpha)\lambda(m) + \alpha\lambda(x_1) = \alpha\lambda(x_1) \leq \lambda(x_1).$$

(v) Uma vez que $x = m$ é um mínimo local então $\lambda^{'}(m) = 0$ e $\lambda^{''}(m) > 0$.

□

Na proposição abaixo adotaremos $\hat{\mu}_X(t)$, $\lambda_X(x)$ e μ_X para denotar, respectivamente, a transformada de Laplace, de Cramér e a lei de probabilidade de uma v.a. X.

Proposição 1.1.3. *Seja $(X_n)_{n\geq 1}$ uma seqüência de v.a's i.i.d. com lei de probabilidade comum μ, transformada de Laplace $\hat{\mu}(t)$ e transformada de Cramér $\lambda(x)$. Tem-se:*

(i) Se $Y = aX_1 + b$, onde $a, b \in \mathbf{R}$, então

$$\lambda_Y(x) = \lambda\left(\frac{x-b}{a}\right), \quad x \in \mathbf{R};$$

(ii) Se $S_n = \sum_{i=1}^n X_i$ então

$$\lambda_{S_n}(x) = n\lambda\left(\frac{x}{n}\right), \quad x \in \mathbf{R};$$

(iii) Se $W = S_n - na$, onde $a \in \mathbf{R}$, então

$$\lambda_W(x) = n\lambda\left(\frac{x}{n} + a\right), \quad x \in \mathbf{R}; \tag{1.1.3}$$

(iv) Se $Z = \frac{S_n - nm}{\sigma\sqrt{n}}$, onde m e σ são, respectivamente, a esperança e variância da v.a. X_1, então

$$\lambda_Z(x) = n\lambda\left(\frac{\sigma x}{\sqrt{n}} + m\right), \quad x \in \mathbf{R}.$$

Prova: Utilizaremos as propriedades da função geradora de momentos (ver Proposição A.1.1 do Apêndice 1).

(i) Sabe-se que $\mu_Y(t) = e^{tb}\mu(at)$. Disso decorre que

$$\begin{aligned}\lambda_Y(x) &= \sup_{t\in\mathbf{R}}\{xt - \ln \mu_Y(t)\} = \sup_{t\in\mathbf{R}}\left\{xt - \ln\left(e^{tb}\hat{\mu}(at)\right)\right\}\\ &= \sup_{t\in\mathbf{R}}\{xt - tb - \ln\hat{\mu}(at)\} = \sup_{t\in\mathbf{R}}\left\{at\left(\frac{x-b}{a}\right) - \ln\hat{\mu}(at)\right\}.\end{aligned}$$

Fazendo a mudança de variável $s = at$, finalmente vem

$$\sup_{s\in\mathbf{R}}\left\{s\left(\frac{x-b}{a}\right) - \ln\hat{\mu}(s)\right\} = \lambda\left(\frac{x-b}{a}\right).$$

(ii)

$$\lambda_{S_n}(x) = \sup_{t\in\mathbf{R}}\{xt - \ln\hat{\mu}_{S_n}(t)\} = \sup_{t\in\mathbf{R}}\left\{n\,\frac{x}{n}\,t - n\ln\hat{\mu}(t)\right\} = n\sup_{t\in\mathbf{R}}\left\{\frac{x}{n}\,t - \ln\hat{\mu}(t)\right\} = n\lambda\left(\frac{x}{n}\right).$$

(iii)

$$\begin{aligned}\lambda_W(x) &= \sup_{t\in\mathbf{R}}\{xt - \ln\hat{\mu}_W(t)\} = \sup_{t\in\mathbf{R}}\left\{xt - \ln(e^{-at}\hat{\mu}(t))^n\right\}\\ &= \sup_{t\in\mathbf{R}}\left\{n\frac{x}{n}t + nat - n\ln\hat{\mu}(t)\right\} = n\sup_{t\in\mathbf{R}}\left\{\left(\frac{x}{n}+a\right)t - \ln\hat{\mu}(t)\right\}\\ &= n\lambda\left(\frac{x}{n}+a\right).\end{aligned}$$

(iv)

$$\begin{aligned}\lambda_Z(x) &= \sup_{t\in\mathbf{R}}\{xt - \ln\hat{\mu}_Z(t)\} = \sup_{t\in\mathbf{R}}\left\{xt - \ln\left(e^{-\frac{tm}{\sigma\sqrt{n}}}\hat{\mu}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n\right\}\\ &= n\sup_{t\in\mathbf{R}}\left\{\frac{t}{\sigma\sqrt{n}}\left(\frac{\sigma x}{\sqrt{n}}+m\right) - \ln\hat{\mu}\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}.\end{aligned}$$

Fazendo a mudança de variável $s = \frac{t}{\sigma\sqrt{n}}$, tem-se

$$n\sup_{s\in\mathbf{R}}\left\{s\left(\frac{\sigma x}{\sqrt{n}}+m\right) - \ln\hat{\mu}(s)\right\} = n\lambda\left(\frac{\sigma x}{\sqrt{n}}+m\right).$$

□

Proposição 1.1.4. *Seja μ uma medida de probabilidade e $\hat{\mu}$ sua respectiva transformada de Laplace. Para a transformada de Cramér λ, tem-se:*

(i) Se $\hat{\mu}(t) = +\infty$ para $\forall t \in \mathbf{R}^*$ então $\lambda(x) = 0$ qualquer que seja $x \in \mathbf{R}$.

(ii) Suponha $\hat{\mu}(t) \geq 1$ para $t \in (-\infty, 0]$. Então $\hat{\mu}(t) = +\infty$ para $t \in (0, +\infty)$ se e somente se $\lambda(x) = 0$ para $x \in [0, +\infty)$. Suponha $\hat{\mu}(t) \geq 1$ para $t \in [0, +\infty)$. Então $\hat{\mu}(t) = +\infty$ para $t \in (-\infty, 0)$ se e somente se $\lambda(x) = 0$ para $x \in (-\infty, 0]$.

(iii) Suponha $\hat{\mu}(t) \geq 1$ para $t \in (-\infty, 0]$. Existe $\tau > 0$ finito tal que $\hat{\mu}(\tau) < +\infty$ se e somente se $\lim_{x \to +\infty} \lambda(x) = +\infty$. Suponha $\hat{\mu}(t) \geq 1$ para $t \in [0, +\infty)$. Existe $\tau < 0$ finito tal que $\hat{\mu}(\tau) < +\infty$ se e somente se $\lim_{x \to -\infty} \lambda(x) = +\infty$.

(iv) Se $m = \int_{-\infty}^{+\infty} x\, d\mu(x)$ for finita então $\hat{\mu}(t) = +\infty$, $\forall\, t > 0$, se e somente se $\lambda(x) = 0$, $\forall\, x \geq m$, e $\hat{\mu}(t) = +\infty$, $\forall\, t < 0$, se e somente se $\lambda(x) = 0$, $\forall\, x \leq m$.

(v) Se $m = \int_{-\infty}^{+\infty} x\, d\mu(x)$ for finita então $\exists\, \tau > 0$ finito tal que $\hat{\mu}(\tau) < +\infty$ se e somente se $\lim_{x \to +\infty} \lambda(x) = +\infty$, e $\exists\, \tau < 0$ finito tal que $\hat{\mu}(\tau) < +\infty$ se e somente se $\lim_{x \to -\infty} \lambda(x) = +\infty$.

Prova: A prova é segundo Azencott **[2]**. Em (ii), (iii), (iv) e (v) provaremos o caso $t > 0$, $x \geq 0$ e $x \to +\infty$, pois para $t < 0$, $x \leq 0$ e $x \to -\infty$ o procedimento é análogo.

(i) Suponha $\hat{\mu}(t) = +\infty$, $\forall\, t \in \mathbf{R}^*$. Então, $tx - \ln \hat{\mu}(t) = -\infty$, $\forall\, t \in \mathbf{R}^*$, seja qual for $x \in \mathbf{R}$. Portanto, $\lambda(x) = 0$.

(ii) Assuma $\hat{\mu}(t) \geq 1$, $\forall\, t \in (-\infty, 0]$. Supondo $\hat{\mu}(t) = +\infty$, $\forall\, t \in (0, +\infty)$, teremos $tx - \ln \hat{\mu}(t) = -\infty$ seja qual for $x \in \mathbf{R}$. Logo, $\lambda(x) = \sup_{t \leq 0}\{tx - \ln \hat{\mu}(t)\}$. Mas pelo fato de $\hat{\mu}(t) \geq 1$, $\forall\, t \in (-\infty, 0]$, tem-se $tx - \ln \hat{\mu}(t) \leq 0$ se $x \in [0, +\infty)$ e assim $\lambda(x) = 0$. Para compreender-se a razão da suposição $\hat{\mu}(t) \geq 1$, $\forall\, t \in (-\infty, 0]$, suponha que exista $\delta \in (-\infty, 0)$ tal que $\hat{\mu}(\delta) < 1$. Sendo assim, podemos encontrar $x \in (0, +\infty)$ de tal maneira que $x\delta - \ln \hat{\mu}(\delta) > 0$, o que contraria a suposição $\hat{\mu}(t) = +\infty$ para todo $t > 0$. Para provar a recíproca suponha que exista $\tau > 0$ tal que $\hat{\mu}(\tau) < +\infty$. Então, da definição de λ temos $\lambda(x) \geq \tau x - \ln \hat{\mu}(\tau)$ e daí $\lim_{x \to +\infty} \lambda(x) \geq \lim_{x \to +\infty}\{\tau x - \ln \hat{\mu}(\tau)\} = +\infty$, ou seja, λ não é identicamente nula para $x \in [0, +\infty)$. Portanto, $\lambda(x) = 0$ para $\forall\, x \in [0, +\infty)$ implica em $\hat{\mu}(t) = +\infty$, $\forall\, t \in (0, +\infty)$.

(iii) Na prova da recíproca de (ii) vimos que se $\exists \tau > 0$ tal que $\hat{\mu}(\tau) < +\infty$ então $\lim_{x \to +\infty} \lambda(x) = +\infty$. Para provar que $\lim_{x \to +\infty} \lambda(x) = +\infty$ implica na existência de $\tau > 0$ finito tal que $\hat{\mu}(\tau) < +\infty$ utilizaremos a demonstração por absurdo. Sendo assim, assuma $\hat{\mu}(t) = +\infty$ para $\forall$, $t \in (0, +\infty)$. Então, pela prova do item (ii), temos que $\lambda(x) = 0$ seja qual for $x \in [0, +\infty)$, isto é, chega-se a uma negação de $\lim_{x \to +\infty} \lambda(x) = +\infty$.

(iv) Suponha que $\hat{\mu}(t) = +\infty$, $\forall t > 0$. Então, $\sup_{t > 0}\{tx - \ln \hat{\mu}(t)\} = -\infty$. Portanto, $\lambda(x) = \sup_{t \leq 0}\{tx - \ln \hat{\mu}(t)\}$. Por outro lado, por (1.1.2), se $x \geq m$ então $tx - \ln \hat{\mu}(t) \leq 0$ para todo $t \leq 0$. Assim, $\lambda(x) = 0$, $\forall\, x \geq m$. Suponha agora que $\lambda(x) = 0$, $\forall\, x \geq m$, e que $\exists\, \tau > 0$ com $\hat{\mu}(\tau) < +\infty$. Sendo assim,

$$\lim_{x \to +\infty} \lambda(x) \geq \lim_{x \to +\infty} (\tau x - \ln \mu(\tau)) = +\infty,$$

que é uma contradição.

(v) Aqui a prova é feita como em (iii).

□

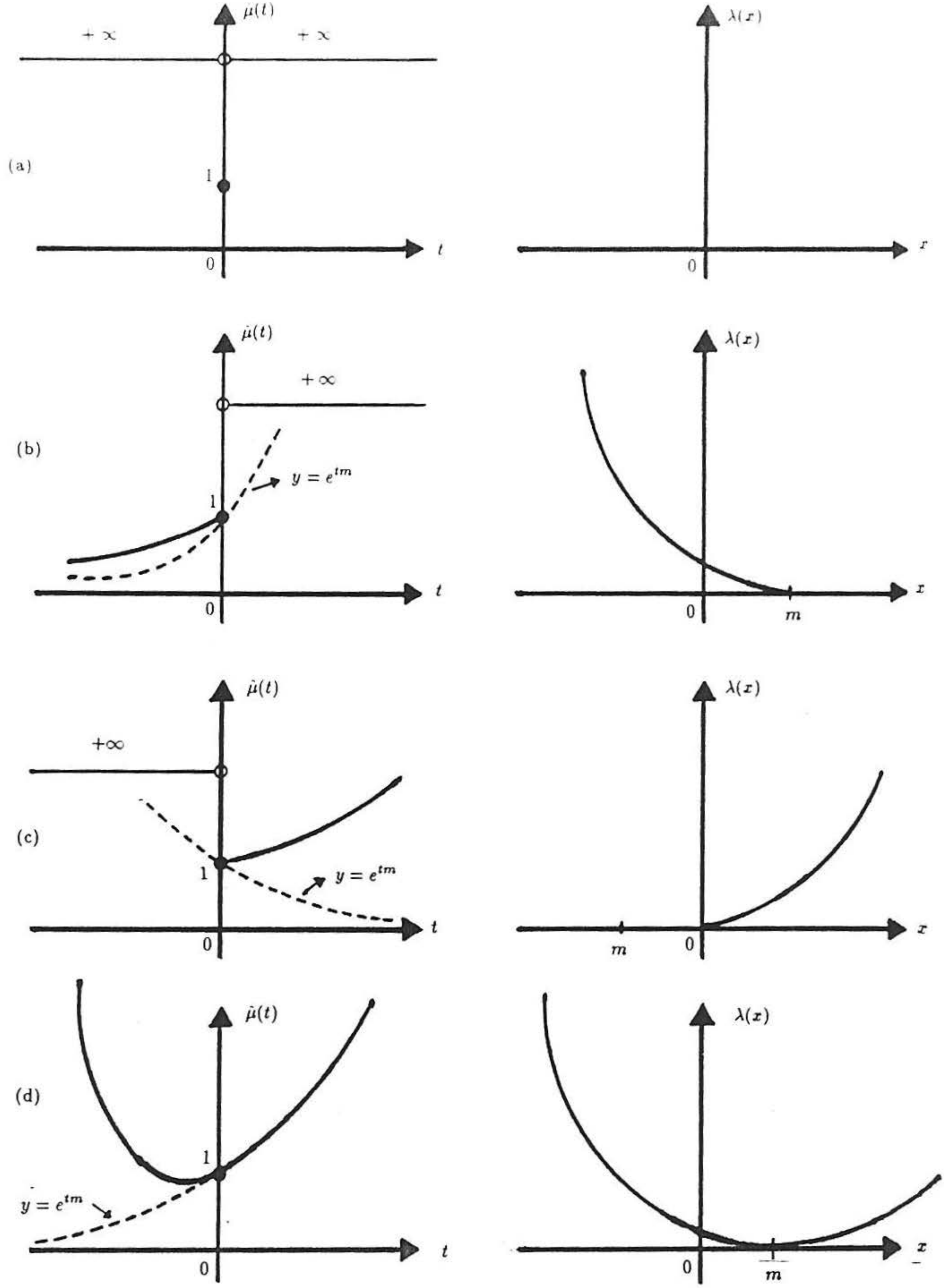

figura 1.1.2 - Relação entre $\hat{\mu}$ e λ

Para compreendermos melhor a transformada de Cramér, bem como suas propriedades, veremos alguns exemplos envolvendo algumas distribuições de probabilidade conhecidas. No cálculo de λ usaremos (1.1.1).

Exemplo 1.1.1. Considere X uma v.a. constante, isto é, $\mu(a) = P(X = a) = 1$, $a \in \mathbf{R}$. A transformada de Laplace é $\hat{\mu}(t) = e^{ta}$, $\forall t \in \mathbf{R}$, e a de Cramér é

$$\lambda(x) = \begin{cases} +\infty, & x \neq a \\ 0, & x = a \end{cases}.$$

Note que λ não é diferenciável para nenhum $x \in \mathbf{R}$.

Exemplo 1.1.2. Seja X uma v.a. com lei de probabilidade $\mu = p\,\delta_v + (1-p)\delta_w$, onde $w < v$ e $0 < p < 1$. As funções δ_w e δ_v são definidas como

$$\delta_w = \begin{cases} 1 & \text{se } X = w \\ 0 & \text{caso contrário} \end{cases} \quad \text{e} \quad \delta_v = \begin{cases} 1 & \text{se } X = v \\ 0 & \text{caso contrário} \end{cases}.$$

A transformada de Laplace é $\hat{\mu}(t) = p\,e^{vt} + (1-p)e^{wt}$, $t \in \mathbf{R}$, a qual é de classe $C^{\infty}(\mathbf{R})$. Minimizando $e^{-tx}\hat{\mu}(t)$ em relação a t obtemos

$$e^{\tau} = \left[\left(\frac{1-p}{p}\right)\left(\frac{x-w}{v-x}\right)\right]^{\frac{1}{v-w}}, \quad x \in (w, v),$$

onde τ é o ponto de mínimo. Logo, após alguns cálculos, obtém-se

$$\begin{aligned} -\ln\left(e^{-\tau x}\hat{\mu}(\tau)\right) &= -\ln\left(\left(\frac{1-p}{p}\right)^{-\left(\frac{x-w}{v-w}\right)}\left(\frac{x-w}{v-x}\right)^{-\left(\frac{x-w}{v-w}\right)}\frac{(1-p)(v-w)}{(v-x)}\right) \\ &= \left(\frac{x-w}{v-w}\right)\ln\left(\frac{x-w}{p}\right) + \left(\frac{v-x}{v-w}\right)\ln\left(\frac{v-x}{1-p}\right) - \ln(v-w). \end{aligned}$$

Se $x = w$,

$$\lambda(w) = -\ln\inf_{t\in\mathbf{R}}\left\{p\,e^{t(v-w)} + (1-p)\right\} = -\ln(1-p) = \ln\left(\frac{1}{1-p}\right),$$

e se $x = v$,

$$\lambda(v) = -\ln\inf_{t\in\mathbf{R}}\left\{p + (1-p)e^{t(w-v)}\right\} = -\ln p = \ln\left(\frac{1}{p}\right).$$

No caso $x < w$, $\lambda(x) = -\lim_{t\to -\infty} \ln\left(p\,e^{t(v-x)} + (1-p)e^{t(w-x)}\right) = +\infty$ e para $x > v$, $\lambda(x) = -\lim_{t\to +\infty} \ln\left(p\,e^{t(v-x)} + (1-p)e^{t(w-x)}\right) = +\infty$.

Assim,

$$\lambda(x) = \begin{cases} \ln\left(\frac{1}{1-p}\right), & x = w \\ \left(\frac{x-w}{v-w}\right)\ln\left(\frac{x-w}{p}\right) + \left(\frac{v-x}{v-w}\right)\ln\left(\frac{v-x}{1-p}\right) - \ln(v-w), & w < x < v \\ \ln\left(\frac{1}{p}\right), & x = v \\ +\infty, & x \notin [w,v] \end{cases}.$$

Sendo $EX = pv + (1-p)w$, por (iii) da Proposição 1.1.2, $\lambda(EX) = 0$. Uma vez que λ é de classe $C^{\infty}((w,v))$ então por (v) dessa mesma proposição temos que $\lambda'(EX) = 0$ e $\lambda''(EX) = \frac{1}{(w-v)^2 p(1-p)} > 0$.

Exemplo 1.1.3. X tem distribuição Normal com parâmetros $m = EX$ e $\sigma^2 = VarX$, isto é, $d\mu(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$, $x \in \mathbf{R}$.

Sabe-se que $\hat{\mu}(t) = e^{mt + \frac{\sigma^2}{2}t^2}$, $t \in \mathbf{R}$, a qual é uma função de classe $C^{\infty}(\mathbf{R})$. Minimizar $e^{-tx}\hat{\mu}(t)$ equivale a minimizar $-tx + mt + \frac{\sigma^2}{2}t^2$. Portanto, obtemos o ponto de mínimo $\tau = \frac{x-m}{\sigma^2}$, $x \in \mathbf{R}$, e isto resulta em

$$\lambda(x) = -\ln\left(e^{-\tau x}\hat{\mu}(\tau)\right) = -\ln e^{-\frac{(x-m)^2}{2\sigma^2}} = \frac{(x-m)^2}{2\sigma^2},$$

para $x \in \mathbf{R}$. É claro que $\lambda(m) = 0$, $\lambda'(m) = 0$ e $\lambda''(m) = \frac{1}{\sigma^2} > 0$, visto que λ é de classe $C^{\infty}(\mathbf{R})$.

Exemplo 1.1.4. X tem distribuição de Cauchy com parâmetros $\theta \in \mathbf{R}$ e $\delta > 0$, ou seja, $d\mu(x) = \frac{\delta}{\pi(\delta^2 + (x-\theta)^2)} dx$, $x \in \mathbf{R}$.

A função geradora de momentos de X é tal que $\hat{\mu}(t) = +\infty$, $\forall t \neq 0$. Logo, pela Proposição 1.1.4 (i), $\lambda(x) = 0$ para $\forall x \in \mathbf{R}$.

Exemplo 1.1.5. Seja X v.a. com distribuição Gama de parâmetros $\alpha > 0$ e $k > 0$. A densidade e a transformada de Laplace de X são respectivamente:

$$\frac{\alpha^k}{\Gamma(k)} e^{-\alpha x} x^{k-1},\ x > 0,\ k > 0,\ \alpha > 0 \qquad \text{e} \qquad \hat{\mu}(t) = \begin{cases} \left(1 - \frac{t}{\alpha}\right)^{-k}, & t < \alpha \\ +\infty, & t \geq \alpha \end{cases}.$$

Observe que $\hat{\mu}$ é de classe $C^{\infty}((-\infty,\alpha))$.

Minimizando $e^{-tx}\hat{\mu}(t)$ em relação a t temos que o ponto de mínimo é $\tau = \alpha - \frac{k}{x}$, $x > 0$. Logo, $\lambda(x) = \alpha x - k + k\ln\left(\frac{k}{\alpha x}\right)$, se $x > 0$ e $\lambda(x) = \lim_{t\to -\infty}\left\{tx - ln\left(1 - \frac{t}{\alpha}\right)^{-k}\right\} = +\infty$, se $x \leq 0$. Sendo $\frac{k}{\alpha}$ a esperança de X, então $\lambda\left(\frac{k}{\alpha}\right) = 0$. Uma vez que λ é de classe $C^{\infty}((0,+\infty))$, $\lambda'\left(\frac{k}{\alpha}\right) = 0$ e $\lambda''\left(\frac{k}{\alpha}\right) = \frac{k}{\alpha^2} > 0$.

Sabe-se que quando $k = 1$, tem-se a distribuição exponencial. Quando $k = \frac{n}{2}$ (n inteiro positivo) e $\alpha = \frac{1}{2}$, tem-se a distribuição qui-quadrado. Logo, através da expressão de λ obtida para a distribuição Gama podemos obter λ para estas duas distribuições.

1.2 - O teorema de Cramér-Chernoff

Nesta seção nos concentraremos no importante teorema de Cramér-Chernoff, cuja demonstração é extraída de Vares [30].

Teorema 1.2.1. *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de v.a's i.i.d. definidas no espaço de probabilidade* $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mu)$, *com* $m = \int_{-\infty}^{+\infty} x\, d\mu(x)$ *finita, transformada de Laplace* $\hat{\mu}$ *e transformada de Cramér* λ. *Seja* $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} = \frac{S_n}{n}$. *Então:*

(i) $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \leq x) = -\lambda(x), \quad x \leq m;$

(ii) $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \geq x) = -\lambda(x), \quad x \geq m.$

Prova: Decorrência do teorema de Cramér-Chernoff, a seguir.

Teorema 1.2.2 (Cramér-Chernoff). *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de v.a's i.i.d. definidas no espaço de probabilidade* $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mu)$, *com* $m = EX_1$, *transformada de Laplace* $\hat{\mu}$ *e transformada de Cramér* λ.

(i) Se $E|X_1| < +\infty$ então

(1.2.1) $$P(\bar{X}_n \leq x) \leq e^{-n\lambda(x)}, \quad \text{se } x \leq m, \quad \text{e}$$

(1.2.2) $$P(\bar{X}_n \geq x) \leq e^{-n\lambda(x)}, \quad \text{se } x \geq m;$$

(ii)

(1.2.3) $$-\lambda(x) \leq \varliminf_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \leq x), \quad \forall x \in \mathbf{R}, \quad \text{e}$$

(1.2.4) $$-\lambda(x) \leq \varliminf_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \geq x), \; \forall x \in \mathbf{R}.$$

Prova:

(i) Sem perda de generalidade podemos supor $x = 0$, pois como veremos adiante é possível reduzir a este caso particular se $x \neq 0$. Primeiro consideraremos o caso $m \geq 0$. Para $t \leq 0$

$$P(\bar{X}_n \leq 0) = P(tS_n \geq 0) = P\left(e^{tS_n} \geq 1\right).$$

Logo, pela desigualdade de Markov e independência das v.a's X_i temos que $P\left(e^{tS_n} \geq 1\right) \leq E\left(e^{tS_n}\right) = (\hat{\mu}(t))^n$. Disso segue que

$$P(\bar{X}_n \leq 0) \leq (\hat{\mu}(t))^n, \forall t \leq 0.$$

Por outro lado, através da desigualdade de Jensen, tem-se que

$$\hat{\mu}(t) = E\left(e^{tX_1}\right) \geq e^{tm} \geq e^{t0} = 1 = \hat{\mu}(0), \quad \text{se} \quad t \geq 0,$$

e isto permite-nos concluir que $\inf_{t \in \mathbf{R}} \hat{\mu}(t) = \inf_{t \leq 0} \hat{\mu}(t)$. Uma outra maneira de ver isto, quando $\hat{\mu}(t)$ é de classe $C^{\infty}(\mathbf{R})$, é observar que como $\hat{\mu}$ é convexa (ver Proposição A.1.2 do Apêndice 1) e $m = \hat{\mu}'(0) \geq 0$ então $\inf_{t \in \mathbf{R}} \hat{\mu}(t)$ ocorre quando $t \leq 0$, conforme a figura 1.2.1(a), na qual é possível ver que $\inf_{t \in \mathbf{R}} \hat{\mu}(t) \in [0,1]$. Logo,

$$P(\bar{X}_n \leq 0) \leq (\inf_{t \in \mathbf{R}} \hat{\mu}(t))^n = e^{n \ln \inf_{t \in \mathbf{R}} \hat{\mu}(t)} = e^{-n\lambda(0)},$$

onde a última igualdade é justificada por (1.1.1).

