Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Convergência em Entropia Relativa e Distribuições Estáveis

Alex Barros Azevedo Bomfim

Brasília

2017

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Convergência em Entropia Relativa e Distribuições Estáveis

Alex Barros Azevedo Bomfim

Orientadora: Prof. Dra. **Cátia Regina Gonçalves**

Brasília

2017

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

BB277c

Barros Azevedo Bomfim, Alex
Convergência em Entropia Relativa e Distribuições Estáveis / Alex Barros Azevedo Bomfim; orientador Cátia Regina Gonçalves. -- Brasília, 2017.
79 p.

Dissertação (Mestrado - Mestrado em Matemática) -- Universidade de Brasília, 2017.

1. Entropia. 2. Entropia Relativa. 3. Teorema do limite central. 4. Leis estáveis. I. Regina Gonçalves, Cátia, orient. II. Título.

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Convergência em Entropia Relativa e Distribuições Estáveis

por

Alex Barros Azevedo Bomfim *

Dissertação apresentada ao Departamento de Matemática da Universidade de Brasília, como parte dos requisitos para obtenção do grau de

MESTRE EM MATEMÁTICA

Brasília, 11 de agosto de 2017.

Comissão Examinadora:

C. Gonçalves

Profa. Dra. Cátia Regina Gonçalves MAT/UnB (Orientadora)

Paulo Henrique P. Costa

Prof. Dr. Paulo Henrique Pereira da Costa – MAT/UnB (Membro)

Profa. Dra. Cristiane Nespoli de Oliveira – FCT/UNESP (Membro)

* O autor foi bolsista do CNPq durante a elaboração desta dissertação.

Dedicatória

Dedico meu título de mestre a minha família.

Agradecimentos

Agradeço a minha família pelo amor e por sempre me apoiar em minhas decisões e me dar todo o suporte para conseguir atingir meus objetivos. Tenho muita sorte tê-los por perto, meu pai Sérgio, minha mãe Sandra e meu irmão Daniel.

Agradeço a minha orientadora Cátia pelas sugestões, pela dedicação em ajudar a sanar minhas dúvidas quando precisei, pelo rigor e principalmente pelo trabalho de tornar esta dissertação mais apresentável. Sou muito grato por ter sido seu aluno e orientando.

Agradeço aos professores Paulo Henrique e Cristiane por aceitarem compor a banca examinadora e pelas sugestões para a versão final deste trabalho.

Agradeço a todos os amigos que fiz durante esses dois anos por tornarem o mestrado mais agradável e pela ajuda mútua nas provas e qualificações. Em especial, alguns dos nomes daqueles que estiveram mais presentes: Alancoc, Carol, Fabian, Filipe, Guillermo, Marta, Matheus, Natália, Wállef, Welinton, Wenison, Wilson... Também, agradeço aos meus amigos que conheci na estatística e com quem compartilhei a decisão de fazer mestrado em matemática. Alguns nomes: Agda, Augusto, Geiziane, Guilherme, Quintino, Thiago e Veloso. Finalmente, aos meus amigos que conservo há mais tempo: Gabs, Gustavo, Marcos, Sorriso, entre outros.

Agradeço aos professores e funcionários do departamento Matemática que colaboraram diretamente para minha formação.

Por fim, agradeço à CAPES e ao CNPq pelo auxílio financeiro.

Resumo

Neste trabalho, baseados em Bobkov, Chistyakov e Götze [4], estudamos a convergência em entropia relativa de somas normalizadas de variáveis aleatórias independentes e identicamente distribuídas para uma variável aleatória estável não-extremal, sob a hipótese de convergência fraca.

Palavras-chave: Entropia. Entropia Relativa. Teorema do limite central. Leis estáveis.

Abstract

In this work, based on Bobkov, Chistyakov and Götze [4], we study the convergence in relative entropy of normalized sums of independent identically distributed random variables to a non-extremal stable law, under the hypothesis of weak convergence.

Keywords: Entropy. Relative entropy. Central limit theorem. Stable laws.

Sumário

Lista de Símbolos

f.d.	Função de distribuição
v.a.	Variável aleatória.
i.i.d.	Independentes e identicamente distribuídas.
$\stackrel{\mathfrak{D}}{\to}$	Convergência em distribuição.
$\stackrel{\mathfrak{D}}{=}$	Igualdade em distribuição.
$f(x) \sim g(x)$ quando $x \to a$	$\lim\limits_{x\to a} \frac{f(x)}{g(x)} = 1$.
$f(x) = o(g(x))$ quando $x \to a$	$\lim\limits_{x\to a} \frac{f(x)}{g(x)} = 0$.
$f * g$	Convolução entre f e g.
$\mathbf{E}X$	Esperança da v.a. X.
$\mathrm{Var}X$	Variância da v.a. X.
$\mathbb{R}$	Conjunto dos números reais.
1_A	Função Indicadora no conjunto A.
$\int_A f(x)dx$	Integral de Lebesgue sobre o conjunto $A \subset \mathbb{R}$.
$\int f(x)dx$	Integral de Lebesgue sobre o domínio de f.
$q.c.$	Quase Certamente.
$\mathcal{B}$	Borelianos da reta.

Introdução

As distribuições estáveis são as únicas possíveis distribuições limites de somas normalizadas de variáveis aleatórias independentes e identicamente distribuídas. Especificamente, uma variável aleatória Z é estável se, e somente se, existem uma sequência de variáveis aleatórias $\{X_n\}_{n\geq 1}$ independentes e identicamente distribuídas (i.i.d.), com função de distribuição comum F, e sequências de constantes reais $\{a_n\}_{n\geq 1}$ e $\{b_n\}_{n\geq 1}$, com $b_n > 0$, tais que a sequência de somas parciais normalizadas (ou estabilizadas)

$$Z_n = \frac{X_1 + \ldots + X_n}{b_n} - a_n, \quad n \geq 1 \tag{1}$$

converge em distribuição para Z, ou seja,

$$Z_n \xrightarrow{\mathfrak{D}} Z. \tag{2}$$

Neste caso, dizemos que a função de distribuição F está no *domínio de atração* da função de distribuição de Z.

Dentre as distribuições estáveis, a mais conhecida é a distribuição Normal, e o Teorema do Limite Central, na sua versão mais simples, garante que: se $\mathbf{E}X_1^2 < +\infty$ e $\mathrm{Var}X_1 > 0$, então (2) é válida com $a_n = \dfrac{\sqrt{n}\mathbf{E}X_1}{\sqrt{\mathrm{Var}X_1}}$, $b_n = \sqrt{n\mathrm{Var}X_1}$ e Z com distribuição Normal padrão $N(0,1)$. No caso $b_n = \sqrt{n\mathrm{Var}X_1}$, a condição $\mathbf{E}X_1^2 < +\infty$ também é uma condição necessária para que uma função de distribuição F esteja no domínio de atração da distribuição Normal.

Nas últimas décadas, as distribuições estáveis distintas da Normal têm sido uma alternativa para a modelagem de dados com variabilidade alta e cauda pesada, coletados nas mais diversas áreas, como por exemplo em economia, finanças, hidrologia, telecomunicações, entre muitas outras. Isso porque as distribuições estáveis não normais possuem variância infinita e caudas (à direita e à esquerda) com decaimento regularmente variante, ou seja, como uma função potência.

As taxas de decaimento das caudas das distribuições estáveis dependem do *índice* *de estabilidade* α, $0 < \alpha \leq 2$. Quanto menor for α, mais lento será o decaimento e mais pesadas serão as caudas. Quando $\alpha = 2$, temos a distribuição Normal, que é a única distribuição estável com variância finita e caudas leves. Para $0 < \alpha < 2$, a variância é infinita e as caudas da distribuição são pesadas, com taxas de decaimento tais que quando $x \to \infty$,

$$P(X > x) \sim C_\alpha \frac{1+\beta}{2} \sigma^\alpha x^{-\alpha} \qquad \text{e} \qquad P(X < -x) \sim C_\alpha \frac{1-\beta}{2} \sigma^\alpha x^{-\alpha},$$

onde C_α é uma constante que depende do parâmetro α, $\beta \in [-1, 1]$ é chamado parâmetro de *simetria* ("skewness") e $\sigma \geq 0$ é conhecido como parâmetro de *escala*.

Toda distribuição estável é completamente determinada pelos parâmetros α, β, σ e um quarto parâmetro $\mu \in \mathbb{R}$, chamado parâmetro de *locação*, através da sua função característica dada por

$$f(t) = \exp\{i\mu t - \sigma^2 |t|^\alpha [1 + i\beta \text{sign}(t) \omega(t, \alpha)]\},$$

em que $\text{sign}(t) = \frac{t}{|t|}$, se $t \neq 0$, $\text{sign}(0) = 0$ e

$$w(t, \alpha) = \begin{cases} \tan \frac{\pi\alpha}{2}, & \text{se } \alpha \neq 1, \\ \frac{2}{\pi} \log |t|, & \text{se } \alpha = 1. \end{cases}$$

Quando $\alpha = 2$, temos a distribuição Normal, quando $\alpha = 1$ temos a distribuição de Cauchy e para $\alpha = \frac{1}{2}$ temos a distruibuição de Levy, que são as únicas distribuições estáveis não-degeneradas cujas densidades são conhecidas e podem ser escritas por meio de funções elementares.

O domínio de atração de uma distribuição α-estável, com $0 < \alpha < 2$, é caracterizado como sendo o conjunto de funções de distribuição F cujas caudas têm o seguinte comportamento quando $x \to \infty$:

$$F(-x) \sim (c_0 + o(1)) x^{-\alpha} B(x)$$

e

$$1 - F(x) \sim (c_1 + o(1)) x^{-\alpha} B(x),$$

onde $B(x)$ é uma função lentamente variante no sentido de Karamata e $c_0, c_1 \geq 0$ são constantes reais que não são ambas simultaneamente nulas.

Um estudo detalhado sobre distribuições α-estáveis pode ser encontrado por exemplo em [10] e [19].

Uma questão que vem sendo analisada na literatura é a possibilidade da convergência de Z_n à Z ocorrer em um sentido mais forte, quando $\{Z_n\}_{n \geq 1}$ é uma sequência de somas parciais tal que $Z_n \xrightarrow{\mathfrak{D}} Z$, ou seja, quando a função de distribuição F está no domínio de atração da função de distribuição de uma variável aleatória α-estável Z. Nosso interesse neste trabalho é a convergência em entropia relativa.

Também chamada de *divergência* ou *distância de Kullback-Leibler*, a *entropia relativa* foi introduzida no contexto estatístico por Kullback e Leibler em 1951, [12], como uma medida de discrepância entre duas densidades de probabilidade. Assim, se X e Y são variáveis aleatórias com densidades de probabilidade p e q, respectivamente, então a entropia relativa de X com respeito a Y (ou de p com respeito a q) é definida por

$$D(X||Y) = \int p(x) \log \frac{p(x)}{q(x)} dx, \tag{3}$$

para p e q tais que $q(x) = 0$ implica $p(x) = 0$ quase certamente (com respeito à medida de Lebesgue). Caso contrário, define-se $D(X||Y) = +\infty$. Como (3) depende somente das densidades p e q e não especificamente de X e Y, costuma-se também denotar $D(p||q)$.

Na verdade, Kullback e Leibler [12] aplicaram os princípios da Teoria de Informação de Shannon [18] para o universo da Estatística. O principal conceito da Teoria de Informação, introduzida por Shannon em 1948, é o conceito de *entropia* que, em linhas gerais, mede a quantidade média de informação contida em um sistema. No contexto da Estatística, em particular em teste de hipóteses, Kullback e Leibler adaptaram o conceito de informação e de entropia definidos por Shannon, considerando $\log \frac{p(x)}{q(x)}$ como sendo a informação em um valor observado x de X para a discriminação entre a hipótese nula $H_0 : X$ tem densidade q e a hipótese alternativa $H_1 : X$ tem densidade p e interpretando $D(X||Y)$ como a informação média por amostra para discriminar em favor de H_1 contra H_0, dado que H_1 é verdadeira. Dessa forma, se uma observação vem de p, a entropia relativa $D(p||q)$ mede o risco médio de usar q no lugar de p.

Embora tenhamos $D(p||q) \geq 0$ para quaisquer duas densidades p e q e $D(p||q) = 0$ se, e somente se, $p = q$, a entropia relativa não é uma métrica, pois não satisfaz a propriedade de simetria e nem a desigualdade triangular. Uma síntese das principais propriedades de entropia e de entropia relativa pode ser encontrada em [11].

Uma propriedade de interesse é que a entropia relativa está relacionada com a norma da variação total através da desigualdade de Pinsker

$$||\nu_p - \nu_q||_{VT} \leq \sqrt{2(D(p||q))}, \tag{4}$$

onde ν_p e ν_q são as medidas de probabilidade na reta absolutamente contínuas com densidades p e q, respectivamente, e

$$||\nu_p - \nu_q||_{VT} = 2 \sup\{|\nu_p(A) - \nu_q(A)| : A \in \mathcal{B}\}.$$

Da mesma forma, a desigualdade (4) é válida se considerarmos a norma em L^1, $||p-q||_{L^1} = \int_{\mathbb{R}} |p(x)-q(x)|dx$, no lugar da norma de variação total. Assim, convergência em entropia relativa é uma propriedade mais forte do que convergência na norma de variação total ou na norma em L^1.

Nesse sentido, sob as hipóteses do Teorema do Limite Central Clássico, com $\mathbf{E}X_1 = 0$ e $\mathrm{Var}X_1 = \sigma^2$, $0 < \sigma^2 < +\infty$, e considerando a sequência de somas normalizadas $\{Z_n\}$ em (1), com $a_n = 0$ e $b_n = \sigma\sqrt{n}$, Barron [2], em 1986, demonstrou que uma condição necessária e suficiente para a convergência em entropia relativa, isto é, para que $D(Z_n||N(0,1)) \underset{n\to\infty}{\longrightarrow} 0$, é a finitude de $D(Z_n||N(0,1))$ para algum $n \geq 1$. Assim, o resultado de Barron implica que se $Z_n \overset{\mathfrak{D}}{\to} N(0,1)$, com $b_n = \sigma\sqrt{n}$, então $D(Z_n||N(0,1)) \underset{n\to\infty}{\longrightarrow} 0$ se, e somente se, $D(Z_n||N(0,1)) < +\infty$ para algum $n \geq 1$.

Em 2013, Bobkov, Chistyakov e Götze [4], obtiveram a extensão do resultado de Barron para o caso em que $Z_n \overset{\mathfrak{D}}{\to} Z$ e Z é uma variável aleatória α-estável não-extremal, ou seja, quando Z é normal ($\alpha = 2$) ou quando Z é α-estável com $0 < \alpha < 2$ e $-1 < \beta < 1$.

Segundo Bobkov et al. [4], o caso em que Z é α-extremal, ou seja, $0 < \alpha < 2$ e $|\beta| = 1$, apresenta dificuldades adicionais pois a densidade ψ de Z tem um comportamento diferente do caso não-extremal. Particularmente, se Z é não-extremal então sua densidade ψ é estritamente positiva em toda a reta e quando Z não é normal, sua densidade satisfaz as relações assintóticas quando $x \to \infty$

$$\psi(-x) \sim c_0 x^{-(1+\alpha)} \qquad \text{e} \qquad \psi(x) \sim c_1 x^{-(1+\alpha)},$$

para determinadas constantes $c_0, c_1 > 0$. Este comportamento assintótico no caso não-extremal foi fundamental para passar da convergência fraca para a convergência em entropia relativa. Por outro lado, quando $|\beta| = 1$ e $0 < \alpha < 1$, tem-se que $\psi(x) > 0$ somente sobre um intervalo I do tipo $(-\infty, x_0)$ ou $(x_0, +\infty)$ e quando x tende a x_0, $\psi(x)$ se aproxima de zero de forma muito rápida. Assim, neste caso, segundo Bobkov et al., para garantir a finitude de $D(Z_n||Z)$ são necessárias hipóteses adicionais de tal forma que o suporte da densidade de Z_n esteja contido em I e uma hipótese sobre o comportamento da densidade de X_1 perto de x_0. Uma situação semelhante ocorre quando $1 \leq \alpha < 2$ e $|\beta| = 1$, pois, embora ψ seja estritamente positiva sobre toda a reta, tem-se que $\psi(x)$ converge a zero muito rapidamente quando $x \to +\infty$ ou $x \to -\infty$.

O principal objetivo deste trabalho é estudar em detalhes o artigo de Bobkov et al. [4].

Para isso, no Capítulo 1 apresentamos os conceitos e resultados preliminares necessários para o desenvolvimento do trabalho.

O Capítulo 2 é reservado para a apresentação dos conceitos de entropia e entropia relativa, de suas propriedades básicas e de propriedades específicas de interesse.

Por fim, no Capítulo 3 apresentamos em detalhes os resultados auxiliares e o teorema principal obtidos por Bobkov et al. [4].

Capítulo 1

Preliminares

1.1 Introdução

Neste capítulo, apresentamos conceitos e resultados preliminares que serão úteis para o desenvolvimento dos capítulos seguintes.

Na Seção 1.2, introduzimos o conceito de função de variação lenta e enunciamos o Teorema da Representação de Karamata, que nos permite obter propriedades importantes para as funções de variação lenta.

Na Seção 1.3, apresentamos a definição de convolução entre funções e suas propriedades básicas, entre elas a decomposição binomial para convoluções.

Na Seção 1.4, definimos as variáveis aleatórias estáveis e apresentamos suas principais propriedades. Além dos teoremas mais importantes, apresentamos alguns resultados necessários para os próximos capítulos e definimos as variáveis estáveis não-extremais, que serão o foco deste trabalho.

Dedicamos a Seção 1.5 exclusivamente para apresentar e demonstrar algumas desigualdades elementares que serão utilizadas nos capítulos 2 e 3.

Finalmente, na Seção 1.6 apresentamos algumas propriedades especiais de função característica que serão úteis no desenvolvimento do Capítulo 3.

1.2 Funções de Variação Lenta

Neste trabalho, estamos interessados na classe de distribuições estáveis, cujas propriedades e caracterizações de seus domínios de atração estão relacionadas com as funções de variação regular, conforme veremos na Seção 1.4.

Assim, nesta seção apresentamos as definições de funções de variação regular e lenta e o Teorema de Karamata, que fornece uma caracterização das funções de variação lenta.

As referências bibliográficas desta seção são [17] e [9].

Definição 1.1 *Seja $g : [A, +\infty) \to (0, +\infty)$, onde $A > 0$. Diz-se que g é de variação regular de índice $\alpha \in \mathbb{R}$ se g for mensurável e, para todo $t > 0$,*

$$\lim_{x\to\infty} \frac{g(tx)}{g(x)} = t^{\alpha}.$$

Definição 1.2 *Uma função h de variação regular de índice $\alpha = 0$ é dita ser de variação lenta, isto é, quando h for mensurável e, para todo $t > 0$,*

$$\lim_{x\to\infty} \frac{h(tx)}{h(x)} = 1.$$

É fácil ver que g é uma função de variação regular se, e somente se, $g(x) = x^{\alpha}h(x)$, onde h é uma função de variação lenta. O próximo teorema apresenta uma caracterização das funções de variação lenta. Sua demonstração pode ser encontrada em [17].

Teorema 1.1 *(Teorema da Representação de Karamata). Uma função $h : [A, +\infty) \to (0, +\infty)$, $A > 0$, é de variação lenta se, e somente se, existir $x_0 \geq A$ tal que*

$$h(x) = c(x) \exp\left(\int_{x_0}^{x} \frac{w(y)}{y} dy\right),$$

onde c é uma função limitada mensurável em $[x_0, +\infty)$ tal que

$$\lim_{x\to\infty} c(x) = c \neq 0$$

e w é uma função contínua em $[x_0, +\infty)$ tal que

$$\lim_{x\to\infty} w(x) = 0.$$

Corolário 1.1 *Se h é de variação lenta, então para qualquer $\alpha > 0$, tem-se*

$$\lim_{x\to\infty} x^{\alpha}h(x) = +\infty$$

e

$$\lim_{x\to\infty} x^{-\alpha}h(x) = 0.$$

Note que a última igualdade afirma que, se h é de variação lenta, então $h(x) = o(x^{\alpha})$ quando $x \to \infty$ para todo $\alpha > 0$.

1.3 Convolução de Funções

Para a obtenção de teoremas de limites de somas parcias de variáveis aleatórias independentes, que estudaremos nos próximos capítulos, apresentamos nesta seção algumas propriedades básicas de convolução de funções. Em particular, a densidade da soma de variáveis aleatórias independentes é obtida como a convolução das respectivas densidades dessas variáveis.

Iniciamos com a definição geral de convolução.

Definição 1.3 *Sejam f e g funções reais. A função $f * g$ dada por*

$$(f * g)(x) = \int_{-\infty}^{\infty} f(y)g(x-y)dy \tag{1.1}$$

é chamada a convolução entre f e g desde que a integral em 1.1 exista e seja finita para todo $x \in \mathbb{R}$.

As propriedades de convoluções, a seguir, são de fácil verificação.

Proposição 1.1 *Sejam f, g e h funções reais. As seguintes propriedades são válidas desde que as convoluções abaixo existam:*

(a) $f * g = g * f$ *(Comutatividade);*

(b) $f * (g * h) = (f * g) * h$ *(Associatividade);*

(c) $f * (g + h) = (f * g) + (f * h)$ *(Distributividade);*

(d) $c(f * g) = (cf * g) = f * (cg)$, *onde* $c \in \mathbb{R}$.

Seja f uma função real. Denote por f^{n*} a convolução de f n vezes, ou seja

$$\begin{aligned} f^{1*} &= f \\ f^{2*} &= f * f \\ &\vdots \\ f^{n*} &= f^{(n-1)*} * f, \quad n > 1. \end{aligned}$$

O teorema seguinte fornece a principal justificativa para estudarmos propriedades de convoluções.

Teorema 1.2 *Se X e Y são variáveis aleatórias independentes e absolutamente contínuas com densidades f_X e f_Y, respectivamente, então a sua soma $Z = X + Y$ possui densidade $f_Z = f_X * f_Y$.*

Destacamos a seguir duas propriedades de convoluções que serão importantes no desenvolvimento dos próximos capítulos. A primeira delas é a decomposição binomial para convoluções.

Proposição 1.2 *Sejam f e g funções reais. Então*

$$(f+g)^{n*} = \sum_{k=0}^{n} \binom{n}{k} f^{k*} * g^{(n-k)*},$$

desde que as convoluções acima existam.

Demonstração. Segue das propriedades (a) à (c) da Proposição 1.1, por indução sobre n. □

Proposição 1.3 *Se f e g são densidades e g é limitada, então $(f * g)$ também é limitada.*

Demonstração. Seja $M > 0$ tal que $g(x) \leq M$. Então

$$(f * g)(x) = \int_{-\infty}^{\infty} f(y)g(x-y)dy \leq \int_{-\infty}^{\infty} f(y)Mdy = M.$$

□

1.4 Distribuições Estáveis

Apresentamos nesta seção as definições e principais resultados acerca das distribuições estáveis. As demonstrações dos teoremas desta seção podem ser encontradas em Ibragimov [10] ou Zolotarev [19].

Definição 1.4 *Sejam X_1 e X_2 cópias independentes de uma variável aleatória X. Dizemos que X é estável (ou possui distribuição estável) se, para quaisquer constantes $A, B > 0$, existem constantes reais $C > 0$ e D tais que*

$$AX_1 + BX_2 \stackrel{\mathfrak{D}}{=} CX + D,$$

onde $\stackrel{\mathfrak{D}}{=}$ *indica igualdade em distribuição. Se* $D = 0$*, dizemos que* X *é estritamente estável.*

O teorema a seguir fornece uma definição equivalente à Definição 1.4.

Teorema 1.3 *Uma variável aleatória* X *é estável se, e somente se, dado* $n \geq 2$*, existem constantes* $C_n > 0$ *e* $D_n \in \mathbb{R}$ *tais que*

$$X_1 + \ldots + X_n \stackrel{\mathfrak{D}}{=} C_n X + D_n,$$

onde $X_1, \ldots, X_n$ *são cópias independentes de* X*.*

Exemplo 1.1 *(Caso Degenerado). Seja* X *uma v.a. tal que* $P(X = \mu) = 1$*. Dados* $A, B > 0$*, considere* $C = A + B$ *e* $D = 0$*. Então* X *é estritamente estável, pois* $AX_1 + BX_2 \stackrel{\mathfrak{D}}{=} CX$*.*

A partir de agora vamos supor que as variáveis aleatórias estáveis são não-degeneradas. Neste caso, elas são absolutamente contínuas, conforme o teorema a seguir (ver [15]).

Teorema 1.4 *Toda variável aleatória estável não-degenerada é absolutamente contínua e sua densidade é contínua e infinitamente diferenciável.*

Na grande maioria dos casos não é possível descrever analiticamente e explicitamente as densidades de variáveis aleatórias estáveis. No entanto, o teorema a seguir mostra que suas funções características possuem uma representação elementar.

Teorema 1.5 *Uma variável aleatória* X *não-degenerada é estável se, e somente se, sua função característica* f *é escrita como*

$$f(t) = \exp\{i\mu t - \sigma^\alpha |t|^\alpha [1 + i\beta \operatorname{sign}(t) w(t, \alpha)]\}, \tag{1.2}$$

onde $0 < \alpha \leq 2$*,* $-1 \leq \beta \leq 1$*,* $\mu \in \mathbb{R}$*,* $\sigma > 0$ *e*

$$w(t,\alpha) = \begin{cases} \tan \frac{\pi\alpha}{2}, & \textit{se } \alpha \neq 1, \\ \frac{2}{\pi} \log |t|, & \textit{se } \alpha = 1 \end{cases}$$

Em particular, $|f(t)| = e^{-\sigma^\alpha |t|^\alpha}$*.*

Note que os parâmetros α, σ, β e μ em (1.2) caracterizam completamente a v.a. estável X. Denotamos

$$X \stackrel{\mathfrak{D}}{=} S_\alpha(\sigma, \beta, \mu)$$

e dizemos que X é α-estável.

O parâmetro α é chamado *índice de estabilidade*, β é chamado *parâmetro de simetria*, σ é o *parâmetro de escala* e μ o *parâmetro de locação.*

Como consequência do Teorema 1.5, podemos obter algumas propriedades importantes de distribuições estáveis, que destacamos nos corolários a seguir.

Corolário 1.2 *A função característica de uma variável aleatória estável não-degenerada é integrável.*

Demonstração. De fato, considere f como em (1.2). Então

$$\int_{-\infty}^{+\infty} |f(t)|dt = \int_{-\infty}^{+\infty} e^{-\sigma^\alpha |t|^\alpha} dt = 2\int_0^{+\infty} e^{-\sigma^\alpha t^\alpha} dt$$

Faça a mudança de variáveis $x = \sigma^\alpha t^\alpha$. Temos $t = x^{1/\alpha}/\sigma$ e $dt = x^{1/\alpha - 1}/(\alpha\sigma)$. Logo,

$$\int_{-\infty}^{+\infty} |f(t)|dt = \frac{2}{\alpha\sigma}\int_0^{+\infty} x^{\frac{1}{\alpha}-1} e^{-x} dx = \frac{2}{\alpha\sigma}\Gamma\left(\frac{1}{\alpha}\right) < +\infty.$$

Portanto f é integrável. □

Corolário 1.3 *A densidade de uma variável aleatória estável é limitada.*

Demonstração. Considere ψ e f a densidade e a função característica de uma v.a. estável, respectivamente. Como f é integrável, podemos usar fórmula da inversão para obter que

$$|\psi(t)| = \frac{1}{2\pi}\left|\int e^{-itx} f(x)dx\right| \leq \frac{1}{2\pi}\int |f(x)|dx = \frac{1}{\alpha\sigma\pi}\Gamma\left(\frac{1}{\alpha}\right).$$

□

A seguir, apresentamos os únicos exemplos conhecidos de variáveis aleatórias estáveis não-degeneradas cujas densidades podem ser representadas em termos de funções elementares.

Exemplo 1.2 *(Distribuição Normal). Seja* $X \stackrel{\mathfrak{D}}{=} N(\mu, \sigma^2)$*, ou seja,* X *tem densidade*

$$\psi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \forall x \in \mathbb{R}.$$

Como a função característica de X *é dada por* $f(t) = \exp\{i\mu t - \frac{t^2\sigma^2}{2}\}$*, então* $X \stackrel{\mathfrak{D}}{=} S_2(\frac{\sigma}{\sqrt{2}}, 0, \mu)$*.*

Exemplo 1.3 *(Distribuição Cauchy). Seja* $X \stackrel{\mathfrak{D}}{=} Cauchy(\mu, \sigma)$*, ou seja,* X *tem densidade*

$$\psi(x) = \frac{1}{\pi\sigma\left[1 + \left(\frac{x-\mu}{\sigma}\right)^2\right]}, \quad \forall x \in \mathbb{R}.$$

Como a função característica de X *é dada por* $f(t) = \exp\{i\mu t - \sigma|t|\}$*, então* $X \stackrel{\mathfrak{D}}{=} S_1(\sigma, 0, \mu)$*.*

Exemplo 1.4 *(Distribuição Lévy). Seja* $X \stackrel{\mathfrak{D}}{=} Levy(\mu, \sigma)$*, ou seja,* X *tem densidade*

$$\psi(x) = \begin{cases} \sqrt{\dfrac{\sigma}{2\pi}} \dfrac{e^{-\frac{\sigma}{2(x-\mu)}}}{(x-\mu)^{3/2}}, & \textit{se } x \in (\mu, \infty), \\ 0, & \textit{caso contrário.} \end{cases}$$

Como a função característica de X *é dada por* $f(t) = \exp\{i\mu t - |\sigma t|^{1/2}(1 - i\ \mathrm{sign}(t))\}$*, então* $X \stackrel{\mathfrak{D}}{=} S_{1/2}(\sigma, -1, \mu)$*.*

Neste trabalho, nosso interesse está voltado para uma classe específica de distribuição estável, a qual definimos a seguir.

Definição 1.5 *Uma variável aleatória estável* X *é dita não-extremal se* X *tem distribuição Normal ou se seus parâmetros* α *e* β *satisfazem* $0 < \alpha < 2$ *e* $-1 < \beta < 1$*.*

Se X é não-extremal e não-normal, então sua densidade é positiva em toda a reta e satisfaz a seguinte relação assintótica:

Proposição 1.4 *Seja* X *uma variável aleatória* α*-estável não-extremal a qual não é normal, com densidade* ψ*. Então*

$$\psi(x) \sim c_0|x|^{-(1+\alpha)} \quad (x \to -\infty), \qquad \psi(x) \sim c_1 x^{-(1+\alpha)} \quad (x \to +\infty),$$

para determinadas constantes $c_0, c_1 > 0$*.*

Como consequência da proposição acima, podemos obter uma limitação inferior para ψ.

Corolário 1.4 *Seja X uma variável aleatória α-estável não-extremal a qual não é normal, com densidade ψ. Então existe uma constante $c > 0$ tal que*

$$c(1+|x|)^{-(1+\alpha)} \leq \psi(x), \quad \forall x \in \mathbb{R}. \tag{1.3}$$

Demonstração. Da Proposição 1.4, temos $\psi(x) \sim c_0|x|^{-(1+\alpha)}$ quando $x \to -\infty$. Então existe $A > 0$ tal que para $x < -A$ temos

$$\left| \frac{\psi(x)}{c_0|x|^{-(1+\alpha)}} - 1 \right| < \frac{1}{2},$$

ou seja,

$$\frac{1}{2} c_0 |x|^{-(1+\alpha)} < \psi(x).$$

Assim, segue que

$$\frac{1}{2} c_0 (1+|x|)^{-(1+\alpha)} < \psi(x), \quad \forall x < -A. \tag{1.4}$$

Também, da Proposição 1.4, temos que $\psi(x) \sim c_1 x^{-(1+\alpha)}$ quando $x \to +\infty$. Então existe $B > 0$ tal que para $x > B$ temos

$$\left| \frac{\psi(x)}{c_1 x^{-(1+\alpha)}} - 1 \right| < \frac{1}{2},$$

de onde segue

$$\frac{1}{2} c_1 (1+|x|)^{-(1+\alpha)} < \psi(x), \quad \forall x > B. \tag{1.5}$$

Por outro lado, como ψ é positiva e contínua no compacto $[-A, B]$, segue que existe $m > 0$ tal que $\psi(x) \geq m$ para todo $x \in [-A, B]$. Assim, podemos obter

$$m(1+|x|)^{-(1+\alpha)} \leq \psi(x), \quad \forall x \in [-A, B]. \tag{1.6}$$

Se considerarmos $c = \min(c_0/2, c_1/2, m)$, de (1.4), (1.5) e (1.6) segue que

$$c(1+|x|)^{-(1+\alpha)} \leq \psi(x), \quad \forall x \in \mathbb{R}.$$

e (1.3) está provada. $\square$

Uma outra caracterização de distribuição estável é dada no teorema a seguir, como sendo distribuição limite de somas normalizadas

Teorema 1.6 *Uma variável aleatória X é α-estável se, e somente se, existem uma sequência de variáveis aleatórias i.i.d. $\{X_n\}_{n\geq 1}$, com função de distribuição comum F, e sequências de constantes $\{a_n\}_{n\geq 1}$ e $\{b_n\}_{n\geq 1}$, com $b_n > 0$, tais que*

$$Z_n = \frac{X_1 + \ldots + X_n}{b_n} - a_n \stackrel{\mathfrak{D}}{\to} X, \tag{1.7}$$

onde $\stackrel{\mathfrak{D}}{\to}$ significa convergência em distribuição. Em caso afirmativo, se X é não-degenerada, então as constantes b_n são da forma $b_n = n^{1/\alpha}h(n)$, onde h é uma função de variação lenta no sentido de Karamata.

Uma sequência $\{Z_n\}_{n\geq 1}$ como no teorema anterior, ou seja, definida por

$$Z_n = \frac{X_1 + \ldots + X_n}{b_n} - a_n, \quad n \geq 1,$$

onde $X_1, X_2, \ldots$ são variáveis aleatórias i.i.d., $b_n > 0$ e $a_n \in \mathbb{R}$, para todo $n \geq 1$, é chamada uma *sequência de somas normalizadas*.

Dada uma variável aleatória α-estável X, com função de distribuição G, o conjunto de todas as funções de distribuição F para as quais temos que o limite em (1.7) é chamado *domínio de atração de G*. Especificamente, definimos

Definição 1.6 *Sejam F e G duas funções de distribuição. Dizemos que F é atraída por G se existem sequências de constantes $\{a_n\}$ e $\{b_n\}$, com $b_n > 0$ tais que*

$$Z_n = \frac{X_1 + \ldots + X_n}{b_n} - a_n \stackrel{\mathfrak{D}}{\to} Z,$$

onde $\{X_n\}_{n\geq 1}$ é uma sequência de variáveis aleatórias i.i.d. com função de distribuição F e Z é uma v.a. com função distribuição G. O conjunto de todas as funções de distribuição que são atraídas por G é dito ser o domínio de atração de G e denotado por $D(G)$.

O Teorema 1.6 diz que G é uma função de distribuição de uma v.a. estável se, e somente se, seu domínio de atração é não-vazio. O próximo teorema apresenta uma condição assintótica necessária e suficiente para que F pertença ao domínio de atração de uma distribuição α-estável, com $0 < \alpha < 2$.

Teorema 1.7 *A fim de que uma função distribuição F pertença ao domínio de atração de uma distribuição α-estável, com $0 < \alpha < 2$, é necessário e suficiente que*

$$F(x) \sim (c_0 + o(1))|x|^{-\alpha}B(|x|) \quad (x \to -\infty)$$

e

$$1 - F(x) \sim (c_1 + o(1))x^{-\alpha} B(x) \quad (x \to +\infty),$$

onde $B(x)$ é uma função lentamente variante no sentido de Karamata e $c_0, c_1 \geq 0$ são constantes que não são ambas nulas.

O domínio de atração da distribuição Normal é caracterizado no seguinte teorema.

Teorema 1.8 *Para que uma função de distribuição F esteja no domínio de atração de uma distribuição Normal padrão, com $b_n = a\sqrt{n}$ sendo $a \in \mathbb{R}$, é necessário e suficiente que ela tenha variância finita.*

Se Z é α-estável, então pode-se mostrar que $\mathbf{E}|Z|^\delta < +\infty$ para todo $0 \leq \delta < \alpha$. Podemos generalizar esse fato com o seguinte teorema.

Teorema 1.9 *Sejam $X_1, X_2, \ldots$ variáveis aleatórias i.i.d. e Z_n uma sequência de somas normalizadas dada por*

$$Z_n = \frac{X_1 + \ldots + X_n}{b_n} - a_n,$$

onde $b_n > 0$ e $a_n \in \mathbb{R}$ para todo $n \geq 1$. Se $Z_n \xrightarrow{\mathfrak{D}} Z$, onde Z é uma v.a. α-estável não-degenerada, então, para todo $0 \leq \delta < \alpha$, temos

(a) $\mathbf{E}|Z|^\delta < +\infty$,

(b) $\mathbf{E}|X_1|^\delta < +\infty$,

(c) $\sup_n \mathbf{E}|Z_n|^\delta < +\infty$.

A seguir, apresentamos o Lema de Brown [6], que como consequência garante, sob a hipótese de variância finita, a integrabilidade uniforme da sequência Z_n^2, com $b_n = \sigma\sqrt{n}$ e que será utilizada na demonstração do teorema principal do Capítulo 3.

Lema 1.1 *Suponha que $X_1, X_2, \ldots$ são v.a.'s i.i.d. com média zero e variância σ^2. Seja $Z_n = \dfrac{X_1 + \ldots + X_n}{\sqrt{n}}$. Então*

$$\lim_{B \to \infty} \mathbf{E}(Z_n^2 I_{(|Z_n| \leq B)}) = \mathbf{E}Z^2$$

uniformemente em n, onde Z tem distribuição Normal de média 0 e variância σ^2.

Para finalizar esta seção, apresentamos uma proposição que estabelece condições necessárias sobre a sequência de funções características de Z_n para que $Z_n \xrightarrow{\mathfrak{D}} Z$, onde Z é α-estável não-degenerada.

Proposição 1.5 *Sejam $X_1, X_2, \ldots$ variáveis aleatórias i.i.d. e Z_n uma sequência de somas normalizadas dada por*

$$Z_n = \frac{X_1 + \ldots + X_n}{b_n} - a_n,$$

onde $b_n > 0$ e $a_n \in \mathbb{R}$ para todo $n \geq 1$. Se $Z_n \xrightarrow{\mathfrak{D}} Z$, onde Z é uma v.a. α-estável não-degenerada, e $f_n(t)$ é a função característica de Z_n, então existem constantes $c > 0$ e $t_0 > 0$ tais que $|f_n(t)| \leq e^{-c|t|^{\alpha/2}}$ para todo $n \geq 1$ e t no intervalo $|t| < t_0 b_n$.

Demonstração. Vide Ibragimov [10], página 133.