Agora trataremos o caso $m \leq 0$. Tomando $t \geq 0$ e aplicando novamente a desigualdade de Markov, obtém-se

$$P(\bar{X}_n \geq 0) = P\left(e^{tS_n} \geq 1\right) \leq E\left(e^{tS_n}\right) = (\hat{\mu}(t))^n.$$

Aplicando a desigualdade de Jensen quando $t \leq 0$, chega-se a $\hat{\mu}(t) \geq 1$, e portanto, $\inf_{t \in \mathbf{R}} \hat{\mu}(t) = \inf_{t \geq 0} \hat{\mu}(t)$. Geometricamente, considerando $\hat{\mu}(t)$ de classe $C^{\infty}(\mathbf{R})$, pela figura 1.2.1(b) vemos que o ínfimo de $\hat{\mu}$ ocorre quando $t \geq 0$. Disso vem que

$$P(\bar{X}_n \leq 0) \leq (\inf_{t \in \mathbf{R}} \hat{\mu}(t))^n = e^{n \ln \inf_{t \in \mathbf{R}} \hat{\mu}(t)} = e^{-n\lambda(0)}.$$

Para $x \neq 0$, com $x \leq m$, $E(\bar{X}_n - x) = m - x \geq 0$ e assim podemos reduzir ao caso particular acima provado, ou seja,

$$P(\bar{X}_n \leq x) = P(S_n - nx \leq 0) \leq \inf_{t\in\mathbb{R}} \hat{\mu}_{S_n - nx}(t) = e^{\ln \inf_{t\in\mathbb{R}} \hat{\mu}_{S_n-nx}(t)} = e^{-\lambda_{S_n - nx}(0)}.$$

Mas por (1.1.3), $e^{-\lambda_{S_n-nx}(0)} = e^{-n\lambda(x)}$, e disso segue $P(\bar{X}_n \leq x) \leq e^{-n\lambda(x)}$. Analogamente $P(\bar{X}_n \geq x) \leq e^{-n\lambda(x)}$, se $x \neq 0$ e $m \leq x$.

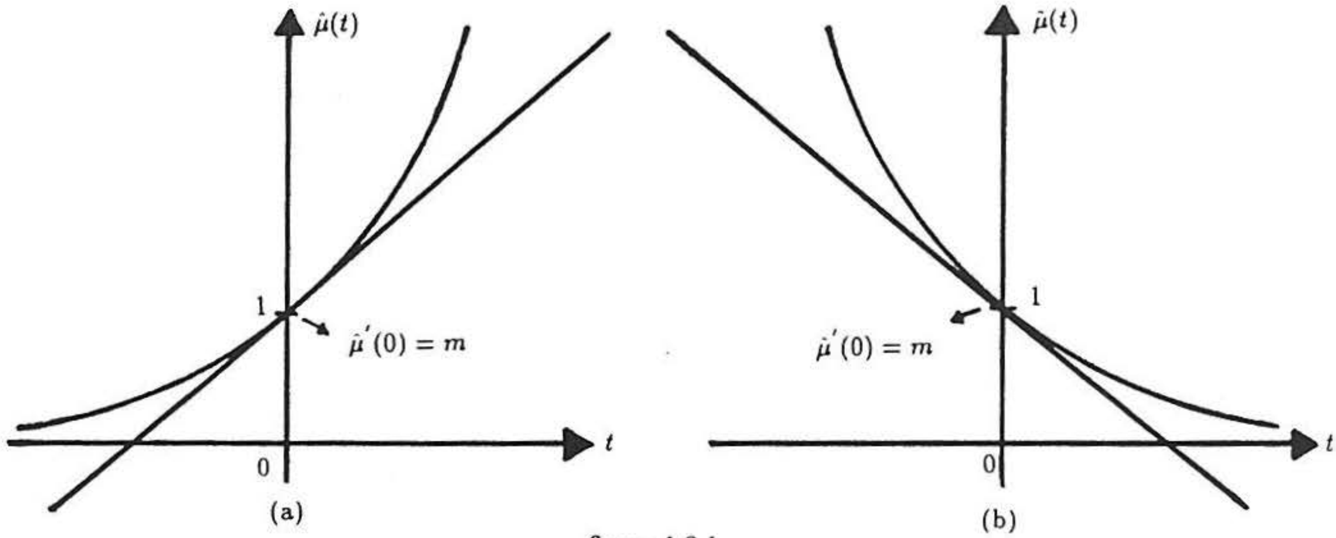

figura 1.2.1

(ii) Se $P(X_1 > 0) = 0$ então $P(\bar{X}_n \leq 0) = 1$ e $\frac{1}{n}\ln P(\bar{X}_n \leq 0) = 0, \forall n \geq 1$, e isto resulta em $\underline{\lim}_{n\to\infty} \frac{1}{n}\ln P(\bar{X}_n \leq 0) \geq -\lambda(0)$. Analogamente, se $P(X_1 < 0) = 0$ então $P(\bar{X}_n \geq 0) = 1$ e $\frac{1}{n}\ln P(\bar{X}_n \geq 0) = 0, \forall n \geq 1$, resultando em $\underline{\lim}_{n\to\infty} \frac{1}{n}\ln P(\bar{X}_n \geq 0) \geq -\lambda(0)$. Logo, nos casos $P(X_1 < 0) = 0$ e $P(X_1 > 0) = 0$, (1.2.3) e (1.2.4) são facilmente verificadas. Por isto vamos supor $P(X_1 < 0) > 0$ e $P(X_1 > 0) > 0$ no que segue. A demonstração será dividida em duas partes: caso onde X_i é v.a. discreta e o caso geral.

Caso Discreto:

Seja $\{X_n\}_{n\geq 1}$ uma seqüência de v.a's i.i.d. cuja distribuição de probabilidade tem suporte em $\{x_1, x_2, \cdots\}$, isto é, $P(X_1 = x_i) = p_i > 0$ para $i \in \{1, 2, \cdots\}$ e $\sum_{i=1}^{\infty} p_i = 1$. Da suposição $P(X_1 < 0) > 0$ e $P(X_1 > 0) > 0$ temos que $\exists r \in \mathbf{N}^*$ tal que $\min_{1\leq i\leq r} x_i < 0 < \max_{1\leq i\leq r} x_i$. Disso, tem-se $\sum_{i=1}^{r} p_i < 1$.

Defina $\varphi_k(t) \equiv \sum_{i=1}^{k} p_i e^{tx_i}$, $k \geq 1$. A seqüência $\{\varphi_k\}_{k\geq 1}$ é tal que $\varphi_1(t) \leq \varphi_2(t) \leq \cdots$, ou seja, é monótona crescente, convergindo pontualmente para $\hat{\mu}(t) = \sum_{i=1}^{\infty} p_i e^{tx_i}$, quando $t \in \mathbf{R}$. Uma vez que existe um conjunto compacto C tal que $C \subseteq \{t : \hat{\mu}(t) < +\infty\}$ (de fato, mesmo que $\hat{\mu}(t) = +\infty$ para todo $t \neq 0$, tem-se $\hat{\mu}(0) = 1$ e daí $C = \{0\}$, o qual é compacto) então pelo Teorema A.2.2 do Apêndice 2 temos que $\{\varphi_k\}_{k\geq 1}$ converge uniformemente para $\hat{\mu}$ em $t \in C$. Logo,

$$(1.2.5) \qquad \lim_{k\to\infty} \inf_{t\in C} \varphi_k(t) = \inf_{t\in C} \lim_{k\to\infty} \varphi_k(t) = \inf_{t\in C} \hat{\mu}(t).$$

Vamos admitir, sem perda de generalidade, que $x_1 = \min_{1\leq i\leq r} x_i$ e $x_r = \max_{1\leq i\leq r} x_i$ Observe que $\lim_{t\to+\infty} \varphi_1(t) = \lim_{t\to+\infty} p_1 e^{x_1 t} = 0$ e $\lim_{t\to-\infty} \varphi_1(t) = \lim_{t\to-\infty} p_1 e^{x_1 t} = +\infty$, isto é, não podemos afirmar que para $k < r$, $\varphi_k(t)$ converge para $+\infty$ quando $|t| \to \infty$. Entretanto, $\varphi_r(t) = p_1 e^{x_1 t} + \cdots + p_r e^{x_r t}$ converge a $+\infty$ quando $|t| \to +\infty$, pois $\lim_{t\to-\infty} p_1 e^{x_1 t} = +\infty$ e $\lim_{t\to+\infty} p_r e^{x_r t} = +\infty$, e assim $\varphi_k(t)$ converge a $+\infty$ quando $|t| \to +\infty$, para $k \geq r$. Também vale que $\hat{\mu}(t)$ converge a $+\infty$ quando $|t| \to +\infty$ (ver Teorema A.1.1 do Apêndice 1).

Fixe $k \geq r \geq 1$ e defina $\alpha_k \equiv \inf_{t\in\mathbb{R}} \varphi_k(t)$. Pelo fato de φ_k e $\hat{\mu}$ serem convexas, estritamente positivas e $\lim_{|t|\to\infty} \varphi_k(t) = \lim_{|t|\to\infty} \hat{\mu}(t) = +\infty$ então existem τ_k e τ finitos tais que $\alpha_k = \varphi_k(\tau_k)$ e $\hat{\mu}(\tau) = \inf_{t\in\mathbb{R}} \hat{\mu}(t)$. Tomando C compacto tal que $\tau_k, \tau \in C$ teremos $\alpha_k = \inf_{t\in C} \varphi_k(t)$ e $\inf_{t\in\mathbb{R}} \hat{\mu}(t) = \inf_{t\in C} \hat{\mu}(t)$. Assim, por (1.2.5)

$$\lim_{k\to\infty} \alpha_k = \inf_{t\in\mathbb{R}} \hat{\mu}(t) = e^{-\lambda(0)}.$$

Sejam $A_1 = \{x_1\}, A_2 = \{x_2\}, \cdots, A_k = \{x_k\}$ e $A_{k+1} = \{x_{k+1}, \cdots\}$. Sendo assim, $\{A_1, \cdots, A_{k+1}\}$ pode ser considerado como o espaço amostral de um experimento aleatório $\mathcal{E}$ com $k+1$ resultados possíveis, onde $P(A_i) = p_i$, $i = 1, \cdots, k$, e $P(A_{k+1}) = 1 - p_1 - \cdots - p_k$. Definindo as v.a's $M_i \equiv \{$ número de vezes que A_i ocorre em n repetições independentes do experimento $\mathcal{E}\}$, $i \in \{1, \cdots, k+1\}$, segue que $(M_1, \cdots, M_{k+1})$ tem distribuição de probabilidade multinomial (ver James **[17]**) com parâmetros $(n, p_1, \cdots, p_k, 1 - p_1 - \cdots - p_k)$. Pelo fato de $[X_1 + \cdots + X_n \leq 0 \,:\, X_i \in \{x_1, \cdots, x_k\},\, i = 1, \cdots, n] \subset [X_1 + \cdots + X_n \leq 0]$ e $[X_1 + \cdots + X_n \leq 0 \,:\, X_i \in \{x_1, \cdots, x_k\},\, i = 1, \cdots, n]$ ser equivalente ao evento $[M_1 x_1 + \cdots + M_k x_k \leq 0,\, M_{k+1} = 0]$, tem-se

$$\begin{aligned}
P(\bar{X}_n \leq 0) &= P(X_1 + \cdots + X_n \leq 0) \geq P(M_1 x_1 + \cdots + M_k x_k \leq 0, M_{k+1} = 0) \\
&= \sum_{(m_1,\cdots,m_k)\in S} P(M_1 = m_1, \cdots, M_k = m_k, M_{k+1} = 0) \\
&\geq P(M_1 = n_1, \cdots, M_k = n_k, M_{k+1} = 0) \\
&= n! \prod_{i=1}^{k} \frac{p_i^{n_i}}{n_i!} = p(n_1, \cdots, n_k),
\end{aligned}$$

onde $S = \left\{(m_1, \cdots, m_k) : \sum_{i=1}^k m_i x_i \leq 0,\, \sum_{i=1}^k m_i = n\right\}$. Note que $(n_1, \cdots, n_k) \in S$.

Pelo Lema A.2.2 do Apêndice 2, $\exists c_k > 0$, $N_k \in \mathbf{N}^*$ tal que para $n \geq N_k$ existem inteiros positivos $n_1, \ldots, n_k$ tais que $\sum_{i=1}^{k} n_i = n$, $\sum_{i=1}^{k} n_i x_i \leq 0$ e $p(n_1, \cdots, n_k) \geq c_k \frac{1}{n^{k/2}} (\alpha_k)^n$. Daí vem que

$$\frac{1}{n} \ln P(\bar{X}_n \leq 0) \geq \ln \left(\sqrt[n]{c_k} \frac{1}{\sqrt[2n]{n^k}} \right) + \ln \alpha_k,$$

e conseqüentemente $\underline{\lim}_{n \to \infty} \frac{1}{n} \ln P(\bar{X}_n \leq 0) \geq \ln \alpha_k$, para todo $k \geq r \geq 1$.

Uma vez que $\ln \alpha_k$ converge para $\ln \inf_{t \in \mathbf{R}} \hat{\mu}(t) = -\lambda(0)$ quando $k \to \infty$, chega-se a $\underline{\lim}_{n \to \infty} \frac{1}{n} \ln P(\bar{X}_n \leq 0) \geq -\lambda(0)$. O procedimento para $P(\bar{X}_n \geq 0)$ é semelhante. Para $x \neq 0$ observe que $\left[\frac{\sum_{i=1}^{n} (X_i - x)}{n} \leq 0 \right]$ e $[\bar{X}_n \leq x]$ são equivalentes. Sendo assim, repetindo o mesmo procedimento acima para a seqüência de v.a's $\{X_n - x\}_{n \geq 1}$, chega-se a $\underline{\lim}_{n \to \infty} \frac{1}{n} \ln P \left(\frac{\sum_{i=1}^{n} (X_i - x)}{n} \leq 0 \right) \geq -\lambda_{X_1 - x}(0)$. Pela Proposição 1.1.3(i), com $a = 1$ e $b = -x$, vem que $\lambda_{X_1 - x}(0) = \lambda(x)$. Portanto, $\underline{\lim}_{n \to \infty} \frac{1}{n} \ln P\left(\bar{X}_n \leq x\right) \geq -\lambda(x)$. Para $P\left(\bar{X}_n \geq x\right)$ o raciocínio é o mesmo.

Caso Geral:

Utilizaremos o caso discreto acima. Assuma $X_1, \cdots, X_n$ v.a's i.i.d. quaisquer e defina

$$X_k^{(s)}(w) \equiv \frac{i}{s} I_{\left(\frac{i-1}{s}, \frac{i}{s}\right]}(X_k(w)) \quad \text{e} \quad \bar{X}_n^{(s)}(w) \equiv \frac{1}{n} \sum_{k=1}^{n} X_k^{(s)}(w),$$

onde $s, k \in \mathbf{N}^*$ e $i \in \mathbf{Z}$. Observe que as v.a's $X_k^{(s)}$ são i.i.d. Na figura 1.2.2 vemos que $X_k^{(s)}(w) - X_k(w) \leq \frac{1}{s}$ e $X_k(w) \leq X_k^{(s)}(w)$.

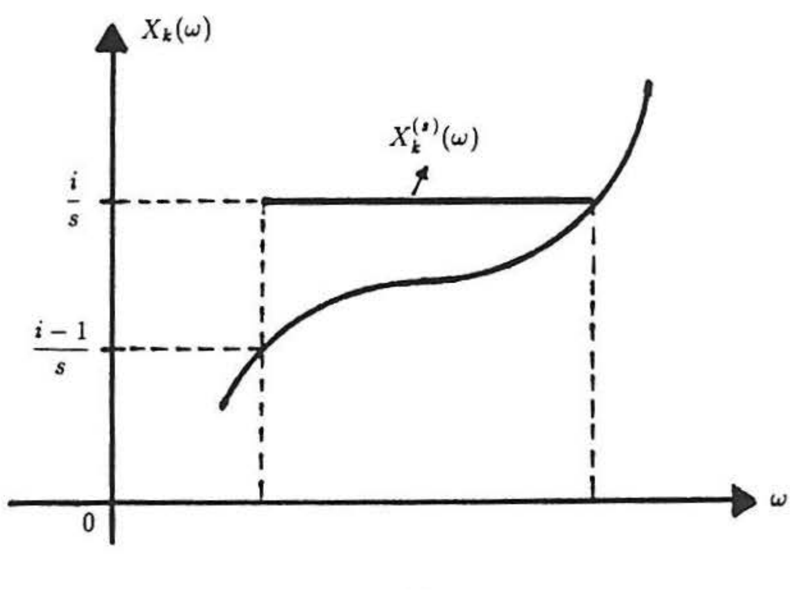

figura 1.2.2

Seja μ_s a lei comum de $\left\{X_n^{(s)}(w)\right\}_{n\geq 1}$. Então

$$tX_k \leq tX_k^{(s)} \leq t\left(X_k + \frac{1}{s}\right), \quad \text{se} \quad t \geq 0 \quad \text{e} \quad tX_k \geq tX_k^{(s)} \geq t\left(X_k + \frac{1}{s}\right), \quad \text{se} \quad t < 0.$$

Assim, $\hat{\mu}(t) \leq \hat{\mu}_s(t) \leq e^{t/s}\hat{\mu}(t)$ e $\hat{\mu}(t)e^{-t/s} \leq \hat{\mu}_s(t)$, para $t \geq 0$. Por outro lado, se $t < 0$ temos $\hat{\mu}(t) \geq \hat{\mu}_s(t) \geq e^{t/s}\hat{\mu}(t)$. Daí vem que $\hat{\mu}_s(t) \geq e^{-|t|/s}\hat{\mu}(t)$. Por (1.1.1),

$$e^{-\lambda_s(1/s)} = \inf_{t\in\mathbf{R}} \left\{e^{-t/s}\hat{\mu}_s(t)\right\} \geq \inf_{t\in\mathbf{R}} \left\{e^{-|t|/s}e^{-|t|/s}\hat{\mu}(t)\right\}, \tag{1.2.6}$$

sendo λ_s a transformada de Cramér para $X_k^{(s)}$.

A seqüência $\left\{e^{-2|t|/s}\hat{\mu}(t)\right\}_{s\geq 1}$ é monótona crescente, e portanto, converge pontualmente para $\hat{\mu}(t)$ quando $s \to \infty$. Considerando novamente o conjunto compacto $C \subseteq \{t : \hat{\mu}(t) < +\infty\}$, pelo Teorema A.2.2 do Apêndice 2, $e^{-2|t|/s}\hat{\mu}(t)$ converge uniformemente para $\hat{\mu}(t)$, $t \in C$, quando $s \to \infty$. Além disso, pelo Teorema A.1.1 do Apêndice 1, existem A, B reais positivos tal que $\hat{\mu}(t) \geq Ae^{B|t|}$. Logo, pelo fato de

$$\lim_{|t|\to\infty} e^{-2|t|/s}\hat{\mu}(t) \geq \lim_{|t|\to\infty} Ae^{|t|(B-2/s)} = +\infty, \quad \text{se} \quad s > \frac{2}{B},$$

e por $e^{-2|t|/s}\hat{\mu}(t)$ ser convexa, então existe $\tau \in \mathbf{R}$ finito tal que $e^{-2|\tau|/s}\hat{\mu}(\tau) = \inf_{t\in\mathbf{R}} e^{-2|t|/s}\hat{\mu}(t)$. Tomando C compacto tal que $\tau \in C$ teremos que $\inf_{t\in C} e^{-2|t|/s}\hat{\mu}(t) = \inf_{t\in\mathbf{R}} e^{-2|t|/s}\hat{\mu}(t)$. Assim,

$$\lim_{s\to\infty} \ln\left(\inf_{t\in\mathbf{R}} e^{-2|t|/s}\hat{\mu}(t)\right) = \ln\left(\inf_{t\in\mathbf{R}} \lim_{s\to\infty} e^{-2|t|/s}\hat{\mu}(t)\right) = \ln\left(\inf_{t\in\mathbf{R}} \hat{\mu}(t)\right) = -\lambda(0),$$

e por (1.2.6) concluimos que $\lim_{s\to\infty} -\lambda_s(\frac{1}{s}) \geq -\lambda(0)$.

De $X_k \leq X_k^{(s)}$ decorre que $\bar{X}_n \leq \bar{X}_n^{(s)}$ e por sua vez $P(\bar{X}_n \leq 0) \geq P(\bar{X}_n^{(s)} \leq 0)$. Uma vez que $X_k^{(s)}$ é v.a. discreta, pelo caso anterior

$$\varliminf_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \leq 0) \geq \varliminf_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n^{(s)} \leq 0) \geq -\lambda_s(\frac{1}{s}),$$

e quando $s \to \infty$, $\varliminf_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \leq 0) \geq -\lambda(0)$. Por outro lado, de $X_k^{(s)} \leq X_k + \frac{1}{s}$ temos que $\bar{X}_n^{(s)} \leq \bar{X}_n + \frac{1}{s}$ e isto implica em $P(\bar{X}_n + \frac{1}{s} \geq 0) \geq P(\bar{X}_n^{(s)} \geq 0)$. Novamente pelo caso discreto,

$$\varliminf_{n\to\infty} \frac{1}{n} \ln P\left(\bar{X}_n \geq -\frac{1}{s}\right) \geq \varliminf_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n^{(s)} \geq 0) \geq -\lambda_s(\frac{1}{s}),$$

e quando $s \to \infty$, $\underline{\lim}_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \geq 0) \geq -\lambda(0)$.

□

Observação 1.2.1. No caso em que $\mu(t) = +\infty$ para $\forall t \neq 0$, o Teorema 1.2.2 continua valendo, mas $\lambda(x) = 0$ seja qual for $x \in \mathbf{R}$ (Proposição 1.1.4 (i)), e por isso $P(\bar{X}_n \leq x) \leq e^{-n\lambda(x)} = 1$, ou seja, não ganhamos nenhuma informação relevante sobre $P(\bar{X}_n \leq x)$.

Observação 1.2.2. De (1.2.1) e (1.2.3), se $x \leq m$, então

$$-\lambda(x) \leq \underline{\lim}_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \leq x) \leq \overline{\lim}_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \leq x) \leq -\lambda(x),$$

ou seja, $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \leq x) = -\lambda(x)$. Para $x \geq m$, também chega-se a $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \geq x) = -\lambda(x)$, por (1.2.2) e (1.2.4). Assim, o Teorema 1.2.2 nos diz que a probabilidade de $\bar{X}_n$ pertencer ao intervalo $(-\infty, x]$, com $x \leq m$, e $[x, +\infty)$, com $x \geq m$, é da ordem $e^{-n\lambda(x)}$, isto é, $\forall \epsilon > 0$, $\exists N > 0$ tal que $\forall n \geq N$

$$e^{-n(\lambda(x)+\epsilon)} < P(\bar{X}_n \leq x) < e^{-n(\lambda(x)-\epsilon)}.$$

Além disso, se $x = m$ então $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \leq x) = -\lambda(m) = 0$, ou seja, a mesma informação que a Lei dos Grandes Números nos fornece.

Observação 1.2.3. Dizemos que existe uma equivalência logarítmica entre $P(\bar{X}_n \geq x)$ e $e^{-n\lambda(x)}$ quando $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \geq x) = -\lambda(x)$ e denotamos por $P(\bar{X}_n \geq x) \approx e^{-n\lambda(x)}$. Analogamente, existe uma equivalência logarítmica entre $P(\bar{X}_n \leq x)$ e $e^{-n\lambda(x)}$ se $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \leq x) = -\lambda(x)$.

Observação 1.2.4. Sob as hipóteses:

(1ª) $E\left(e^{tX_1}\right) < +\infty$, se $|t| < \delta$ para algum $\delta > 0$;

(2ª) X_1 com lei de probabilidade μ absolutamente contínua em relação à medida de Lebesgue,

Cramér [8] mostrou que $\lim_{n\to\infty} \frac{P(\bar{X}_n \leq x)}{e^{-n\lambda(x)}} = 1$, se $x \leq m$, ou seja, com estas hipóteses tem-se uma equivalência "exata" para a probabilidade do evento $[\bar{X}_n \leq x]$. Contudo, estas duas hipóteses restringem o universo de v.a's. Já a equivalência logarítmica é um resultado mais abrangente, embora forneça $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \leq x)$, $x \leq m$.

Observação 1.2.5. A equivalência exata é um resultado mais "forte" do que a equivalência logarítmica, pois se tivermos, por exemplo, $\lim_{n\to\infty} \frac{P(\bar{X}_n \leq x)}{e^{-n\lambda(x)}} = 1$, $x \leq m$, então isto implica em $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \leq$

$x) = \lambda(x)$. Além disso, a equivalência exata nos dá uma estimativa de $P(\bar{X}_n \leq x)$ quando $n \to \infty$, enquanto a logarítmica estima $\ln P(\bar{X}_n \leq x)$.

No exemplo abaixo, extraído de Chernoff [7], apresentamos uma interessante aplicação em Estatística Matemática.

Exemplo 1.2.1. Sejam $X_1, \cdots, X_n$ v.a's i.i.d. com distribuição de Bernoulli de parâmetro $0 < p < 1$. Supondo p desconhecido é natural formular as hipóteses $H_0 : p = p_0$ e $H_1 : p = p_1$, com $0 < p_0 < p_1 < 1$. O método da razão de verossimilhança para testar hipóteses simples nos diz que a regra de decisão é: decidir por H_0 se $\bar{X}_n < x$, caso contrário decidir por H_1, sendo $0 < x < 1$. Seja α_n e β_n as probabilidades dos erros tipo I e II respectivamente, isto é,

$$\alpha_n = P(\bar{X}_n \geq x | H_0) \qquad \text{e} \qquad \beta_n = P(\bar{X}_n < x | H_1).$$

Uma maneira de estimar α_n e β_n para n fixo é utilizar a aproximação normal, uma vez que pelo Teorema Central do Limite, $\sqrt{n}\left(\frac{\bar{X}_n - p}{\sqrt{p(1-p}}\right)$ converge em distribuição para uma normal padrão quando $n \to \infty$. Assim, para n fixo

$$\alpha_n \cong 1 - \Phi\left(\sqrt{n}\left(\frac{x - p_0}{\sqrt{p_0(1-p_0)}}\right)\right), 0 < x < 1, \qquad \text{e} \qquad \beta_n \cong \Phi\left(\sqrt{n}\left(\frac{x - p_1}{\sqrt{p_1(1-p_1)}}\right)\right), 0 < x < 1.$$

Entretanto, quando $n \to \infty$ os quocientes $\sqrt{n}\left(\frac{x-p_0}{\sqrt{p_0(1-p_0)}}\right)$ e $\sqrt{n}\left(\frac{x-p_1}{\sqrt{p_1(1-p_1)}}\right)$ tendem a $+\infty$, e portanto, as probabilidades a serem obtidas correspondem a uma área muito extrema sob a curva normal padrão, a qual não fornece uma aproximação adequada.

Outra alternativa para estimar α_n e β_n é utilizar a equivalência logarítmica. Note que não podemos utilizar a equivalência exata, pois X_1 não tem lei de probabilidade absolutamente contínua. Sob H_0 o evento $[\bar{X}_n \geq x]$, $x > p_0$, é uma situação de Grande Desvio em relação à esperança p_0, pois pela Lei dos Grandes Números $\lim_{n \to \infty} P(\bar{X}_n \geq x) = 0$. Por outro lado, sob H_1 o evento $[\bar{X}_n \leq x]$, $x < p_1$, é uma situação de Grande Desvio em relação à esperança p_1.

Seja $\lambda(x,p)$ denotando a transformada de Cramér para uma distribuição de Bernoulli com parâmetro p desconhecido. Pelo Exemplo 1.1.2, com $w = 0$ e $v = 1$, $\lambda(x,p) = x \ln\left(\frac{x}{p}\right) + (1-x)\ln\left(\frac{1-x}{1-p}\right)$. A equivalência logarítmica nos diz que

$$P(\bar{X}_n \geq x | H_0) \approx e^{-n\lambda(x,p_0)}, \quad x > p_0, \qquad \text{e} \qquad P(\bar{X}_n \leq x - \delta | H_1) \approx e^{-n\lambda(x-\delta,p_1)}, \quad x < p_1 \text{ e } \delta > 0.$$

Quando $\delta \to 0$ temos que $P(\bar{X}_n \leq x - \delta | H_1)$ converge para $P(\bar{X}_n < x | H_1)$, e além disso,

$$\lim_{\delta \to 0} \lambda(x - \delta, p) = \lim_{\delta \to 0} \left[(x - \delta) \ln\left(\frac{x - \delta}{p}\right) + (1 - x + \delta) \ln\left(\frac{1 - x + \delta}{1 - p}\right) \right] = x \ln\left(\frac{x}{p}\right) + (1 - x) \ln\left(\frac{1 - x}{1 - p}\right).$$

Então, $P(\bar{X}_n < x | H_1) \approx e^{-n\lambda(x, p_1)}$, $x < p_1$.

Suponha que se deseje $\lambda(x, p_0) = \lambda(x, p_1)$, ou seja, que a taxa de convergência para zero das probabilidades dos erros tipo I e II sejam iguais. Obtendo x tal que $\lambda(x, p_0) = \lambda(x, p_1)$:

$$x \ln\left(\frac{x}{p_0}\right) + (1 - x) \ln\left(\frac{1 - x}{1 - p_0}\right) = x \ln\left(\frac{x}{p_1}\right) + (1 - x) \ln\left(\frac{1 - x}{1 - p_1}\right)$$

$$x \left(\ln\left(\frac{p_1}{p_0}\right) + \ln\left(\frac{1 - p_0}{1 - p_1}\right) \right) = \ln\left(\frac{1 - p_0}{1 - p_1}\right)$$

$$x(p_0, p_1) = \frac{\ln\left(\frac{1 - p_0}{1 - p_1}\right)}{\ln\left(\frac{p_1}{p_0}\right) + \ln\left(\frac{1 - p_0}{1 - p_1}\right)}.$$

Logo, as probabilidades dos erros tipo I e II são da ordem $e^{-n\lambda(x(p_0, p_1), p_0)}$. Se $\lambda(x(p_0, p_1), p_0)$ for grande então haverá uma chance maior de decidir corretamente entre p_0 e p_1.

Nas duas proposições que seguem mostraremos a importância do Teorema de Cramér-Chernoff para a prova de algumas propriedades adicionais de λ. A prova destas duas proposições é baseada em Azencott **[2]**.

Proposição 1.2.1. *Seja $\{X_n\}_{n \geq 1}$ uma seqüência de v.a's i.i.d. com $\mu(A) = P(X_1 \in A)$, para A boreliano de* $\mathbf{R}$. *Seja $[z, v] \subset [-\infty, +\infty]$ cobertura convexa fechada do suporte de μ. Então:*

(i) $\lambda(x) = +\infty$ para $x \notin [z, v]$, enquanto $\lambda(x)$ é finita e contínua em $x \in (z, v)$;

(ii) Se z é finito, $\lambda(z) = +\infty$ equivale a $P(X_1 \leq z) = 0$ e se v é finito, $\lambda(v) = +\infty$ equivale a $P(X_1 \geq v) = 0$;

(iii) λ é contínua à esquerda em v, se v é finito, e contínua à direita em z, se z é finito.

Prova:

(i) Primeiro observe que $m = EX_1 \in [z, v]$ e $\bar{X}_n \in [z, v]$ com probabilidade 1. Logo, $P(\bar{X}_n \leq x) = 0$ para $\forall n \geq 1$, se $x < z$, e então por (1.2.3), $\lambda(x) = +\infty$. Por outro lado, $P(\bar{X}_n \geq x) = 0$ para $\forall n \geq 1$, se $x > v$, e por (1.2.4), $\lambda(x) = +\infty$. É claro que λ deve ser finita em $x \in (z, v)$, pois $P(\bar{X}_n \leq x) > 0$ e $P(\bar{X}_n \geq x) > 0$.

Pela Proposição 1.1.2 (ii), λ é convexa em $\mathbf{R}$ e, em particular, em $[z, v]$. Segue que uma função convexa em um intervalo (fechado ou aberto) é contínua no interior desse intervalo (Corolário A.2.1 do Apêndice 2). Portanto, λ é contínua em (z, v).

(ii) $P(X_1 \leq z) = 0$ resulta em $\frac{1}{n} \ln P(\bar{X}_n \leq z) = -\infty$, $\forall n \geq 1$, e por (1.2.3), $\lambda(z) = +\infty$. Reciprocamente, se $\lambda(z) = +\infty$ então $P(\bar{X}_n \leq z) = 0$, por (1.2.1). Para v o raciocínio é análogo usando (1.2.2) e (1.2.4).

(iii) Faremos a prova da continuidade à direita em z, pois a continuidade à esquerda em v é provada de forma semelhante. Seja $L = \lim_{x \to z^+} \lambda(x)$. Há dois casos possíveis: $P(X_1 \leq z) = 0$ e $P(X_1 \leq z) > 0$. Quando $P(X_1 \leq z) = 0$, por (ii) tem-se $\lambda(z) = +\infty$. Pela semicontinuidade inferior de λ segue que $\lambda(z) \leq \underline{\lim}_{n \to \infty} \lambda(x_n) = L$, onde $\{x_n\}_{n \geq 1} \in (z, +\infty)$ e $x_n \to z$. Isto implica em $+\infty = \underline{\lim}_{n \to \infty} \lambda(x_n) = L$, e portanto, $\lambda(z) = L$. No caso $P(X_1 \leq z) > 0$, $P(\bar{X}_n \leq z) > 0$ para $\forall n \geq 1$. Assim, $\lambda(z)$ é finita, por (1.2.1). Pela Proposição 1.1.2(iv), temos que λ é não crescente para $z \leq x \leq m$, logo, $\lambda(z) \geq L$. Agora suponha $\lambda(z) > L$ e tome uma seqüência $\{x_n\}_{n \geq 1} \in (z, +\infty)$, $x_n \to z$. Uma vez que $\underline{\lim}_{n \to \infty} \lambda(x_n) = L$, isto leva-nos a uma contradição do fato de λ ser s.c.i. em $\mathbf{R}$ (ver o item (i) da Proposição 1.1.2). Logo, $\lambda(z) = L$, permitindo-nos concluir que λ é contínua à direita em z.

$\square$

Na figura 1.2.3 temos algumas das formas que λ pode assumir, de acordo com a Proposição 1.2.1. Em (a), $P(X = z) = 1$, em (b) $P(X = z) > 0$ e $P(X = v) > 0$. Já em (c), $P(X = z) = P(X = v) = 0$, e no caso (d), não existem $z, v \in \mathbf{R}$ tais que $P(X \leq z) = P(X \geq v) = 0$.

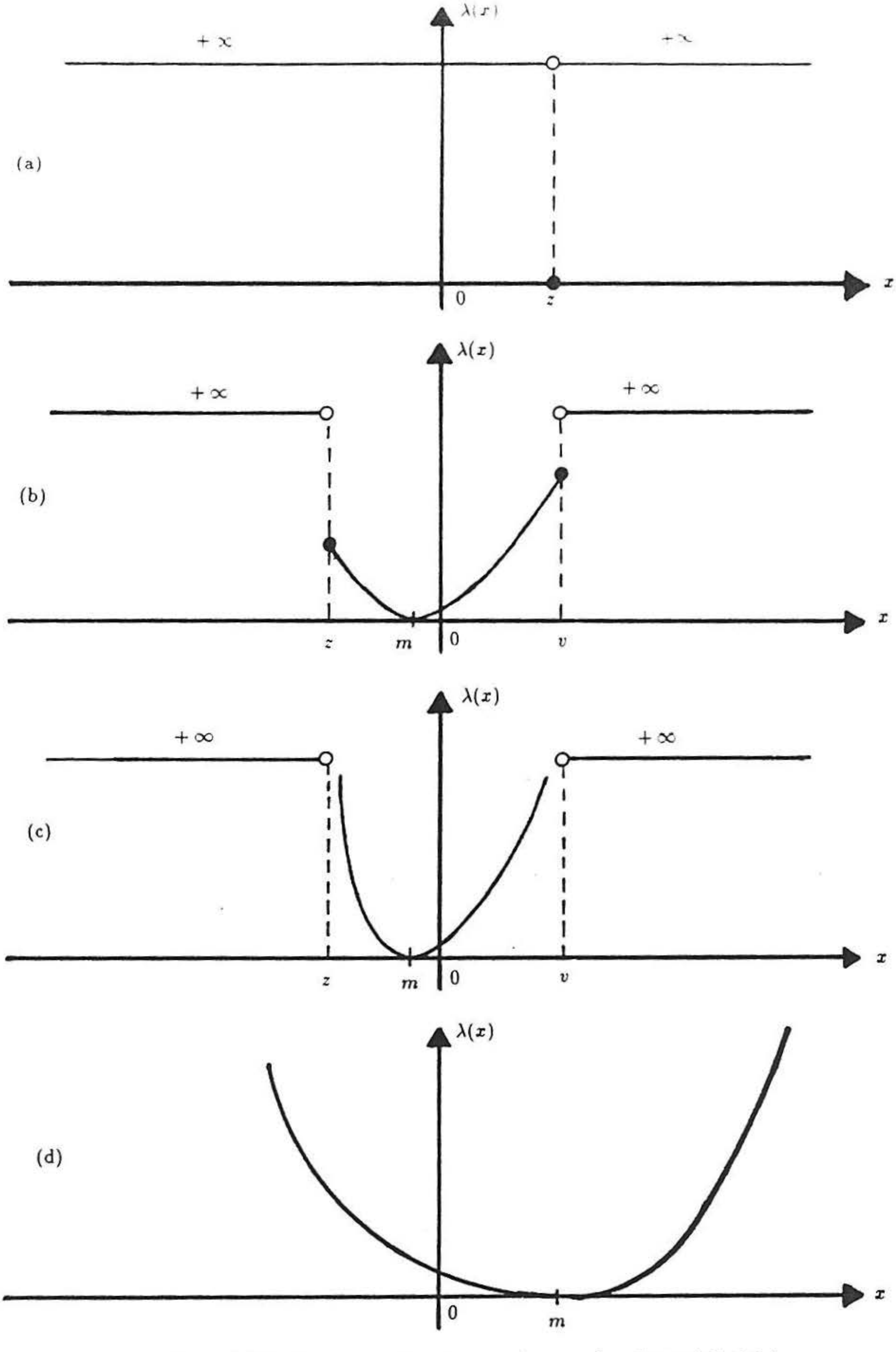

figura 1.2.3 - Formas que λ pode assumir, segundo a Proposição 1.2.1

Exemplo 1.2.2. Com relação ao Exemplo 1.1.1, $\lambda(x) = 0$ se $x = a$ e $\lambda(x) = +\infty$ se $x \neq a$, pois $P(X < a) = 0$ e $P(X > a) = 0$.

Exemplo 1.2.3. No Exemplo 1.1.2, $P(X \leq z) = 1 - p$ e $P(X \geq v) = p$. Logo, pela Proposição 1.2.1, pode-se constatar que λ é contínua e finita em (z, v), contínua à esquerda em z, contínua à direita em v e $\lambda(x) = +\infty$ se $x \notin [z, v]$.

Exemplo 1.2.4. No Exemplo 1.1.3 não existem z, v finitos tal que $P(X \geq v) = 0$ e $P(X \leq z) = 0$. Por isso, pelo item (iii) da Proposição acima, λ não assume $+\infty$ seja qual for $x \in \mathbf{R}$.

Exemplo 1.2.5. X com distribuição Gama (Exemplo 1.1.5) é tal que $z = 0$, $P(X \leq z) = 0$ e então $\lambda(z) = +\infty$. Contudo, não existe v finito tal que $P(X \geq v) = 0$. Logo, λ assume $+\infty$ somente para $x \leq 0$.

Proposição 1.2.2. *Seja μ uma medida de probabilidade em $\mathbf{R}$ tal que $\int_{-\infty}^{+\infty} |x| d\mu(x)$ seja finita. Tem-se:*

(i) $\int_{-\infty}^{+\infty} e^{s\lambda(x)} d\mu(x)$ é finita para todo $s < 1$;

(ii) $\hat{\mu}(t)$ é finita para todo $t > 0$ se e somente se $\lim_{x \to +\infty} \frac{\lambda(x)}{x} = +\infty$ e $\hat{\mu}(t)$ é finita para todo $t < 0$ se e somente se $\lim_{x \to -\infty} \frac{\lambda(x)}{x} = -\infty$.

Prova:

(i) Primeiro consideremos o caso $\mu((-\infty, z)) = \mu((v, +\infty)) = 0$, com z, v finitos e $\mu(\{z\})) > 0$, $\mu(\{v\} > 0$ (ver a figura 1.2.3(b)). Pela Proposição 1.2.1, verifica-se que

$$\int_{-\infty}^{+\infty} e^{s\lambda(x)} d\mu(x) = \int_{z}^{v} e^{s\lambda(x)}\, d\mu(x) < +\infty.$$

No caso de z e v serem finitos, com pelo menos um deles de medida nula, utilizaremos a fórmula de integração por partes (ver Billingsley [4]). Suponhamos que $\mu(\{z\}) > 0$ e $\mu(\{v\}) = 0$. Então

$$\int_{z}^{v} e^{s\lambda(x)}\, d\mu(x) = \int_{z}^{m} e^{s\lambda(x)}\, d\mu(x) + \int_{m}^{v} e^{s\lambda(x)}\, d\mu(x),$$

sendo $m = \int_{-\infty}^{+\infty} x\, d\mu(x)$ finita.

Já sabemos que $\int_{z}^{m} e^{s\lambda(x)}\, d\mu(x) < +\infty$. Para mostrar que a segunda integral é finita, notemos que $\int_{m}^{v} e^{s\lambda(x)}\, d\mu(x) = \int_{m}^{v} e^{s\lambda(x)}\, dF(x)$, onde $F(x) = \mu(-\infty, x]$. Agora defina $G(x) = \mu([x, +\infty))$ e observe que

$F(x) = 1 - G(x) + \mu(\{x\})$. Note que $\mu(\{x\}) > 0$ no máximo em um número contável de pontos. Verifica-se, então, que $\int_m^v e^{s\lambda(x)}\,dF(x) < +\infty$ se e somente se $\int_m^v e^{s\lambda(x)}\,dG(x) < +\infty$.

Pela fórmula de integração por partes,

$$(1.2.7) \qquad \int_m^v e^{s\lambda(x)}\,dG(x) = e^{s\lambda(x)}\mu([x,+\infty))\Big|_m^v - \int_m^v \mu([x,+\infty))s\lambda'(x)e^{s\lambda(x)}\,dx,$$

onde $\lambda'(x)$ é a derivada de $\lambda(x)$. Note que pela convexidade de $\lambda(x)$, $\lambda'(x)$ existe (exceto talvez em um conjunto contável de pontos).

Se $\hat{\mu}(t) = +\infty$, $\forall t > 0$ (figura 1.1.2(b)), temos que $\lambda(x) \equiv 0$ para $x \geq m$ (Proposição 1.1.4(iv)), e trivialmente $\int_m^v e^{s\lambda(x)}\,d\mu(x) < +\infty$. Por outro lado, se $\exists\,\tau > 0$ tal que $\hat{\mu}(\tau) < +\infty$ (figura 1.1.2(c)), então, pela Proposição 1.1.4(v), $\lim_{x\to+\infty}\lambda(x) = +\infty$. Além disso, pelo Teorema de Cramér-Chernoff, $e^{s\lambda(x)}\mu([x,+\infty) \leq e^{(s-1)\lambda(x)}$, $x \geq m$. Destes dois fatos, segue que $\lim_{x\to+\infty} e^{s\lambda(x)}\mu([x,+\infty) = 0$, permitindo-nos concluir que o primeiro termo de (1.2.7) é finito. Pelos mesmos fatos acima referidos,

$$\int_m^v \mu([x,+\infty))s\lambda'(x)e^{s\lambda(x)}\,dx \leq \int_m^v s\lambda'(x)e^{(s-1)\lambda(x)}\,dx < +\infty.$$

Os outros casos: $\mu(\{z\}) = 0$, z ou v infinitos, são analisados de forma semelhante.

(ii) Considere $\hat{\mu}(t)$ finita para $t > 0$. Então $\frac{\lambda(x)}{x} \geq t - \frac{1}{x}\ln\hat{\mu}(t)$, e portanto, $\lim_{x\to+\infty}\frac{\lambda(x)}{x} \geq t$. Como t é arbitrário, $\lim_{x\to+\infty}\frac{\lambda(x)}{x} = +\infty$. Reciprocamente, vamos supor $\lim_{x\to+\infty}\frac{\lambda(x)}{x} = +\infty$. Para $t > 0$ arbitrário e fixo, podemos encontrar um x_t suficientemente grande tal que $\frac{\lambda(x)}{x} \geq 2t$, para todo $x > x_t$. Agora observe que $\hat{\mu}(t) = \int_{-\infty}^{x_t} e^{tx}d\mu(x) + \int_{x_t}^{+\infty} e^{tx}d\mu(x)$. Uma vez que $tx \leq \frac{1}{2}\lambda(x)$,

$$\int_{x_t}^{+\infty} e^{tx}d\mu(x) \leq \int_{x_t}^{+\infty} e^{\frac{1}{2}\lambda(x)}d\mu(x) < +\infty,$$

sendo a última desigualdade justificada por (i). Por outro lado, $\int_{-\infty}^{x_t} e^{tx}d\mu(x) < +\infty$, e então $\hat{\mu}(t) < +\infty$. Na prova com $t < 0$ e $x \to -\infty$ o procedimento é o mesmo.

□

1.3 - Alguns conceitos gerais

Definição 1.3.1. *Uma função $I : \mathbf{R} \longrightarrow [0,+\infty]$ é denominada função taxa se:*

(i) $I \not\equiv +\infty$, ou seja, I não é identicamente $+\infty$;

(ii) I é semicontínua inferiormente;

(iii) para qualquer $L \in [0, +\infty)$, o conjunto $\{x : I(x) \leq L\}$ é compacto.

Definição 1.3.2. *Uma família de medidas de probabilidade* $\{\mu_\epsilon\}_{\epsilon>0}$, *no espaço mensurável* $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, *satisfaz o princípio de Grandes Desvios com taxa* I *se:*

(i) $\overline{\lim}_{\epsilon \downarrow 0} \epsilon \ln \mu_\epsilon(F) \leq -\inf_{x \in F} I(x)$, para todo fechado $F \subset \mathbf{R}$;

(ii) $\underline{\lim}_{\epsilon \downarrow 0} \epsilon \ln \mu_\epsilon(G) \geq -\inf_{x \in G} I(x)$, para todo aberto $G \subset \mathbf{R}$ e $G \neq \emptyset$.

As definições 1.3.1 e 1.3.2 podem ser estendidas para espaços gerais, que sejam completos e separáveis. Na verdade, a reta real é um espaço métrico completo e separável (ver Lima [22]). Mostraremos que a transformada de Cramér satisfaz as duas definições acima, utilizando o roteiro da prova de Stroock [28].

Lema 1.3.1. *A transformada de Cramér é uma função taxa.*

Prova: É preciso verificar se λ satisfaz (i), (ii) e (iii) da Definição 1.3.1.

(i) λ não é identicamente igual a $+\infty$: Por (iii) da Proposição 1.1.2, $\lambda(m) = 0$, sendo $m = \int_{-\infty}^{+\infty} x\, d\mu(x)$.

(ii) λ é semicontínua inferiormente: ver item (i) da Proposição 1.1.2.

(iii) $K_L = \{x : \lambda(x) \leq L\}$ é compacto se $L \in [0, +\infty)$:

Seja $x \in K_L$. Para $t = 1$ temos $x - \ln \hat{\mu}(1) \leq L$ e para $t = -1$, $-x - \ln \hat{\mu}(-1) \leq L$. Segue que

$$K_L \subseteq \{-L - \ln \hat{\mu}(-1), L + \ln \hat{\mu}(1)\},$$

isto é, K_L é limitado.

Seja $[z, v] \subset [-\infty, +\infty]$ cobertura convexa fechada do suporte da lei de probabilidade μ, na qual λ está definida. Seja $[a, b]$ o fecho de K_L. Se $[a, b] \subset [z, v]$ então pela Proposição 1.2.1 (i), $\lambda(a) \leq L$ e $\lambda(b) \leq L$, ou seja, $a, b \in K_L$. Por outro lado, se $[a, b] = [z, v]$, com z e v finitos, temos $\lambda(x) \leq L$ para $\forall\, x \in (z, v)$. Então, segue da Proposição 1.2.1(iii) que $\lambda(a) \leq L$ e $\lambda(b) \leq L$. Logo, $[a, b] = K_L$, e isto significa que K_L é fechado.

Portanto, concluímos que K_L é compacto, pois é limitado e fechado.

□

Teorema 1.3.1. *Seja* $\{X_n\}_{n \geq 1}$ *uma seqüência de v.a's i.i.d. definidas no espaço de probabilidade* $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mu)$ *tal que* $\hat{\mu}(t) < +\infty$, $\forall t \in \mathbf{R}$. *Seja* μ_n *a lei de probabilidade de* $\bar{X}_n$. *A seqüência* $\{\mu_n : n \geq 1\}$ *satisfaz o princípio de Grandes Desvios da Definição 1.3.2 com função taxa* λ *e* $\epsilon = \frac{1}{n}$.

Prova: É preciso verificar se $\{\mu_n : n \geq 1\}$ satisfaz (i) e (ii) da Definição 1.3.2.

(i) Seja F um conjunto fechado de $\mathbf{R}$. Se $F = \emptyset$ então $\mu_n(F) = 0$ e por sua vez $\ln \mu_n(F) = -\infty$, $\forall n \geq 1$, logo, não há nada para provar. Se $m = EX_1 \in F$, pela Lei Fraca dos Grandes Números $\mu_n(F)$ converge a 1 quando $n \to \infty$, e então $\overline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(F) = 0 = \lambda(m)$, isto é, também não há nada para provar. Assim, assumiremos $F \neq \emptyset$ e $m \notin F$.

Primeiro consideraremos $F \subset (m, +\infty)$. Definindo $a \equiv \inf\{y : y \in F\}$, tem-se

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \ln \mu_n(F) \leq \overline{\lim_{n\to\infty}} \frac{1}{n} \ln \mu_n([a, +\infty)) \leq -\lambda(a),$$

onde a última desigualdade é justificada por (1.2.2), pois $a > m$ e $a \in F$ (observe que F é fechado e limitado inferiormente por m). Pela Proposição 1.1.2 (iv),

$$\lambda(a) = \inf_{y\in[a,+\infty)} \lambda(y) = \inf_{y\in F} \lambda(y),$$

uma vez que $a > m$ e $a \in F$. Portanto,

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \ln \mu_n(F) \leq -\inf_{y\in F} \lambda(y).$$

Para $F \subset (-\infty, m)$, definimos $b \equiv \sup\{y : y \in F\}$, e de maneira análoga mostra-se que $\overline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(F) \leq -\inf_{y\in F} \lambda(y)$, utilizando (1.2.1) e a Proposição 1.1.2 (iv).

Agora considere F um conjunto fechado qualquer tal que $m \notin F$, $F_1 = (-\infty, m) \cap F \neq \emptyset$ e $F_2 = (m, +\infty) \cap F \neq \emptyset$. Seja $a \equiv \inf\{y : y \in F_2\}$ e $b \equiv \sup\{y : y \in F_1\}$. Novamente pela Proposição 1.1.2 (iv), $\lambda(b) = \inf_{y\in(-\infty,b]} \lambda(y)$ e $\lambda(a) = \inf_{y\in[a,+\infty)} \lambda(y)$, visto que $b < m < a$. Segue que $\mu_n((-\infty, b]) \leq e^{-n\lambda(b)}$, por (1.2.1), e $\mu_n([a, +\infty)) \leq e^{-n\lambda(a)}$, por (1.2.2). Como $F \subseteq (-\infty, b] \cup [a, +\infty)$, então

$$\mu_n(F) \leq \mu_n((-\infty, b]) + \mu_n([a, +\infty)) \leq e^{-n\lambda(b)} + e^{-n\lambda(a)} \leq 2e^{-n \min\{\lambda(a), \lambda(b)\}},$$

e isto resulta em $\overline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(F) \leq -\min\{\lambda(a), \lambda(b)\} = -\inf_{y\in F} \lambda(y)$.

(ii) Mostraremos que para $\forall x \in G$, $\underline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(G) \geq -\lambda(x)$. No caso $\lambda(x) = +\infty$, $x \in G$, temos $\underline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(G) \geq -\infty$, ou seja, não há nada para provar. Portanto, assumiremos que $x \in G$ é tal que $\lambda(x) < +\infty$.

Suponha que não existe $t \in \mathbf{R}$ tal que $\lambda(x) = tx - \ln \hat{\mu}(t)$. Então $x \neq m$, uma vez que $\lambda(m) = 0 = tm - \ln \hat{\mu}(t)$ quando $t = 0$. Suponha $x > m$. Por (1.1.2), $tx - \ln \hat{\mu}(t) \leq t(x - m) \leq 0$, se $t \leq 0$, e então $\lambda(x) = \sup_{t>0}\{tx - \ln \hat{\mu}(t)\}$. Logo, existe uma seqüência $\{t_n\}_{n\geq 1} \in (0, +\infty)$ tendendo a ∞ tal que $t_n x - \ln \hat{\mu}(t_n)$ converge para $\lambda(x)$ quando $n \to \infty$, e conseqüentemente

$$\lim_{n\to\infty} e^{-(t_n x - \ln \hat{\mu}(t_n))} = e^{-\lambda(x)} < +\infty.$$

Por outro lado,

$$e^{-(t_n x - \ln \hat{\mu}(t_n))} = \int_{-\infty}^{+\infty} e^{t_n(y-x)} d\mu(y),$$

e portanto,

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} e^{t_n(y-x)} d\mu(y) = e^{-\lambda(x)}. \tag{1.3.1}$$

A seqüência $\left\{e^{t_n(y-x)}\right\}_{n\geq 1}$, com $y \in (-\infty, x)$, converge uniformemente para a função $f(y) = 0$ quando $n \to \infty$, logo,

$$\lim_{n\to\infty} \int_{(-\infty,x)} e^{t_n(y-x)} d\mu(y) = \int_{(-\infty,x)} \lim_{n\to\infty} e^{t_n(y-x)} d\mu(y) = 0.$$

Por outro lado, para que

$$\lim_{n\to\infty} \int_{(x,+\infty)} e^{t_n(y-x)} d\mu(y) < +\infty$$

é preciso que $\mu((x, +\infty)) = 0$. Para ver isto, suponha que $\mu(\{x\}) = \mu(\{z\}) = \frac{1}{2}$, $z > x$. Logo,

$$\int_{[x,+\infty)} e^{t_n(y-x)} d\mu(y) = \frac{1}{2} + \frac{1}{2} e^{t_n(z-x)},$$

e então, $\lim_{n\to\infty} \left(\frac{1}{2} + \frac{1}{2} e^{t_n(z-x)}\right) = +\infty$. Portanto, $\mu((x, +\infty)) = 0$ e

$$\mu(\{x\}) = \lim_{n\to\infty} \int_{[x,+\infty)} e^{t_n(y-x)} d\mu(y).$$

Mas por (1.3.1) temos que $e^{-\lambda(x)} = \mu(\{x\})$, resultando em

$$\mu_n(G) \geq \mu_n(\{x\}) \geq (\mu(\{x\}))^n = e^{-n\lambda(x)},$$

e daí $\underline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(G) \geq -\lambda(x)$. Similarmente, prova-se para $x < m$.