1.5 Desigualdades Elementares

Nos próximos capítulos, diversas desigualdades elementares serão necessárias a fim de demonstrar o resultado principal deste trabalho. Nesta seção, enunciamos e demonstramos todas as desigualdades que serão utilizadas posteriormente.

Proposição 1.6 *Se $x_1, \ldots, x_n \geq 0$ e $0 < s < 1$, então*

$$(x_1 + \ldots + x_n)^s \leq x_1^s + \ldots + x_n^s. \tag{1.8}$$

Demonstração. Defina a função

$$f(x) = 1 + x^s - (1+x)^s, \quad x \geq 0.$$

Mostremos que $f(x) \geq 0$. De fato, $f(0) = 0$ e

$$f'(x) = sx^{s-1} - s(1+x)^{s-1} > 0, \quad \forall x > 0. \tag{1.9}$$

Dado $x > 0$, pelo Teorema do Valor Médio, existe $c \in (0, x)$ tal que

$$f(x) - f(0) = f'(c) \cdot x.$$

Então, de (1.9) segue que

$$f(x) > 0, \quad \forall x > 0.$$

Portanto, $f(x) \geq 0$ para todo $x \geq 0$. Ou seja,

$$(1+x)^s \leq 1 + x^s, \quad \forall x \geq 0. \tag{1.10}$$

Agora, utilizando (1.10), provaremos (1.8) por indução sobre n. Se $n = 2$, então temos

$$(x_1 + x_2)^s = x_1^s \left(1 + \frac{x_2}{x_1}\right)^s \leq x_1^s \left(1 + \left(\frac{x_2}{x_1}\right)^s\right) = x_1^s + x_2^s, \quad \forall x_1, x_2 > 0. \tag{1.11}$$

Suponha que para algum $n \geq 2$ temos

$$(x_1 + \ldots + x_{n-1})^s \leq x_1^2 + \ldots + x_{n-1}^s, \quad \forall x_i > 0, \quad i = 1, \ldots, n-1. \tag{1.12}$$

Então de (1.11) temos

$$(x_1 + \ldots + x_n)^s \leq (x_1 + \ldots + x_{n-1})^s + x_n^s$$

e da hipótese de indução (1.12) o resultado segue. □

Proposição 1.7 *Se* $x_1, \ldots, x_n \geq 0$ *e* $s \geq 1$, *então*

$$(x_1 + \ldots + x_n)^s \leq n^{s-1}(x_1^s + \ldots + x_n^s). \tag{1.13}$$

Demonstração. Considere a função

$$f(x) = x^s, \quad x \geq 0.$$

Note que f é convexa, pois $f''(x) = s(s-1)x^{s-2} \geq 0$. Logo, para quaisquer $x_1, \ldots, x_n \geq 0$, tem-se

$$f\left(\frac{1}{n}x_1 + \ldots \frac{1}{n}x_n\right) \leq \frac{1}{n}f(x_1) + \cdots + \frac{1}{n}f(x_n)$$

e assim obtemos (1.13). □

Observação 1.1 *Combinando (1.8) e (1.13) podemos concluir que se* $x_1, \ldots, x_n \geq 0$ *e* $s > 0$, *então*

$$(x_1 + \ldots + x_n)^s \leq n^s(x_1^s + \ldots + x_n^s). \tag{1.14}$$

Proposição 1.8 *Para todo* $x > 0$, *temos* $1 - \dfrac{1}{x} \leq \log x$ *e* $\log x \leq x - 1$. *As igualdades ocorrem se, e somente se,* $x = 1$.

Demonstração. Considere a função $f : (0, +\infty) \to \mathbb{R}$ dada por

$$f(x) = \log x + \frac{1}{x} - 1, \quad \forall x > 0.$$

Como pela regra de L'Hôpital temos $\lim\limits_{x \to 0^+} x \log x = 0$, então temos

$$\lim_{x \to 0^+} f(x) = \lim_{x \to 0^+} \frac{x \log x + 1 - x}{x} = +\infty.$$

Logo, existe $0 < \varepsilon < 1$ tal que

$$f(x) > 0, \quad \forall x \in (0, \varepsilon]. \tag{1.15}$$

Além disso, como $\lim\limits_{x \to +\infty} f(x) = +\infty$, existe $c > 1$ tal que

$$f(x) > 0, \quad \forall x \geq c. \tag{1.16}$$

Por outro lado, pela continuidade de f no compacto $[\varepsilon, c]$, existe $x_0 \in [\varepsilon, c]$ que minimiza f nesse intervalo, ou seja,

$$f(x) \geq f(x_0), \quad \forall x \in [\varepsilon, c].$$

Note que $f(1) = 0$, de modo que $x_0 \in (\varepsilon, c)$. Logo $f'(x_0) = 0$. Assim,

$$f'(x_0) = \frac{1}{x_0} - \frac{1}{x_0^2} = 0,$$

ou seja,

$$x_0 = 1.$$

Portanto,

$$f(x) \geq f(1) = 0, \quad \forall x \in [\varepsilon, c]. \tag{1.17}$$

Então, de (1.15), (1.16) e (1.17), segue que

$$f(x) \geq 0, \quad \forall x > 0,$$

ou seja, $1 - \frac{1}{x} \leq \log x$, para todo $x > 0$. Para a segunda desigualdade, basta aplicar a desigualdade anterior com $\frac{1}{x}$ no lugar de x e obtemos

$$1 - x \leq \log \frac{1}{x}, \quad \forall x > 0,$$

ou seja,

$$\log x \leq x - 1, \quad \forall x > 0.$$

$\square$

Proposição 1.9 *Dado $\varepsilon \in (0,1]$, existe $C_\varepsilon \geq 1$ tal que*

$$\log x \leq C_\varepsilon (x-1)^\varepsilon, \quad \forall x \geq 1.$$

Demonstração. Se $\varepsilon = 1$, basta considerar $C_\varepsilon = 1$ e usar a Proposição 1.8. Suponha $\varepsilon \in (0,1)$ e considere a função

$$f(x) = \begin{cases} \frac{\log x}{(x-1)^\varepsilon}, & \text{se } x > 1, \\ 0, & \text{se } x = 1. \end{cases}$$

Então, pela regra de L'Hôpital,

$$\lim_{x \to 1^+} f(x) = \lim_{x \to 1^+} \frac{\frac{1}{x}}{\varepsilon (x-1)^{\varepsilon - 1}} = \lim_{x \to 1^+} \frac{(x-1)^{1-\varepsilon}}{\varepsilon x} = 0.$$

Logo f é contínua. Mais ainda, usando novamente a regra de L'Hôpital,

$$\lim_{x \to +\infty} f(x) = \lim_{x \to +\infty} \frac{\frac{1}{x}}{\varepsilon (x-1)^{\varepsilon - 1}} = \lim_{x \to +\infty} \frac{(x-1)^{1-\varepsilon}}{\varepsilon x} = \lim_{x \to +\infty} \frac{(1-\frac{1}{x})^{1-\varepsilon}}{\varepsilon x^\varepsilon} = 0.$$

Então existe $x_0 > 1$ tal que

$$f(x) \leq 1, \quad \forall x > x_0.$$

Como f é contínua no compacto $[1, x_0]$, f assume um valor máximo M neste intervalo. Denote

$$C_\varepsilon = \max\{1, M\}.$$

Então $f(x) \leq C_\varepsilon$ para todo $x \geq 1$, isto é,

$$\log x \leq C_\varepsilon (x-1)^\varepsilon, \quad \forall x \geq 1.$$

$\square$

Proposição 1.10 *Considere a função*

$$L(x) = \begin{cases} x \log x, & \textit{se } x > 0, \\ 0, & \textit{se } x = 0. \end{cases} \tag{1.18}$$

(a) Para quaisquer $u, v \geq 0$ e $0 \leq \varepsilon \leq 1$, as seguintes propriedades são satisfeitas

(a.1) $L(uv) = uL(v) + vL(u)$

(a.2) $L(u+v) \geq L(u) + L(v)$

(a.3) $L(u) \geq -\frac{1}{e}$

(a.4) $L((1-\varepsilon)u + \varepsilon v) \leq (1-\varepsilon)L(u) + \varepsilon L(v)$

(a.5) $L((1-\varepsilon)u + \varepsilon v) \geq (1-\varepsilon)L(u) + \varepsilon L(v) + uL(1-\varepsilon) + vL(\varepsilon)$

(a.6) $L((1-\varepsilon)u + \varepsilon v) \geq (1-\varepsilon)L(u) - \frac{1}{e}u - \frac{1}{e}$

(a.7) $L(u) \leq (u-1) + (u-1)^2$

(b) Dado $0 < \varepsilon \leq 1$, *existe* $C_\varepsilon \geq 1$ *tal que*

$$L(u) \leq (u-1) + C_\varepsilon |u-1|^{1+\varepsilon}, \quad \forall u \geq 0.$$

Demonstração.
(a.1) Se $u = 0$ ou $v = 0$, então a igualdade segue trivialmente. Suponha $u \neq 0$ e $v \neq 0$, então

$$L(uv) = uv \log uv = uv \log u + uv \log v = uL(v) + vL(u).$$

(a.2) Se $u = 0$ ou $v = 0$, então a desigualdade segue trivialmente. Suponha $u \neq 0$ e $v \neq 0$, então

$$\begin{aligned} L(u+v) - L(u) - L(v) &= (u+v)\log(u+v) - u \log u - v \log v \\ &= u(\log(u+v) - \log u) + v(\log(u+v) - \log u) \geq 0, \end{aligned}$$

onde a última desigualdade segue da monotonicidade crescente da função $\log x,\ x > 0$.

(a.3) Note que $L(0) = L(1) = 0$ e $L(u) < 0$ se $0 < u < 1$. Pela continuidade de L no compacto $[0,1]$, existe $u_0 \in (0,1)$ tal que

$$L(u) \geq L(u_0), \quad \forall u \in [0,1].$$

Como L é derivável em $(0,1)$, então $L'(u_0) = 0$. Assim, $\log u_0 + 1 = 0$, o que implica que $u_0 = \frac{1}{e}$ e $L(u_0) = -\frac{1}{e}$. Logo,

$$L(u) \geq -\frac{1}{e}, \quad \forall u \in [0,1].$$

Como $L(u) > 0$ se $u > 1$, então segue que $L(u) \geq -\frac{1}{e},\ \forall u \geq 0$.

(a.4) Note que $L''(u) = \frac{1}{u} > 0$ se $u > 0$. Logo L é convexa e segue o resultado.

(a.5) Pelos itens *(a.1)* e *(a.2)* desta Proposição, segue que

$$\begin{aligned} L((1-\varepsilon)u+\varepsilon v) &\geq L((1-\varepsilon)u)+L(\varepsilon v) \\ &= (1-\varepsilon)L(u)+uL(1-\varepsilon)+\varepsilon L(v)+vL(\varepsilon). \end{aligned}$$

(a.6) Pelos itens *(a.1)*, *(a.2)* e *(a.3)*, temos

$$\begin{aligned} L((1-\varepsilon)u+\varepsilon v) &\geq L((1-\varepsilon)u)+L(\varepsilon v) \\ &\geq L((1-\varepsilon)u)-\frac{1}{e} \\ &= (1-\varepsilon)L(u)+uL(1-\varepsilon)-\frac{1}{e} \\ &\geq (1-\varepsilon)L(u)-\frac{1}{e}u-\frac{1}{e}. \end{aligned}$$

(a.7) Se $u=0$, então a desigualdade é trivial. Suponha $u\neq 0$. Então, pela Proposição 1.8, segue que $\log u \leq u-1,\ \forall u>0$ e

$$L(u)\leq u^2-u=(u-1)+(u-1)^2, \quad \forall u>0.$$

(b) Seja $0<\varepsilon\leq 1$. Por (a.7), se $0\leq u\leq 1$, então

$$L(u)\leq (u-1)+(1-u)^2\leq (u-1)+(1-u)^{1+\varepsilon}.$$

Neste caso, basta tomar $C_\varepsilon=1$. Agora suponha $u>1$. Pela Proposição 1.9, existe $C_\varepsilon\geq 1$ tal que

$$\log u\leq C_\varepsilon(u-1)^\varepsilon.$$

Logo

$$(u-1)\log u\leq C_\varepsilon(u-1)^{\varepsilon+1}$$

e usando a Proposição 1.8, segue

$$L(u)\leq \log u+C_\varepsilon(u-1)^{\varepsilon+1}\leq (u-1)+C_\varepsilon(u-1)^{\varepsilon+1},$$

Portanto, para todo $u>0$, temos

$$L(u)\leq (u-1)+C_\varepsilon|u-1|^{1+\varepsilon}.$$

□

Proposição 1.11 *Para quaisquer $u,v\geq 0$, temos*

(a) $\log(1+u)\leq \frac{u}{v}+\log(1+v)$.

(b) $\log(1+u+v) \leq \log(1+u) + \log(1+v)$.

Demonstração.
(a) Pela Proposição 1.8,

$$\log \frac{1+u}{1+v} \leq \frac{1+u}{1+v} - 1 = \frac{u-v}{1+v} \leq \frac{u}{v}$$

e, assim, segue que

$$\log(1+u) \leq \frac{u}{v} + \log(1+v).$$

(b) Basta notar que $1+u+v \leq (1+u)(1+v)$ e podemos obter

$$\log(1+u+v) \leq \log[(1+u)(1+v)] = \log(1+u) + \log(1+v).$$

□

1.6 Propriedades Auxiliares de Funções Características

Nesta seção, apresentamos resultados envolvendo funções características que serão utilizados no Capítulo 3.

O primeiro resultado é o Teorema de Riemann-Lebesgue, que afirma que a função característica $f(t)$ de uma variável absolutamente contínua tende a zero quando t tende a $+\infty$ ou $-\infty$.

Teorema 1.10 *(Teorema de Riemann-Lebesgue). Seja f a função característica de uma variável aleatória absolutamente contínua. Então*

$$\lim_{t \to \pm\infty} |f(t)| = 0.$$

Além do Teorema de Riemann-Lebesgue, um outro resultado importante é que a função característica $f(t)$ de uma variável absolutamente contínua tem módulo menor que 1 fora de $t = 0$, que é consequência da próxima proposição.

Proposição 1.12 *Seja X uma variável aleatória com função característica $f(t)$. Então existe $t_0 \neq 0$ tal que $|f(t_0)| = 1$ se, e somente se, X é discreta assumindo valores em $\{C + 2\frac{k\pi}{t_0},\ k \in \mathbb{Z}\}$, para algum $C \in \mathbb{R}$.*

Demonstração. Suponha que existe $t_0 \neq 0$ tal que $|f(t_0)| = 1$. Como $f(t_0) \in \mathbb{C}$ e $|f(t_0)| = 1$, então existe $C \in \mathbb{R}$ tal que $f(t_0) = e^{it_0C}$. Assim, se P_X é a medida de probabilidade induzida por X, então

$$e^{it_0C} = \int_{-\infty}^{+\infty} e^{it_0x} dP_X(x).$$

Logo temos $\int_{-\infty}^{+\infty} e^{it_0(x-C)} dP_X(x) = 1$. Assim, pela propriedade de que $e^{it_0(x-C)} = \cos[t_0(x-C)] + i\text{sen}[t_0(x-C)]$, temos

$$\int_{-\infty}^{+\infty} \cos[t_0(x-C)]dP_X(x) + i\int_{-\infty}^{+\infty} \text{sen}[t_0(x-C)]dP_X(x) = 1.$$

Logo $\int_{-\infty}^{+\infty} \cos[t_0(x-C)]dP_X(x) = 1$, ou seja, $\int_{-\infty}^{+\infty} \{1 - \cos[t_0(x-C)]\}dP_X(x) = 0$. Como $1 - \cos[t_0(x-C)] \geq 0$, por propriedades de integral de Lebesgue, segue que

$$1 - \cos[t_0(x-C)] = 0 \quad \text{para q.t. } x \ [P_X].$$

Logo $P_X(\{x : \cos[t_0(x-C)] = 1\}) = 1$, ou seja, $P_X(\{x : x = C + 2\frac{k\pi}{t_0},\ k \in \mathbb{Z}\}) = 1$. Concluímos então que X é uma v.a. discreta assumindo valores $C + 2\frac{k\pi}{t_0}$, onde $k \in \mathbb{Z}$. Podemos mostrar a recíproca usando o mesmo argumento na ordem inversa. □

Por fim, enunciamos a Fórmula de Plancherel, que estabelece que a integral do módulo de uma função ao quadrado é igual à integral do módulo de sua transformada de Fourier ao quadrado.

Teorema 1.11 *(Fórmula de Plancherel). Seja f uma função em $L_1(\mathbb{R})$ e $L_2(\mathbb{R})$. Então a transformada de Fourier de f dada por*

$$\hat{f}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x)e^{-itx}dx$$

está em $L_2(\mathbb{R})$ e satisfaz

$$\int_{-\infty}^{\infty} |\hat{f}(t)|^2 dt = \int_{-\infty}^{\infty} |f(x)|^2 dx.$$

Capítulo 2

Entropia

2.1 Introdução

Em 1948, o matemático e engenheiro Claude Shannon introduziu, no seu artigo "A mathematical Theory of Communication" [18], os conceitos fundamentais que deram origem à hoje conhecida Teoria da Informação, a qual basicamente estuda a quantificação, armazenagem e comunicação de informação. Desde então, a teoria introduzida por Shannon tem sido aplicada em diferentes áreas, tais como transmissão de dados, criptografia, computação digital, codificação de sinais, assim como em inferência estatística, linguística, fonética, psicologia, neurobiologia, física, entre muitas outras áreas.

O conceito de informação é bastante amplo e, nesse trabalho, Shannon propôs uma definição matemática de informação. Nesse contexto, se A é um evento que ocorre com probabilidade $P(A)$, então a quantidade de informação $I(A)$ que se ganha com o conhecimento da ocorrência de A é definida como

$$I(A) = -\log_2 P(A) = \log_2 \left(\frac{1}{P(A)}\right).$$

Assim, quanto menor é a probabilidade de ocorrer um determinado evento, maior é a informação obtida ao saber de sua ocorrência. A escolha da função logaritmo pode ser justificada por algumas de suas propriedades: por ser uma função crescente, permite a interpretação descrita acima, satisfaz $\log 1 = 0$, ou seja, eventos certos (com $P(A) = 1$) não carregam nenhuma informação e a propriedade $\log xy = \log x + \log y$ que traduz a aditividade da quantidade de informação, ou seja, $I(A \cap B) = I(A) + I(B)$, quando A e B são eventos independentes. Em [11] podemos encontrar uma justificativa mais detalhada, baseada nos axiomas de Rényi. A escolha do log na base 2 expressa a informação em "bits", mas outras bases são usadas, já que a troca de base pode ser obtida por $I_b(A) = (\log_b a) I_a(A)$. Neste trabalho vamos assumir log na base e e vamos indicar simplesmente por $\log x$.

Um dos conceitos chaves da Teoria da Informação é a *entropia*, ou *entropia de Shannon*, que em linhas gerais mede o grau médio de incerteza associado a diversas fontes de informação presentes em um sistema, ou, em outras palavras, é a quantidade média de informação contida em um sistema. Especificamente, considerando um sistema com N resultados possíveis e se p_i é a probabilidade de ocorrência do i-ésimo resultado, $i = 1, \ldots, N$, com $\sum_{i=1}^{N} p_i = 1$, então a entropia do sistema é calculada por

$$H(p) = -\sum_{i=1}^{N} p_i \log p_i.$$

Ou ainda, se considerarmos X uma variável aleatória assumindo N valores possíveis $\{x_1, \ldots, x_n\}$ com probabilidades $p_i = P(X = x_i) = p_X(x_i),\ i = 1, \ldots, N$, então a entropia de X é dada por

$$H(X) = -\sum_{i=1}^{N} p_X(x_i) \log p_X(x_i),$$

ou seja,

$$H(X) = \mathbf{E}(-\log p_X(X))$$

e mede o grau médio de incerteza do valor obtido pela variável aleatória X.

Para exemplificarmos de maneira simples o comportamento da entropia, consideremos por exemplo o lançamento de uma moeda, não necessariamente honesta, sendo p a probabilidade de ocorrer cara e $q = 1 - p$ a probabilidade de ocorrer coroa, com $0 \leq p \leq 1$. Seja X a variável aleatória representando o resultado possível do lançamento da moeda, com valores possíveis 0 (cara) e 1 (coroa) e função de probabilidade $p_X(0) = p$, $p_X(1) = q = 1 - p$. Neste caso,

$$H(X) = -p \log(p) - q \log(q).$$

Logo, a entropia de X é maximizada se a moeda é honesta, ou seja, $p = q = \frac{1}{2}$ e $H(X) = \log 2$. Esta é a situação de incerteza máxima, já que é a situação mais difícil de se prever o resultado do próximo lançamento. Agora, se a moeda não é honesta, ou seja, $p \neq q$, então $H(X) < \log 2$ e existe menos incerteza, já que cada vez que jogamos a moeda, uma face é mais provável de ocorrer do que a outra. No outro extremo, se a moeda tem duas caras (ou duas coroas), ou seja, $p = 1$ e $q = 0$ (ou $p = 0$ e $q = 1$), então $H(X) = 0$ e não existe incerteza, já que o resultado de cada lançamento da moeda é sempre certo e nenhuma informação é liberada a cada lançamento.

Cabe observar que, na verdade, o conceito de entropia teve origem primeiramente em Termodinâmica e em Mecânica Estatística, no século XIX, como sendo uma medida do grau de desorganização das partículas em um determinado sistema físico, ou seja, uma medida do grau de aleatoriedade do sistema. Em [18], Seção 1.1.2, podemos encontrar uma síntese da ligação da entropia de Shannon com a entropia termodinâmica.

O uso da Teoria da Informação em Estatística foi introduzido por Kullback and Leibler em 1951 [12]. No contexto de Estatística, a entropia é interpretada como uma medida de incerteza ou de risco. Enquanto na teoria da comunicação o foco é a entropia de variáveis aleatórias discretas, na Estatística existe, frequentemente, um grande interesse em variáveis aleatórias contínuas. Neste caso, se X é uma variável aleatória contínua com densidade p, a entropia de X, chamada de *entropia diferencial de* X, é definida por

$$H(X) = H(p) = -\mathbf{E}\log p(X) = -\int p(x)\log p(x)dx \tag{2.1}$$

e algumas das propriedades de entropia para variáveis aleatórias discretas não são mantidas no caso contínuo, como, por exemplo, a não-negatividade.

Com o objetivo estatístico de discriminar o quanto uma distribuição de probabilidade se diferencia da outra, Kullback e Leibler [12] introduziram o conceito de *entropia relativa*, também conhecida como *risco* ou *divergência de Kullback-Leibler*, de uma densidade q relativo a p, como sendo

$$D(p||q) = \int p(x)\log\frac{p(x)}{q(x)}dx, \tag{2.2}$$

se $q(x) = 0$ implicar $p(x) = 0$ quase-certamente e no caso contrário, define-se $D(p||q) = +\infty$. Em outras palavras, se X é uma variável aleatória com densidade $p(x)$, então

$$D(p||q) = \mathbf{E}\left[\log\frac{p(X)}{q(X)}\right].$$

Dessa forma, se uma observação vem de p, $D(p||q)$ mede o risco de usar q no lugar de p. Assim, no contexto de teste de hipóteses, se temos o interesse em testar as hipóteses

$$H_0 : X \text{ tem densidade } q$$
$$H_1 : X \text{ tem densidade } p$$

então Kullback e Leibler definiram $\log\frac{p(x)}{q(x)}$ como a *informação* em um valor observado x, de X, para a discriminação entre H_0 e H_1 e $D(p||q)$ pode ser interpretada como a informação média por amostra para discriminar em favor de H_1 contra H_0, dado que H_1 é verdadeira. Vale notar que o conhecido Lema de Neyman-Pearson estabelece que o teste da razão de log-verossimilhança é o teste de hipóteses mais poderoso para testar H_0 contra H_1, pois quanto maior é $p(x)$ com relação à $q(x)$, mais informação temos para rejeitar a hipótese H_0.

Embora a entropia relativa seja positiva semi-definida, ou seja, $D(p||q) \geq 0$ para quaisquer duas densidades p e q e $D(p||q) = 0$ se, e somente se, $p = q$, ela não é uma métrica, pois não satisfaz a propriedade de simetria e nem a desigualdade triangular. Apesar disso, é também comum ser chamada de distância de Kullback-Leibler.

A entropia relativa está relacionada com outras normas, como por exemplo, a norma da variação total, através da seguinte desigualdade, conhecida como desigualdade de Pinsker

$$||\nu_p - \nu_q||_{VT} \leq \sqrt{2(D(p||q))}, \tag{2.3}$$

onde ν_p e ν_q são as medidas de probabilidade na reta absolutamente contínuas com densidades p e q, respectivamente, e

$$||\nu_p - \nu_q||_{VT} = 2\sup\{|\nu_p(A) - \nu_q(A)| : A \in \mathcal{B}\} = 2\sup\left\{\left|\int_A p(x) - q(x)dx\right| : A \in \mathcal{B}\right\}.$$

Também em particular, temos a desigualdade obtida por Kullback-Leibler [12], estabelecendo a relação com a norma em L^1,

$$||p - q||_{L^1} = \int_{\mathbb{R}} |p(x) - q(x)|dx \leq \sqrt{2D(p||q)}.$$

Consequentemente, convergência em entropia relativa é uma propriedade mais forte do que convergência na norma de variação total ou na norma em L^1.

Neste capítulo, apresentamos os conceitos e propriedades de entropia diferencial e entropia relativa que serão importantes para o desenvolvimento do próximo capítulo.

Na Seção 2.2, apresentamos o conceito de entropia diferencial, bem como as definições de entropia conjunta e entropia condicional para vetores aleatórios absolutamente contínuos. Algumas propriedades úteis para o decorrer do trabalho são enunciadas e demonstradas.

Na Seção 2.3, introduzimos o conceito de entropia relativa para variáveis aleatórias quaisquer e o caso particular onde as variáveis são absolutamente contínuas. Também, mostramos a não-negatividade da Entropia Relativa e obtemos uma cota superior para a Entropia Diferencial de uma v.a. com variância finita.

Por fim, na Seção 2.4, apresentamos uma condição necessária e suficiente para a finitude da Entropia Relativa $D(X||Z)$ quando Z é α-estável não-extremal. Finalizamos essa seção explorando as consequências desses resultados.

2.2 Entropia Diferencial

A principal referência bibliográfica utilizada nesta seção é Johnson [11].

O conceito de entropia, inicialmente definido por Shannon [18], está associado a variáveis aleatórias discretas assumindo um conjunto finito de valores. Podemos estender a definição para variáveis aleatórias discretas com valores em um conjunto enumerável.

Definição 2.1 *Se X é uma variável aleatória assumindo valores $\{x_1, x_2, \ldots\} \subset \mathbb{R}$, com função de probabilidade $p_i = p_X(x_i) = P(X = x_i)$, $i \geq 1$, a entropia de X é definida por*

$$H(X) = \mathbf{E}[-\log X] = -\sum_{i=1}^{\infty} p_i \log p_i.$$

Note que, neste caso quando a série diverge temos $H(X) = +\infty$.

Como a entropia definida acima não depende dos valores $\{x_i,\ i = 1, 2, \ldots\}$ mas somente do vetor de probabilidades $p = (p_1, p_2, \ldots)$, costuma-se também referir-se à entropia da distribuição de probabilidade p e denotá-la por $H(p)$.

Observação 2.1 *Algumas das principais propriedades para entropias discretas são listadas a seguir. Para maiores detalhes e resultados adicionais, vide [11].*

(a) $H(X) \geq 0$ *e* $H(X) = 0$ *se, e somente se,* $p_i = 1$ *para algum* i *(ou seja, se, e só se,* X *é degenerada).*

(b) H *é invariante por translação e por escala, ou seja,* $\forall a, b \in \mathbb{R},\ a \neq 0$, *temos*

$$H(aX + b) = H(X)$$

(c) Se X *é uma variável aleatória assumindo* N *valores, então a entropia de* X *é maximizada pela entropia da distribuição uniforme discreta,* $p_i = \frac{1}{N}$, $i = 1, \ldots, N$, *ou seja,*

$$0 \leq H(X) \leq \log N.$$

A Definição 2.1 pode ser generalizada para o caso em que X é uma variável aleatória absolutamente contínua e que é o caso estudado neste trabalho.

Definição 2.2 *A entropia diferencial de uma v.a.* X *absolutamente contínua com densidade* p *é definida por*

$$H(X) = H(p) = -\mathbf{E} \log p(X) = -\int_{-\infty}^{+\infty} p(x) \log p(x) dx,$$

desde que a integral exista e adota-se a convenção $0 \log 0 = 0$.

Note que, ao contrário da entropia discreta, a entropia diferencial de X não é necessariamente não-negativa e pode assumir $-\infty$ ou $+\infty$.

Como exemplo, vamos calcular a entropia diferencial de uma v.a. com distribuição Normal. Na Seção 2.3, utilizaremos este cálculo para mostrar que toda v.a. X absolutamente contínua com $\text{Var}X = \sigma^2 < +\infty$ tem entropia diferencial limitada superiormente pela entropia da Normal com variância σ^2.

Exemplo 2.1 *Se $X \sim N(\mu, \sigma^2)$, então $H(X) = \frac{1}{2}\log(2\pi e\sigma^2)$.*

De fato, seja ψ a densidade de X. Então

$$
\begin{aligned}
h(X) &= -\int_{-\infty}^{+\infty} \psi(x) \log \psi(x) \mathrm{d}x \\
&= -\int_{-\infty}^{+\infty} \psi(x) \log \left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \right) \mathrm{d}x \\
&= -\int_{-\infty}^{+\infty} \psi(x) \log \left(\frac{1}{\sqrt{2\pi}\sigma} \right) \mathrm{d}x + \int \psi(x) \left(\frac{(x-\mu)^2}{2\sigma^2} \right) \mathrm{d}x \\
&= \log(\sqrt{2\pi}\sigma) + \frac{\sigma^2}{2\sigma^2} \\
&= \frac{1}{2}\log(2\pi e\sigma^2).
\end{aligned}
$$

Na próxima proposição mostramos que a entropia diferencial não preserva a propriedade dada na Observação 2.1(b) do caso discreto.

Proposição 2.1 *Para quaisquer $a > 0$ e $b \in \mathbb{R}$, temos a igualdade*

$$H(aX + b) = H(X) + \log a.$$

Demonstração. Seja $Y = aX + b$. Se f e g são as densidade de X e Y, respectivamente, então

$$g(x) = \frac{1}{a} f\left(\frac{x-b}{a} \right).$$

Logo,

$$
\begin{aligned}
H(Y) &= -\int_{-\infty}^{+\infty} g(x) \log g(x) dx \\
&= -\int_{-\infty}^{+\infty} \frac{1}{a} f\left(\frac{x-b}{a} \right) \log \left[\frac{1}{a} f\left(\frac{x-b}{a} \right) \right] dx.
\end{aligned}
$$

Fazendo a mudança de variáveis $y = (x - b)/a$, temos $dy = dx/a$ e

$$\begin{aligned} H(Y) &= -\int_{-\infty}^{+\infty} f(y)\log\left[\frac{1}{a}f(y)\right]dy \\ &= -\int_{-\infty}^{+\infty} \log\frac{1}{a}f(y)dy - \int_{-\infty}^{+\infty} f(y)\log f(y)dy \\ &= \log a + H(X). \end{aligned}$$

□

De maneira análoga, podemos também definir entropia para vetores aleatórios absolutamente contínuos.

Definição 2.3 *Seja (X, Y) um vetor aleatório absolutamente contínuo com densidade conjunta f. A entropia conjunta $H(X, Y)$ é definida como*

$$H(X,Y) = -\mathbf{E}\log f(X,Y) = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\log f(x,y)dxdy,$$

desde que a integral exista.

A definição acima pode ser generalizada para vetores n-dimensionais. A proposição a seguir afirma que a entropia conjunta nunca supera a soma das entropias individuais.

Proposição 2.2 *Se (X, Y) é um vetor aleatório absolutamente contínuo tal que $H(X)$ e $H(Y)$ existem e não assumem $+\infty$, então $H(X, Y)$ existe, não assume $+\infty$ e*

$$H(X,Y) \leq H(X) + H(Y),$$

onde a igualdade ocorre se X e Y são independentes. Se $H(X, Y)$ é finita, então a igualdade é válida se, e somente se, X e Y são independentes.

Demonstração. Sejam U e V v.a.'s independentes tais que $U \stackrel{\mathfrak{D}}{=} X$ e $V \stackrel{\mathfrak{D}}{=} Y$. Se $f_{X,Y}$, f_X e f_Y são as densidades de (X, Y), X e Y, respectivamente, então defina a variável aleatória

$$\xi(U,V) = \frac{f_{X,Y}(U,V)}{f_X(U)f_Y(V)}.$$

Por independência, a densidade conjunta de (U, V) é dada pelo produto $f_{U,V}(u, v) = f_X(u)f_Y(v)$. Assim, temos $\mathbf{E}\xi = 1$ e, da convexidade da função

$L(t) = t \log t$ e da desigualdade de Jensen, segue que

$$\begin{aligned} \iint f_{X,Y}(x,y) \log \frac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)} dxdy &= \mathbf{E}\xi \log \xi \\ &\geq \mathbf{E}\xi \log \mathbf{E}\xi \\ &= 0. \end{aligned} \tag{2.4}$$

Podemos escrever

$$f_{X,Y} \log \frac{f_{X,Y}}{f_X f_Y} + f_{X,Y} \log \frac{1}{f_{X,Y}} = f_{X,Y} \log \frac{1}{f_X} + f_{X,Y} \log \frac{1}{f_Y}. \tag{2.5}$$

Como por hipótese $H(X)$ e $H(Y)$ existem e não assumem $+\infty$, de (2.4) e (2.5) segue que $H(X,Y)$ existe e temos

$$H(X,Y) = H(X) + H(Y) - \iint f_{X,Y}(x,y) \log \frac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)} dxdy. \tag{2.6}$$

Portanto, por (2.4), temos $H(X,Y) \leq H(X) + H(Y)$. Se X e Y são independentes, então $f_{X,Y} = f_X f_Y$ e claramente vale a igualdade. Por outro lado, de (2.6) temos que se $H(X,Y)$ é finita e $H(X,Y) = H(X) + H(Y)$, então

$$\iint_{\mathbb{R}^2} f_{X,Y}(x,y) \log \frac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)} dxdy = 0.$$

Logo, podemos obter

$$\iint_{\mathbb{R}^2} f_{X,Y}(x,y) \left[\left(\frac{f_X(x) f_Y(y)}{f_{X,Y}(x,y)} - 1 \right) - \log \frac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)} \right] dxdy = 0.$$

Agora como $\log x \leq x - 1, \ \forall x > 0$ (Proposição 1.8), então temos que o integrando acima é uma função não-negativa e das propriedades de integração segue que

$$\frac{f_X(x) f_Y(y)}{f_{X,Y}(x,y)} - 1 = \log \frac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)} \quad q.c.$$

Como $x - 1 = \log x$ se, e só se, $x = 1$, obtemos que

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

e assim X e Y são independentes. $\square$

Um outro conceito importante para obter informação sobre o vetor (X,Y) é a entropia condicional, que em linhas gerais é uma medida de incerteza do valor de X

dado o conhecimento do valor de Y.

Definição 2.4 *Seja (X,Y) um vetor aleatório absolutamente contínuo, com densidade conjunta $f_{X,Y}$. Se f_Y é a densidade de Y, definimos a entropia condicional $H(X|Y)$ como*

$$H(X|Y) = -\mathbf{E}\left(\log \frac{f_{X,Y}(X,Y)}{f_Y(Y)}\right),$$

desde que a esperança exista.

Note que, se $H(X,Y)$ e $H(Y)$ existem e $H(X,Y) - H(Y)$ está bem definida, então

$$H(X|Y) = H(X,Y) - H(Y).$$

Proposição 2.3 *Seja (X,Y) um vetor aleatório absolutamente contínuo. Se $H(X|Y)$ existe e não assume $-\infty$, então $H(X)$ existe e satisfaz*

$$H(X|Y) \leq H(X),$$

com a igualdade ocorrendo se X e Y são independentes. Se $H(X)$ é finita, então a igualdade é válida se, e somente se, X e Y são independentes.

Demonstração. Se $f_{X,Y}$, f_X e f_Y são as densidades de (X,Y), X e Y, respectivamente, então podemos escrever

$$f_{X,Y} \log \frac{f_{X,Y}}{f_X f_Y} + f_{X,Y} \log \frac{f_Y}{f_{X,Y}} = f_{X,Y} \log \frac{1}{f_X},$$

ou seja,

$$-f_{X,Y} \log f_X = f_{X,Y} \log \frac{f_Y}{f_{X,Y}} + f_{X,Y} \log \frac{f_{X,Y}}{f_X f_Y}. \tag{2.7}$$

Agora como por hipótese $H(X|Y) > -\infty$, então temos

$$\iint_{\mathbb{R}^2} \left[f_{X,Y}(x,y) \log \frac{f_Y}{f_{X,Y}} \right]^- dxdy < +\infty$$

e juntamente com (2.4), segue de (2.7) que $H(X)$ existe, não assume $-\infty$ e

$$H(X) = H(X|Y) + \iint_{\mathbb{R}^2} f_{X,Y}(x,y) \log \frac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)} dxdy. \tag{2.8}$$

Novamente de (2.4), segue que $H(X|Y) \leq H(X)$. Se X e Y são independentes, então $f_{X,Y} = f_X f_Y$ e a igualdade segue da definição de entropia condicional. Por outro lado, se $H(X)$ é finita e $H(X|Y) = H(X)$, de (2.8) segue

$$\iint_{\mathbb{R}^2} f_{X,Y}(x,y) \log \frac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)} dxdy = 0.$$

e repetindo o mesmo argumento do final da prova da Proposição 2.2, segue que X e Y são independentes. $\square$

Usando as desigualdades anteriores de entropia conjunta e condicional, podemos obter a seguinte desigualdade para a entropia da soma de duas variáveis aleatórias independentes.