Consideremos agora que existe $\tau \in \mathbf{R}$ tal que $\lambda(x) = \tau x - \ln \hat{\mu}(\tau)$, $x \in G$. Assumiremos $x \geq m$, o qual implica em $\tau \geq 0$, pois $\lambda(x) \geq 0$ e $\tau x - \ln \mu(\tau) \leq \tau(x - m)$, por (1.1.2). Pelo fato de G ser aberto, existe $\delta > 0$ tal que $(x - \delta, x + \delta) \subset G$. Notemos que

$$\left[\frac{X_1 + \cdots + X_n}{n} \in (x - \delta, x + \delta)\right] \quad \text{e} \quad \left[\left|\frac{\sum_{i=1}^{n}(X_i - x)}{n}\right| < \delta\right]$$

são equivalentes. Defina $S = \left\{y_i,\ i = 1, \cdots, n : \left|\frac{\sum_{i=1}^{n}(y_i - x)}{n}\right| < \delta\right\}$. Daí

$$\mu_n(G) \geq \mu_n((x - \delta, x + \delta)) = \int_S d\mu(y_1) \cdots d\mu(y_n).$$

Mas de $\frac{\sum_{i=1}^{n}(y_i - x)}{n} < \delta$ vem que $e^{\tau \sum_{i=1}^{n} y_i - n\tau(x+\delta)} < 1$, e assim,

$$\int_S d\mu(y_1) \cdots d\mu(y_n) > e^{-n\tau(x+\delta)} \int_S e^{\tau y_1} \cdots e^{\tau y_n} d\mu(y_1) \cdots d\mu(y_n) = e^{-n\tau(x+\delta)}(\hat{\mu}(\tau))^n \int_S d\nu(y_1) \cdots d\nu(y_n),$$

onde $d\nu(y) = \frac{e^{\tau y}}{\hat{\mu}(\tau)} d\mu(y)$. Pela hipótese $\hat{\mu}(t) < +\infty$, $\forall t \in \mathbf{R}$, e pela Proposição A.1.1(v) do Apêndice 1, tem-se que $\hat{\mu}(t)$ é de classe $C^\infty(\mathbf{R})$. Sendo assim,

$$\frac{d}{dt} \int_{-\infty}^{+\infty} e^{ty} d\mu(y) = \int_{-\infty}^{+\infty} \left(\frac{d}{dt} e^{ty}\right) d\mu(y) = \int_{-\infty}^{+\infty} y\, e^{ty} d\mu(y),$$

ou seja, $\hat{\mu}'(t) = \int_{-\infty}^{+\infty} y\, e^{ty} d\mu(y)$. Logo,

$$\int_{-\infty}^{+\infty} y d\nu(y) = \int_{-\infty}^{+\infty} y \frac{e^{\tau y}}{\hat{\mu}(\tau)} d\mu(y) = \frac{\hat{\mu}'(\tau)}{\hat{\mu}(\tau)} = x,$$

onde a última igualdade é justificada pela equação $\frac{d}{d\tau}(\tau x - \ln \hat{\mu}(\tau)) = \frac{d}{d\tau}\lambda(x)$. Logo, concluimos que x é a esperança de uma v.a. cuja lei de probabilidade é ν. Agora assuma que $\{Y_n\}_{n \geq 1}$ é uma seqüência de v.a's i.i.d. com lei de probabilidade ν. Seja $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Então $P(\bar{Y}_n \in (x - \delta, x + \delta)) = \int_S d\nu(y_1) \cdots d\nu(y_n)$, e pela Lei dos Grandes Números, $\lim_{n \to \infty} P(\bar{Y}_n \in (x - \delta, x + \delta)) = 1$, uma vez que $EY_1 = x$. Assim, $\lim_{n \to \infty} \int_S d\nu(y_1) \cdots d\nu(y_n) = 1$. Das considerações acima segue que

$$\mu_n(G) > e^{-n\tau(x+\delta)}(\hat{\mu}(\tau))^n \int_S d\nu(y_1) \cdots d\nu(y_n),$$

resultando em

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(G) \geq -(\tau x - \ln \hat{\mu}(\tau)) - \tau\delta = -\lambda(x) - \tau\delta.$$

Como $\delta > 0$ é arbitrário, vem que $\underline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(G) \geq -\lambda(x)$. No caso $x < m$ o procedimento é análogo.

□

Observação 1.3.1. Dembo & Zeitouni [9] (Teorema 2.2.3), mostram que o princípio de Grandes Desvios da Definição 1.3.2 é válido para qualquer medida de probabilidade μ, ou seja, não são necessárias as restrições $\int_{-\infty}^{+\infty} |x|\, d\mu(x) < +\infty$ ou $\int_{-\infty}^{+\infty} e^{tx}\, d\mu(x) < +\infty$.

Observação 1.3.2. Uma vez que A° é aberto e $\bar{A}$ é fechado, com $A \subseteq \mathbf{R}$, pelo teorema acima podemos concluir que

$$-\inf_{x\in A^\circ} \lambda(x) \leq \underline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(A^\circ) \leq \underline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(A) \leq \overline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(A) \leq \overline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(\bar{A}) \leq -\inf_{x\in \bar{A}} \lambda(x),$$

tendo-se igualdade se $\inf_{x\in A^\circ} \lambda(x) = \inf_{x\in \bar{A}} \lambda(x)$.

Observação 1.3.3. Pela Observação 1.3.1, pode-se obter (i) e (ii) do Teorema 1.2.1. Para mostrar isto, considere $x \leq m$. Então

$$\overline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n((-\infty, x]) \leq -\inf_{x\in(-\infty,x]} \lambda(x) = -\lambda(x).$$

Tomando $\delta > 0$,

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n((-\infty, x+\delta]) \geq \underline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n((x-\delta, x+\delta)) \geq -\inf_{y\in(x-\delta,x+\delta)} \lambda(y).$$

Como δ é arbitrário, $\underline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n((-\infty, x]) \geq -\lambda(x)$. De forma semelhante, mostra-se para $x \geq m$.

CAPÍTULO 2

GRANDES DESVIOS PARA A MÉDIA AMOSTRAL DE VARIÁVEIS ALEATÓRIAS I.I.D. EM $\mathbb{R}^d$

2.1 - Estabelecendo um princípio de Grandes Desvios para a média amostral

O princípio de Grandes Desvios de Cramér-Chernoff possui uma extensão para a média amostral de v.a's i.i.d. assumindo valores em $\mathbf{R}^d$.

Inicialmente iremos enunciar e demonstrar dois lemas, os quais serão necessários para definir a transformada de Cramér em $\mathbf{R}^d$ e para provar a Proposição 2.1.2 adiante. A prova destes dois lemas é feita de acordo com Stroock **[28]**.

Lema 2.1.1. *Seja* $f : \mathbf{Z}^+ \longrightarrow [-\infty, 0]$ *superaditiva, ou seja,* $f(m+n) \geq f(m) + f(n)$, *para todo* $m, n \in \mathbf{Z}^+$. *Se para algum* $n_0 \in \mathbf{Z}^+$, $f(n) > -\infty$ *para todo* $n \geq n_0$, *então* $\lim_{n\to\infty} \frac{f(n)}{n} = \sup_{m\geq n_0} \left\{\frac{f(m)}{m}\right\}$ *é finito.*

Prova: Seja $n_0 \in \mathbf{Z}^+$ tal que $f(n) > -\infty, \forall n \geq n_0$. Defina $\alpha \equiv \sup\left\{\frac{f(n)}{n} : n \geq n_0\right\}$. Uma vez que $-\infty < f(n) \leq 0$ para $n \geq n_0$ então $\frac{f(n)}{n}$ é finito, e por isso α é finito. Dado $m \geq n_0$, defina $q_n \equiv \left[\frac{n}{m}\right]$ como sendo o maior inteiro de $\frac{n}{m}$ e $r_n \equiv n - \left[\frac{n}{m}\right] m$, onde $n \geq m$. Para $n \geq 2m$, utilizando a propriedade de superaditividade:

$$
\begin{aligned}
\frac{f(n)}{n} &= \frac{f(q_n m + r_n)}{n} = \frac{f((q_n-1)m + m + r_n)}{n} \geq \frac{f((q_n-1)m) + f(m+r_n)}{n} \\
&\geq \frac{(q_n-1)m}{n}\frac{f(m)}{m} + \frac{f(m+r_n)}{n} \geq \frac{(q_n-1)m}{n}\frac{f(m)}{m} + \min_{m\leq k\leq 2m-1}\frac{f(k)}{n}.
\end{aligned}
$$

Note que tomamos $n \geq 2m$ para que $(q_n - 1) \geq 1$, pois f está definida em $\mathbf{Z}^+$. Também observe que $0 \leq r_n \leq m-1$, pois $0 \leq r_n < m$ e r_n é inteiro.

Uma vez que $\underline{\lim}_{n\to\infty} \min_{m\leq k\leq 2m-1} \frac{f(k)}{n} = 0$ e $\underline{\lim}_{n\to\infty} \frac{(q_n-1)m}{n} = 1$, tem-se $\underline{\lim}_{n\to\infty} \frac{f(n)}{n} \geq \frac{f(m)}{m}$ para $m \geq n_0$ e daí vem que $\underline{\lim}_{n\to\infty} \frac{f(n)}{n} \geq \sup_{m\geq n_0}\left\{\frac{f(m)}{m}\right\} = \alpha$. Por outro lado, como $\alpha = \sup_{m\geq n_0}\left\{\frac{f(m)}{m}\right\} \geq \inf_{n\in\mathbf{N}^*} \sup_{k\geq n}\left\{\frac{f(k)}{k}\right\} = \overline{\lim}_{n\to\infty}\frac{f(n)}{n}$, então $\lim_{n\to\infty}\frac{f(n)}{n} = \alpha$. Logo, concluimos que $\lim_{n\to\infty}\frac{f(n)}{n}$ existe, é finito e igual a $\sup_{m\geq n_0}\left\{\frac{f(m)}{m}\right\}$.

$\square$

Lema 2.1.2. *Seja $\{X_n\}_{n\geq 1}$ uma seqüência de v.a's i.i.d. definidas em $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d), \mu)$, sendo $\mathcal{B}(\mathbf{R}^d)$ a $\sigma-$álgebra de Borel em $\mathbf{R}^d$. Defina $f(n) \equiv \ln P(\bar{X}_n \in C)$, onde C é conjunto convexo. Então f é superaditiva e se $f(n) \equiv -\infty$ ou $f(n) > -\infty$ para todo $n \geq n_0$, $n_0 \in \mathbf{Z}^+$, tem-se que $\lim_{n\to\infty} \frac{f(n)}{n}$ existe em $[-\infty, 0]$. Além disso, $\lim_{n\to\infty} \frac{f(n)}{n}$ existe para todo aberto e convexo C.*

Prova: Considere C convexo. Então para $m, n \in \mathbf{Z}^+$

$$P(\bar{X}_n \in C)P(\bar{X}_m \in C) = P\left(\frac{X_1 + X_2 + \cdots + X_n}{n} \in C\right) P\left(\frac{X_{n+1} + \cdots + X_{n+m}}{m} \in C\right)$$
$$= P\left(\frac{X_1 + X_2 + \cdots + X_n}{n} \in C,\ \frac{X_{n+1} + \cdots + X_{n+m}}{m} \in C\right),$$

onde a última igualdade justifica-se pela independência das v.a's. Vamos definir $Z \equiv \frac{X_1+X_2+\cdots+X_n}{n}$ e $W \equiv \frac{X_{n+1}+\cdots+X_{n+m}}{m}$. Disso decorre que $\bar{X}_{n+m} = \frac{n}{n+m}Z + \frac{m}{n+m}W$. Por C ser convexo e $\bar{X}_{n+m}$ ser uma combinação linear convexa de Z e W, tem-se que $[Z \in C, W \in C]$ implica em $[\bar{X}_{n+m} \in C]$. Portanto, $P(Z \in C, W \in C) \leq P(\bar{X}_{n+m} \in C)$. Daí vem que $P(\bar{X}_n \in C)P(\bar{X}_m \in C) \leq P(\bar{X}_{n+m} \in C)$ e aplicando logaritmo obtemos $f(n) + f(m) \leq f(n+m)$, ou seja, f é superaditiva.

Agora suponha que $\exists\, n_0 \in \mathbf{Z}^+$ tal que $f(n) > -\infty, \forall n \geq n_0$. Então, pelo Lema 2.1.1, $\lim_{n\to\infty} \frac{f(n)}{n} = \sup_{m\geq n_0}\left\{\frac{f(m)}{m}\right\} > -\infty$. Por outro lado se $P(\bar{X}_n \in C) = 0, \forall n \geq 1$, então $f(n) \equiv -\infty$, e portanto, $\lim_{n\to\infty} \frac{f(n)}{n} = -\infty$. Em ambos os casos diremos que $\lim_{n\to\infty} \frac{f(n)}{n}$ existe, uma vez que $f(n)$ pode assumir $-\infty$.

Resta ainda provar que se C é aberto convexo então $f(n) \equiv -\infty$ ou $f(n) > -\infty$ para n suficientemente grande. Seja K cobertura convexa fechada do suporte de μ. Se $C \cap K = \emptyset$ então $P(\bar{X}_n \in C) = 0$ para $\forall n \geq 1$, e portanto, $\lim_{n\to\infty} \frac{f(n)}{n} = -\infty$. No caso em que $P(\bar{X}_m \in C) > 0$, $m \in \mathbf{Z}^+$, prova-se (ver Deuschel & Stroock [10], Lema 3.1.2) que $\exists\, N \in \mathbf{N}^*$ tal que para $n \geq N$, $f(n) > -\infty$.

□

Definição 2.1.1. *Seja $\{X_n\}_{n\geq 1}$ uma seqüência de v.a's i.i.d. definidas no espaço de probabilidade $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d), \mu)$. Seja*

$$l(C) \equiv \lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in C), \quad \text{para } C \text{ convexo.} \tag{2.1.1}$$

Define-se $\lambda : \mathbf{R}^d \longrightarrow [0, +\infty]$ como

(2.1.2) $\lambda(x) = -\inf\{l(C) : C \text{ aberto convexo}, x \in C\} = \sup\{-l(C) : C \text{ aberto convexo}, x \in C\}.$

Definição 2.1.2. *Seja $A \in \mathcal{B}(\mathbf{R}^d)$. Definimos*

$$\underline{l}(A) \equiv \varliminf_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in A) \qquad \text{e} \qquad \bar{l}(A) \equiv \varlimsup_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in A).$$

Quando $\underline{l}(A) = \bar{l}(A)$, denotamos por $l(A)$ o valor comum, conforme (2.1.1).

Lema 2.1.3. *A função λ, definida em (2.1.2) é:*

(i) semicontínua inferiormente;

(ii) convexa.

Prova: A prova é extraída de Vares **[30]**.

(i) Provaremos que $\lambda(x) \leq \varliminf_{y\to x} \lambda(y)$ para $\forall x \in \mathbf{R}^d$. Seja $x \in \mathbf{R}^d$. Então, por (2.1.2), tem-se que para $\forall \epsilon > 0, \exists C$ aberto e convexo, com $x \in C$, tal que $-l(C) > \lambda(x) - \epsilon$. Por outro lado, se $y \in C$ então -$l(C) \leq \lambda(y)$. Assim, $\lambda(x) - \epsilon < -l(C) \leq \lambda(y)$. Logo, $\lambda(x) - \epsilon < \varliminf_{y\to x} \lambda(y)$, e como ϵ é arbitrário vem que $\lambda(x) \leq \varliminf_{y\to x} \lambda(y)$, isto é, λ é semicontínua inferiormente em x.

(ii) É preciso provar que $\lambda(\alpha x + (1-\alpha)y) \leq \alpha\lambda(x) + (1-\alpha)\lambda(y)$ para $\forall x, y \in \mathbf{R}^d$ e todo $\alpha \in [0,1]$. Utilizando o Teorema A.2.4 do Apêndice 2, basta provar que $\lambda\left(\frac{x+y}{2}\right) \leq \frac{1}{2}(\lambda(x) + \lambda(y))$, $\forall x, y \in \mathbf{R}^d$. Sejam $x, y \in \mathbf{R}^d$ e C um aberto convexo tal que $\frac{x+y}{2} \in C$. Afirmamos que existe C_1, C_2, abertos convexos com $x \in C_1$ e $y \in C_2$, tal que $\frac{1}{2}(C_1 + C_2) = \left\{\frac{z+w}{2} : z \in C_1, w \in C_2\right\} \subseteq C$. Para verificar isto, tome $C_1 = B_\delta(x)$ e $C_2 = B_\delta(y)$, com $\delta > 0$, isto é, assuma que C_1 e C_2 são bolas abertas em $\mathbf{R}^d$. Agora tome $z \in C_1$ e $w \in C_2$. Então $|z - x| < \delta$ e $|w - y| < \delta$, e portanto,

$$\left|\frac{z+w}{2} - \frac{x+y}{2}\right| = \left|\frac{z-x}{2} - \frac{w-y}{2}\right| \leq \frac{1}{2}(|z-x| + |w-y|) < \delta.$$

Logo, $\frac{z+w}{2} \in B_\delta\left(\frac{x+y}{2}\right)$ e conseqüentemente $\frac{z+w}{2} \in C$, para $\delta > 0$ suficientemente pequeno. Assim, $z \in C_1$ e $w \in C_2$ implicam em $\frac{z+w}{2} \in C$, ou seja, $\frac{1}{2}(C_1 + C_2) \subseteq C$. Utilizando este fato,

$$\left[\frac{X_1 + \cdots + X_n}{n} \in C_1\right] \cap \left[\frac{X_{n+1} + \cdots + X_{2n}}{n} \in C_2\right] \subseteq \left[\frac{X_1 + \cdots + X_{2n}}{2n} \in C\right],$$

resultando em $P(\bar{X}_n \in C_1)P(\bar{X}_n \in C_2) \leq P(\bar{X}_{2n} \in C)$, e assim,

$$\frac{1}{2}\left(\frac{1}{n}\ln P(\bar{X}_n \in C_1) + \frac{1}{n}\ln P(\bar{X}_n \in C_2)\right) \leq \frac{1}{2n}\ln P(\bar{X}_{2n} \in C), \quad \text{para} \, \forall n \geq 1.$$

Disso resulta em $\frac{1}{2}\left(-l(C_1) - l(C_2)\right) \geq -l(C)$ e conseqüentemente $\frac{1}{2}\left(\lambda(x) + \lambda(y)\right) \geq -l(C)$, para todo aberto convexo C com $\frac{x+y}{2} \in C$. Logo, $\frac{1}{2}\left(\lambda(x) + \lambda(y)\right) \geq \lambda\left(\frac{x+y}{2}\right)$.

□

Proposição 2.1.1. *Para $A, B \in \mathcal{B}(\mathbf{R}^d)$, tem-se:*

$$\max\{\underline{l}(A), \underline{l}(B)\} \leq \underline{l}(A \cup B) \leq \bar{l}(A \cup B) \leq \max\{\bar{l}(A), \bar{l}(B)\}. \tag{2.1.3}$$

Prova: A prova é feita de acordo com Azencott [2]. Claramente $\underline{l}(A) \leq \underline{l}(A \cup B)$ e $\underline{l}(B) \leq \underline{l}(A \cup B)$, e disto decorre a primeira desigualdade. Para provar a terceira desigualdade de (2.1.3), tome $c > \max\{\bar{l}(A), \bar{l}(B)\}$. Como $c > \bar{l}(A) = \inf_{n \in \mathbf{N}} \sup_{k \geq n} \frac{1}{k}\ln P(\bar{X}_k \in A)$, então $\exists n_0 \in \mathbf{N}$ tal que $c > \sup_{k \geq n_0} \frac{1}{k}\ln P(\bar{X}_k \in A)$ e disso decorre que $c > \frac{1}{k}\ln P(\bar{X}_k \in A)$, para $k \geq n_0$. Daí, $e^{ck} > P(\bar{X}_k \in A)$, $k \geq n_0$. Analogamente, $\exists m_0 \in \mathbf{N}^*$ tal que $e^{ck} > P(\bar{X}_k \in B)$ para todo $k \geq m_0$. Segue que para k suficientemente grande

$$P(\bar{X}_k \in A \cup B) \leq P(\bar{X}_k \in A) + P(\bar{X}_k \in B) < e^{ck} + e^{ck},$$

e portanto, $\bar{l}(A \cup B) \leq c$. Agora suponha que $\max\{\bar{l}(A), \bar{l}(B)\} < \bar{l}(A \cup B)$. Sendo c arbitrário, podemos tomá-lo de tal forma que $\max\{\bar{l}(A), \bar{l}(B)\} < c < \bar{l}(A \cup B)$. Mas isto contradiz a conclusão anterior de que $\bar{l}(A \cup B) \leq c$. Logo, $\bar{l}(A \cup B) \leq \max\{\bar{l}(A), \bar{l}(B)\}$.

□

Proposição 2.1.2. *Seja $\bar{A}$ o fecho de A e A° o interior de A, com $A \in \mathcal{B}(\mathbf{R}^d)$. Tem-se:*

(i) $-\inf_{x \in A^\circ} \lambda(x) \leq \underline{l}(A)$;

(ii) se $\bar{A}$ for compacto, $\bar{l}(A) \leq -\inf_{x \in \bar{A}} \lambda(x)$;

(iii) se $\bar{A}$ for compacto e $\inf_{x \in A^\circ} \lambda(x) = \inf_{x \in \bar{A}} \lambda(x)$ então $l(A)$ existe e $l(A) = -\inf_{x \in A} \lambda(x)$;

(iv) se A é uma união finita de abertos convexos então $l(A)$ existe e $l(A) = -\inf_{x \in A} \lambda(x)$.

Prova: A prova é extraída de Vares [30].

(i) Inicialmente suponha A aberto ($A = A^\circ$). Tome $x \in A$. Como A é aberto, $\exists \delta > 0$ tal que $B_\delta(x) = \{y : |y - x| < \delta\} \subseteq A$. Por outro lado, sendo $B_\delta(x)$ aberto convexo temos pelo Lema 2.1.2 que $\underline{l}(B_\delta(x)) = \overline{l}(B_\delta(x))$. Como $B_\delta(x) \subseteq A$ implica em $\underline{l}(B_\delta(x)) \leq \underline{l}(A)$ vem que

$$-\lambda(x) = \inf\{l(K) : K \text{ aberto convexo}, x \in K\} \leq l(B_\delta(x)) = \underline{l}(B_\delta(x)) \leq \underline{l}(A),$$

ou seja, $-\lambda(x) \leq \underline{l}(A)$ para $\forall x \in A$. Logo $\sup_{x\in A}\{-\lambda(x)\} \leq \underline{l}(A)$, o qual é equivalente a $-\inf_{x\in A}\lambda(x) \leq \underline{l}(A)$. No caso de A não ser aberto, usamos o fato de A° ser aberto e $A^\circ \subseteq A$. Assim, $-\inf_{x\in A^\circ}\lambda(x) \leq \underline{l}(A^\circ)$ e conseqüentemente $-\inf_{x\in A^\circ}\lambda(x) \leq \underline{l}(A)$.

(ii) Suponha A compacto ($A = \bar{A}$). Seja

$$a > -\inf_{x\in A}\lambda(x) = \sup_{x\in A}\inf\{l(C) : C \text{ aberto convexo}, \ x \in C\}.$$

Então, para cada $x \in A$, $\exists C_x$ aberto convexo, contendo x, tal que $a > l(C_x)$ e também $A \subseteq \bigcup_{x\in A} C_x$. Uma vez que A é compacto, pode-se obter uma subcobertura finita para A, isto é, $\exists\, m \geq 1$ e $x_1, \cdots, x_m \in A$ tal que $A \subseteq \bigcup_{i=1}^m C_{x_i}$. Sendo C_{x_i} aberto convexo, pelo Lema 2.1.2, $\underline{l}(C_{x_i}) = l(C_{x_i})$, e assim

$$\overline{l}(A) \leq \overline{l}\left(\bigcup_{i=1}^m C_{x_i}\right) \leq \max_{i=1,\cdots,m} l(C_{x_i}) < a,$$

onde a segunda desigualdade justifica-se por (2.1.3).

Suponha que $-\inf_{x\in A}\lambda(x) < \max_{i=1,\cdots,m} l(C_{x_i})$. Então, como a é arbitrário podemos tomá-lo de tal forma que $-\inf_{x\in A}\lambda(x) < a \leq \max_{i=1,\cdots,m} l(C_{x_i})$. Mas isto contradiz o fato de $\max_{i=1,\cdots,m} l(C_{x_i}) < a$. Logo,

$$\overline{l}(A) \leq \overline{l}\left(\bigcup_{i=1}^m C_{x_i}\right) \leq \max_{i=1,\cdots,m} l(C_{x_i}) \leq -\inf_{x\in A}\lambda(x),$$

isto é, $\overline{l}(A) \leq -\inf_{x\in A}\lambda(x)$.

No caso de A não ser compacto utilizamos o fato $A \subset \bar{A}$, o qual implica em $\overline{l}(A) \leq \overline{l}(\bar{A})$, completando a prova.

(iii) De (i) e (ii), $-\inf_{x\in A^\circ}\lambda(x) \leq \underline{l}(A) \leq \overline{l}(A) \leq -\inf_{x\in \bar{A}}\lambda(x)$. Logo, $\underline{l}(A) = \overline{l}(A) = l(A) = -\inf_{x\in A}\lambda(x)$.

(iv) Suponha $A = \bigcup_{i=1}^m C_i$, onde $C_1, \cdots, C_m$ são abertos convexos ($m \geq 1$). Pelo Lema 2.1.2 $\underline{l}(C_i) = \overline{l}(C_i) = l(C_i)$. Assim, pela Proposição 2.1.1,

$$\max_{i=1,\cdots,m} l(C_i) \leq \underline{l}\left(\bigcup_{i=1}^m C_i\right) \leq \overline{l}\left(\bigcup_{i=1}^m C_i\right) \leq \max_{i=1,\cdots,m} l(C_i),$$

e portanto, $l(A) = \max_{i=1,\cdots,m} l(C_i)$.