Proposição 2.4 *Seja (X, Y) um vetor aleatório absolutamente contínuo, onde X e Y são independentes. Se $H(X) > -\infty$, então $H(X+Y)$ existe e satisfaz*

$$H(X) \leq H(X+Y).$$

Demonstração. Como X e Y são independentes, segue pelas Proposições 2.2 e 2.3 que $H(X,Y) = H(X) + H(Y)$ e $H(X|Y) = H(X)$. Mostremos que $H(X+Y|Y) = H(X|Y)$. De fato, sejam $f_{X,Y}$ e $f_{X+Y,Y}$ as densidades conjuntas de (X,Y) e $(X+Y,Y)$, respectivamente. Usando o método do Jacobiano, podemos obter que a densidade conjunta de $X+Y$ e Y é dada por

$$f_{X+Y,Y}(u,v) = f_{X,Y}(u-v,v).$$

Assim,

$$\begin{aligned}
H(X+Y|Y) &= -\iint f_{X+Y,Y}(u,v) \log \frac{f_{X+Y,Y}(u,v)}{f_Y(v)} dudv \\
&= -\iint f_{X,Y}(u-v,v) \log \frac{f_{X,Y}(u-v,v)}{f_Y(v)} dudv \\
&= -\iint f_{X,Y}(x,y) \log \frac{f_{X,Y}(x,y)}{f_Y(y)} dxdy \\
&= H(X|Y).
\end{aligned}$$

Agora, novamente pela Proposição 2.3, segue que

$$H(X) = H(X|Y) = H(X+Y|Y) \leq H(X+Y).$$

$\square$

2.3 Entropia Relativa

Nesta seção apresentamos a definição e propriedades básicas da entropia relativa, ou divergência de Kullback-Leibler, de variáveis aleatórias absolutamente contínuas. A principal referência bibliográfica utilizada foi [11].

Apresentamos inicialmente o conceito geral de entropia relativa de X com respeito à Y, onde X e Y são v.a.'s definidas sobre $(\Omega, \mathcal{F}, P)$ tais que a medida de probabilidade induzida por X é absolutamente contínua com respeito à medida de probabilidade induzida por Y, ou seja, se $B \in \mathcal{B}$ é tal que $P_Y(B) = P(Y \in B) = 0$, então $P_X(B) = P(X \in B) = 0$.

Definição 2.5 *Sejam X e Y variáveis aleatórias com probabilidades induzidas μ e ν, respectivamente. A entropia relativa de X com relação a Y é definida por*

$$D(X||Y) = D(\mu||\nu) = \int \log \frac{d\mu}{d\nu} d\mu, \tag{2.9}$$

desde que μ seja absolutamente contínua com respeito a ν. Caso contrário, definimos $D(X||Y) = +\infty$.

Neste trabalho, temos interesse no caso em que X e Y são v.a.'s absolutamente contínuas. A seguinte proposição mostra como calcular a entropia relativa para esse caso.

Proposição 2.5 *Sejam X e Y variáveis aleatórias absolutamente contínuas sobre $(\Omega, \mathcal{F}, P)$ com densidades p e q, respectivamente. Então a entropia relativa de X com relação a Y é dada por*

$$D(X||Y) = D(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx, \tag{2.10}$$

desde que $q(x) = 0$ implique $p(x) = 0$ q.c. com respeito à medida de Lebesgue m. Caso contrário, $D(X||Y) = +\infty$.

Note que por (2.10) temos que a entropia relativa $D(X||Y)$ depende somente das densidades p e q e não das variáveis X e Y, por isso também nos referimos à entropia da densidade p com relação à q e denotamos $D(p||q)$.

Demonstração. Suponha que $q(x) = 0$ implica $p(x) = 0$ quase certamente e sejam μ e ν as probabilidades induzidas pelas variáveis aleatórias X e Y, respectivamente. Como

X e Y são absolutamente contínuas, então μ e ν são dados por

$$\mu(A) = \int_A p(x)dx, \quad \forall A \in \mathcal{F}.$$

e

$$\nu(A) = \int_A q(x)dx, \quad \forall A \in \mathcal{F}.$$

Se $\nu(N) = 0$ para algum $N \in \mathcal{F}$, então $q(x) = 0$ q.c. sobre N. Usando a hipótese, temos $p(x) = 0$ q.c. sobre N e assim $\mu(N) = 0$. Logo μ é absolutamente contínua com repeito à ν. Pelo Teorema de Radon-Nikodym, $\dfrac{d\mu}{d\nu} : \Omega \to \mathbb{R}$ é uma função $\mathcal{F}$-mensurável tal que

$$\mu(A) = \int_A \frac{d\mu}{d\nu} d\nu, \quad \forall A \in \mathcal{F}.$$

Por outro lado,

$$\mu(A) = \int_A p(x)dx = \int_A \frac{p(x)}{q(x)} q(x)dx = \int_A \frac{p(x)}{q(x)} d\nu(x).$$

Pela unicidade quase certa da derivada de Radon-Nikodym,

$$\frac{d\mu}{d\nu}(x) = \frac{p(x)}{q(x)} \quad q.c.$$

com respeito à medida de Lebesgue sobre $\mathbb{R}$. Logo, da Definição 2.5, temos

$$\begin{aligned} D(X||Y) &= \int \log \frac{d\mu}{d\nu} d\mu \\ &= \int \log \frac{p(x)}{q(x)} d\mu(x) \\ &= \int p(x) \log \frac{p(x)}{q(x)} dx. \end{aligned}$$

Por outro lado, se $q(x) = 0$ não implicar $p(x) = 0$ q.c., então o conjunto

$$B = \{x \in \mathbb{R} : q(x) = 0 \text{ e } p(x) \neq 0\}$$

não tem medida de Lebesgue nula. Logo, temos

$$\nu(B) = \int_B q(x)dx = 0,$$

mas

$$\mu(B) = \int_B p(x)dx > 0.$$

Portanto μ não é absolutamente contínua com respeito à ν. Neste caso, por definição, segue que $D(X||Y) = +\infty$. $\square$

Observação 2.2 *(a) Note que, a rigor, a integral em (2.10) é definida sobre o conjunto $\{x \in \mathbb{R} : p(x) \neq 0$ e $q(x) \neq 0\}$, já que $m(\{x \in \mathbb{R} : p(x) \neq 0$ e $q(x) = 0\}) = 0$.*

(b) Analogamente ao que foi feito na proposição anterior, podemos mostrar que no caso de X e Y serem variáveis aleatórias discretas com funções de probabilidade $p = (p_1, p_2, \ldots)$ e $q = (q_1, q_2, \ldots)$, respectivamente, a Definição 2.5 reduz-se a

$$D(X||Y) = D(p||q) = \sum_i p_i \log \frac{p_i}{q_i},$$

desde que $p_i = 0$ implique $q_i = 0$. No caso contrário, $D(X||Y) = +\infty$ por definição.

Daqui para frente, restringiremos nosso estudo ao caso em que X e Y são absolutamente contínuas e a entropia relativa $D(X||Y)$ é, então, definida como na Proposição 2.5.

Ao contrário da entropia diferencial, a entropia relativa está sempre bem definida e é maior ou igual a zero. É o que provamos no próximo teorema.

Teorema 2.1 *Sejam X e Y duas variáveis aleatórias absolutamente contínuas quaisquer com densidades p e q, respectivamente. Então*

$$D(X||Y) = D(p||q) \geq 0.$$

A igualdade é válida se, e somente se, $p = q$ quase certamente.

Demonstração. Denote $A = \{x \in \mathbb{R} : p(x) \neq 0\}$ e $B = \{x \in \mathbb{R} : q(x) \neq 0\}$. Se $A \backslash B = A \cap B^{\mathsf{c}}$ não tem medida nula, então $D(X||Y) = +\infty$ e o resultado segue. Suponha então que $A \backslash B$ tem medida nula, ou seja, $m(\{x \in \mathbb{R} : p(x) \neq 0$ e $q(x) = 0\}) = 0$. Neste caso, $D(X||Y)$ é dada por (2.10) e como $\log x \geq 1 - \frac{1}{x}$, $\forall x > 0$, (Proposição 1.8), segue que

$$\begin{aligned} D(X||Y) &= \int p(x) \log \left(\frac{p(x)}{q(x)} \right) \mathrm{d}x \\ &\geq \int p(x) \left(1 - \frac{q(x)}{p(x)} \right) \mathrm{d}x \\ &= 0. \end{aligned}$$

Se $D(X||Y) = 0$, como $\int p(x)\left(1 - \frac{q(x)}{p(x)}\right) dx = 0$, podemos escrever

$$\int p(x)\left[\log\frac{p(x)}{q(x)} - \left(1 - \frac{q(x)}{p(x)}\right)\right] dx = 0.$$

Como o integrando da integral anterior é uma função não-negativa, pois $p(x) > 0$ e $\log\frac{p(x)}{q(x)} \geq 1 - \frac{q(x)}{p(x)}$, segue pelas propriedades da integral de Lebesgue que

$$\log\frac{p(x)}{q(x)} - \left(1 - \frac{q(x)}{p(x)}\right) = 0$$

para quase-todo x. Novamente, pela Proposição 1.8, temos $\log x = 1 - \frac{1}{x}$ se, e somente se, $x = 1$, então segue

$$p(x) = q(x) \text{ para quase-todo } x.$$

Reciprocamente, se $p = q$ q.c. $[m]$, então segue diretamente de (2.10) que $D(X||Y) = 0$. $\square$

Como uma aplicação do Teorema 2.1, podemos mostrar que a entropia de uma v.a. com distribuição $N(\mu, \sigma^2)$ é o máximo das entropias de todas as v.a.'s absolutamente contínuas com variância σ^2, ou seja, temos o seguinte princípio de entropia máxima.

Corolário 2.1 *Se p é a densidade de uma variável aleatória X com variância σ^2 ($0 < \sigma^2 < +\infty$) e ψ_{σ^2} é a densidade $N(0, \sigma^2)$, então*

$$H(p) \leq H(\psi_{\sigma^2}) = \frac{\log(2\pi e\sigma^2)}{2}. \tag{2.11}$$

A igualdade vale se, e somente se, p é uma densidade Normal.

Demonstração. Podemos assumir que X tem esperança $\mu = 0$, pois caso contrário, basta considerar $Y = X - \mathbf{E}X$ e usar a Proposição 2.1 para obter $H(X) = H(Y)$.

Assumindo que $\mathbf{E}X = 0$ então $\mathrm{Var}X = \int x^2 p(x)dx = \sigma^2$. Pelo Teorema 2.1, temos $D(p||\psi_{\sigma^2}) \geq 0$. Então, como $\psi_{\sigma^2} = \dfrac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}$, $\forall x \in \mathbb{R}$, obtemos

$$\begin{aligned}
0 &\leq \int p(x)\log\frac{p(x)}{\psi_{\sigma^2}(x)}dx \\
&= \int p(x)[\log p(x) - \log\psi_{\sigma^2}(x)]dx \\
&= \int p(x)\left[\log p(x) + \frac{\log(2\pi\sigma^2)}{2} + \frac{x^2}{2\sigma^2}\right] dx.
\end{aligned}$$

Como $H(p) = -\int p(x)\log p(x)dx$ e $\sigma^2 = \int x^2 p(x)dx$, segue que

$$H(p) \leq \frac{\log(2\pi\sigma^2)+1}{2} = \frac{\log(2\pi\sigma^2 e)}{2}.$$

Pelo Exemplo 2.1 temos $H(\psi_{\sigma^2}) = \dfrac{\log(2\pi\sigma^2 e)}{2}$ e então temos (2.11). □

Na verdade a desigualdade (2.11) pode ser estendida para outras distribuições, além da Normal, como mostramos na proposição a seguir.

Proposição 2.6 *Seja X uma variável aleatória contínua com densidade p. Considere que ψ seja uma função densidade de probabilidade tal que $\psi(x) = 0$ implica $p(x) = 0$ q.c. e $\mathbf{E}\log^+(\frac{1}{\psi(X)}) < +\infty$. Então $H(X)$ existe, $H(X) < +\infty$ e satisfaz*

$$H(X) \leq \mathbf{E}\log\frac{1}{\psi(X)}. \tag{2.12}$$

Demonstração. Seja Z uma variável aleatória contínua com densidade ψ. Note que, se $\psi(x) \neq 0$, então

$$p(x)\log\frac{p(x)}{\psi(x)} + p(x)\log\frac{1}{p(x)} = p(x)\log\frac{1}{\psi(x)}. \tag{2.13}$$

Por hipótese, $\mathbf{E}\log^+(\frac{1}{\psi(X)}) = \int p(x)\log^+(\frac{1}{\psi(x)})\mathrm{d}x < +\infty$. Além disso, pelo Teorema 2.1, segue que $\mathbf{E}\log^-(\frac{p(X)}{\psi(X)}) = \int p(x)\log^-(\frac{p(x)}{\psi(x)})\mathrm{d}x < +\infty$. Pela equação (2.13), temos

$$p(x)\log^+\left(\frac{1}{p(x)}\right) \leq p(x)\log^+\left(\frac{1}{\psi(x)}\right) + p(x)\log^-\left(\frac{p(x)}{\psi(x)}\right).$$

Logo $\mathbf{E}\log^+(\frac{1}{p(X)}) = \int p(x)\log^+(\frac{1}{p(x)})\mathrm{d}x < +\infty$ e, dessa forma, $H(X)$ existe e não assume $+\infty$. Novamente, pela equação (2.13), temos

$$\begin{aligned} &p(x)\log^+\left(\frac{p(x)}{\psi(x)}\right) + p(x)\log^+\left(\frac{1}{p(x)}\right) + p(x)\log^-\left(\frac{1}{\psi(x)}\right) \\ &= p(x)\log^-\left(\frac{p(x)}{\psi(x)}\right) + p(x)\log^-\left(\frac{1}{p(x)}\right) + p(x)\log^+\left(\frac{1}{\psi(x)}\right). \end{aligned} \tag{2.14}$$

Integrando os termos em (2.14) e reordenando-os, obtemos

$$H(X) = \mathbf{E}\log\frac{1}{\psi(X)} - D(X||Z). \tag{2.15}$$

Como $D(X||Z) \geq 0$ pelo Teorema 2.1, segue que

$$H(X) \leq \mathbf{E}\log\frac{1}{\psi(X)},$$

e como $\mathbf{E}\log^+(\frac{1}{\psi(X)}) < +\infty$, segue que $H(X) < +\infty$. $\square$

2.4 Propriedades Auxiliares

Nesta seção apresentamos propriedades especiais de entropia relativa que foram obtidas por *Bobkov et al.* em [4] e que auxiliarão no desenvolvimento dos resultados do Capítulo 3 sobre convergência em entropia relativa para variáveis aleatórias estáveis. Basicamente, os resultados desta seção referem-se à estabelecer condições para a finitude da entropia relativa.

Proposição 2.7 *Sejam X e Z variáveis aleatórias absolutamente contínuas com densidade p e ψ, respectivamente. Suponha que*

(a) $\psi(x) = 0 \Rightarrow p(x) = 0$ *q.c.;*

(b) $\mathbf{E}\log^+(\frac{1}{\psi(X)}) < +\infty$*;*

(c) H(X) é finita.

Então $D(X||Z)$ é finita e dada por

$$D(X||Z) = \mathbf{E}\log\frac{1}{\psi(X)} - H(X).$$

Reciprocamente, se ψ é limitada, existe $0 < \gamma < 1$ *tal que* $\int \psi(x)^\gamma dx < +\infty$ *e $D(X||Z)$ é finita, então valem (a), (b) e (c).*

Demonstração. Suponha que (a), (b) e (c) são válidos. Pela Proposição 2.6 e por (2.15), $H(X)$ existe e é dada por

$$H(X) = \mathbf{E}\log\frac{1}{\psi(X)} - D(X||Z).$$

Se $D(X||Z)$ não é finita, então $D(X||Z) = +\infty$ e $H(X) = -\infty$, contradizendo (c). Logo $D(X||Z)$ é finita e $D(X||Z) = \mathbf{E}\log\frac{1}{\psi(X)} - h(X)$. Reciprocamente, suponha que ψ é limitada, existe $0 < \gamma < 1$ tal que $\int \psi(x)^\gamma dx < +\infty$ e $D(X||Z)$ é finita. Pela definição de entropia relativa, a condição (a) é necessária para que $D(X||Z)$ seja finita. Para mostrar (b), divida a reta real em dois conjuntos $A = \{x \in \mathbb{R} : p(x) \geq \psi(x)^{\gamma'}\}$ e

$B = \{x \in \mathbb{R} : p(x) < \psi(x)^{\gamma'}\}$, onde $\gamma < \gamma' < 1$. Então

$$\begin{aligned}
\int p(x) \log^+ \left(\frac{p(x)}{\psi(x)}\right) dx &\geq \int_A p(x) \log^+ \left(\frac{p(x)}{\psi(x)}\right) dx \\
&\geq \int_A p(x) \log^+ \left(\frac{1}{\psi(x)^{1-\gamma'}}\right) dx \\
&= (1-\gamma') \int_A p(x) \log^+ \left(\frac{1}{\psi(x)}\right) dx.
\end{aligned}$$

Logo $\int_A p(x) \log^+ \left(\frac{1}{\psi(x)}\right) dx < +\infty$. Por outro lado, pela Proposição 1.9, existe uma constante $C \geq 1$ tal que $t^{\gamma'} \log\left(\frac{1}{t}\right) \leq Ct^{\gamma}$ para todo $t \in (0, 1]$. Então

$$\begin{aligned}
\int_B p(x) \log^+ \left(\frac{1}{\psi(x)}\right) dx &\leq \int_B \psi(x)^{\gamma'} \log^+ \left(\frac{1}{\psi(x)}\right) dx \\
&= \int_B \psi(x)^{\gamma'} \log \left(\frac{1}{\psi(x)}\right) 1_{\{\psi(x) \geq 1\}} dx \\
&\leq C \int_B \psi(x)^{\gamma} 1_{\{\psi(x) \geq 1\}} dx \\
&\leq C \int_{-\infty}^{+\infty} \psi(x)^{\gamma} dx < +\infty.
\end{aligned}$$

Portanto vale (b). Como ψ é limitada, existe $k > 0$ tal que $\psi(x) \leq k$ para todo $x \in \mathbb{R}$. Logo

$$\mathbf{E} \log \frac{1}{\psi(X)} = \int p(x) \log \frac{1}{\psi(x)} dx \geq \int p(x) \log \frac{1}{k} dx = \log \frac{1}{k} > -\infty.$$

Dessa forma, $\mathbf{E} \log \frac{1}{\psi(X)}$ é finita e, por (2.15), concluímos (c). □

Como consequência, podemos mostrar que, em particular, se Z é uma v.a. não-extremal as condições (a), (b) e (c) são necessárias e suficientes para a finitude da entropia relativa com respeito à ψ.

Corolário 2.2 *Sejam X e Z variáveis aleatórias absolutamente contínuas com densidade p e ψ, respectivamente, e suponha que Z é estável não-extremal. Então $D(X||Z)$ é finito se, e somente se, valem (a), (b) e (c) da Proposição 2.7. Neste caso,*

$$D(X||Z) = \mathbf{E} \log \frac{1}{\psi(X)} - H(X).$$

Demonstração. Pela Proposição 2.7 as condições (a), (b) e (c) são suficientes para a finitude de $D(X||Z)$, para qualquer Z absolutamente contínua. Para provar a suficiência, no caso de Z estável não-extremal, é suficiente mostrar que, ψ é limitada e existe $0 < \gamma < 1$ tal que $\int \psi(x)^{\gamma} dx < +\infty$. De fato, considere $f(t) = \mathbf{E}e^{itZ}$ a função característica de ψ. Pelo Corolário 1.3, ψ é limitada. Logo existe $M > 0$ tal que $\psi(t) \leq M$. Se Z tem distribuição normal, então para todo $0 < \gamma < 1$,

$$\int \psi(x)^\gamma dx = (2\pi\sigma^2)^{-\frac{\gamma}{2}} \int \exp\left(-\frac{\gamma(x-\mu)^2}{2\sigma^2}\right) dx < +\infty.$$

Se Z não tem distribuição normal, então, pela Proposição 1.4, ψ satisfaz as relações assintóticas

$$\psi(x) \sim c_0|x|^{-(1+\alpha)} \quad (x \to -\infty), \qquad \psi(x) \sim c_1 x^{-(1+\alpha)} \quad (x \to +\infty),$$

para determinadas constantes $c_0, c_1 > 0$. Tome $\frac{1}{1+\alpha} < \gamma < 1$. Por continuidade, segue que

$$\psi(x)^\gamma \sim c_0^\gamma |x|^{-\gamma(1+\alpha)} \quad (x \to -\infty), \qquad \psi(x)^\gamma \sim c_1^\gamma x^{-\gamma(1+\alpha)} \quad (x \to +\infty).$$

Logo, existe $A > 0$ tal que, para todo $x < -A$, temos

$$\left|\frac{\psi(x)^\gamma}{c_0^\gamma |x|^{-\gamma(1+\alpha)}} - 1\right| < \frac{1}{2}$$

e assim,

$$\psi(x)^\gamma < \frac{3}{2} c_0^\gamma |x|^{-\gamma(1+\alpha)}, \quad \forall x < -A. \tag{2.16}$$

Também, existe $B > 0$ tal que, para todo $x > B$, temos

$$\left|\frac{\psi(x)^\gamma}{c_1^\gamma x^{-\gamma(1+\alpha)}} - 1\right| < \frac{1}{2}$$

e então,

$$\psi(x)^\gamma < \frac{3}{2} c_1^\gamma x^{-\gamma(1+\alpha)}, \quad \forall x > B. \tag{2.17}$$

De (2.16) e (2.17), segue que

$$\begin{aligned} \int_{-\infty}^{+\infty} \psi(x)^\gamma dx &= \int_{-\infty}^{-A} \psi(x)^\gamma dx + \int_{-A}^{B} \psi(x)^\gamma dx + \int_{B}^{+\infty} \psi(x)^\gamma dx \\ &\leq \int_{-\infty}^{-A} \frac{3}{2} c_0^\gamma |x|^{-\gamma(1+\alpha)} dx + (A+B)M^\gamma + \int_{B}^{+\infty} \frac{3}{2} c_1^\gamma x^{-\gamma(1+\alpha)} < +\infty, \end{aligned}$$

o que conclui a prova do corolário. □

Na verdade, no caso da densidade ψ ser não-extremal, é possível obter condições necessárias e suficientes para a finitude da entropia relativa com respeito à ψ, mais específicas e simples, tanto para o caso normal, quanto para o caso não-normal

Primeiramente, como consequência do Corolário 2.2 podemos provar que a finitude do segundo momento e da entropia de X é necessária e suficiente para a finitude da entropia relativa de X com respeito a uma v.a. Normal.

Proposição 2.8 *Se Z tem distribuição normal, então $D(X||Z) < +\infty$ se, e somente se, $\mathbf{E}X^2 < +\infty$ e $H(X)$ é finita.*

Demonstração. Seja ψ a densidade de Z. Se $D(X||Z) < +\infty$, então pelo Corolário 2.2, $\mathbf{E}\log^+(\frac{1}{\psi(X)}) < +\infty$ e $H(X)$ é finita. Note que para $\psi(x)$ a densidade normal temos

$$\begin{aligned}
\mathbf{E}\log\frac{1}{\psi(X)} &= -\int p(x)\log\psi(x)\mathrm{d}x \\
&= -\int p(x)\log\left(\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right)\mathrm{d}x \\
&= -\int p(x)\log\left(\frac{1}{\sqrt{2\pi}\sigma}\right)\mathrm{d}x + \int p(x)\left(\frac{(x-\mu)^2}{2\sigma^2}\right)\mathrm{d}x \\
&= \log(\sqrt{2\pi}\sigma) + \frac{1}{2\sigma^2}\mathbf{E}(X-\mu)^2. \qquad (2.18)
\end{aligned}$$

Logo, como $\mathbf{E}\log^+(\frac{1}{\psi(X)}) < +\infty$, temos $\mathbf{E}(X-\mu)^2 < +\infty$, o que implica $\mathbf{E}X^2 < +\infty$. Reciprocamente, suponha que $\mathbf{E}X^2 < +\infty$ e $H(X)$ é finita. Pelo Corolário 2.2, basta mostrar (a), (b) e (c) para concluir que $D(X||Z) < +\infty$. Claramente, (a) é válido, pois $\psi(x) > 0$ para todo $x \in \mathbb{R}$. De (2.18), segue que $\mathbf{E}\log\frac{1}{\psi(X)}$ existe e é finito, de modo que (b) é verdadeiro. Por fim, vale (c) por hipótese. □

Observação 2.3 *Pela Proposição 1.7 e por (2.18), segue que*

$$\mathbf{E}\log\frac{1}{\psi(X)} \leq \log(\sqrt{2\pi}\sigma) + \frac{1}{\sigma^2}(\mathbf{E}X^2 + \mu^2).$$

Para o caso em que Z é estável não-extremal e não-normal, para obter condições alternativas para a finitude de $D(X||Z)$, necessitamos de um lema auxiliar.

Lema 2.1 *Seja X uma variável aleatória absolutamente contínua. Dadas constantes $a \geq 0$ e $b > 0$, então $\mathbf{E}\log^+(|X|) < +\infty$ se, e somente se, $\mathbf{E}\log^+(a + b|X|) < +\infty$.*

Demonstração. Denote por p a densidade de X e suponha que $\mathbf{E}\log^+(|X|) < +\infty$. Sejam $A = \{x \in \mathbb{R} : |x| > 1\}$ e $B = \{x \in \mathbb{R} : |x| \leq 1\}$. Note que, se $x > 1$, então $(a+b)x \geq a + bx$. Logo $\log(a+bx) \leq \log(x) + \log(a+b)$, $\forall x > 1$. Então, usando esta desigualdade podemos obter

$$\begin{aligned}
\mathbf{E}\log^+(a+b|X|) &= \int_A p(x)\log^+(a+b|x|)dx + \int_B p(x)\log^+(a+b|x|)dx \\
&\leq \int_A p(x)(\log|x| + \log(a+b))1_{\{a+b|x|\geq 1\}}dx \\
&+ \int_B p(x)\log(a+b)1_{\{a+b|x|\geq 1\}}dx \\
&\leq \int p(x)\log(|x|)dx + 2\log(a+b) < +\infty.
\end{aligned}$$

Reciprocamente, suponha que $\mathbf{E}\log^+(a+b|X|) < +\infty$. Então

$$\mathbf{E}\log^+(|X|) = \int p(x)\log^+(|x|)dx = \int p(x)\log(|x|)1_{\{|x|\geq 1\}}dx.$$

Denote $C = \log b \int p(x)1_{\{|x|\geq 1\}}dx$. Logo

$$\begin{aligned}
\mathbf{E}\log^+(|X|) &= \int p(x)\log(b|x|)1_{\{|x|\geq 1\}}dx - C \\
&\leq \int p(x)\log(a+b|x|)1_{\{|x|\geq 1\}}dx - C \\
&\leq \int p(x)\log^+(a+b|x|)dx - C < +\infty.
\end{aligned}$$

□

Observação 2.4 *Em particular, se $a = b = 1$, então $\mathbf{E}\log^+(|X|) < +\infty$ se, e somente se, $\mathbf{E}\log(1+|X|) < +\infty$.*

Como consequência do Corolário 2.2 e usando o Lema 2.1 podemos mostrar que a finitude do momento logaritmico $\mathbf{E}\log(1+|X|)$ e a finitude da entropia de X são condições necessárias e suficientes para a finitude da entropia relativa de X com respeito à Z estável não-extremal e não-normal.

Proposição 2.9 *Se Z tem uma distribuição estável não-extremal, a qual não é normal, então $D(X||Z) < +\infty$ se, e somente se $\mathbf{E}\log(1+|X|) < +\infty$ e $H(X)$ é finita.*

Demonstração. Seja ψ a densidade de Z. Se $D(X||Z) < +\infty$, então pelo Corolário 2.2, $\mathbf{E}\log^+(\frac{1}{\psi(X)}) < +\infty$ e $H(X)$ é finita. Pela Proposição 1.4, existem constantes $c_0, c_1 > 0$ tais que

$$\psi(x) \sim c_0|x|^{-(1+\alpha)} \quad (x \to -\infty), \qquad \psi(x) \sim c_1 x^{-(1+\alpha)} \quad (x \to +\infty).$$

Logo existe uma constante $C > 1$ tal que, para $x < -C$, temos

$$\left|\frac{\psi(x)}{c_0|x|^{-(1+\alpha)}} - 1\right| < \frac{1}{2}$$

e para $x > C$, temos

$$\left|\frac{\psi(x)}{c_1 x^{-(1+\alpha)}} - 1\right| < \frac{1}{2}.$$

Logo,

$$\frac{2}{3c_0}|x|^{(1+\alpha)} < \frac{1}{\psi(x)}, \quad \forall x < -C \tag{2.19}$$

e

$$\frac{2}{3c_1}x^{(1+\alpha)} < \frac{1}{\psi(x)}, \quad \forall x > C. \tag{2.20}$$

Por um lado, usando (2.19), obtemos

$$\begin{aligned}
\int_{-\infty}^{-C} p(x)\log^+(|x|)dx &= \int_{-\infty}^{-C} p(x)\log(|x|)dx \qquad (2.21)\\
&= \frac{1}{\alpha+1}\int_{-\infty}^{-C} p(x)\log(|x|^{(1+\alpha)})dx\\
&\leq \frac{1}{\alpha+1}\int_{-\infty}^{-C} p(x)\log\left(\frac{\frac{3}{2}c_0}{\psi(x)}\right)dx\\
&\leq \frac{1}{\alpha+1}\log^+\left(\frac{3c_0}{2}\right) + \frac{1}{\alpha+1}\int_{-\infty}^{-C} p(x)\log\left(\frac{1}{\psi(x)}\right)dx < +\infty.
\end{aligned}$$

Por outro lado, usando (2.20) podemos obter

$$\begin{aligned}
\int_{C}^{+\infty} p(x)\log^+(|x|)dx &= \int_{C}^{+\infty} p(x)\log(x)dx \qquad (2.22)\\
&= \frac{1}{\alpha+1}\int_{C}^{+\infty} p(x)\log(x^{(1+\alpha)})dx\\
&\leq \frac{1}{\alpha+1}\int_{C}^{+\infty} p(x)\log\left(\frac{\frac{3}{2}c_1}{\psi(x)}\right)dx\\
&\leq \frac{1}{\alpha+1}\log^+\left(\frac{3c_1}{2}\right) + \frac{1}{\alpha+1}\int_{C}^{+\infty} p(x)\log\left(\frac{1}{\psi(x)}\right)dx < +\infty.
\end{aligned}$$

Além disso, temos que

$$\int_{-C}^{C} p(x)\log^+(|x|)dx \leq \int_{-C}^{C} p(x)\log(C) \leq \log(C) < +\infty. \tag{2.23}$$

Assim, por (2.21), (2.22) e (2.23), $\mathbf{E}\log^+(|X|) < +\infty$, o que implica $\mathbf{E}\log(1 + |X|) < +\infty$ pela Observação 2.4. Reciprocamente, suponha que

$\mathbf{E}\log(1+|X|) < +\infty$ e $H(X)$ é finita. Pelo Corolário 2.2, basta mostrar (a), (b) e (c) para concluir que $D(X||Z) < +\infty$. Claramente, (a) é válido, pois $\psi(x) > 0$ para todo $x \in \mathbb{R}$. Por hipótese, vale (c). Pelo Corolário 1.4, existe uma constante $c > 0$ tal que

$$c(1+|x|)^{-(1+\alpha)} \leq \psi(x), \quad \forall x \in \mathbb{R}.$$

Logo

$$\begin{aligned} \mathbf{E}\log^+\left(\frac{1}{\psi(X)}\right) &= \int p(x)\log^+\left(\frac{1}{\psi(x)}\right)dx \\ &\leq \log^+\frac{1}{c} + (1+\alpha)\int p(x)\log(1+|x|)dx < +\infty. \end{aligned} \tag{2.24}$$

Portanto vale (b) e a proposição está demonstrada. □

Observação 2.5 *Por (2.24), segue que existem constantes A e B tais que*

$$\mathbf{E}\log^+\left(\frac{1}{\psi(X)}\right) \leq A + B\mathbf{E}\log(1+|X|).$$

Finalizamos esta seção aplicando os resultados anteriores para a análise da finitude da entropia relativa da soma parcial normalizada Z_n de variáveis aleatórias independentes com respeito a uma variável aleatória estável não-extremal Z. Os resultados a seguir serão fundamentais para a obtenção do teorema principal do próximo capítulo que estabelece condições para a convergência em entropia relativa de Z_n a Z, sob a hipótese que $Z_n \xrightarrow{\mathfrak{D}} Z$.

Assim, no restante da seção, consideramos a soma parcial normalizada

$$Z_n = \frac{S_n}{b_n} - a_n, \quad n \geq 1,$$

em que $S_n = X_1 + \ldots + X_n$, com $X_1, \ldots, X_n$ variáveis aleatórias i.i.d. e a_n e b_n são constantes reais com $b_n > 0$.

O primeiro resultado estabelece condições necessárias e suficientes para $D(Z_n||Z) < +\infty$, para cada $n \geq 1$, quando $Z_n \xrightarrow{\mathfrak{D}} Z$.

Proposição 2.10 *Suponha que Z_n converge em distribuição para uma v.a. Z estável não-extremal, a qual não é normal. Então, para cada $n \geq 1$, $D(Z_n||Z) < +\infty$ se, e somente se, $H(Z_n)$ é finita.*

Demonstração. Se $D(Z_n||Z) < +\infty$, então $H(Z_n)$ é finita pelo Corolário 2.2. Reciprocamente, suponha que $H(Z_n)$ é finita. Como $Z_n \xrightarrow{\mathfrak{D}} Z$ e Z é α-estável, então pelo

Teorema 1.9 segue que $\mathbf{E}|Z_n|^s < +\infty$ para todo $0 < s < \alpha$. Tome $0 < s < \min\{1, \alpha\}$. Então, pela Proposição 1.9, existe $C \geq 1$ tal que

$$\mathbf{E}\log(1 + |Z_n|) \leq C\mathbf{E}|Z_n|^s < +\infty.$$

Usando a Proposição 2.9, concluímos a demonstração. $\square$

Observe que uma conclusão análoga à da Proposição 2.10 pode ser obtida para o caso em que $Z_n \xrightarrow{\mathfrak{D}} Z$ onde Z tem distribuição Normal padrão e $b_n \sim \sqrt{n}$, pois neste caso, pelo Teorema 1.8 temos $\mathbf{E}X_1^2 < +\infty$ e a conclusão segue da Proposição 2.8.

Proposição 2.11 *Suponha que Z_n converge em distribuição para uma v.a. Z estável não-extremal. Se a entropia relativa $D(Z_n||Z)$ é finita para algum $n = n_0$, então ela será finita para todo $n \geq n_0$.*

Demonstração. Suponha que $D(Z_{n_0}||Z) < +\infty$. Pela Proposição 2.1 temos $H(S_{n_0}) = H(Z_{n_0}) + \log b_{n_0}$ e então pelo Corolário 2.2 segue que $H(S_{n_0})$ é finita. Considere primeiro o caso em que Z não é normal. Seguindo os mesmos passos da demonstração da Proposição 2.10, como $Z_n \xrightarrow{\mathfrak{D}} Z$ e Z é α-estável, então pelo Teorema 1.9 temos que $\mathbf{E}|Z_n|^s < +\infty$ para todo $0 < s < \alpha$. Pela Proposição 1.9, se escolhermos $0 < s < \min\{1, \alpha\}$, então existe $C \geq 1$ tal que

$$\mathbf{E}\log(1 + |Z_n|) \leq C\mathbf{E}|Z_n|^s < +\infty, \quad \forall n \geq 1$$

Logo, se ψ é a densidade de Z, de acordo com a Observação 2.5 segue que $\mathbf{E}\log^+\left(\frac{1}{\psi(Z_n)}\right) < +\infty$ e pela Proposição 2.6, obtemos $H(Z_n) < +\infty$ para todo $n \geq 1$. Como $X_1, X_2, \ldots$ são independentes, segue da Proposição 2.4 que

$$H(S_n) \geq H(S_{n_0}) > -\infty,$$

para todo $n \geq n_0$. Assim, $H(Z_n) = H(S_n) - \log b_n$ é finita sempre que $n \geq n_0$. Da Proposição 2.10, segue então que $D(Z_n||Z)$ é finita para todo $n \geq n_0$.

Agora, suponha que Z é normal. Pela Proposição 2.8, $\mathbf{E}Z_{n_0}^2 < +\infty$, o que implica que

$$\mathbf{E}S_{n_0}^2 = b_{n_0}^2\mathbf{E}(Z_{n_0} + a_{n_0})^2 \leq 2b_{n_0}^2(\mathbf{E}Z_{n_0}^2 + a_{n_0}^2) < +\infty$$

e, como $X_1, \ldots, X_n$ são i.i.d., segue que

$$\mathbf{E}X_1^2 \leq n_0\mathbf{E}X_1^2 + n_0(n_0 - 1)(\mathbf{E}X_1)^2 = \mathbf{E}S_{n_0}^2 < +\infty.$$

Logo, para todo $n \geq 1$, pela Proposição 1.7, temos

$$\mathbf{E}S_n^2 \leq n\mathbf{E}(X_1^2 + \ldots + X_n^2) = n^2\mathbf{E}X_1^2 < +\infty$$

e podemos obter

$$\mathbf{E}Z_n^2 \leq b_n^2\mathbf{E}S_n^2 + a_n^2 < +\infty.$$

Analogamente aos argumentos acima, para o caso não normal, da Observação 2.3 e pela Proposição 2.6, segue que $H(S_n) < +\infty$. Também, a independência das variáveis $X_1, X_2, \ldots$ garante que

$$H(S_n) \geq H(S_{n_0}) > -\infty,$$

para todo $n \geq n_0$. Logo $H(Z_n) = H(S_n) - \log b_n$ é finito para todo $n \geq n_0$. Pela Proposição 2.8, concluímos que $D(Z_n||Z) < +\infty$ para todo $n \geq n_0$. $\square$

Capítulo 3

Convergência em Entropia Relativa

3.1 Introdução

Seja Z uma variável aleatória estável não-degenerada. Pelo Teorema 1.6, existem uma sequência de variáveis aleatórias i.i.d. $\{X_n\}_{n\geq 1}$, com função de distribuição comum F, e sequências de constantes $\{a_n\}_{n\geq 1}$ e $\{b_n\}_{n\geq 1}$, com $b_n > 0$, tais que

$$Z_n \overset{\mathfrak{D}}{\to} Z, \tag{3.1}$$

onde $\{Z_n\}_{n\geq 1}$ é a sequência de somas normalizadas

$$Z_n = \frac{X_1 + \ldots + X_n}{b_n} - a_n, \quad n \geq 1. \tag{3.2}$$

Os teoremas 1.7 e 1.8 fornecem condições necessárias e suficientes para que uma função de distribuição F esteja no domínio de atração de uma distribuição α-estável, ou seja, para que (3.1) seja válida.

Assumindo que a convergência em (3.1) é satisfeita, uma questão de grande interesse na literatura é se essa convergência é válida em um sentido mais forte. Resultados sobre a convergência na norma em L^1 e a norma de variação total podem ser encontrados em [10]. Como observamos na introdução do Capítulo 2, convergência em entropia relativa é mais forte do que na norma de variação total. Neste sentido, Barron [2] obteve a convergência de Z_n à Z em entropia relativa para o caso em que Z tem distribuição normal. Especificamente, seu principal resultado é o seguinte:

Teorema 3.1 *Sejam $X_1, X_2, \ldots$ variáveis aleatórias i.i.d. com média zero e variância $0 < \sigma^2 < +\infty$ e seja a sequência de somas normalizadas Z_n, como em (3.2), com $a_n = 0$ e $b_n = \sigma\sqrt{n}$. Então $D(Z_n||N(0,1)) \underset{n\to\infty}{\longrightarrow} 0$ se, e somente se, $D(Z_n||N(0,1)) < +\infty$ para algum $n \geq 1$, onde $N(0,1)$ indica a variável aleatória com distribuição normal padrão.*

É claro que o resultado também é válido no caso em que as variáveis X_i's têm média μ não-nula. Neste caso, $a_n = n\mu$.