Vamos definir $\Lambda(A) \equiv \inf_{x\in A}\lambda(x)$ e $\Lambda(C_i) \equiv \inf_{x\in C_i}\lambda(x)$. Observemos que $\Lambda(A) = \min_{i=1,\cdots,m}\{\Lambda(C_i)\}$ e equivalentemente $-\Lambda(A) = \max_{i=1,\cdots,m}\{-\Lambda(C_i)\}$. Quer-se provar que

$$\max_{i=1,\cdots,m}\{-\Lambda(C_i)\} = \max_{i=1,\cdots,m}\{l(C_i)\} = l(A).$$

Para isto, basta verificar que $l(C_i) = -\Lambda(C_i)$, onde $i = 1,\cdots,m$. No caso $l(C_i) = -\infty$, a igualdade é facilmente verificada, pois $l(C_i) = -\infty$ implica em $\lambda(x) = +\infty$, $\forall\, x \in C_i$, e conseqüentemente $-\inf_{x\in C_i}\lambda(x) = -\infty$. Agora suponha $l(C_i) > -\infty$. De (i) temos $-\Lambda(C_i) \le l(C_i)$ para $i = 1,\cdots,m$. Falta-nos ainda provar $l(C_i) \le -\Lambda(C_i)$. Pelo Lema 2.1.2, existe $\lim_{n\to\infty}\frac{1}{n}\ln\mu_n(C_i)$, sendo $\mu_n(C_i) = P(\bar{X}_n \in C_i)$. Isto quer dizer que dado $\epsilon > 0$, $\exists\, n_\epsilon \in \mathbf{N}^*$ tal que se $N \ge n_\epsilon$ então $\left|\frac{1}{N}\ln\mu_N(C_i) - l(C_i)\right| < \epsilon$. Assim,

$$-\epsilon + l(C_i) < \frac{1}{n_\epsilon}\ln\mu_{n_\epsilon}(C_i). \tag{2.1.4}$$

Sendo C_i aberto convexo, existe um conjunto K_ϵ, compacto e convexo, tal que $K_\epsilon \subset C_i$ e $\mu_{n_\epsilon}(K_\epsilon) \ge (1-\epsilon)\mu_{n_\epsilon}(C_i)$ para $\epsilon \in (0,1)$ (ver Deuschel & Stroock **[10]**, Lema 3.1.1). Deste fato e por (2.1.4),

$$\ln\mu_{n_\epsilon}(K_\epsilon) \ge \ln\mu_{n_\epsilon}(C_i) + \ln(1-\epsilon) > n_\epsilon(l(C_i)-\epsilon) + \ln(1-\epsilon).$$

Defina $f(n) \equiv \ln\mu_n(K_\epsilon)$. Pelo Lema 2.1.2, $f(m+n) \ge f(m) + f(n)$, visto que K_ϵ é convexo, e assim

$$f(mn_\epsilon) \ge mf(n_\epsilon) \ge mn_\epsilon(l(C_i)-\epsilon) + m\ln(1-\epsilon) \quad \text{para} \quad m \in \mathbf{N}^*.$$

Logo,

$$\frac{f(m\,n_\epsilon)}{m\,n_\epsilon} \ge l(C_i) - \left(\epsilon + \frac{|\ln(1-\epsilon)|}{n_\epsilon}\right),$$

resultando em

$$l(K_\epsilon) = \lim_{m\to\infty}\frac{f(m\,n_\epsilon)}{m\,n_\epsilon} \ge l(C_i) - \left(\epsilon + \frac{|\ln(1-\epsilon)|}{n_\epsilon}\right).$$

Por outro lado, como $K_\epsilon \subseteq C_i$, tem-se $\sup_{x\in C_i}\{-\lambda(x)\} \ge \sup_{x\in K_\epsilon}\{-\lambda(x)\}$, ou seja, $-\Lambda(C_i) \ge -\Lambda(K_\epsilon)$. Portanto, por (ii)

$$-\Lambda(C_i) \ge -\Lambda(K_\epsilon) \ge l(K_\epsilon) \ge l(C_i) - \left(\epsilon + \frac{|\ln(1-\epsilon)|}{n_\epsilon}\right).$$

Como $\epsilon > 0$ é arbitrário, tomando-se ϵ suficientemente pequeno e concomitantemente n_ϵ suficientemente grande, vem que $\left(\epsilon + \frac{|\ln(1-\epsilon)|}{n_\epsilon}\right)$ pode tornar-se arbitrariamente pequeno, e portanto, $-\Lambda(C_i) \geq l(C_i)$. Isto prova que $l(C_i) = -\Lambda(C_i)$ para $i = 1, 2, \cdots, m$, permitindo-nos concluir que $l(A) = -\Lambda(A)$.

□

Observação 2.1.1. A Proposição 2.1.2 permite-nos enunciar um Princípio Fraco de Grandes Desvios (segundo a definição de Dembo & Zeitouni [9]) para a família $\{\mu_n\}_{n\geq 1}$, $\mu_n(A) = P(\bar{X}_n \in A)$, isto é:

(i) $\overline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(F) \leq -\inf_{x\in F} \lambda(x)$, para $F \subset \mathbf{R}^d$ e F compacto;

(ii) $\underline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(G) \geq -\inf_{x\in G} \lambda(x)$, para $G \subset \mathbf{R}^d$ e G aberto,

sendo λ definida como em (2.1.2).

Observação 2.1.2. Para a Proposição 2.1.2 ser uma generalização do Teorema 1.3.1 é preciso eliminar a hipótese de compacidade no item (ii). Pelo fato de $\{\mu_n : n \geq 1\}$ ser uma família de medidas de probabilidade exponencialmente "tight", isto é, $\{\mu_n : n \geq 1\}$ é tal que para todo $\alpha < +\infty$ existe um compacto $K_\alpha \subset \mathbf{R}^d$ tal que $\overline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(K_\alpha^c) < -\alpha$, e por (i) da Observação 2.1.1, é possível provar que $\overline{\lim}_{n\to\infty} \frac{1}{n} \ln \mu_n(F) \leq -\inf_{x\in F} \lambda(x)$, para qualquer fechado $F \subset \mathbf{R}^d$ (ver Dembo & Zeitouni [9], Teorema 2.2.30).

2.2 - Generalizando a transformada de Cramér para $\mathbf{R}^d$

Analizando a expressão (2.1.2) não é de imediato deduzir que se trata de uma generalização da Definição 1.1.2. Na proposição abaixo obteremos uma extensão "natural" da Definição 1.1.2.

Definição 2.2.1. *Seja μ uma medida de probabilidade em $\mathbf{R}^d$. A transformada de Laplace $\hat{\mu} : \mathbf{R}^d \longrightarrow (0, +\infty]$ para μ é definida por*

$$\hat{\mu}(t) = \int_{\mathbf{R}^d} e^{\langle t,x\rangle} d\mu(x)\,, \quad t \in \mathbf{R}^d,$$

onde $t = (t_1, \cdots, t_d)$, $x = (x_1, \cdots, x_d)$ e $\langle t, x\rangle = t_1x_1 + \cdots + t_dx_d$.

Proposição 2.2.1. *Se $E|X_1| = \int_{\mathbf{R}^d} |x| d\mu(x) < \infty$ então*

$$\lambda(x) = -\inf\{l(C) : C \,\text{aberto convexo}\,, x \in C\} = \sup_{t\in\mathbf{R}^d} \{\langle t, x\rangle - \ln \hat{\mu}(t)\}\,.$$

Prova: Segundo a prova de Vares [30], dividiremos em duas partes: caso $d=1$ e $d>1$.

Caso $d=1$:
Sem perda de generalidade assuma $EX_1=0$. Defina $\tilde{\lambda}(x)\equiv\sup_{t\in\mathbb{R}}\{tx-\ln\mu(t)\}$. A função $\tilde{\lambda}:\mathbb{R}\longrightarrow[0,+\infty]$ é semicontínua inferiormente e convexa (Proposição 1.1.2) e $\lambda:\mathbb{R}\longrightarrow[0,+\infty]$ também é semicontínua inferiormente e convexa (Lema 2.1.3). Além disso,

$$\lambda(x)=-\inf\{l(C): C \text{ aberto convexo}, x\in C\}=-\lim_{\epsilon\to 0}l((x-\epsilon,x+\epsilon)).$$

Então, $\lambda(0)=-\lim_{\epsilon\to 0}l((-\epsilon,\epsilon))$. Mas $l((-\epsilon,\epsilon))=\lim_{n\to\infty}\frac{1}{n}\ln P(\bar{X}_n\in(-\epsilon,\epsilon))$, e pela Lei Fraca dos Grandes Números, $\forall\epsilon>0$, $P\left(|\bar{X}_n|\geq\epsilon\right)$ converge a zero quando $n\to\infty$. Assim, $P\left(|\bar{X}_n|<\epsilon\right)$ converge a 1 quando $n\to\infty$. Logo, $l((-\epsilon,\epsilon))=0$, e disso temos $\lambda(0)=0$. Por outro lado, $\tilde{\lambda}(0)=0$, pela Proposição 1.1.2 (iii).

Supondo que $\tilde{\lambda}$ assume $+\infty$ temos pela Proposição 1.2.1 (i) que $\exists\, r\in(-\infty,0]$ tal que $\tilde{\lambda}(x)=+\infty$ se $x<r$ e $\exists\, p\in[0,+\infty)$ tal que $\tilde{\lambda}(x)=+\infty$ se $x>p$. Além disso, $\tilde{\lambda}$ é finita e contínua em (r,p). Também pelo item (iii) dessa mesma proposição, $\tilde{\lambda}$ é contínua à esquerda em p e à direita em r. Analogamente à prova da Proposição 1.2.1, mostra-se que se λ assume $+\infty$ então existem $s\in(-\infty,0]$ e $q\in[0,+\infty)$ tal que $\lambda(x)=+\infty$ se $x\notin[s,q]$, λ é contínua e finita em (s,q), contínua à direita em s contínua à esquerda em q.

No que segue utilizaremos as notações:

$$\lim_{y\to x^-}\tilde{\lambda}(y)\equiv\tilde{\lambda}(x^-),\ \lim_{y\to x^+}\tilde{\lambda}(y)\equiv\tilde{\lambda}(x^+),\ \lim_{y\to x^-}\lambda(y)\equiv\lambda(x^-)\ \text{ e }\ \lim_{y\to x^+}\lambda(y)\equiv\lambda(x^+).$$

Seja $x>EX_1=0$ e $\epsilon>0$ suficientemente pequeno de maneira que $x-\epsilon\geq 0$. Pelo Teorema 1.2.1(ii),

$$\begin{aligned}l((x-\epsilon,x+\epsilon))&=\lim_{n\to\infty}\frac{1}{n}\ln P(\bar{X}_n\in(x-\epsilon,x+\epsilon))\leq\lim_{n\to\infty}\frac{1}{n}\ln P(\bar{X}_n\in[x-\epsilon,+\infty))\\&=l([x-\epsilon,+\infty))\leq-\tilde{\lambda}(x-\epsilon),\end{aligned}$$

e assim $-\lambda(x)=\lim_{\epsilon\to 0}l((x-\epsilon,x+\epsilon))\leq\lim_{\epsilon\to 0}-\tilde{\lambda}(x-\epsilon)=-\tilde{\lambda}(x^-)$, ou seja,

$$\lambda(x)\geq\tilde{\lambda}(x^-),\quad\text{se } x>0. \tag{2.2.1}$$

Seja $x\geq 0$. Por (1.2.4) e (2.1.3),

$$-\tilde{\lambda}(x)\leq\underline{l}([x,+\infty))\leq\bar{l}([x,+\infty))\leq\max\left\{\bar{l}(\{x\}),\bar{l}((x,+\infty))\right\}.$$

Pelo Lema 2.1.2, $\bar{l}((x,+\infty)) = l((x,+\infty))$, e pela Proposição 2.1.2 (iv),

$$l((x,+\infty)) = -\inf_{y>x} \lambda(y).$$

Além disso, pelo item (ii) dessa mesma proposição, $\bar{l}(\{x\}) \leq -\inf_{y\in\{x\}} \lambda(y) = -\lambda(x)$. Daí vem que

$$\max\left\{\bar{l}(\{x\}), \bar{l}((x,+\infty))\right\} \leq -\min\left\{\lambda(x), \inf_{y>x} \lambda(y)\right\},$$

e portanto, $-\tilde{\lambda}(x) \leq -\min\{\lambda(x), \inf_{y>x}\lambda(y)\}$. Como $x \geq 0$, então $\inf_{y>x}\lambda(y) = \lim_{\epsilon\to 0^+}\lambda(x+\epsilon) = \lambda(x^+)$, pois $\lambda(\cdot)$ é convexa e $\lambda(y) \geq \lambda(0) = 0$ para $y > x$. Logo,

(2.2.2) $$\min\{\lambda(x), \lambda(x^+)\} \leq \tilde{\lambda}(x), \quad \text{se } x \geq 0.$$

Supondo $p < q$, iremos discutir os seguintes casos:

(1º) $0 \leq a < p < q$:

Pelo fato de $\tilde{\lambda}$ ser contínua em (r,p), $\tilde{\lambda}(a^-) = \tilde{\lambda}(a)$ e por (2.2.1) $\tilde{\lambda}(a) \leq \lambda(a)$. Por outro lado, como λ é contínua em (s,q), $\lambda(a^+) = \lambda(a)$ e por (2.2.2) $\lambda(a) \leq \tilde{\lambda}(a)$.

(2º) $0 < p < q < a$:

Quando $x \notin [r,p]$, $\tilde{\lambda}(x) = +\infty$, e quando $y \notin [s,q]$, $\lambda(y) = +\infty$. Assim, $\tilde{\lambda}(a^-) = \lambda(a) = +\infty$ e $\lambda(a^+) = \tilde{\lambda}(a) = +\infty$.

(3º) $0 < p < a < q$:

Neste caso $\tilde{\lambda}(a^-) = +\infty$, $\lambda(a) < +\infty$ e $\lambda(a^+) = \lambda(a) < \tilde{\lambda}(a) = +\infty$.

No (1º) e (2º) casos temos $\tilde{\lambda}(a) = \lambda(a)$, mas o (3º) caso não satisfaz (2.2.1). Assim, vemos que não é possível $p < q$. Supondo $q < p$ e fazendo uma discussão análoga, conclui-se que este caso também não é possível. Logo, $p = q$, e portanto $\lambda(x) = \tilde{\lambda}(x)$ para $x \geq 0$. Seguindo o mesmo raciocínio acima, chega-se a $r = s$ e $\lambda(x) = \tilde{\lambda}(x)$ para $x \leq 0$. No Exemplo 1.1.3, $\tilde{\lambda}$ e λ não assumem $+\infty$, motivo pelo qual não é necessário fazer a discussão acima para p e q.

Caso $d > 1$:

A prova consistirá em utililizar o caso unidimensional. Antes provaremos

(2.2.3) $$\lambda(x) = -\inf\{l(H) : H \text{ semiespaço aberto}, x \in H\} = \sup\{-l(H) : H \text{ semiespaço aberto}, x \in H\}.$$

Sabe-se que um semiespaço é convexo, logo,

$$\lambda(x) = \sup\{-l(C) : C \text{ aberto convexo}, x \in C\} \geq \sup\{-l(H) : H \text{ semiespaço aberto}, x \in H\}.$$

Para provar a desigualdade recíproca vamos supor $\lambda(x) > 0$, pois

$$0 = \lambda(x) \geq \sup\{-l(H) : H \text{ semiespaço aberto}, x \in H\} \geq 0,$$

visto que $-l(H) \geq 0$.

Defina $C \equiv \{y : \lambda(y) \leq \lambda(x) - \epsilon\}$, para $\epsilon > 0$ fixo. A seguir mostraremos que C é convexo e fechado.

C é convexo:

Sejam $y_1, y_2 \in C$ e $\alpha \in [0,1]$. Disso, tem-se

$$\alpha\lambda(y_1) \leq \alpha(\lambda(x) - \epsilon) \qquad \text{e} \qquad (1-\alpha)\lambda(y_2) \leq (1-\alpha)(\lambda(x) - \epsilon),$$

e assim, $\alpha\lambda(y_1) + (1-\alpha)\lambda(y_2) \leq \lambda(x) - \epsilon$. Sendo λ convexa, $\lambda(\alpha y_1 + (1-\alpha)y_2) \leq \alpha\lambda(y_1) + (1-\alpha)\lambda(y_2) \leq \lambda(x) - \epsilon$, isto é, C é convexo.

C é fechado:

Provaremos que o complementar de C é aberto, ou seja, $C^c = \{z : \lambda(z) > \lambda(x) - \epsilon\}$ é aberto. Uma vez que λ é s.c.i. em x temos que para $\forall\epsilon > 0$, $\exists\delta > 0$, $\forall z \in \mathbf{R}^d$, com $|z - x| < \delta$, $\lambda(z) > \lambda(x) - \epsilon$. Logo, $\forall z \in \mathbf{R}^d$, $|z - x| < \delta$, tem-se $z \in C^c$. Como também $x \in C^c$, $B_\delta(x) \subset C^c$. Seja $v \in C^c$. Logo, $\lambda(v) > \lambda(x) - \epsilon$. Por outro lado, como λ é s.c.i. em v, $\forall\epsilon > 0$, $\exists\gamma > 0$, $\forall y \in \mathbf{R}^d$, com $|y - v| < \gamma$, $\lambda(y) > \lambda(v) - \epsilon > \lambda(x) - 2\epsilon$, e portanto, $y \in C^c$. Assim, concluimos que para $\forall v \in C^c$ podemos obter uma bola aberta contida em C^c. Logo, C^c é aberto e por conseguinte C é fechado.

Agora provaremos que o fato de $x \notin C$ implica na existência de um hiperplano, o qual separa estritamente x e C.

O ponto $p_C(x) \in C$ tal que $||x - p_C(x)|| \leq ||x - y||$, $\forall y \in C$, denomina-se de projeção de x sobre C, e além disso $p_C(x)$ é único (ver Teorema A.2.5 do Apêndice 2).

Seja $u = x - p_C(x)$ e defina $h(y) = \langle y, u\rangle$. Então, $h(x) - h(p_C(x)) = \langle x - p_C(x), x - p_C(x)\rangle = ||u||^2 > 0$. Logo, $h(x) > h(p_C(x))$. Por outro lado,

$$h(y) - h(p_C(x)) = \langle y, u\rangle - \langle p_C(x)), u\rangle = \langle y - p_C(x), x - p_C(x)\rangle,$$

e pela Proposição A.2.2 do Apendice 2, $\langle y - p_C(x), x - p_C(x)\rangle \leq 0$ para $\forall y \in C$. Por conseguinte, tem-se $h(y) \leq h(p_C(x)) < h(x)$, $\forall y \in C$, ou seja, $M = \sup_{y\in C} h(y) < h(x)$. Portanto, $h(y) = h(p_C(x))$ é um hiperplano o qual separa estritamente x e C.

Prosseguindo, seja $H \equiv \{z : \langle z, u\rangle > M\}$. Note que H é um semiespaço aberto e $x \in H$ se e somente se $h(x) > M$. Suponha $H \cap C \neq \emptyset$. Então, para $w \in H \cap C$, tem-se $\langle w, u\rangle > M$ e $\langle w, u\rangle \leq M$, a qual é uma contradição. Logo, $H \cap C = \emptyset$, e por isso $H \cap C^c \neq \emptyset$, $H \subset C^c$. Tendo em vista que $H \subset C^c$ e utilizando a Proposição 2.1.2(iv), temos

$$-l(H) = \inf_{z\in H} \lambda(z) > \lambda(x) - \epsilon,$$

pois para todo $z \in H$, $\lambda(z) > \lambda(x) - \epsilon$. Isto significa que

$$\sup\{-l(H) : H \text{ semiespaço aberto, } x \in H\} \geq \lambda(x) - \epsilon$$

para $\epsilon > 0$ arbitrário, provando (2.2.3).

Todo semiespaço que contém x pode ser expresso como

$$H(t,r) \equiv \left\{z \in \mathbf{R}^d : \langle z, t\rangle < r\right\}, \quad \text{onde } t \in \mathbf{R}^d \text{ e } r \in \mathbf{R}.$$

Portanto, $\langle x, t\rangle < r$. Seja $\{\langle t, X_n\rangle\}_{n\geq 1}$ seqüência de v.a's i.i.d. com lei de probabilidade comum μ_t e respectiva transformada de Cramér λ_{μ_t}. Denotemos $\lambda_t(\cdot) \equiv \lambda_{\mu_t}(\cdot)$. Note que o domínio de $\lambda_t(\cdot)$ é $\mathbf{R}$. Seja $l_t(I) \equiv \lim_{n\to\infty} \frac{1}{n} \ln P(\langle t, \bar{X}_n\rangle \in I)$ para I intervalo em $\mathbf{R}$. Por (2.2.3), com $d = 1$,

$$\lambda_t(s) = \max\left\{\sup_{r>s}\{-l_t(-\infty, r)\}, \sup_{r<s}\{-l_t(r, +\infty)\}\right\}, \quad s \in \mathbf{R}.$$

Mas $\langle t, \bar{X}_n\rangle \in (-\infty, r)$ equivale a $\bar{X}_n \in \left\{y \in \mathbf{R}^d : \langle t, y\rangle < r\right\} = H(t,r)$ e $\langle t, \bar{X}_n\rangle \in (r, +\infty)$ equivale a $\bar{X}_n \in \left\{z \in \mathbf{R}^d : \langle -t, z\rangle < -r\right\} = H(-t,-r)$. Disso vem que

$$l_t(-\infty, r) = \lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in H(t,r)) = l(H(t,r)) \text{ e}$$

$$l_t(r, +\infty) = \lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in H(-t,-r)) = l(H(-t,-r)).$$

Segue que

$$\lambda_t(\langle t, x\rangle) = \max\left\{\sup_{r>\langle t,x\rangle}\{-l(H(t,r))\}, \sup_{r<\langle t,x\rangle}\{-l(H(-t,-r))\}\right\},$$

onde usamos o fato de $\langle -t, x\rangle < -r$ e $\langle t, x\rangle > r$ serem equivalentes.

Lembrando que μ é a lei de $P(\bar{X}_n \in A)$ e λ é a transformada de Cramér para μ, temos por (2.2.3) que $\lambda(x) = \sup_{t\in\mathbf{R}^d} \lambda_t(\langle t, x\rangle)$. Pelo caso $d = 1$ decorre que

$$\lambda_t(\langle t, x\rangle) = \sup_{s\in\mathbf{R}} \{s\langle t, x\rangle - \ln \hat{\mu}_t(s)\}.$$

Notando que

$$\hat{\mu}_t(s) = Ee^{s\langle t, X_1\rangle} = Ee^{\langle st, X_1\rangle} = \hat{\mu}(st)$$

temos

$$\lambda(x) = \sup_{t\in\mathbf{R}^d} \sup_{s\in\mathbf{R}} \{\langle ts, x\rangle - \ln \hat{\mu}(st)\} = \sup_{t\in\mathbf{R}^d} \{\langle t, x\rangle - \ln \hat{\mu}(t)\}.$$

□

Através da Proposição 2.1.2, Observação 2.1.2 e da Proposição 2.2.1 agora podemos enunciar o teorema a seguir, o qual generaliza o Teorema 1.3.1.

Teorema 2.2.1. *Para* $\{\mu_n : n \geq 1\}, \mu_n(A) = P(\bar{X}_n \in A)$, *tem-se:*

(i) $\overline{\lim}_{n\to\infty} \frac{1}{n}\mu_n(F) \leq -\inf_{x\in F} \lambda(x)$, para qualquer fechado $F \subset \mathbf{R}^d$;

(ii) $\underline{\lim}_{n\to\infty} \frac{1}{n}\mu_n(G) \geq -\inf_{x\in G} \lambda(x)$, para qualquer aberto $G \subset \mathbf{R}^d$,

sendo $\lambda(x) = \sup_{t\in\mathbf{R}^d} \{\langle t, x\rangle - \ln \hat{\mu}(t)\}$.

2.3 - Alguns exemplos

Exemplo 2.3.1. Seja $Z = (X, Y)$ um vetor aleatório com distribuição Normal com parâmetros $m = (m_X, m_Y)$, $\sigma = (\sigma_X, \sigma_Y)$ e $|\rho| < 1$, onde ρ é o coeficiente de correlação entre X e Y.

A densidade do vetor Z é

$$\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{Q(u, v)}{2}\right\},$$

onde

$$Q(u,v) = \frac{1}{1-\rho^2}\left[\left(\frac{u-m_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{u-m_X}{\sigma_X}\right)\left(\frac{v-m_Y}{\sigma_Y}\right) + \left(\frac{v-m_Y}{\sigma_Y}\right)^2\right]$$

e $z = (u,v) \in \mathbf{R}^2$.

A extensão da Proposição 1.1.1 é $\lambda(z) = -\ln \inf_{t\in\mathbf{R}^2}\left\{e^{-\langle t,z\rangle}\hat{\mu}(t)\right\}$. Sendo

$$\hat{\mu}(t) = \exp\left\{m_X t_1 + m_Y t_2 + \frac{\sigma_X^2}{2}t_1^2 + \rho\sigma_X\sigma_Y t_1 t_2 + \frac{\sigma_Y^2}{2}t_2^2\right\}, \quad t = (t_1,t_2) \in \mathbf{R}^2,$$

tem-se $e^{-\langle t,z\rangle}\hat{\mu}(t) \equiv g(t_1,t_2)$, com

$$g(t_1,t_2) = \exp\left\{-t_1 u - t_2 v + m_X t_1 + m_Y t_2 + \frac{\sigma_X^2}{2}t_1^2 + \frac{\sigma_Y^2}{2}t_2^2 + \rho\sigma_X\sigma_Y t_1 t_2\right\}.$$

As derivadas segunda de g são: $g_{t_1,t_1} = \sigma_X^2 > 0$, $g_{t_2,t_2} = \sigma_Y^2 > 0$ e $g_{t_1,t_2} = \rho\sigma_X\sigma_Y > 0$. Como $g_{t_1,t_1}g_{t_2,t_2} - (g_{t_1,t_2})^2 = \sigma_X^2\sigma_Y^2(1-\rho^2) > 0$ e $g_{t_1,t_1} > 0$ segue que $g(t_1,t_2)$ tem um ponto τ, o qual é um mínimo absoluto. Analogamente ao Exemplo 1.1.3, para obter τ minimizamos

$$t_1(m_X - u) + t_2(m_Y - v) + \frac{\sigma_X^2}{2}t_1^2 + \frac{\sigma_Y^2}{2}t_2^2 + \rho\sigma_X\sigma_Y t_1 t_2$$

em relação a t_1 e t_2. Mas isto resulta no sistema de equações

$$\begin{cases} -u + m_X + \sigma_X^2\tau_1 + \rho\sigma_X\sigma_Y\tau_2 = 0 \\ -v + m_Y + \sigma_Y^2\tau_2 + \rho\sigma_X\sigma_Y\tau_1 = 0 \end{cases},$$

cuja solução é

$$\tau_1 = \frac{\sigma_Y^2(u-m_X) - \rho\sigma_X\sigma_Y(v-m_Y)}{\sigma_X^2\sigma_Y^2(1-\rho^2)} \qquad \text{e} \qquad \tau_2 = \frac{\sigma_X^2(v-m_Y) - \rho\sigma_X\sigma_Y(u-m_X)}{\sigma_X^2\sigma_Y^2(1-\rho^2)}.$$

Assim,

$$\lambda(z) = -\ln\left(e^{-\langle\tau,z\rangle}\hat{\mu}(\tau)\right) = \frac{\sigma_Y^2(u-m_X)^2 + \sigma_X^2(v-m_Y)^2 - 2\rho\sigma_X\sigma_Y(u-m_X)(v-m_Y)}{2\sigma_X^2\sigma_Y^2(1-\rho^2)},$$

$0 \le \rho < 1$.

Suponha $A = \left\{z = (u,v) \in \mathbf{R}^2 : -\infty < u < x \text{ e } -\infty < v < y\right\}$, onde $x, y \in \mathbf{R}$. Gostaríamos de calcular $\lim_{n\to\infty}\frac{1}{n}\ln P(\bar{X}_n \in A)$. Analogamente ao caso $d = 1$, a transformada de Cramér em $\mathbf{R}^2$ é tal que

$\lambda(m_X, m_Y) = 0$ e $\lambda(u, v) > 0$ se $(u, v) \neq (m_X, m_Y)$. No caso de uma distribuição normal bidimensional, o gráfico de λ é um parabolóide elíptico centrado em (m_X, m_Y). Se $x \geq m_X$ e $y \geq m_Y$ então

$$\inf_{z \in A^\circ} \lambda(z) = \inf_{z \in \bar{A}} \lambda(z) = 0 = \lambda(m_X, m_Y),$$

e assim $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in A) = 0$. Por outro lado, se $x < m_X$ e $y < m_Y$ então

$$\inf_{z \in A^\circ} \lambda(z) = \inf_{z \in \bar{A}} \lambda(z) = \lambda(x, y) > 0,$$

e portanto, $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in A) = -\lambda(x, y)$.

Quando $\rho = 0$ e $X = Y$, $\lambda(u, v) = 2\frac{(u-m)^2}{2\sigma^2}$, isto é, o mesmo resultado que chegaríamos utilizando o item (ii) da Proposição 1.1.3.

Neste exemplo vemos como a Proposição 2.2.1 fornece-nos uma maneira prática de calcular λ em $\mathbf{R}^d$. Já a expressão (2.1.2) não tem a mesma operacionalidade, embora seja um instrumento valioso na prova da Proposição 2.1.2.

Exemplo 2.3.2 (extraído de Bucklew [5], exemplo 6 - pag. 21). Seja Z um vetor aleatório d-dimensional com distribuição normal de esperança $m = (m_1, \cdots, m_d)$ e matriz de covariâncias $V = \{v_{i,j}\}_{i,j=1}^d$. Assumindo que V possui inversa, a densidade de Z é

$$\frac{\sqrt{det(V^{-1})}}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}(z-m)^T V^{-1}(z-m)\right\},\ z \in \mathbf{R}^d,$$

onde $(z-m)^T$ indica o vetor transposto de $(z-m)$. A transformada de Laplace é

$$\hat{\mu}(t) = \exp\left\{\langle t, m\rangle + \frac{1}{2}t^T V t\right\},\ t = (t_1, \cdots, t_d).$$

Quer-se calcular

$$\lambda(z) = \sup_{t\in\mathbf{R}^d} \{\langle t, z\rangle - \ln\hat{\mu}(t)\} = \sup_{t\in\mathbf{R}^d} \left\{\langle t, z\rangle - \langle t, m\rangle - \frac{1}{2}t^T V t\right\}$$
$$= \sup_{t\in\mathbf{R}^d} \left\{\langle t, z-m\rangle - \frac{1}{2}t^T V t\right\} = \sup_{t\in\mathbf{R}^d} \left\{t^T(z-m) - \frac{1}{2}t^T V t\right\}.$$

Para isto, defina

$$g(t) = t^T(z-m) - \frac{1}{2}t^T V t = \sum_{i=1}^{d} t_i(z_i - m_i) - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} t_i t_j v_{i,j}.$$

A matriz hessiana de g é $H(g) = \left\{\frac{\partial^2 g}{\partial t_i t_j}\right\}_{i,j=1}^{d} = \{-v_{i,j}\}_{i,j=1}^{d}$, ou seja, $H(g) = -V$. Uma vez que V é matriz positiva definida, então todos seus autovalores são reais positivos (ver Strang **[27]**). Mas isto implica que os autovalores de $-V$ são todos reais negativos. Sendo assim, o índice da matriz hessiana de g (definido como o número de autovalores positivos) é zero, e por isso g tem um ponto de máximo absoluto τ (ver Lima **[23]**).

Para calcular τ, note que o gradiente de $t^T(z-m) - \frac{1}{2}t^T V t$ é $(z-m) - Vt$. Então τ é obtido através da solução da equação $(z-m) - Vt = 0$, ou seja, $\tau = V^{-1}(z-m)$, $z \in \mathbf{R}^d$. Segue que

$$\lambda(z) = \tau^T(z-m) - \frac{1}{2}\tau^T V \tau = (z-m)^T V^{-1}(z-m) - \frac{1}{2}(z-m)^T V^{-1}(z-m) = \frac{1}{2}(z-m)^T V^{-1}(z-m).$$

Nos exemplos do Capítulo 1 podemos observar que $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in A)$ sempre existe quando A é uma semireta. Nos exemplos 2.3.1 e 2.3.2 este limite também existe quando A é um semiplano. No exemplo a seguir veremos um caso em que $\lim_{n\to\infty} \frac{1}{n} \ln P(\bar{X}_n \in A)$ não existe, mesmo para um semiplano.

Exemplo 2.3.3 (extraído de Dembo & Zeitouni [9], exercício 2.2.37). Seja $\{X_n\}_{n\geq 1}$ uma seqüência de v.a's i.i.d. a valores em $\mathbf{R}^2$ tal que $P(X_1 = (1,0)) = P(X_1 = (0,1)) = \frac{1}{2}$. Considere o semiplano $F = \left(-\infty, \frac{1}{2}\right] \times \left(-\infty, \frac{1}{2}\right]$.

Suponha $n = 2$. Então $[\bar{X}_n \in F]$ se e somente se $[X_1 = (1,0), X_2 = (0,1)]$ ou $[X_1 = (0,1), X_2 = (1,0)]$ e daí vem que $P(\bar{X}_n \in F) = 2\frac{1}{2^2}$. Se por exemplo a amostra aleatória tivesse resultado em $x_1 = (1,0)$ e $x_2 = (1,0)$, então $\bar{x}_n = (1,0) \notin F$. Para $n = 4$, $[\bar{X}_n \in F]$ se e somente se ocorrer um dos seguintes eventos:

$$[X_1 = (1,0), X_2 = (1,0), X_3 = (0,1), X_4 = (0,1)],$$
$$[X_1 = (1,0), X_2 = (0,1), X_3 = (1,0), X_4 = (0,1)],$$
$$[X_1 = (1,0), X_2 = (0,1), X_3 = (0,1), X_4 = (1,0)],$$
$$[X_1 = (0,1), X_2 = (1,0), X_3 = (1,0), X_4 = (0,1)],$$
$$[X_1 = (0,1), X_2 = (0,1), X_3 = (1,0), X_4 = (1,0)] \quad \text{ou}$$
$$[X_1 = (0,1), X_2 = (1,0), X_3 = (1,0), X_4 = (0,1)].$$

Logo, $P(\bar{X}_n \in F) = 6\frac{1}{2^4}$. Generalizando, para um n par, $P(\bar{X}_n \in F) = \binom{n}{\frac{n}{2}}\frac{1}{2^n}$.

Se n for ímpar, $P(\bar{X}_n \in F) = 0$. Para ver isto, assuma por exemplo $n = 3$. Então, $\frac{(1,0)+(1,0)+(1,0)}{3} = (1,0) \notin F$, $\frac{(0,1)+(0,1)+(0,1)}{3} = (0,1) \notin F$, $\frac{(0,1)+(0,1)+(1,0)}{3} = (\frac{1}{3}, \frac{2}{3}) \notin F$ e $\frac{(1,0)+(1,0)+(0,1)}{3} = (\frac{2}{3}, \frac{1}{3}) \notin F$. Portanto,

$$(2.3.1) \qquad P(\bar{X}_n \in F) = \begin{cases} \binom{n}{\frac{n}{2}} \frac{1}{2^n} & \text{se } n \text{ é par} \\ 0 & \text{se } n \text{ é ímpar} \end{cases}.$$

Uma vez que $\inf_{k \geq n} \frac{1}{k} \ln P(\bar{X}_k \in F) = -\infty$, $\forall n \in \mathbf{N}^*$, então

$$\varliminf_{n \to \infty} \frac{1}{n} \ln P(\bar{X}_n \in F) = \sup_{n \in \mathbf{N}^*} \inf_{k \geq n} \frac{1}{k} \ln P(\bar{X}_k \in F) = -\infty.$$

Por outro lado, a subseqüência $\left\{\frac{1}{n} \ln P(\bar{X}_n \in F) \, : \, n \text{ é par}\right\}$ converge para zero. Para ver isto, recorde a aproximação de Stirling: $\lim_{n \to \infty} \frac{n!}{\sqrt{2\pi} n^{n+1/2} e^{-n}} = 1$, cuja notação é $n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n}$. Logo, pela aproximação de Stirling,

$$\binom{n}{\frac{n}{2}} \frac{1}{2^n} \sim \frac{1}{\sqrt{n/2\pi}},$$

e como $\lim_{n \to \infty} \frac{\ln\left(\frac{1}{\sqrt{n/2\pi}}\right)}{n} = 0$ então $\left\{\frac{1}{n} \ln P(\bar{X}_n \in F) \, : \, n \text{ é par}\right\}$ converge para zero. Por conseguinte,

$$\varlimsup_{n \to \infty} \frac{1}{n} \ln P(\bar{X}_n \in F) = \inf_{n \in \mathbf{N}^*} \sup_{k \geq n} \frac{1}{k} \ln P(\bar{X}_k \in F) = 0.$$

Logo, concluímos que o limite não existe. De fato, por (2.3.1), vemos que as hipóteses do Lema 2.1.2 não são satisfeitas.

CAPÍTULO 3

ALGUMAS EXTENSÕES E APLICAÇÕES

No Capítulo 2 apresentamos um princípio de Grandes Desvios para médias de v.a's i.i.d. a valores em $\mathbf{R}^d$, o qual é uma extensão do teorema de Cramér-Chernoff. Outras extensões dizem respeito basicamente ao espaço onde as v.a's estão definidas e à existência de alguma interação entre elas.

O objetivo deste capítulo é apresentar de maneira suscinta algumas extensões e aplicações do teorema de Cramér-Chernoff, dando-se uma maior ênfase à interpretação destes resultados do que as suas respectivas provas laboriosas.

3.1 - O teorema de Sanov para a medida empírica de v.a's i.i.d. em $\mathbf{R}^d$

No decorrer desta seção consideraremos $\{X_n\}_{n\geq 1}$ uma seqüência de v.a's i.i.d. a valores em $\mathbf{R}^d$ com lei de probabilidade μ.

Considere a medida de Dirac δ em X_i, isto é,

$$\delta_{X_i(\omega)}(A) \equiv \begin{cases} 1 & \text{se } X_i(\omega) \in A \\ 0 & \text{caso contrário} \end{cases},$$

$A \in \mathcal{B}(\mathbf{R}^d)$ e $\omega \in \Omega$, onde $\mathcal{B}(\mathbf{R}^d)$ é a σ-álgebra de Borel em $\mathbf{R}^d$ e Ω é tal que $X_i : \Omega \longrightarrow \mathbf{R}^d$. A medida empírica de $\{X_n\}_{n\geq 1}$ é

$$L_n(\omega, A) \equiv \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(\omega)}(A), \quad n \in \mathbf{N}^*.$$

Observação 3.1.1. Se $A = \emptyset$ então $L_n(\omega, A)$ é por definição igual a zero.

Observação 3.1.2. Quando ω está fixo, $L_n(\omega, \cdot)$ define uma medida de probabilidade sobre o espaço mensurável $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$. Por outro lado, se $A \in \mathcal{B}(\mathbf{R}^d)$ está fixo então $L_n(\cdot, A)$ é uma variável aleatória.

Observação 3.1.3. As variáveis aleatórias $\delta_{X_i(\omega)}(A)$, $i = 1, \cdots, n$ com $A \in \mathcal{B}(\mathbf{R}^d)$, são i.i.d. e também $E\left(\delta_{X_i(\omega)}(A)\right) = \mu(A)$. Além disso, $E\left(L_n(\cdot, A)\right) = \mu(A)$. Sendo assim, pela Lei dos Grandes Números, para $\forall \epsilon > 0$

$$\lim_{n\to\infty} P\{\omega \in \Omega : |L_n(\omega, A) - \mu(A)| > \epsilon\} = 0, \quad A \in \mathcal{B}(\mathbf{R}^d),$$

isto é, o evento $\{\omega \in \Omega : |L_n(\omega, A) - \mu(A)| > \epsilon\}$ representa uma situação de Grande Desvio em relação à Lei dos Grandes Números para a medida empírica. Note ainda que para cada $\omega \in \Omega$, $L_n(\omega, \cdot)$ é um elemento de $\mathcal{M}_1(\mathbf{R}^d)$, onde

$$\mathcal{M}_1(\mathbf{R}^d) \equiv \{\eta : \mathcal{B}(\mathbf{R}^d) \longrightarrow [0,1] \mid \eta \text{ é probabilidade } \sigma-\text{aditiva}\}. \tag{3.1.1}$$

Observação 3.1.4. Por simplicidade de notação utilizaremos L_n para denotar a medida empírica da seqüência $\{X_n\}_{n\geq 1}$.

Definição 3.1.1. *A informação de Kullback de ν em relação a μ é definida como*

$$K(\mu, \nu) = \begin{cases} \int_{\mathbf{R}^d} f(x) \ln f(x)\, d\mu(x) & \text{se } \nu \ll \mu \\ +\infty & \text{caso contrário} \end{cases},$$

onde $f(x) = \frac{d\nu(x)}{d\mu(x)}$ e " $\ll$ " denota que ν é absolutamente contínua em relação μ.

Teorema 3.1.1 (Sanov). *Seja $\{X_n\}_{n\geq 1}$ uma seqüência de v.a's i.i.d. a valores em $\mathbf{R}^d$ com lei de probabilidade μ. Então:*

(i) Para todo $B \in \mathcal{B}\left(\mathcal{M}_1(\mathbf{R}^d)\right)$, onde $\mathcal{B}\left(\mathcal{M}_1(\mathbf{R}^d)\right)$ é a $\sigma-$álgebra de Borel de $\mathcal{M}_1(\mathbf{R}^d)$, tem-se que

$$-\inf_{\nu \in B^\circ} K(\mu,\nu) \leq \varliminf_{n\to\infty} \frac{1}{n} \ln \mu_n(B) \leq \varlimsup_{n\to\infty} \frac{1}{n} \ln \mu_n(B) \leq -\inf_{\nu\in \bar{B}} K(\mu,\nu),$$

sendo $\mu_n(B) = P(L_n \in B)$;

(ii) Se B for união finita de abertos convexos de $\mathcal{M}_1(\mathbf{R}^d)$, $\lim_{n\to\infty} \frac{1}{n} \ln \mu_n(B) = -\inf_{\nu\in B} K(\mu,\nu)$.

Prova: A prova é omitida nesta dissertação, mas pode ser encontrada em Azencott [2] e Vares [30].

Observação 3.1.5. A família de medidas de probabilidade $\{\mu_n\}_{n\geq 1}$ é tal que $\mu_n : \mathcal{B}\left(\mathcal{M}_1(\mathbf{R}^d)\right) \longrightarrow [0,1]$.

Para uma seqüência $\{X_n\}_{n\geq 1}$ de v.a's i.i.d. a valores em $\mathbf{R}^d$ com lei de probabilidade μ considere $g : \mathbf{R}^d \longrightarrow \mathbf{R}$ e defina $\bar{S}_n \equiv \frac{1}{n}\sum_{i=1}^n g(X_i)$. Também neste caso podemos indagar sobre o comportamento assintótico de $\bar{S}_n$.

Observe que $\bar{S}_n$ pode ser escrito como

$$\bar{S}_n = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbf{R}^d} g(y)\,\delta_{X_i}(y)\,dy = \int_{\mathbf{R}^d} g(y)\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}(y)\right)dy = \int_{\mathbf{R}^d} g(y)\,dL_n(y),$$

ou seja, $\bar{S}_n$ é o valor esperado de $g(\cdot)$ sob a lei de probabilidade definida por L_n.

Pela Lei dos Grandes Números, sabe-se que para $\forall\,\epsilon > 0$, $\lim_{n\to\infty} P(|\bar{S}_n - m| > \epsilon) = 0$, onde $m = \int_{\mathbf{R}^d} g(y)\,d\mu(y)$, permitindo-nos concluir que o evento $\left[|\bar{S}_n - m| > \epsilon\right]$ é uma situação de Grande Desvio em relação à esperança m. Em particular, $\left[\bar{S}_n > m + \epsilon\right]$ também é um evento do tipo Grande Desvio.

Além da informação dada pela Lei dos Grandes Números gostaríamos de saber qual é a taxa de convergência de $\ln P(\bar{S}_n > m+\epsilon)$ para $-\infty$ quando $n \to \infty$. Esta informação é obtida pelo seguinte corolário do Teorema 3.1.1:

Corolário 3.1.1. *Seja $(X_n)_{n\geq 1}$ uma seqüência de v.a's i.i.d. a valores em $\mathbf{R}^d$ com lei de probabilidade μ. Considere a transformação $g : \mathbf{R}^d \longrightarrow \mathbf{R}$ e $\bar{S}_n \equiv \frac{1}{n}\sum_{i=1}^{n} g(X_i)$. Então para o evento $\left[\bar{S}_n > m+\epsilon\right]$, $\epsilon > 0$, tem-se que*

$$-\inf_{\nu\in A^\circ} K(\mu,\nu) \leq \varliminf_{n\to\infty}\frac{1}{n}\ln P\left(\int_{\mathbf{R}^d} g(y)\,dL_n(y) > m+\epsilon\right) \leq \varlimsup_{n\to\infty}\frac{1}{n}\ln P\left(\int_{\mathbf{R}^d} g(y)\,dL_n(y) > m+\epsilon\right)$$
$$\leq -\inf_{\nu\in\bar{A}} K(\mu,\nu),$$

sendo A° e $\bar{A}$ o interior e o fecho de A, respectivamente, e

$$A = \left\{\nu \in \mathcal{M}_1(\mathbf{R}^d) : \int_{\mathbf{R}^d} g(y)\,d\nu(y) > m+\epsilon\right\}.$$

3.2 - Grandes Desvios em Cadeias de Markov finitas

Considere uma Cadeia de Markov $\{X_n\}_{n\geq 1}$ com espaço de estados $S = \{1,\cdots,N\}$, espaço de parâmetros $T = \{0,1,\cdots\}$ e matriz de transição $\mathbf{P} = \{p(i,j)\}_{i,j=1}^{N}$. Assuma também que essa cadeia é irredutível e aperiódica, e portanto, ergódica. A medida empírica dessa cadeia é definida como $L_n \equiv (L_n(1),\cdots,L_n(N))$, onde

$$L_n(i) = \frac{1}{n}\sum_{k=1}^{n}\delta_{X_k}(i) \qquad \text{e} \qquad \delta_{X_k}(i) = \begin{cases} 1 & \text{se } X_k = i \\ 0 & \text{caso contrário} \end{cases}, \quad i \in S.$$

$L_n(i)$, $i \in S$, tem como interpretação a proporção de tempo em que a Cadeia assume o estado i nas n primeiras transições.

Seja a seqüência $\{Q_{n,j}\}_{n\geq 1}$ tal que

$$(3.2.1) \qquad Q_{n,j}(B) = P_j\left[(L_n(1), \cdots, L_n(N)) \in B\right],$$

onde $j \in S$ é o estado inicial da cadeia e $B \in \mathcal{B}(\mathcal{M}_1(S))$, sendo $\mathcal{M}_1(S) \equiv \{$ espaço de todas medidas de probabilidade sobre $S\}$ e $\mathcal{B}(\mathcal{M}_1(S))$ sua respectiva $\sigma-$álgebra de Borel. Note que $Q_{n,j} : \mathcal{B}(\mathcal{M}_1(S)) \longrightarrow [0,1]$.

Agora considere o evento $[(L_n(1), \cdots, L_n(N)) \in B_\epsilon]$, onde $\epsilon > 0$, $B_\epsilon = \{(\nu_1, \cdots, \nu_N) \in \mathcal{B}(\mathcal{M}_1(S)) : |\nu_i - \pi_i| > \epsilon, i = 1, \cdots, N\}$ e $\pi = (\pi_1, \cdots, \pi_N)$ é a medida invariante da cadeia. Então, pela Lei dos Grandes Números para Cadeias de Markov (ver Teorema A.1.6 do Apêndice 1), $\lim_{n\to\infty} L_n(i) = \pi_i$, $i \in S$, e conseqüentemente $\lim_{n\to\infty} Q_{n,j}(B_\epsilon) = 0$ para $\forall\epsilon > 0$ e $\forall j \in S$. Portanto, o evento $[(L_n(1), \cdots, L_n(N)) \in B_\epsilon]$ é uma situação de Grande Desvio em relação à medida invariante π.

A seguir enunciamos o Teorema de Sanov para a medida empírica de Cadeias de Markov finitas.

Teorema 3.2.1. *Seja $\{X_n\}_{n\geq 1}$ uma Cadeia de Markov irredutível e aperiódica com espaço de estados $S = \{1, \cdots, N\}$ e matriz de transição $\mathbf{P} = \{p(i,j)\}_{i,j=1}^N$. Seja $L_n = (L_n(1), \cdots, L_n(N))$ a medida empírica da cadeia. Para cada $t = (t_1, \cdots, t_N) \in \mathbf{R}^N$ defina $\mathbf{P}_t = \{p_t(i,j)\}_{i,j=1}^N$, onde $p_t(i,j) = p(i,j)e^{t_j}$. Portanto, para a seqüência $\{Q_{n,j}\}_{n\geq 1}$, conforme (3.2.1), tem-se*

$$-\inf_{q\in B^\circ} \mathbf{I}(q) \leq \varliminf_{n\to\infty} \frac{1}{n} \ln Q_{n,j}(B) \leq \varlimsup_{n\to\infty} \frac{1}{n} \ln Q_{n,j}(B) \leq -\inf_{q\in \bar{B}} \mathbf{I}(q)$$

para $\forall B \in \mathcal{B}(\mathcal{M}_1(S))$ e $\forall j \in S$, sendo $\mathbf{I}(q) = \sup_{t\in\mathbf{R}^N}\{\langle t, q\rangle - \ln\rho(\mathbf{P}_t)\}$, $q \in \mathcal{M}_1(S)$ e $\rho(\mathbf{P}_t)$ denotando o maior autovalor positivo da matriz $\mathbf{P}_t$.

Prova: Dembo & Zeitouni [9].

Observação 3.2.1. Uma expressão alternativa para $\mathbf{I}(\cdot)$ é dada por

$$\mathbf{I}(q) = \begin{cases} \sup_{u\gg 0} \sum_{j=1}^N q_j \ln\left(\frac{\pi_j}{(\mathbf{uP})_j}\right) & \text{se } q \in \mathcal{M}_1(S) \\ +\infty & \text{caso contrário} \end{cases},$$

onde $u \gg 0$ significa que o vetor u deve ter todas componentes estritamente positivas e $(u\mathbf{P})_j$ é a j-ésima componente do vetor $u\mathbf{P}$.

3.3 - Aplicações

As aplicações aqui apresentadas foram extraídas de Bucklew [5].

Aplicação 3.3.1. *Considere $\{X_n\}_{n\geq 1}$ uma seqüência de v.a's i.i.d. a valores em $\mathbf{R}$ com densidade de probabilidade $p(x,\theta)$, $\theta \in \Theta$. Seja $g(\cdot)$ uma transformação tal que $g : \Theta \longrightarrow \mathbf{R}$. Suponha que se deseje estimar $g(\theta)$. Assuma que $T_n(X)$, $X = (X_1, \cdots, X_n)$, é um estimador consistente para $g(\theta)$, ou seja, para* $\forall\, \epsilon > 0$

$$\lim_{n\to\infty} \int_{\{x\in\mathbf{R}^n \,:\, |T_n(x)-g(\theta)|>\epsilon\}} p_n(x,\theta)\, dx = 0, \tag{3.3.1}$$

onde $p_n(x,\theta) = p(x_1,\theta).\cdots.p(x_n,\theta)$ e $x = (x_1, \cdots, x_n)$.

Em (3.3.1) vemos que o evento $[x \in \mathbf{R}^n : |T_n(x) - g(\theta)| > \epsilon]$ é uma situação de Grande Desvio em relação a $g(\theta)$. Note que não estamos exatamente diante da Lei dos Grandes Números, pois $T_n(X)$ não é necessariamente uma média aritmética das n observações.

Para $\epsilon > 0$ defina:

$$a_n(\epsilon,\gamma) \equiv \int_{\{x\in\mathbf{R}^n \,:\, |T_n(x)-g(\gamma)|>\epsilon\}} p_n(x,\gamma)\, dx,$$

$$\Delta(\epsilon,\gamma) \equiv \{\delta \,:\, \delta \in \Theta,\ |g(\delta) - g(\gamma)| > \epsilon\},$$

$$b(\epsilon,\gamma) \equiv \begin{cases} \inf\{K(p(\cdot,\delta),p(\cdot,\gamma)) \,:\, \delta \in \Delta(\epsilon,\gamma)\} & \text{se } \Delta(\epsilon,\gamma) \neq \emptyset \\ +\infty & \text{caso contrário} \end{cases},$$

sendo

$$K(p(\cdot,\delta),p(\cdot,\gamma)) = \int_{\mathbf{R}} p(x,\delta) \ln \frac{p(x,\delta)}{p(x,\gamma)}\, dx.$$

Note que $K(p(\cdot,\delta).p(\cdot,\gamma))$ é a informação de Kullback de $p(x,\delta)$ em relação a $p(x,\gamma)$, conforme a Definição 3.1.1. Agora podemos enunciar o teorema de Bahadur-Zabell-Gupta (cuja prova encontra-se em Bucklew [5], pág. 160), o qual diz:

$$(3.3.2) \qquad \varliminf_{n\to\infty} \frac{1}{n} \ln a_n(\epsilon,\gamma) \geq -b(\epsilon,\gamma).$$

Através de (3.3.1), sabe-se que a probabilidade do evento $\{x \in \mathbf{R}^n : |T_n(x) - g(\theta)| > \epsilon\}$, $\epsilon > 0$, tende a zero quando $n \to \infty$. Além dessa informação, a desigualdade (3.3.2) fornece uma cota superior para a taxa de decaimento exponencial de $a_n(\epsilon,\gamma)$ para zero quando $n \to \infty$ (no sentido descrito em (0.1)).