Observe que a hipótese $0 < \sigma^2 < +\infty$ implica, pelo Teorema do Limite Central, que $Z_n \stackrel{\mathfrak{D}}{\to} N(0,1)$. Assim, o teorema de Barron nos diz que se $Z_n \stackrel{\mathfrak{D}}{\to} N(0,1)$, com $b_n = \sigma\sqrt{n}$, então Z_n converge em entropia relativa para $N(0,1)$ se, e somente se, a entropia relativa entre Z_n e $N(0,1)$ é finita para algum n.

Uma extensão do resultado de Barron foi obtida por Bobkov et al. [4] para o caso em que $Z_n \stackrel{\mathfrak{D}}{\to} Z$ e Z é estável não-extremal, ou seja, Z é normal ou Z é α-estável, com $0 < \alpha < 2$ e com parâmetro de simetria $-1 < \beta < 1$.

O objetivo deste capítulo é apresentar em detalhes os resultados obtidos em [4].

Assim, nas seções 3.2 e 3.3 apresentamos basicamente os resultados auxiliares obtidos por Bobkov et al. que serão utilizados na demonstração do teorema principal, que será apresentada na Seção 3.4.

3.2 Decomposição Binomial das Convoluções

Sejam $X_1, X_2, \ldots$ variáveis aleatórias i.i.d. e $\{a_n\}_{n\geq 1}$ e $\{b_n\}_{n\geq 1}$ sequências de números reais com $b_n > 0$. Considere a sequência de somas parciais
$S_n = X_1 + \ldots + X_n,\ n \geq 1$ e a sequência de somas normalizadas Z_n, como em (3.2), ou seja,

$$Z_n = \frac{S_n}{b_n} - a_n, \quad n \geq 1.$$

Vamos assumir que as variáveis $X_n,\ n \geq 1$ tem densidade comum $p(x)$ e Z_n tem densidade $p_n(x)$ para cada $n \geq 1$.

O objetivo desta seção é obter uma aproximação para $p_n(x)$ por uma densidade limitada $\widetilde{p}_n(x)$.

Para isso, seja um número real fixo $0 < b < \frac{1}{2}$ e considere dois conjuntos de Borel da reta H_0 e H_1, de tal forma que para alguma constante $M > 0$ temos

$$p(x) \leq M, \quad \forall x \in H_1$$

e

$$1 - b = \int_{H_1} p(x)dx.$$

Observe que a existência de H_0 e H_1 é garantida pelo fato de $p(x)$ ser uma densidade de probabilidade. Defina, para todo $x \in \mathbb{R}$,

$$\rho_0(x) = \frac{p(x)}{b} 1_{H_0}(x) \quad \text{e} \quad \rho_1(x) = \frac{p(x)}{1-b} 1_{H_1}(x)$$

as restrições normalizadas de p nos conjuntos H_0 e H_1, respectivamente. Então ρ_0, ρ_1 são densidades e temos

$$p(x) = (1-b)\rho_1 + b\rho_0. \tag{3.3}$$

Por (3.3) e pelas Proposições 1.1 e 1.2, podemos escrever a convolução n vezes de p da forma

$$p^{n*} = \sum_{k=0}^{n} \binom{n}{k} (1-b)^k b^{n-k} \rho_1^{k*} * \rho_0^{(n-k)*}.$$

Pelo Teorema 1.2, temos que p^{n*} é a densidade de S_n. Para $n \geq 2$, podemos escrever

$$p^{n*} = \rho_{n0} + \rho_{n1}, \tag{3.4}$$

com

$$\rho_{n1} = \sum_{k=2}^{n} \binom{n}{k} (1-b)^k b^{n-k} \rho_1^{k*} * \rho_0^{(n-k)*}, \tag{3.5}$$

$$\begin{aligned} \rho_{n0} &= \sum_{k=0}^{1} \binom{n}{k} (1-b)^k b^{n-k} \rho_1^{k*} * \rho_0^{(n-k)*} \\ &= b^n \rho_0^{*n} + n(1-b)b^{n-1}\rho_1 * \rho_0^{*(n-1)}. \end{aligned} \tag{3.6}$$

Então, como $P(Z_n \leq x) = P(S_n \leq a_n b_n + b_n x)$, derivando em relação à x e usando (3.4) obtemos

$$p_n(x) = b_n \rho_{n1}(a_n b_n + b_n x) + b_n \rho_{n0}(a_n b_n + b_n x). \tag{3.7}$$

Agora, para $\varepsilon_n = \int \rho_{n0}(x)dx$, temos

$$\varepsilon_n = b^n + n(1-b)b^{n-1} < nb^{n-1} \tag{3.8}$$

e

$$1 - \varepsilon_n = \sum_{k=2}^{n} \binom{n}{k} (1-b)^k b^{n-k}. \tag{3.9}$$

Então, por (3.7), (3.8) e (3.9), as funções definidas por

$$\widetilde{p}_n(x) = \frac{b_n}{1-\varepsilon_n}\rho_{n1}(a_n b_n + b_n x), \qquad p_{n0}(x) = \frac{b_n}{\varepsilon_n}\rho_{n0}(a_n b_n + b_n x) \tag{3.10}$$

são densidades tais que

$$p_n(x) = (1-\varepsilon_n)\widetilde{p}_n(x) + \varepsilon_n p_{n0}(x). \tag{3.11}$$

Como ρ_1 é limitada, segue pela Proposição 1.3 que ρ_{n1} e $\widetilde{p}_n$ são limitadas. Além disso, por (3.11), temos

$$|\widetilde{p}_n(x) - p_n(x)| = \varepsilon_n |\widetilde{p}_n(x) - p_{n0}(x)| \leq \varepsilon_n(\widetilde{p}_n(x) + p_{n0}(x)). \tag{3.12}$$

Assim, de (3.8), segue que

$$\int |\widetilde{p}_n(x) - p_n(x)| dx \leq 2\varepsilon_n < 2nb^{n-1} = o(2^{-n}).$$

Logo, para n grande o suficiente, temos

$$\int |\widetilde{p}_n(x) - p_n(x)| dx < 2^{-n}, \tag{3.13}$$

ou seja, a medida que n cresce, $\widetilde{p}_n(x)$ é uma boa aproximação de p_n. Agora, se considerarmos as funções características

$$\widetilde{f}_n(t) = \int e^{itx} \widetilde{p}_n(x) dx$$

e

$$f_n(t) = \int e^{itx} p_n(x) dx,$$

então por (3.13), podemos obter para n suficientemente grande a seguinte estimativa para as funções características

$$|\widetilde{f}_n(t) - f_n(t)| \leq \int |\widetilde{p}_n(x) - p_n(x)| dx < 2^{-n}. \tag{3.14}$$

Veremos no lema a seguir que a desigualdade (3.13) pode ser refinada, usando a distância L^1 ponderada por uma função peso tipo potência.

Lema 3.1 *Se* $\mathbf{E}|X_1|^s < +\infty$ $(s > 0)$ *e* $|a_n| b_n + 1/b_n = o(n^\gamma)$ *para algum* $\gamma > 0$*, então para todo* n *suficientemente grande temos*

$$\int |x|^s |\widetilde{p}_n(x) - p_n(x)| dx < 2^{-n}. \tag{3.15}$$

Demonstração. Por (3.12) e (3.10), temos

$$|\widetilde{p}_n(x) - p_n(x)| \leq \varepsilon_n(\widetilde{p}_n(x) + p_{n0}(x)) = \varepsilon_n \left(\frac{b_n}{1 - \varepsilon_n} \rho_{n1}(a_n b_n + b_n x) + \frac{b_n}{\varepsilon_n} \rho_{n0}(a_n b_n + b_n x) \right).$$

Logo,

$$\int |x|^s |\widetilde{p}_n(x) - p_n(x)| \leq \frac{\varepsilon_n}{1 - \varepsilon_n} \int |x|^s b_n \rho_{n1}(a_n b_n + b_n x) dx + \int |x|^s b_n \rho_{n0}(a_n b_n + b_n x) dx.$$

Fazendo a mudança de variáveis $y = a_n b_n + b_n x$, obtemos $dy = b_n dx$ e assim,

$$\begin{aligned}\int |x|^s |\widetilde{p}_n(x) - p_n(x)| \leq \frac{\varepsilon_n}{1-\varepsilon_n} b_n^{-s} \int |y - a_n b_n|^s \rho_{n1}(y) dy \\ + b_n^{-s} \int |y - a_n b_n|^s \rho_{n0}(y) dy.\end{aligned} \tag{3.16}$$

Para obter (3.15), vamos obter limitantes superiores para as duas integrais do lado direito de (3.16). Para isso, sejam U e V variáveis aleatórias com densidades ρ_1 e ρ_0, respectivamente. Considere $U_1, U_2, \ldots$ cópias independentes de U e $V_1, V_2, \ldots$ cópias independentes de V (as quais também são independentes de $U_1, U_2, \ldots$). Como $\mathbf{E}|X_1|^s < +\infty$, por (3.3), temos

$$\begin{aligned}\mathbf{E}|X_1|^s &= \int |x|^s p(x) dx \\ &= \int |x|^s [(1-b)\rho_1(x) + b\rho_0(x)] dx \\ &= (1-b) \int |x|^s \rho_1(x) dx + b \int |x|^s \rho_0(x) dx \\ &= (1-b)\mathbf{E}|U|^s + b\mathbf{E}|V|^s.\end{aligned} \tag{3.17}$$

Denotando $\beta_s = \mathbf{E}|X_1|^s$, como b foi fixado tal que $0 < b < 1/2$, então podemos obter

$$\frac{\beta_s}{b} = \frac{1-b}{b}\mathbf{E}|U|^s + \mathbf{E}|V|^s \geq \mathbf{E}|U|^s + \mathbf{E}|V|^s \geq \max\{\mathbf{E}|U|^s, \mathbf{E}|V|^s\}.$$

Assim, $\mathbf{E}|U|^s \leq \beta_s/b$ e $\mathbf{E}|V|^s \leq \beta_s/b$. Agora, defina as somas

$$S_{0,n} = V_1 + \ldots + V_n$$

e

$$S_{k,n} = U_1 + \ldots + U_k + V_1 + \ldots + V_{n-k}, \quad 1 \leq k \leq n.$$

Por um lado, se $s \geq 1$, então pela desigualdade de Minkovski no espaço L^s com norma $||\xi||_s = (\mathbf{E}|\xi|^s)^{1/s}$, temos

$$||S_{k,n}||_s \leq \sum_{j=1}^{k} ||U_j||_s + \sum_{j=k+1}^{n} ||V_j||_s = k||U||_s + (n-k)||V||_s \leq n(\beta_s/b)^{1/s}.$$

Logo $\mathbf{E}|S_{k,n}|^s \leq \frac{\beta_s}{b} n^s$ e, pela Proposição 1.7, segue que

$$\mathbf{E}|S_{k,n} - a_n b_n|^s \leq \mathbf{E}(|S_{k,n}| + |a_n| b_n)^s \leq 2^{s-1}(\mathbf{E}|S_{k,n}|^s + |a_n|^s b_n^s) \leq 2^{s-1}\left(\frac{\beta_s}{b} n^s + |a_n|^s b_n^s\right).$$

Por outro lado, se $0 < s < 1$, então pela Proposição 1.6, segue que

$$|S_{k,n}|^s \leq |U_1|^s + \ldots + |U_k|^s + |V_1|^s + \ldots + |V_{n-k}|^s,$$

o que implica que $\mathbf{E}|S_{k,n}|^s \leq n\beta_s/b$ e, usando novamente a Proposição 1.6, temos

$$\mathbf{E}|S_{k,n} - a_n b_n|^s \leq \mathbf{E}(|S_{k,n}| + |a_n| b_n)^s \leq \mathbf{E}|S_{k,n}|^s + |a_n|^s b_n^s \leq n\beta_s/b + |a_n|^s b_n^s.$$

Logo, para todo $s > 0$, temos

$$\mathbf{E}|S_{k,n} - a_n b_n|^s \leq 2^s \left(\frac{\beta_s}{b} n^{\max(s,1)} + |a_n|^s b_n^s \right). \tag{3.18}$$

Agora, por (3.5) e (3.9) podemos escrever

$$\begin{aligned}
\int |x - a_n b_n|^s \rho_{n1}(x) dx &= \int |x - a_n b_n|^s \sum_{k=2}^{n} \binom{n}{k} (1-b)^k b^{n-k} (\rho_1^{k*} * \rho_0^{(n-k)*})(x) dx \\
&= \sum_{k=2}^{n} \binom{n}{k} (1-b)^k b^{n-k} \int |x - a_n b_n|^s (\rho_1^{k*} * \rho_0^{(n-k)*})(x) dx.
\end{aligned}$$

Como $\rho_1^{k*} * \rho_0^{(n-k)*}$ é a densidade de $S_{k,n}$ e usando (3.18) segue que

$$\begin{aligned}
\int |x - a_n b_n|^s \rho_{n1}(x) dx &= \sum_{k=2}^{n} \binom{n}{k} (1-b)^k b^{n-k} \mathbf{E}|S_{k,n} - a_n b_n|^s \\
&\leq 2^s \left(\frac{\beta_s}{b} n^{\max(s,1)} + |a_n|^s b_n^s \right) (1 - \varepsilon_n). \qquad (3.19)
\end{aligned}$$

Analogamente, por (3.6) e (3.8) e (3.18) podemos obter

$$\begin{aligned}
\int |x - a_n b_n|^s \rho_{n0}(x) dx &= \int |x - a_n b_n|^s \sum_{k=0}^{1} \binom{n}{k} (1-b)^k b^{n-k} (\rho_1^{k*} * \rho_0^{(n-k)*})(x) dx \\
&= \sum_{k=0}^{1} \binom{n}{k} (1-b)^k b^{n-k} \int |x - a_n b_n|^s (\rho_1^{k*} * \rho_0^{(n-k)*})(x) dx \\
&= \sum_{k=0}^{1} \binom{n}{k} (1-b)^k b^{n-k} \mathbf{E}|S_{k,n} - a_n b_n|^s \\
&\leq 2^s \left(\frac{\beta_s}{b} n^{\max(s,1)} + |a_n|^s b_n^s \right) \varepsilon_n. \qquad (3.20)
\end{aligned}$$

Usando as desigualdades (3.19) e (3.20) em (3.16), obtemos

$$\int |x|^s |\widetilde{p}_n(x) - p_n(x)| \leq \frac{2^{s+1}}{b_n^s} \left(\frac{\beta_s}{b} n^{\max(s,1)} + |a_n|^s b_n^s \right) \varepsilon_n.$$

Assim, aplicando as hipóteses iniciais sobre os coeficientes a_n, b_n e (3.8), concluímos que

$$\int |x|^s |\widetilde{p}_n(x) - p_n(x)| = o(2^{-n}).$$

Portanto, para n grande o suficiente, segue que

$$\int |x|^s |\widetilde{p}_n(x) - p_n(x)| < 2^{-n}.$$

$\square$

Observação 3.1 *Se assumirmos que* $Z_n \xrightarrow{\mathfrak{D}} Z$, *onde* Z *é uma v.a. não-degenerada, então pelo Teorema 1.6,* Z *é uma variável* α*-estável e* $b_n \sim n^{1/\alpha} B(n)$ *quando* $n \to \infty$, *onde* B *é uma função lentamente variante no sentido de Karamata. Usando argumentos padrões (ver [13]), pode-se mostrar também que* $a_n = o(n)$. *Portanto, sob a hipótese que* $Z_n \xrightarrow{\mathfrak{D}} Z$, *as condições do Lema 3.1 sobre os coeficientes* a_n *e* b_n *são satisfeitas.*

O lema a seguir apresenta uma estimativa envolvendo a integral da função característica $\widetilde{f}_n(x)$ associada à densidade $\widetilde{p}_n(x)$.

Lema 3.2 *Para cada* $t_0 > 0$, *existem constantes positivas* c *e* C *tais que, para todo* $n \geq 2$,

$$\int_{|t| \geq t_0 b_n} |\widetilde{f}_n(x)| < C b_n e^{-cn}.$$

Demonstração. Por (3.10) temos

$$\widetilde{f}_n(x) = \int e^{itx} \widetilde{p}_n(x) dx = \int e^{itx} \frac{b_n}{1 - \varepsilon_n} \rho_{n1}(a_n b_n + b_n x) dx.$$

Fazendo a mudança de variáveis $y = a_n b_n + b_n x$, então $dy = b_n dx$ e segue que

$$\begin{aligned} \widetilde{f}_n(x) &= \frac{1}{1 - \varepsilon_n} e^{-ita_n} \int e^{i\left(\frac{t}{b_n}\right)y} \rho_{n1}(y) dy \\ &= \frac{1}{1 - \varepsilon_n} e^{-ita_n} \hat{\rho}_{n1}\left(\frac{t}{b_n}\right), \end{aligned}$$

onde $\hat{\rho}_{n1}(t) = \int e^{itx}\rho_{n1}(x)dx$. Assim, para cada $t_0 > 0$, temos

$$\begin{aligned} \int_{|t|\geq t_0 b_n} |\widetilde{f}_n(t)|dt &= \int_{|t|\geq t_0} b_n|\widetilde{f}_n(b_n t)|dt \\ &= \frac{b_n}{1-\varepsilon_n}\int_{|t|\geq t_0} |\hat{\rho}_{n1}(t)|dt. \end{aligned} \tag{3.21}$$

Para obter uma estimativa para $|\hat{\rho}_{n1}(t)|$, repetindo as mesmas notações usadas na demonstração do Lema 3.1, considere U e V v.a.'s independentes com densidades ρ_1 e ρ_0, respectivamente, $U_1, U_2 \dots$ cópias independentes de U e $V_1, V_2, \dots$ cópias independentes de V. Dado $n \geq 1$, defina

$$S_{0,n} = V_1 + \ldots + V_n,$$

$$S_{k,n} = U_1 + \ldots + U_k + V_1 + \ldots + V_{n-k}, \quad 1 \leq k \leq n,$$

e denote por d_k a sua respectiva densidade para cada $0 \leq k \leq n$. Logo,

$$d_k = \rho_1^{k*} * \rho_0^{(n-k)*}, \quad 1 \leq k \leq n, \tag{3.22}$$

e se $\hat{d}_k$, $\hat{\rho}_1$ e $\hat{\rho}_0$ são as funções características de d_k, ρ_1 e ρ_0, respectivamente, então temos

$$\hat{d}_k(t) = \mathbf{E}e^{itS_{k,n}} = (\mathbf{E}e^{itU})^k(\mathbf{E}e^{itV})^{n-k} = \hat{\rho}_1(t)^k\hat{\rho}_0(t)^{n-k}, \quad \forall t \in \mathbb{R}. \tag{3.23}$$