Aplicação 3.3.2. Seja $(X_1, \cdots, X_n)$ uma amostra aleatória de n observações i.i.d. a valores em $\mathbf{R}$. Sendo assim, as seguintes hipóteses podem ser formuladas:

H_0 : a amostra vem de uma população cuja densidade é $q(\cdot)$;

H_1 : a amostra vem de uma população cuja densidade é $p(\cdot)$.

O lema de Neyman-Pearson tem como princípio fixar a probabilidade do erro tipo I (α) e maximizar o poder do teste ($1-\beta$). Uma forma alternativa do Lema de Neyman-Pearson é fixar o poder do teste e minimizar a probabilidade do erro tipo I. Logo, pelo Lema de Neyman-Pearson, o teste de hipótese ótimo, o qual minimiza α_n (mantendo $1-\beta$ fixo), é da forma:

rejeita-se H_0 se $\frac{1}{n}\sum_{i=1}^{n} \ln \frac{p(X_i)}{q(X_i)} < k_n$;

não rejeita-se H_0 se $\frac{1}{n}\sum_{i=1}^{n} \ln \frac{p(X_i)}{q(X_i)} \geq k_n$,

onde k_n é escolhido de tal maneira que

$$1-\beta \equiv P\left(\frac{1}{n}\sum_{i=1}^{n} \ln \frac{p(X_i)}{q(X_i)} < k_n \quad | \quad H_1 \text{ é verdadeira}\right).$$

Uma vez que $\ln \frac{p(X_1)}{q(X_1)}, \cdots, \ln \frac{p(X_n)}{q(X_n)}$ são v.a's i.i.d. com esperança $m_1 = \int_{\mathbf{R}} p(x) \ln \frac{p(x)}{q(x)}\, dx$, sob H_1, e esperança $m_0 = \int_{\mathbf{R}} q(x) \ln \frac{p(x)}{q(x)}\, dx$, sob H_0, então pela Lei dos Grandes Números

$$(3.3.3) \qquad \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} \ln \frac{p(X_i)}{q(X_i)} - m_0\right| > \epsilon \quad | \quad H_0\right) = 0, \quad \forall \epsilon > 0, \qquad \text{e}$$

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} \ln \frac{p(X_i)}{q(X_i)} - m_1\right| > \epsilon \quad | \quad H_1\right) = 0, \quad \forall \epsilon > 0.$$

Dado $n \in \mathbb{N}^*$, denotemos o erro tipo I por

$$\alpha_n \equiv P\left(\frac{1}{n}\sum_{i=1}^{n}\ln\frac{p(X_i)}{q(X_i)} < k_n \mid H_0\right).$$

Por (3.3.3), se $k_n < m_0$ então $\lim_{n\to\infty}\alpha_n = 0$, levando-nos a concluir que o evento $\left[\frac{1}{n}\sum_{i=1}^{n}\ln\frac{p(X_i)}{q(X_i)} < k_n \mid H_0\right]$ é uma situação de Grande Desvio em relação à esperança m_0. Portanto, pelo lema de Stein (ver Bucklew [5], pág. 93)

$$\lim_{n\to\infty}\frac{1}{n}\ln\alpha_n = -K(q,p), \tag{3.3.4}$$

sendo $K(q,p) = \int_{\mathbb{R}} p(x)\ln\frac{p(x)}{q(x)}\,dx$ (ver a Definição 3.1.1).

Além da informação obtida por (3.3.3), (3.3.4) nos fornece uma estimativa de $\ln\alpha_n$ e também uma idéia da velocidade de convergência para $-\infty$ quando $n\to\infty$.

Aplicação 3.3.3. Considere o problema de discernir entre duas Cadeias de Markov com espaço de estados $S = \{1,\cdots,N\}$, isto é, quer-se discernir entre as hipóteses:

H_0 : $q(\cdot|\cdot)$, $q(\cdot)$;

H_1 : $p(\cdot|\cdot)$, $p(\cdot)$,

onde $p(\cdot|\cdot)$, $p(\cdot)$ e $q(\cdot|\cdot)$, $q(\cdot)$ são as probabilidades de transição e a distribuição inicial sob H_1 e H_0, respectivamente.

O teste da razão de verossimilhança tem a seguinte forma:

rejeita-se H_0 se $T_n(X_1,\cdots,X_n) < k$;

não rejeita-se H_0 se $T_n(X_1,\cdots,X_n) \geq k$,

onde $T_n(X_1,\cdots,X_n) = \ln\frac{p(X_1)\prod_{i=1}^{n-1}p(X_{i+1}|X_i)}{q(X_1)\prod_{i=1}^{n-1}q(X_{i+1}|X_i)}$. Mas

$$T_n(X_1,\cdots,X_n) = \ln\frac{p(X_1)}{q(X_1)} + \sum_{i=1}^{n-1}\ln\frac{p(X_{i+1}|X_i)}{q(X_{i+1}|X_i)},$$

e como $\frac{p(X_1)}{q(X_1)}$ é limitado e independe de n então podemos considerar a estatística $\bar{T}_n(X_1,\cdots,X_n) \equiv \sum_{i=1}^{n-1}\ln\frac{p(X_{i+1}|X_i)}{q(X_{i+1}|X_i)}$ para calcular os erros tipo I e II.

Denotemos por

$$\alpha_n \equiv P\left(\bar{T}_n(X_1,\cdots,X_n) < k \quad | \quad H_0\right), \quad k < m_0 = E\left(\ln \frac{p(X_{i+1}|X_i)}{q(X_{i+1}|X_i)} \quad | \quad H_0\right).$$

como o erro tipo I e por

$$\beta_n \equiv P\left(\bar{T}_n(X_1,\cdots,X_n) \geq k \quad | \quad H_1\right), \quad k > m_1 = E\left(\ln \frac{p(X_{i+1}|X_i)}{q(X_{i+1}|X_i)} \quad | \quad H_1\right)$$

o erro tipo II. Segue que (segundo Bucklew [5], pág. 95)

$$\text{(3.3.5)} \qquad \lim_{n\to\infty} \frac{1}{n} \ln \alpha_n = -\sup_{t\in\mathbb{R}}\{t\,k - \ln\phi_0(t)\},$$

onde $\phi_0(t)$ é o maior autovalor de $M^T Q$ (M^T indica a matriz transposta de M), com $M = \left\{\left(\frac{p(i|j)}{q(i|j)}\right)^t\right\}_{i,j=1}^N$ e $Q = \{q(i|j)\}_{i,j=1}^N$. No caso do erro tipo II,

$$\text{(3.3.6)} \qquad \lim_{n\to\infty} \frac{1}{n} \ln \beta_n = -\sup_{t\in\mathbb{R}}\{t\,k - \ln\phi_1(t)\},$$

onde $\phi_1(t)$ é o maior autovalor de $M^T P$ e $P = \{p(i|j)\}_{i,j=1}^N$.

A interpretação de (3.3.5) e (3.3.6) está relacionada com a taxa de decaimento exponencial para zero dos erros tipo I e tipo II, respectivamente (ver a interpretação de (0.1) na Introdução).

APÊNDICE 1

RESULTADOS DE PROBABILIDADE E PROCESSOS ESTOCÁSTICOS

Definição A.1.1. *Seja Ω um conjunto e $\mathcal{A}$ uma $\sigma -$ álgebra em Ω. Denominamos de espaço mensurável ao par $(\Omega, \mathcal{A})$.*

Definição A.1.2. *Seja $(\Omega, \mathcal{A})$ e $(\Omega', \mathcal{A}')$ espaços mensuráveis. A aplicação $\psi : \Omega \longrightarrow \Omega'$ é dita $\mathcal{A} - \mathcal{A}'$ mensurável se*

$$\psi^{-1}(A') = \left\{\omega \in \Omega : \psi(\omega) \in \mathcal{A}'\right\} \in \mathcal{A}, \text{para} \forall A' \in \mathcal{A}'.$$

Quando $\Omega' = \mathbf{R}$ e $\mathcal{A}' = \mathcal{B}(\mathbf{R})$ dizemos que ψ é $\mathcal{A}-$mensurável. Se também $\Omega = \mathbf{R}$ e $\mathcal{A} = \mathcal{B}(\mathbf{R})$ então diremos simplesmente que ψ é mensurável.

Definição A.1.3. *Seja $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade. Dizemos que $X : \Omega \longrightarrow \mathbf{R}$ é variável aleatória neste espaço se X é $\mathcal{A}-$mensurável, ou seja, $X^{-1}(B) \in \mathcal{A}$ para $\forall B \in \mathcal{B}(\mathbf{R})$.*

Desigualdade de Markov. *Seja X uma v.a. definida no espaço de probabilidade $(\Omega, \mathcal{A}, P)$. Seja ψ uma função mensurável não negativa neste espaço. Se $E(\psi(X))$ for finita, então para qualquer $c > 0$*

$$P(\psi(X) \geq c) \leq \frac{E(\psi(X))}{c}. \tag{A.1.1}$$

Prova: Seja $Z = c\, I_{[\psi(X) \geq c]}(\omega)$, onde

$$I_{[\psi(X)\geq c]}(\omega) = \begin{cases} 1 & \text{se } \psi(X(\omega)) \geq c \\ 0 & \text{caso contrário} \end{cases}.$$

Então Z é v.a. discreta e $0 \leq Z \leq \psi(X)$. Logo, $E(Z) \leq E(\psi(X))$. Sendo

$$E(Z) = 0P(Z = 0) + c\,P(Z = c) = c\,P(\psi(X) \geq c)$$

vem que $c\,P(\psi(X) \geq c) \leq E(\psi(X))$.

□

Desigualdade de Chebychev. *Seja X v.a. definida no espaço de probabilidade $(\Omega, \mathcal{A}, P)$. Se X tem variância finita, então para qualquer real $\alpha > 0$,*

$$P(|X - EX| \geq \alpha) \leq \frac{VarX}{\alpha^2}.$$

Prova: $P(|X-EX|\geq \alpha)=P(|X-EX|^2\geq \alpha^2)$. Tomando $\psi(X)=|X-EX|^2$ e $c=\alpha^2$, temos por (A.1.1)

$$P(|X-EX|\geq \alpha)\leq \frac{VarX}{\alpha^2}.$$

□

Desigualdade de Jensen. *Seja X v.a. tal que $E(X)$ é finita e $\psi(X):\mathbf{R}\longrightarrow\mathbf{R}$ é função mensurável. Se ψ é convexa então*

$$E(\psi(X))\geq \psi(EX)$$

e se ψ é côncava

$$E(\psi(X))\leq \psi(EX).$$

Prova: James **[16]**.

Desigualdade de Hölder. *Sejam X,Y v.a's tais que $E(|X|^p)<+\infty$, $E(|Y|^q)<+\infty$ e $p,q\in(1,+\infty)$ satisfazendo $\frac{1}{p}+\frac{1}{q}=1$. Então $E(|XY|)$ existe e*

$$E(|XY|)\leq (E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}.$$

Prova: Billingsley **[4]**.

Definição A.1.4. *A função geradora de momentos $M_X:\mathbf{R}\longrightarrow(0,+\infty]$ para a v.a. X é definida como*

$$M_X(t)=E\left(e^{tX}\right),\quad t\in\mathbf{R}.$$

Proposição A.1.1. *A função geradora de momentos $M_X(t)$ tem as seguintes propriedades:*

(i) $M_X(0)=1$;

(ii) Seja X uma v.a. Então a função geradora de momentos de $Y=aX+b$, com $a,b\in\mathbf{R}$, é

$$M_Y(t)=e^{tb}M_X(at);$$

(iii) Se X,Y são v.a's independentes com funções geradoras $M_X(t)$ e $M_Y(t)$, respectivamente, então

$$M_{X+Y}(t) = M_X(t)M_Y(t);$$

(iv) Sejam $X_1, X_2, \cdots, X_n$ v.a's i.i.d. com função geradora de momentos comum $M_X(t)$. A função geradora de $S_n = \sum_{i=1}^n X_i$ é

$$M_{S_n}(t) = (M_X(t))^n;$$

(v) Se $M_X(t) < +\infty$ para $t \in (-\delta, \delta)$, $\delta > 0$, então $M_X(t)$ é de classe $C^\infty((-\delta, \delta))$.

(vi) Se $M_X(t) < +\infty$ para $t \in (-\delta, \delta)$, $\delta > 0$, então $M_X(t) = \sum_{k=0}^\infty \frac{t^k EX^k}{k!}$. Além disso, a n-ésima derivada de $M_X(t)$ aplicada em $t = 0$ dá o momento ordinário de ordem n, ou seja,

$$\left.\frac{d^n}{dt^n} M_X(t)\right|_{t=0} = EX^n.$$

Prova: Billingsley [4].

Proposição A.1.2. *A função geradora de momentos $M(t)$ é convexa.*

Prova: Suponha que $M(t) < +\infty$ para $t \in (-\delta, \delta)$, $\delta > 0$. Dividiremos a prova nos seguintes casos:

(1º) $t_1, t_2 \in (-\delta, \delta)$ e $\alpha \in [0, 1]$:

Provaremos que $\ln M(t)$ é convexa. Pela definição de função geradora de momentos temos que

$$M(\alpha t_1 + (1-\alpha)t_2) = E\left(\exp\{\alpha t_1 X + (1-\alpha)t_2 X\}\right)$$
$$M(\alpha t_1 X) = E\left(\exp\{\alpha t_1 X\}\right)$$
$$M((1-\alpha)t_2) = E\left(\exp\{(1-\alpha)t_2 X\}\right)$$

Aplicando a desigualdade de Hölder com $\frac{1}{p} = \alpha$, $\frac{1}{q} = 1 - \alpha$, $\exp\{\alpha t_1 X\}$ e $\exp\{(1-\alpha)t_2 X\}$ temos

$$\ln E\left(\exp\{\alpha t_1 X + (1-\alpha)t_2 X\}\right) \leq \alpha \ln E\left(\exp\{t_1 X\}\right) + (1-\alpha)\ln E\left(\exp\{t_2 X\}\right).$$

Portanto, $M(t)$ é convexa para $t \in (-\delta, \delta)$.

(2º) $t_1 < -\delta$ e $t_2 < -\delta$, $\alpha \in [0, 1]$:

$M(\alpha t_1 + (1-\alpha)t_2) = +\infty$, pois $\alpha t_1 + (1-\alpha)t_2 \in (-\infty, -\delta]$, e $M(t_1) = M(t_2) = +\infty$. Logo,

$$M(\alpha t_1 + (1-\alpha)t_2) = \alpha M(t_1) + (1-\alpha)M(t_2) = +\infty.$$

(3º) $t_1 > \delta$ e $t_2 > \delta$, $\alpha \in [0, 1]$:

Analogamente a (2º) $M(\alpha t_1 + (1-\alpha)t_2) = \alpha M(t_1) + (1-\alpha)M(t_2) = +\infty$.

(4º) $t_1 < -\delta$, $t_2 > \delta$ e $\alpha \in [0, 1]$:

Se $\alpha t_1 + (1-\alpha)t_2 \in (-\delta, \delta)$ então estaremos diante do (1º) caso, e se $\alpha t_1 + (1-\alpha)t_2 \notin (-\delta, \delta)$ então estaremos nos casos (2º) ou (3º).

□

Teorema A.1.1 . *Seja X v.a. tal que $P(X < 0) > 0$ e $P(X > 0) > 0$. Então existem A, B reais positivos tais que $M(t) \geq A\, e^{B|t|}$, e conseqüentemente $\lim_{|t| \to \infty} M(t) = +\infty$.*

Prova: Pelo fato $P(X > 0) > 0$, $\exists\, a \in (0, +\infty)$ tal que $P(X \geq a) > 0$. Tomando $t \geq 0$, $P(X \geq a) = P(e^{tX} \geq e^{ta})$. Por (A.1.1), com $\psi(X) = e^{tX}$ e $c = e^{ta}$,

$$e^{ta} P(e^{tX} \geq e^{ta}) \leq E\left(e^{tX}\right) = M(t).$$

Denotando $A = P(X \geq a)$ e $B = |a|$, vem que $Ae^{B|t|} \leq M(t)$.

De $P(X < 0) > 0$, tem-se que $\exists\, a \in (-\infty, 0)$ tal que $P(X \leq a) > 0$. Aplicando novamente (A.1.1), com $t < 0$,

$$P(X \leq a) = P(e^{tX} \geq e^{ta}) \leq \frac{M(t)}{e^{ta}}.$$

Observando que $t\,a = |t|\,|a|$, então $Ae^{B|t|} \leq M(t)$, onde $A = P(X \leq a)$ e $B = |a|$. Finalmente $\lim_{|t| \to \infty} Ae^{B|t|} = +\infty$, e assim, $\lim_{|t| \to \infty} M(t) = +\infty$.

□

Definição A.1.5. *Seja $X, X_1, X_2, \cdots$ v.a's definidas em um mesmo espaço de probabilidade. A seqüência $\{X_n\}_{n \geq 1}$ converge para X em probabilidade se para todo $\epsilon > 0$*

$$P(|X_n - X| \geq \epsilon) \longrightarrow 0, \quad \text{quando} \quad n \to \infty.$$

Definição A.1.6. *Seja $X, X_1, X_2, \cdots$ v.a's definidas em um mesmo espaço de probabilidade. A seqüência $\{X_n\}_{n \geq 1}$ converge para X quase certamente se*

$$P(\lim_{n \to \infty} X_n(w) = X(w)) = 1.$$

Definição A.1.7. *Sejam $X, X_1, X_2, \cdots$ v.a's com respectivas funções de distribuição $F, F_1, F_2, \cdots$. Dizemos que X_n converge em distribuição para X se $\lim_{n\to\infty} F_n(x) = F(x)$, para todo x ponto de continuidade de F.*

Teorema A.1.2. *Considere a seqüência $\{X_n\}_{n\geq 1}$ de v.a's i.i.d. integráveis com $EX_n = m$. Então $\frac{S_n}{n}$ converge para m quase certamente.*

Prova: James [17].

Teorema A.1.3. *Nas mesmas hipóteses do Teorema A.1.2, $\frac{S_n}{n}$ converge para m em probabilidade.*

Prova: É conseqüência do Teorema A.1.2 (ver James [17]).

Teorema A.1.4 (Teorema Central do Limite). *Seja $\{X_n\}_{n\geq 1}$ seqüência de v.a's i.i.d., com esperança m e variância σ^2 finita e não nula. Então a função de distribuição de $\sqrt{n}\left(\frac{\bar{X}_n - m}{\sigma}\right)$ converge em distribuição para a função de distribuição de uma normal padrão.*

Prova: Billingsley [4] ou James [17].

Definição A.1.8. *Sejam os espaços de medida $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d), \mu)$ e $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d), \nu)$. Dizemos que a medida μ é absolutamente contínua em relação à ν, denotando-se $\mu \ll \nu$, se $\mu(A) = 0$ para todo $A \in \mathcal{B}(\mathbf{R}^d)$ no qual $\nu(A) = 0$.*

Definição A.1.9. *Uma Cadeia de Markov $\{X_n\}_{n\geq 1}$ com espaço de estados $S = \{1, \cdots, N\}$ e matriz de transição $\mathbf{P} = \{p(i,j)\}_{i,j=1}^N$ é dita irredutível se para qualquer par (i,j), com $i, j \in S$, existe $m = m(i,j) \in \mathbf{N}^*$ tal que $p^m(i,j) > 0$, sendo $p^m(i,j)$ a entrada (i,j) da m-ésima potência da matriz $\mathbf{P}$.*

Teorema A.1.5 (Teorema Ergódico). *Seja $\{X_n\}_{n\geq 1}$ uma Cadeia de Markov, com espaço de estados $S = \{1, \cdots, N\}$ e matriz de transição $\mathbf{P} = \{p(i,j)\}_{i,j=1}^N$, irredutível e aperiódica. Então:*

(i) $\lim_{n\to\infty} p^n(i,j) = \lim_{n\to\infty} p^n(j,j) = \pi(j), \quad \forall\, j \in S$;

(ii) $\pi = (\pi(1), \cdots, \pi(N))$ é tal que $\pi(j) > 0$, $\pi(j) = \sum_{i\in S} \pi(i)$ e $\sum_{j\in S} \pi(j) = 1$;

(iii) a distribuição de probabilidade π, em (i) e (ii), é única.

Prova: Karlin & Taylor [18].

Definição A.1.10. *Denomina-se de medida invariante de uma Cadeia de Markov a distribuição de probabilidade π satisfazendo (i) e (ii) do Teorema A.1.5.*

Teorema A.1.6 (Lei dos Grandes Números para Cadeias de Markov). *Considere $\{X_n\}_{n\geq 1}$ uma Cadeia de Markov irredutível e aperiódica, com espaço de estados $S = \{1, \cdots, N\}$ e matriz de transição $\mathbf{P} = \{p(i,j)\}_{i,j=1}^{N}$. Seja*

$$L_n = (L_n(1), \cdots, L_n(N)), \quad \text{onde } L_n(i) = \frac{1}{n}\sum_{i=1}^{N} \delta_{X_k}(i) \text{ e } i \in S,$$

a medida empírica da Cadeia de Markov e $\pi = (\pi(1), \cdots, \pi(N))$ a medida invariante. Então, para qualquer distribuição inicial $\alpha = (\alpha(1), \cdots, \alpha(N))$, tem-se:

(i) $\lim_{n\to\infty} E_j\left(L_n(i)\right) = \pi(i)$ $i \in S$, onde E_j denota a esperança condicionada ao estado inicial $j \in S$;

(ii) para qualquer $\epsilon > 0$, $\lim_{n\to\infty} P_j\left(|L_n(i) - \pi(i)| > \epsilon\right) = 0, \quad \forall\, i, j \in S$.

Prova: Kemeny & Snell **[19]**.

APÊNDICE 2

RESULTADOS DE ANÁLISE REAL

Definição A.2.1. *Seja $I \subset \mathbf{R}$. A função $f : I \longrightarrow \mathbf{R}$ é dita semicontínua inferiormente (s.c.i.) em $a \in I$ se para $\forall \epsilon > 0$, $\exists \delta > 0$, $x \in I, |x - a| < \delta$ tal que $f(a) - \epsilon < f(x)$.*

Proposição A.2.1. *A função $f : I \longrightarrow \mathbf{R}$, com $I \subset \mathbf{R}$, é semicontínua inferiormente em $a \in I$ se e somente se para qualquer seqüência $\{x_n\}_{n\geq 1}$ em I, onde $x_n \longrightarrow a$, $f(a) \leq \underline{\lim}_{n\to\infty} f(x_n)$.*

Prova: Suponha f s.c.i.. Pela Definição A.2.1, para $\forall \epsilon > 0$, $\exists \delta > 0$, $x \in I$, $|x - a| < \delta$ tal que $f(a) - \epsilon < f(x)$. Seja $\{x_n\}_{n\geq 1} \in I$ uma seqüência qualquer tal que $x_n \longrightarrow a$. Então, $\forall \gamma > 0$, $\exists\, n_0 \in \mathbf{N}$ tal que se $n \geq n_0$, $|x_n - a| < \gamma$. Tome $\gamma = \delta$. Segue que $\forall n \geq n_0, f(a) < f(x_n) + \epsilon$, e portanto, $f(a) \leq \underline{\lim}_{n\to\infty} f(x_n)$. Para provar a recíproca, suponha que $\exists \epsilon > 0$ tal que $\forall n \in \mathbf{N}, \exists x_n \in I$, $|x_n - a| < \frac{1}{n}$ e $f(a) \geq f(x_n) + \epsilon$. Então

$$f(a) \geq \underline{\lim}_{n\to\infty} f(x_n) + \epsilon \geq \underline{\lim}_{n\to\infty} f(x_n),$$

ou seja, chega-se a uma contradição da Definição A.2.1.

□

Definição A.2.2. *Uma função é dita convexa no intervalo I se e somente se*

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)\,,$$

para todo $x, y \in I$ e todo real $\alpha \in [0, 1]$.

Teorema A.2.1. *Seja $f : I \longrightarrow \mathbf{R}$ convexa. Então existem as derivadas laterais $f'_+(c)$ e $f'_-(c)$ em todo $c \in I^\circ$.*

Prova: Acker & Dickstein [1] ou Lima [21].

Corolário A.2.1. *Uma função convexa $f : I \longrightarrow \mathbf{R}$ é contínua no interior de I.*

Prova: Corolário do Teorema A.2.1 (ver Lima [21]).

Outras propriedades das funções convexas podem ser consultadas nas referencias [1] e [21].

Teorema A.2.2 (Dini). *Se a seqüência $\{f_n\}_{n\geq 1}$ de funções contínuas, com $f_n : I \longrightarrow \mathbf{R}$ e I compacto, converge monotonicamente para a função $f : I \longrightarrow \mathbf{R}$ então a convergência é uniforme.*

Prova: Lima [21].

Lema A.2.1. *Para $A > 0$ fixo e y variando em $I \subset \mathbb{R}$, onde I é um intervalo limitado fixo, existe uma constante real $K > 0$ tal que*

$$\left(\frac{v}{Av + y}\right)^{Av+y} \geq K\left(\frac{v}{Av}\right)^{Av}, \quad \text{para todo } v \geq 1 \text{ com } Av + y \geq 1.$$

Prova: Quer-se mostrar que existe uma constante $K > 0$ tal que $\left(\frac{v}{Av+y}\right)^{Av+y} A^{Av} \geq K$, $v \geq 1$ e $Av+y \geq 1$.

Vamos supor que $\bar{I} = [a, b]$. Façamos a mudança de variável $u = Av + y$. Então $\left(\frac{v}{Av+y}\right)^{Av+y} A^{Av} = \left(\frac{u-y}{u}\right)^u A^{-y}$, com $u \geq 1$ e $\frac{u-y}{A} \geq 1$.

Se $A < 1$ então uma cota inferior para A^{-y} é A^{-a}. No caso de $A \geq 1$, a cota inferior será A^{-b}.

A fim de encontrarmos uma cota inferior para $f(u, y) = \left(\frac{u-y}{u}\right)^u$, na região $S = \left\{(u, y) : u \geq 1, \frac{u-y}{A} \geq 1, y \in [a, b]\right\}$, vamos considerar os seguintes casos:

(1º) $A > 1$ e $a \leq 1 - A$:

Observe que neste caso $u \geq 1$, se $a \leq y \leq 1 - A$, e $u \geq A + y$, se $1 - A < y \leq b$.

Para $y \in [a, 1 - A]$ fixo, $f(u, y)$ é crescente em u. Logo, $\min_{u \geq 1} f(u, y) = 1 - y$, e portanto, $f(u, y) \geq 1 - y \geq A$. Se $a \leq y \leq b \leq 1 - A$ então também teríamos A como cota inferior para $f(u, y)$.

Para $y > 1 - A$ temos que $u \geq A + y$. Assim, $\min_{u \geq A+y} f(u, y) = \left(\frac{A}{A+y}\right)^{A+y}$. Portanto, quando $y > 1 - A$, $f(u, y) \geq \left(\frac{A}{A+y}\right)^{A+y}$ para $\forall u \geq A + y$. Uma vez que $g(y) = \left(\frac{A}{A+y}\right)^{A+y}$ tem um único ponto crítico, o qual é um máximo, então o mínimo de g está nos extremos do intervalo $[1 - A, b]$. Logo, $f(u, y)A^{-y} \geq K$, $(u, y) \in S$, sendo

$$K = \min\left\{A, \left(\frac{A}{A+b}\right)^{A+b}\right\} A^{-b}.$$

(2º) $A > 1$ e $1 - A < a$:

Para $a \leq y \leq b$, temos $u \geq A + y$, e portanto, $\min_{u \geq A+y} f(u, y) = \left(\frac{A}{A+y}\right)^{A+y}$. Como o único ponto crítico de $g(y)$ é um ponto de máximo, então

$$\min_{a \leq y \leq b} g(y) = \min\left\{\left(\frac{A}{A+a}\right)^{A+a}, \left(\frac{A}{A+b}\right)^{A+b}\right\}.$$

Assim,

$$K = \min\left\{\left(\frac{A}{A+a}\right)^{A+a}, \left(\frac{A}{A+b}\right)^{A+b}\right\} A^{-b}.$$

(3º) $0 < A \leq 1$ e $a \leq 1 - A$:

Neste caso, verifica-se que $f(u,y) \geq \left(\frac{A}{A+b}\right)^{A+b}$, $(u,y) \in S$. Logo, $K = \left(\frac{A}{A+b}\right)^{A+b} A^{-a}$.

(4º) $0 < A \leq 1$ e $1 - A < a$:

Nesta situação $g(y)$ é decrescente em $1 - A < a \leq y \leq b$. Logo, $K = \left(\frac{A}{A+b}\right)^{A+b} A^{-a}$.

□

Lema A.2.2. *Sejam x_i, p_i reais, onde $i \in \{1, \cdots, r\}$, tais que $p_i > 0$ e $\min_i\{x_i\} < 0 < \max_i\{x_i\}$. Defina $m(t) \equiv \sum_{i=1}^{r} p_i e^{tx_i}$, $t \in \mathbf{R}$, e $\alpha \equiv \inf_{t \in \mathbf{R}} m(t)$. Então existem $c > 0$, $N \in \mathbf{N}^*$, tais que para $n \geq N$, pode-se obter números $n_1, \cdots, n_r$ inteiros positivos tais que:*

(A.2.1)
$$\sum_{i=1}^{r} n_i = n;$$

(A.2.2)
$$\sum_{i=1}^{r} n_i x_i \leq 0;$$

(A.2.3)
$$p(n_1, \cdots, n_r) \equiv n! \prod_{i=1}^{r} \frac{p_i{}^{n_i}}{n_i!} \geq c n^{-\frac{r}{2}} \alpha^n.$$

Prova:

Seja $\Gamma(z)$ a função gama para $z \in [0, +\infty)$ e $s(z) \equiv \sqrt{2\pi} z^z z^{\frac{1}{2}} e^{-z}$. A aproximação de Stirling (ver Mitrinović **[25]**) diz que $\lim_{z \to +\infty} \frac{\Gamma(z+1)}{s(z)} = 1$, e quando $n \in \mathbf{N}^*$, $\lim_{n \to \infty} \frac{n!}{s(n)} = 1$, pois $\Gamma(n+1) = n!$.

Dado $n \in \mathbf{N}^*$, $\exists z_1, \cdots, z_r$ reais positivos tais que $n = z_1 + \cdots + z_r$. Sendo n e $z \in \mathbf{R}^+$ suficientemente grandes vale a aproximação de Stirling, isto é,

$$\forall \epsilon > 0, \exists N \in \mathbf{N}^*, \forall n \geq N, (1-\epsilon)s(n) < \Gamma(n+1) < (1+\epsilon)s(n) \text{ e}$$
$$\forall \epsilon > 0, \exists A \in \mathbf{R}^+, \forall z \geq A, (1-\epsilon)s(z) < \Gamma(z+1) < (1+\epsilon)s(z).$$

Logo, $\forall \epsilon > 0$, $\exists N \in \mathbf{N}^*$ e $A \in \mathbf{R}^+$ tal que para $\forall n \geq N$ e $\forall z_i \geq A$, com $n = z_1 + \cdots + z_r$, tem-se

$$(1-\epsilon)s(n) < \Gamma(n+1) < (1+\epsilon)s(n) \text{ e } (1-\epsilon)s(z_i) < \Gamma(z_i+1) < (1+\epsilon)s(z_i).$$

Segue que

$$p(z_1, \cdots, z_r) = \Gamma(n+1)\prod_{i=1}^{r} \frac{p_i^{z_i}}{\Gamma(z_i+1)} > (1-\epsilon)s(n)\prod_{i=1}^{r} \frac{p_i^{z_i}}{(1+\epsilon)s(z_i)} = \frac{(1-\epsilon)}{(1+\epsilon)^r}\frac{n^{\frac{1}{2}}}{(\sqrt{2\pi})^{r-1}}\prod_{i=1}^{r}\frac{(np_i)^{z_i}}{z_i^{z_i} z_i^{\frac{1}{2}}}.$$

Como $n^{\frac{1}{2}} \geq z_i^{\frac{1}{2}}$, visto que $z_i \leq n$ para todo $i \in \{1, \cdots, r\}$, então

$$\frac{(1-\epsilon)}{(1+\epsilon)^r}\frac{n^{\frac{1}{2}}}{(\sqrt{2\pi})^{r-1}}\prod_{i=1}^{r}\frac{(np_i)^{z_i}}{z_i^{z_i} z_i^{\frac{1}{2}}} \geq \frac{(1-\epsilon)}{(1+\epsilon)^r}\frac{n^{\frac{1}{2}}}{(\sqrt{2\pi})^{r-1}}\frac{1}{n^{\frac{r}{2}}}\prod_{i=1}^{r}\left(\frac{np_i}{z_i}\right)^{z_i}.$$

Uma vez que $\frac{(1-\epsilon)}{(1+\epsilon)^r}\frac{n^{\frac{1}{2}}}{(\sqrt{2\pi})^{r-1}} \geq 1$ para n suficientemente grande e ϵ é arbitrário, tem-se $p(z_1, \cdots, z_r) \geq \frac{1}{n^{\frac{r}{2}}}\prod_{i=1}^{r}\left(\frac{np_i}{z_i}\right)^{z_i}$.

Defina $q(z_1, \cdots, z_r) \equiv \frac{1}{n^{\frac{r}{2}}}\prod_{i=1}^{r}\left(\frac{np_i}{z_i}\right)^{z_i}$. Pela Definição 1.1.1, vemos que $m(t)$ é a transformada de Laplace de uma medida μ concentrada no conjunto finito $\{x_1, \cdots, x_r\}$, sendo $\mu(x_i) = p_i$. Além disso, $m(t)$ é função convexa (a prova é análoga à Proposição A.1.2 do Apêndice 1) e estritamente positiva, pois $\mu(x_i) > 0$ para $i \in \{1, \cdots, r\}$. Pela hipótese $\min_i\{x_i\} < 0 < \max_i\{x_i\}$ temos que $\mu((-\infty, 0]) > 0$ e $\mu([0, +\infty)) > 0$, e portanto, por um argumento análogo à prova do Teorema A.1.1 do Apêndice 1, $m(t)$ converge para $+\infty$ quando $|t| \to \infty$. Das considerações acima segue que $\exists \tau \in \mathbf{R}$ tal que $\alpha = m(\tau) = \sum_{i=1}^{r} p_i e^{\tau x_i}$. Mas por τ ser ponto de mínimo absoluto, $m'(\tau) = 0$.

Através de multiplicadores de Lagrange, mostra-se que existem $z_1, \cdots, z_r$ satisfazendo as restrições $\sum_{i=1}^{r} z_i = n$ e $\sum_{i=1}^{r} z_i x_i = 0$. Seja $g : \mathbf{R}^r \longrightarrow \mathbf{R}$ e $h : \mathbf{R}^r \longrightarrow \mathbf{R}$, onde $g(z_1, \cdots, z_r) = \sum_{i=1}^{r} z_i - n$ e $h(z_1, \cdots, z_r) = \sum_{i=1}^{r} z_i x_i$. Assim, as restrições ficam $g(z_1, \cdots, z_r) = 0$ e $h(z_1, \cdots, z_r) = 0$, e o problema consiste em obter a solução de

$$\frac{\partial q}{\partial z_i} + \lambda_1\frac{\partial g}{\partial z_i} + \lambda_2\frac{\partial h}{\partial z_i} = 0, \quad i = 1, \cdots, r,$$

onde $\lambda_1, \lambda_2 \in \mathbf{R}$. Uma vez que $q(\cdot)$ é não negativa, então podemos substituir por $\ln q(\cdot)$ na equação acima. Assim, o problema fica

$$\frac{\partial \ln q}{\partial z_i} + \lambda_1\frac{\partial g}{\partial z_i} + \lambda_2\frac{\partial h}{\partial z_i} = 0.$$

Como

$$\frac{\partial \ln q}{\partial z_i} = \ln\left(\frac{np_i}{z_i}\right) - 1, \quad \frac{\partial g}{\partial z_i} = 1 \quad \text{e} \quad \frac{\partial h}{\partial z_i} = x_i,$$

tem-se $\ln\left(\frac{np_i}{z_i}\right) - 1 + \lambda_1 + \lambda_2 x_i = 0$ e disso decorre que $z_i = np_i e^{\lambda_1+\lambda_2 x_i - 1}$. Substituindo a expressão obtida para z_i nas equações $g(z_1, \cdots, z_r) = 0$ e $h(z_1, \cdots, z_r) = 0$ obtemos

$$\sum_{i=1}^{r} p_i e^{\lambda_1+\lambda_2 x_i - 1} = 1 \quad \text{e} \quad \sum_{i=1}^{r} x_i p_i e^{\lambda_1+\lambda_2 x_i - 1} = 0.$$

Tomando $\lambda_1 = -\ln\alpha + 1$ e $\lambda_2 = \tau$, tem-se:

$$\sum_{i=1}^{r} p_i e^{\lambda_1+\lambda_2 x_i - 1} = \sum_{i=1}^{r} p_i e^{-\ln\alpha + \tau x_i} = \frac{1}{\alpha}\sum_{i=1}^{r} p_i e^{\tau x_i} = \frac{1}{\alpha} m(\tau) = \frac{\alpha}{\alpha} = 1 \quad \text{e}$$

$$\sum_{i=1}^{r} x_i p_i e^{\lambda_1+\lambda_2 x_i - 1} = \sum_{i=1}^{r} x_i p_i e^{-\ln\alpha + \tau x_i} = \frac{1}{\alpha}\sum_{i=1}^{r} x_i p_i e^{\tau x_i} = \frac{1}{\alpha} m'(\tau) = 0.$$

Logo, concluimos que $z_i = \frac{np_i}{\alpha} e^{\tau x_i}$ é a solução da equação $\frac{\partial q}{\partial z_i} + \lambda_1 \frac{\partial g}{\partial z_i} + \lambda_2 \frac{\partial h}{\partial z_i} = 0$ com as restrições $g(z_1, \cdots, z_r) = 0$ e $h(z_1, \cdots, z_r) = 0$. Substituindo a solução em $q(\cdot)$ obtemos, após algumas simplificações, $q(z_1, \cdots, z_r) = \frac{1}{n^{r/2}} \alpha^n$.

Sem perda de generalidade suponha $x_1 = \min_i x_i < 0 < \max_i x_i = x_r$. Defina $n_j = [z_j]$, $j = 2, \cdots, r$, sendo $[z_j]$ o maior inteiro de z_j e $n_1 = n - \sum_{i=2}^{r} n_i$. Como n e $\sum_{i=2}^{r} n_i$ são inteiros então n_1 também deve ser um inteiro e além disso $n = \sum_{i=1}^{r} n_i$, ou seja, satisfaz (A.2.1). Para provar (A.2.2) basta que verifiquemos a desigualdade $x_1 n_1 \leq -\sum_{i=2}^{r} x_i n_i$. Pela definição de n_1 e por $x_1 = \min_i\{x_i\}$,

$$x_1 n_1 = x_1 z_1 + x_1 \sum_{i=2}^{r} (z_i - n_i) \leq x_1 z_1 + \sum_{i=2}^{r} x_i (z_i - n_i) = \sum_{i=1}^{r} x_i z_i - \sum_{i=1}^{r} x_i n_i.$$

Lembrando que tínhamos a restrição $\sum_{i=1}^{r} x_i z_i = 0$, (A.2.2) está provada.

Aplicando o Lema A.2.1, com $z_i = \frac{np_i}{\alpha} e^{\tau x_i}$ e $i \in \{1, \cdots, r\}$, temos:

$y = n_i - z_i$, $\; y \in I$;

$I = (-1, r-1)$, pois $-1 < n_i - z_i \leq 0$ se $i = 2, \cdots, r$ e $0 \leq n_1 - z_1 = (z_2 - n_2) + \cdots + (z_r - n_r) < r - 1$;

$A_i = \frac{z_i}{np_i} = \frac{e^{\tau x_i}}{\alpha} > 0$;

$v = np_i \geq 1$ e $A_i v + y = n_i \geq 1$ para n suficientemente grande;

$A_i v = z_i$.

Assim, $\exists K_i > 0$ tal que $\left(\frac{np_i}{n_i}\right)^{n_i} \geq K_i \left(\frac{np_i}{z_i}\right)^{z_i}$ para todo $n \geq N$. Tomando $K = \min_i\{K_i\} > 0$, então $\left(\frac{np_i}{n_i}\right)^{n_i} \geq K \left(\frac{np_i}{z_i}\right)^{z_i}$ para todo $i \in \{1, \cdots, r\}$ e $n \geq N$. Portanto,

$$q(n_1, \cdots, n_r) = \frac{1}{n^{r/2}} \prod_{i=1}^{r} \left(\frac{np_i}{n_i}\right)^{n_i} \geq (K)^r \frac{1}{n^{r/2}} \prod_{i=1}^{r} \left(\frac{np_i}{z_i}\right)^{z_i} = cq(z_1, \cdots, z_r) = c\frac{1}{n^{r/2}}\alpha^n.$$

Como $p(z_1, \cdots, z_r) \geq \frac{1}{n^{r/2}} \prod_{i=1}^{r} \left(\frac{np_i}{z_i}\right)^{z_i} = q(z_1, \cdots, z_r)$ temos $p(n_1, \cdots, n_r) \geq q(n_1, \cdots, n_r) \geq c\frac{1}{n^{r/2}}\alpha^n$.

□

Definição A.2.3. *Uma função $f : I \longrightarrow \mathbb{R}$, com $I \subset \mathbb{R}$, é dita convexa no sentido de Jensen (ou simplesmente J-convexa) se para quaisquer $x, y \in I$ vale a desigualdade*

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x) + f(y)}{2}.$$

Teorema A.2.3. *Suponha f J-convexa no intervalo $I \subset \mathbb{R}$. Para quaisquer pontos $x_1, x_2, \cdots, x_n \in I$ e quaisquer racionais não negativos $r_1, r_2, \cdots, r_n$ temos*

$$f\left(\sum_{i=1}^{n} r_i x_i\right) \leq \sum_{i=1}^{n} r_i f(x_i). \tag{A.2.4}$$

Prova: A prova a seguir é extraída de Mitrinović [25].

(1º) $r_i = \frac{1}{n}$, $i = 1, 2, ..., n$, $n \in \mathbb{N}^*$:

Provaremos por indução. Para $n = 2$, (A.2.4) vale pela própria Definição A.2.3. Suponha que (A.2.4) vale para $n = 2^k$, onde $k \in \mathbb{N}^*$. Para $x_1, x_2, \cdots, x_m \in I$ e $m = 2^{k+1} = 2^k 2 = 2n$, tem-se

$$f\left(\frac{x_1 + x_2 + \cdots + x_m}{m}\right) \leq f\left(\frac{\frac{1}{n}\sum_{j=1}^{n} x_j + \frac{1}{n}\sum_{j=1}^{n} x_{j+n}}{2}\right) \leq \frac{f\left(\frac{1}{n}\sum_{j=1}^{n} x_j\right) + f\left(\frac{1}{n}\sum_{j=1}^{n} x_{j+n}\right)}{2}$$

$$\leq \frac{\sum_{j=1}^{n} f(x_j) + \sum_{j=1}^{n} f(x_{j+n})}{2n} = \frac{\sum_{j=1}^{2n} f(x_j)}{2n}.$$

Assim, (A.2.4) vale para todo natural $n \in \{2, 2^2, 2^3, \cdots\}$. Falta provar que se (A.2.4) vale para $n \geq 2$ então também vale para $n - 1$. Seja $(x_1, x_2, \cdots, x_{n-1}) \in I$ e $x_n = \frac{1}{n-1}(x_1 + x_2 + \cdots + x_{n-1})$. Pelo que provamos acima temos que

$$f\left(\frac{x_1+x_2+\cdots+x_{n-1}+x_n}{n}\right)\leq\frac{f(x_1)+f(x_2)+\cdots+f(x_{n-1})+f(x_n)}{n}.$$

Mas $x_1+x_2+\cdots+x_{n-1}+\left(\frac{x_1+x_2+\cdots+x_{n-1}}{n-1}\right)=\frac{x_1+x_2+\cdots+x_{n-1}}{n-1}$, e portanto,

$$\begin{aligned}
f\left(\frac{x_1+x_2+\cdots+x_{n-1}}{n-1}\right)&\leq\frac{1}{n}\left(f(x_1)+f(x_2)+\cdots+f(x_{n-1})\right)+\frac{1}{n}f\left(\frac{x_1+x_2+\cdots+x_{n-1}}{n-1}\right)\\
\left(\frac{n-1}{n}\right)f\left(\frac{x_1+x_2+\cdots+x_{n-1}}{n-1}\right)&\leq\frac{1}{n}\left(f(x_1)+f(x_2)+\cdots+f(x_{n-1})\right)\\
f\left(\frac{x_1+x_2+\cdots+x_{n-1}}{n-1}\right)&\leq\frac{f(x_1)+f(x_2)+\cdots+f(x_{n-1})}{n-1}.
\end{aligned}$$

(2º) (r_i é racional não negativo:

Uma vez que $r_1,\cdots,r_n$ são racionais não negativos, existe um número natural m e inteiros não negativos $p_1,\cdots,p_n$ tal que $m=p_1+\cdots+p_n$ e $r_i=\frac{p_i}{m}$, $i=1,\cdots,n$. Pelo (1º) caso,

$$\begin{aligned}
f\left(\frac{p_1}{m}x_1+\cdots+\frac{p_n}{m}x_n\right)&=f\left(\frac{(x_1+\cdots+x_1)+\cdots+(x_n+\cdots+x_n)}{m}\right)\\
&\leq\frac{(f(x_1)+\cdots+f(x_1))+\cdots+(f(x_n)+\cdots+f(x_n))}{m}\\
&=\frac{p_1}{m}f(x_1)+\cdots+\frac{p_n}{m}f(x_n).
\end{aligned}$$

□

Teorema A.2.4. *Seja $f:I\longrightarrow\mathbf{R}$ J-convexa, onde I é intervalo contido em $\mathbf{R}$. Se f é semicontínua inferiormente em I então f é convexa.*

Prova: Seja $x,y\in I$ e $\alpha\in[0,1]$. Por f ser s.c.i. em $\alpha x+(1-\alpha)y\in I$, para $\forall\epsilon>0$, $\exists\delta>0$, $z\in I$, $|z-(\alpha x+(1-\alpha)y)|<\delta$ tal que $f(\alpha x+(1-\alpha)y)-\epsilon<f(z)$. Seja $r_1,r_2,\cdots,r_n$ e $s_1,s_2,\cdots,s_m$ racionais não negativos tais que $\sum_{i=1}^n r_i+\sum_{i=1}^m s_i=1$ e $|\alpha x+(1-\alpha)y-(\sum_{i=1}^n r_i)x-(\sum_{i=1}^m s_i)y|<\delta$. Então

$$f(\alpha x+(1-\alpha)y)-\epsilon<f\left(\left(\sum_{i=1}^n r_i\right)x+\left(\sum_{i=1}^m s_i\right)y\right).$$

Pelo Teorema A.2.3,

$$f\left(\left(\sum_{i=1}^{n} r_i\right)x+\left(\sum_{i=1}^{m} s_i\right)y\right) \leq \left(\sum_{i=1}^{n} r_i\right)f(x)+\left(\sum_{i=1}^{m} s_i\right)f(y).$$

Uma vez que para n, m suficientemente grandes podemos tomar $\sum_{i=1}^{n} r_i$ e $\sum_{i=1}^{m} s_i$ arbitrariamente próximos de α e $(1-\alpha)$, respectivamente, e que ϵ é arbitrário, então $f(\alpha x+(1-\alpha)y) \leq \alpha f(x)+(1-\alpha)f(y)$.

□

Teorema A.2.5. *Seja $C \subset \mathbf{R}^d$ convexo e fechado. Para todo $x \in \mathbf{R}^d$ existe um único ponto $p_C(x) \in C$ tal que*

$$\|x-p_C(x)\| \leq \|x-y\|, \quad \forall\, y \in C. \tag{A.2.5}$$

Prova: Acker & Dickstein [1].

Definição A.2.4. *Denomina-se a projeção de x sobre o conjunto $C \subset \mathbf{R}^d$, convexo e fechado, ao ponto $p_C(x) \in C$ satisfazendo (A.2.5).*

Proposição A.2.2. *Se $x \in \mathbf{R}^d$ e $x_0 \in C$, com $C \subset \mathbf{R}^d$ convexo e fechado, então $x_0 = p_C(x)$ se e somente se $\langle x-x_0, y-x_0\rangle \leq 0$, $\forall y \in C$.*

Prova: Acker & Dickstein [1].

REFERÊNCIAS

[1] **Acker, F. & Dickstein, F. (1983)** . *Uma Introdução à Análise Convexa.* 14^{o} Colóquio Brasileiro de Matemática . Instituto de Matemática Pura e Aplicada (IMPA) - CNPq. Poços de Caldas: 14 a 22/7/83.

[2] **Azencott, R. (1980)**. *Grandes Déviations et Applications.* École d'Éte de Probabilités de Saint-Flour VIII - 1978, Lecture Notes in Math., 774, 12-20. Springer-Verlag, Berlin.

[3] **Bahadur, R. R. & Zabell, S. L. (1979)**. *Large Deviations of the sample mean in general vectors spaces.* Ann. Probab. 7, 587-621.

[4] **Billingsley, P. (1986)**. *Probability and Measure* . John Wiley & Sons, 2nd edition, New York.

[5] **Bucklew, J. A. (1990)**. *Large Deviations Techniques in Decision, Simulation and Estimation.* Wiley, New York.

[6] **Cassandro, M. , Galves, A. , Olivieri, E. ,** Vares, **M. E. (1984)** . *Metastable behavior of stochastic dynamical systems: a pathwise approach* . Journal of Statist. Phys., 35, 603-634.

[7] **Chernoff, H (1952)**. *A measure of asymptotic efficiency for tests of a hipothesis based on sums of observations.* Ann. Math. Statist. 23, 493-507.

[8] **Cramér, H. (1937)** . *Sur un nouveau théorème-limite de la théorie des probabilités.* In Actualités Scientifiques et Industrielles, 736, 5-23. Colloque Consacré à la Théorie des Probabilités. Hermann, Paris.

[9] **Dembo, A. & Zeitouni, O. (1993)** . *Large Deviations: Techniques and Applications.* Jones and Bartlett Publishers, London.

[10] **Deuschel, J. D. & D. W. Stroock (1989)** . *Large Deviations.* Academic Press, San Diego.

[11] **Donsker, M. D. & Varadhan, S. R. S. (1975).** *Large Deviations of Markov processes and the asymptotic evaluation of certain Markov processes expectations of large time.* Lecture Notes in Math. 451. Springer-Verlag, Berlin.

[12] **Ellis, R. S. (1985)**. *Entropy, Large Deviations and Statistical Mechanics.* Springer-Verlag. New York.

[13] **Freidlin, M. I. & Wentzell, A. D. (1984)**. *Random Perturbations of Dynamical Systems.* Springer-Verlag, New York.

[14] **Gärtner, J. (1977)**. *On Large Deviations from the invariant measure.* Th. Prob. Appl., 22, 24-39.

[15] **Gärtner, J. (1982)** . *Location for wave fronts for the multidimensional KPP equations and Brownian first exit densities* . Math. Nachr. Vol. 105, 317-351.

[16] **Holley, R. A. & Stroock, D. W. (1976)**. *Applications of the stochastic Ising model to the Gibbs states.* Commun. Math. Phys. 48, 249-265.

[17] **James, B. (1981)** . *Probabilidade: Um Curso em Nível Intermediário* . Instituto de Matemática Pura e Aplicada (IMPA) - CNPq. Rio de Janeiro.

[18] **Karlin, S. & Taylor, H. (1975)**. *A First Course in Stochastic Processes* . Second Edition, Academic Press, New York.

[19] **Kemeny, J. G. & Snell, J. L. (1960)** . *Finite Markov Chains* . Van Nostrand, Princeton, New Jersey.

[20] **Lanford, O. (1971)**. *Entropy and equilibrium states in classical statistical mechanics.* Lecture Notes in Phys. 20, 1-113.

[21] **Lima, E. L. (1982)**. *Curso de Análise* , Vol. 1. Instituto de Matemática Pura e Aplicada (IMPA) - CNPq. Rio de Janeiro.

[22] **Lima, E. L. (1983)** . *Espaços Métricos* . Instituto de Matemática Pura e Aplicada (IMPA) - CNPq. Rio de Janeiro.

[23] **Lima, E. L. (1985)**. *Curso de Análise* , Vol. 2. Instituto de Matemática Pura e Aplicada (IMPA) - CNPq. Rio de Janeiro.

[24] **Mckean, H. P. (1975)** . *Applications of Brownian Motion to the KPP equations* . Comm. Pur. Appl. Math. Vol. 28, 323-331.

[25] **Mitrinović, D. S. (1970)** . *Analytic Inequalities* . Springer-Verlag, Berlin.

[26] **Sanov, I. N. (1961)**. *On the probability of Large Deviations of random variables.* Selected trans. in Math. Statist. and Prob. I, 213-244.

[27] **Strang, G. (1986)**. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, Wellesley.

[28] **Stroock, D. W. (1984)** . *An Introduction to the Theory of Large Deviations.* Springer-Verlag, New York.

[29] **Varadhan, S. R. S. (1984)**. *Large Deviations and Applications* . SIAM, Philadelphia.

[30] **Vares, M. E. (1985)** . *Grandes Desvios em Processos Markovianos* . Instituto de Matemática Pura e Aplicada (IMPA) - CNPq. Poços de Caldas: 22 a 26/7/85.