Agora, pelo Teorema 1.10, de Riemann-Lebesgue, temos que

$$\lim_{t\to\pm\infty} |\hat{\rho}_j(t)| = 0, \quad j = 0, 1.$$

Também, como $\hat{\rho}_1(t)$ e $\hat{\rho}_0(t)$ são funções características associadas à variáveis aleatórias absolutamente contínuas, temos pela Proposição 1.12 que, para todo $t_0 > 0$, $|\hat{\rho}_j(t)| < 1$ para quaisquer $j = 0, 1$ e $|t| \geq t_0$. Logo, por continuidade, existe uma constante $c > 0$ tal que

$$|\hat{\rho}_j(t)| \leq e^{-c}, \quad \forall |t| \geq t_0.$$

Então, para $2 \leq k \leq n$ e $|t| \geq t_0$, de (3.22) segue que

$$\begin{aligned} |\hat{d}_k(t)| &= |\hat{\rho}_1(t)|^k|\hat{\rho}_0(t)|^{n-k} \\ &= |\hat{\rho}_1(t)|^2|\hat{\rho}_1(t)|^{k-2}|\hat{\rho}_0(t)|^{n-k} \\ &\leq |\hat{\rho}_1(t)|^2(e^{-c})^{k-2}(e^{-c})^{n-k} \\ &= A|\hat{\rho}_1(t)|^2, \end{aligned}$$

onde $A = e^{-c(n-2)} > 0$. Por construção, $\rho_1(x) \leq M$ para todo $x \in \mathbb{R}$. Então, pela fórmula de Plancherel, dada no Teorema 1.11, temos

$$\int_{-\infty}^{+\infty} |\hat{\rho}_1(t)|^2 dt = 2\pi \int_{-\infty}^{+\infty} \rho_1(x)^2 dx$$

e, assim, segue que

$$\int_{|t|\geq t_0} |\hat{d}_k(t)| dt \leq \int_{|t|\geq t_0} A|\hat{\rho}_1(t)|^2 dt < \int_{-\infty}^{+\infty} A|\hat{\rho}_1(t)|^2 dt = 2\pi A \int_{-\infty}^{+\infty} \rho_1(x)^2 dx \leq 2\pi AM. \tag{3.24}$$

Agora, por (3.9) e (3.22) podemos escrever

$$\begin{aligned} \int_{|t|\geq t_0} |\hat{\rho}_{n1}(t)| dt &\leq \int_{|t|\geq t_0} \sum_{k=2}^{n} \binom{n}{k} (1-b)^k b^{n-k} |\hat{d}_k(t)| dt \\ &\leq \sum_{k=2}^{n} \binom{n}{k} (1-b)^k b^{n-k} \int_{|t|\geq t_0} |\hat{d}_k(t)| dt \end{aligned}$$

e de (3.24) e (3.9), segue que

$$\begin{aligned} \int_{|t|\geq t_0} |\hat{\rho}_{n1}(t)| dt &< \sum_{k=2}^{n} \binom{n}{k} (1-b)^k b^{n-k} 2\pi AM \\ &= 2\pi AM(1-\varepsilon_n). \end{aligned}$$

Portanto, usando a desigualdade acima em (3.21), obtemos

$$\begin{aligned} \int_{|t|\geq t_0 b_n} |\widetilde{f}_n(t)| dt &< 2\pi AM b_n \\ &= C b_n e^{-cn}, \end{aligned}$$

onde $C = 2\pi M e^{2c}$ e o lema está provado. □

Observação 3.2 *Do Lema 3.2, segue que $\widetilde{f}_n(t)$ é integrável. De fato,*

$$\int_{-\infty}^{+\infty} |\widetilde{f}_n(t)| dt = \int_{|t|<t_0 b_n} |\widetilde{f}_n(t)| dt + \int_{|t|\geq t_0 b_n} |\widetilde{f}_n(t)| dt < 2t_0 b_n + C b_n e^{-cn} < +\infty.$$

3.3 Convergência Uniforme para Distribuições Estáveis

Seja Z uma v.a. α-estável com densidade $\psi(x)$ e considere a sequência de somas normalizadas

$$Z_n = \frac{X_1 + \ldots + X_n}{b_n} - a_n, \ n \geq 1$$

com $a_n \in \mathbb{R}$, $b_n > 0$ e $X_1, X_2, \ldots$ v.a.'s i.i.d., tal que

$$Z_n \xrightarrow{\mathfrak{D}} Z$$

Assumindo que, para $n \geq n_0$ as v.a.'s Z_n possuem densidades $p_n(x)$, consideremos, respectivamente, as densidades $\widetilde{p}_n(x)$ construídas como na Seção 3.2 para o caso $n_0 = 1$. Dessa forma, podemos obter que a convergência fraca de Z_n à Z implica na convergência uniforme de $\widetilde{p}_n$ à ψ. É o que provaremos na proposição a seguir.

Proposição 3.1 *Suponha que $Z_n \xrightarrow{\mathfrak{D}} Z$, onde Z é uma variável aleatória absolutamente contínua com densidade ψ. Se as variáveis aleatórias Z_n têm distribuições absolutamente contínuas, para $n \geq n_0$, com respectivas densidades p_n, então*

$$\sup_x |\widetilde{p}_n(x) - \psi(x)| \to 0 \quad (n \to \infty),$$

onde $\widetilde{p}_n(x)$ são construídas como em (3.11).

Demonstração. Considere as funções características $f_n(t)$, $\widetilde{f}_n(t)$ e $f(t)$ associadas às densidades $p_n(x)$, $\widetilde{p}_n(t)$ e $\psi(x)$ respectivamente. Da Observação 3.2 e do Corolário 1.2, segue que $\widetilde{f}_n(t)$ e $f(t)$ são integráveis. Podemos então expressar as densidades $\widetilde{p}_n(x)$ e $\psi(x)$ pela fórmula da inversão, ou seja,

$$\widetilde{p}_n(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \widetilde{f}_n(t) dt$$

e

$$\psi(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} f(t) dt.$$

Logo,

$$\widetilde{p}_n(x) - \psi(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} (\widetilde{f}_n(t) - f(t)) dt.$$

Dividiremos a integral acima em duas regiões $\{|t| \leq T_n\}$ e $\{|t| > T_n\}$, onde $0 < T_n \leq t_0 b_n$ para um dado $t_0 > 0$. Então, podemos escrever

$$\widetilde{p}_n(x) - \psi(x) = \frac{1}{2\pi} \int_{|t| \leq T_n} e^{itx} (\widetilde{f}_n(t) - f(t)) dt + \int_{|t| > T_n} e^{itx} (\widetilde{f}_n(t) - f(t)) dt. \qquad (3.25)$$

Por um lado, da hipótese $Z_n \xrightarrow{\mathfrak{D}} Z$ segue do Teorema de Levy que $f_n(t) \to f(t)$ uniformemente em t sobre qualquer intervalo limitado quando $n \to \infty$. Assim, se $T_n \to \infty$ de forma suficientemente lenta, então

$$\lim_{n\to\infty} T_n \max_{|t|\leq T_n} |f_n(t) - f(t)| = 0. \tag{3.26}$$

Por outro lado, considerando $0 < T_n \leq t_0 b_n$, por (3.14) obtemos para n suficientemente grande que

$$0 \leq T_n \max_{|t|\leq T_n} |\widetilde{f}_n(t) - f_n(t)| \leq t_0 b_n 2^{-n}.$$

Mas, pelo Teorema 1.6 temos que $b_n \sim n^{1/\alpha}B(n)$, onde B é uma função lentamente variante. Então, para $0 < T_n \leq t_0 b_n$, segue que

$$\lim_{n\to\infty} T_n \max_{|t|\leq T_n} |\widetilde{f}_n(t) - f_n(t)| = 0. \tag{3.27}$$

Logo, usando a desigualdade triangular, (3.26) e (3.27), obtemos

$$\begin{aligned} 0 &\leq \lim_{n\to\infty} T_n \max_{|t|\leq T_n} |\widetilde{f}_n(t) - f(t)| \\ &\leq \lim_{n\to\infty} (T_n \max_{|t|\leq T_n} |\widetilde{f}_n(t) - f_n(t)| + T_n \max_{|t|\leq T_n} |f_n(t) - f(t)|) = 0. \end{aligned}$$

Dessa forma, como

$$\begin{aligned} \left| \frac{1}{2\pi} \int_{|t|\leq T_n} e^{-itx}(\widetilde{f}_n(t) - f(t))dt \right| &\leq \frac{1}{2\pi} \int_{|t|\leq T_n} |\widetilde{f}_n(t) - f(t)|dt \\ &\leq \frac{1}{2\pi} \int_{|t|\leq T_n} \max_{|t|\leq T_n} |\widetilde{f}_n(t) - f(t)|dt \\ &= \frac{1}{\pi} T_n \max_{|t|\leq T_n} |\widetilde{f}_n(t) - f(t)| \end{aligned}$$

segue que $\lim_{n\to\infty} \frac{1}{2\pi} \int_{|t|\leq T_n} e^{-itx}(\widetilde{f}_n(t) - f(t))dt = 0$. Assim, podemos escrever (3.25) como

$$\widetilde{p}_n(x) - \psi(x) = \frac{1}{2\pi} \int_{|t|>T_n} e^{itx}(\widetilde{f}_n(t) - f(t))dt + o(1). \tag{3.28}$$

Além disso, como f é integrável, temos

$$\lim_{n\to\infty} \int_{|t|>T_n} |f(t)|dt = 0.$$

Logo, aplicando a desigualdade triangular em (3.28) podemos obter

$$\begin{aligned}
\sup_x |\widetilde{p}_n(x) - \psi(x)| &\leq \frac{1}{2\pi}\int_{|t|>T_n} |\widetilde{f}_n(t) - f(t)|dt + o(1) \\
&\leq \frac{1}{2\pi}\int_{|t|>T_n} |\widetilde{f}_n(t)|dt + \frac{1}{2\pi}\int_{|t|>T_n} |f(t)|dt + o(1) \\
&= \frac{1}{2\pi}\int_{|t|>T_n} |\widetilde{f}_n(t)|dt + o(1). \qquad (3.29)
\end{aligned}$$

A integral acima será divida em duas regiões $\{T_n < |t| < t_0 b_n\}$ e $\{|t| \geq t_0 b_n\}$, isto é,

$$\int_{|t|>T_n} |\widetilde{f}_n(t)|dt = \int_{T_n<|t|<t_0 b_n} |\widetilde{f}_n(t)|dt + \int_{|t|\geq t_0 b_n} |\widetilde{f}_n(t)|dt.$$

Pelo Lema 3.2, a segunda integral do lado direito da igualdade tende a zero quando n tende a infinito, pois b_n possui crescimento no máximo polinomial, e assim $b_n e^{-cn} \to 0$ quando $n \to \infty$. Usando em (3.29) esse fato, a desigualdade triangular e (3.14), obtemos

$$\begin{aligned}
\sup_x |\widetilde{p}_n(x) - \psi(x)| &\leq \frac{1}{2\pi}\int_{T_n<|t|<t_0 b_n} |\widetilde{f}_n(t)|dt + o(1) \\
&\leq \frac{1}{2\pi}\int_{T_n<|t|<t_0 b_n} |\widetilde{f}_n(t) - f_n(t)|dt + \frac{1}{2\pi}\int_{T_n<|t|<t_0 b_n} |f_n(t)|dt + o(1) \\
&\leq \frac{1}{2\pi}\int_{T_n<|t|<t_0 b_n} 2^{-n}dt + \frac{1}{2\pi}\int_{T_n<|t|<t_0 b_n} |f_n(t)|dt + o(1) \\
&\leq \frac{1}{\pi}2^{-n}t_0 b_n + \frac{1}{2\pi}\int_{T_n<|t|<t_0 b_n} |f_n(t)|dt + o(1) \\
&= \frac{1}{2\pi}\int_{T_n<|t|<t_0 b_n} |f_n(t)|dt + o(1).
\end{aligned}$$

Finalmente, pela Proposição 1.5 existem constantes $c > 0$ e $t_0 > 0$ tais que $|f_n(t)| \leq e^{-c|t|^{\alpha/2}}$ para todo $n \geq 1$ e $|t| < t_0 b_n$. Como $g(t) = e^{-c|t|^{\alpha/2}}$ é integrável, segue que

$$\int_{T_n<|t|<t_0 b_n} |f_n(t)|dt \leq \int_{T_n<|t|<t_0 b_n} e^{-c|t|^{\alpha/2}}dt \to 0$$

e a Proposição 3.1 está demonstrada.

□

3.4 Convergência em Entropia Relativa

Nesta seção vamos apresentar o principal resultado deste trabalho, devido à Bobkov et al. [4], que estende o Teorema 3.1 de Barron [2], para distribuição α-estável não-extremal.

Assim, sejam Z uma v.a. α-estável, com $\alpha \in (0,2]$, não-extremal, com densidade $\psi(x)$, $X_1, X_2, \ldots$ v.a.'s i.i.d., $\{a_n\}_{n\geq 1}$ e $\{b_n\}_{n\geq 1}$ sequências de constantes reais, com $b_n > 0$, e seja a sequência de somas normalizadas

$$Z_n = \frac{X_1 + \ldots + X_n}{b_n} - a_n, \quad n \geq 1.$$

Vamos assumir que as variáveis Z_n possuem densidade $p_n(x)$, $\forall n \geq 1$. Os mesmos resultados podem ser obtidos se Z_n possuir densidade para $n \geq n_0$, para algum $n_0 > 1$.

O teorema principal a ser demonstrado nesta seção estabelece condições necessárias e suficientes para que, sob a hipótese que $Z_n \xrightarrow{\mathfrak{D}} Z$, tenhamos a convergência em entropia relativa, ou seja, $D(Z_n||Z) \xrightarrow[n\to\infty]{} 0$.

Para isso, necessitamos ainda de mais um lema auxiliar. Por facilidade, denote

$$D_n = D(Z_n||Z) = D(p_n||\psi) = \int p_n(x) \log \frac{p_n(x)}{\psi(x)} dx \tag{3.30}$$

e para $\widetilde{p}_n(x)$, a densidade modificada construída na Seção 3.2, em (3.11), seja

$$\widetilde{D}_n = D(\widetilde{p}_n||\psi) = \int \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\psi(x)} dx. \tag{3.31}$$

O lema a seguir, de certa forma, estende a estimativa obtida no Lema 3.1 para as entropias relativas D_n e $\widetilde{D}_n$ e implica que quando $n \to \infty$ temos $|D_n - \widetilde{D}_n| \to 0$.

Lema 3.3 *Assuma que Z é uma v.a. estável não-extremal com densidade $\psi(x)$. Se D_n é finito para todo $n \geq n_0$, e $|a_n| + \log b_n + 1/b_n = o(n^\gamma)$ para algum $\gamma > 0$, então*

$$|\widetilde{D}_n - D_n| < 2^{-n},$$

para todo n grande o suficiente, onde D_n e $\widetilde{D}_n$ são dadas por (3.30) e (3.31).

Demonstração. Para simplificar as notações, assumiremos $n_0 = 1$, $a_1 = 0$ e $b_1 = 1$ sem perda de generalidade. Então $D_1 = D(X_1||Z)$ é finita e assim, do Corolário 2.2, $H(X_1)$ é finita e $\mathbf{E} \log \frac{1}{\psi(X_1)} < +\infty$. Denote

$$D_{n0} = \int p_{n0}(x) \log \frac{p_{n0}(x)}{\psi(x)} dx.$$

Considerando $L(u) = u \log u, \ u > 0$, pela definição de $\widetilde{p}_n$ em (3.11) podemos escrever

$$\begin{aligned} D_n &= \int \psi(x) L\left(\frac{p_n(x)}{\psi(x)}\right) dx \\ &= \int \psi(x) L\left((1-\varepsilon_n)\frac{\widetilde{p}_n(x)}{\psi(x)} + \varepsilon_n \frac{p_{n0}(x)}{\psi(x)}\right). \end{aligned} \tag{3.32}$$

Agora, pela convexidade da função $L(u)$, provada na Proposição 1.10(a.4), segue que

$$\begin{aligned} D_n &\leq \int \psi(x) \left[(1-\varepsilon_n) L\left(\frac{\widetilde{p}_n(x)}{\psi(x)}\right) + \varepsilon_n L\left(\frac{p_{n0}(x)}{\psi(x)}\right)\right] \\ &= (1-\varepsilon_n)\widetilde{D}_n + \varepsilon_n D_{n0}. \end{aligned}$$

Por outro lado, aplicando a desigualdade da Proposição 1.10(a.5) em (3.32) também obtemos

$$\begin{aligned} D_n &\geq \int \psi(x) \left[(1-\varepsilon_n) L\left(\frac{\widetilde{p}_n(x)}{\psi(x)}\right) + \varepsilon_n L\left(\frac{p_{n0}(x)}{\psi(x)}\right) + \frac{p_{n0}(x)}{\psi(x)} L(1-\varepsilon_n) + \frac{p_{n0}(x)}{\psi(x)} L(\varepsilon_n)\right] \\ &= (1-\varepsilon_n)\widetilde{D}_n + \varepsilon_n D_{n0} + L(\varepsilon_n) + L(1-\varepsilon_n). \end{aligned}$$

Das duas desigualdades acima, segue

$$D_n - \widetilde{D}_n \leq -\varepsilon_n \widetilde{D}_n + \varepsilon_n D_{n0}$$

e

$$\widetilde{D}_n - D_n \leq \varepsilon_n \widetilde{D}_n - \varepsilon_n D_{n0} - L(\varepsilon_n) - L(1-\varepsilon_n).$$

Como $\widetilde{D}_n, D_{n0} \geq 0$ e $L(\varepsilon_n), L(1-\varepsilon_n) \leq 0$, podemos obter

$$|\widetilde{D}_n - D_n| \leq \varepsilon_n \widetilde{D}_n + \varepsilon_n D_{n0} - L(\varepsilon_n) - L(1-\varepsilon_n). \tag{3.33}$$

Queremos obter uma desigualdade a partir de (3.33) onde não apareça o termo ε_n. Da Proposição 1.8, temos

$$-L(1-\varepsilon_n) = (1-\varepsilon_n)\log\frac{1}{1-\varepsilon_n} \leq (1-\varepsilon_n)\left(\frac{1}{1-\varepsilon_n} - 1\right) = \varepsilon_n.$$

Então, usando esta desigualdade em (3.33) e lembrando que $L(\varepsilon_n) = \varepsilon_n \log \varepsilon_n$, podemos escrever

$$|\widetilde{D}_n - D_n| \leq \varepsilon_n(\widetilde{D}_n + D_{n0} - \log \varepsilon_n + 1).$$

Mas, por (3.8) temos $b^n < \varepsilon_n < nb^{n-1}$ sempre que $n \geq 2$. Então segue que

$$\begin{aligned} |\widetilde{D}_n - D_n| &\leq \varepsilon_n(\widetilde{D}_n + D_{n0} - \log \varepsilon_n + 1) \\ &< n(\widetilde{D}_n + D_{n0} + Cn)b^{n-1} \\ &\leq Cn(\widetilde{D}_n + D_{n0} + n)b^{n-1}, \end{aligned} \tag{3.34}$$

onde $C = \log 1/b + 1$.

Para obter estimativas do lado direito da desigualdade (3.34), consideremos, assim como fizemos na demonstração do Lema 3.1, variáveis aleatórias independentes U e V com densidades p_1 e p_0 respectivamente e sejam $U_1, U_2, \ldots$ cópias independentes de U e $V_1, V_2, \ldots$ cópias independentes de V, sendo $\{U_n\}_{n\geq 1}$ independentes de $\{V_n\}_{n\geq 1}$. Para cada $n \geq 1$, defina

$$R_{k,n} = \frac{S_{k,n}}{b_n} - a_n, \quad 0 \leq k \leq n,$$

onde $S_{0,n} = V_1 + \ldots + V_n$ e $S_{k,n} = \sum_{i=1}^{k} U_i + \sum_{i=1}^{n-k} V_i,\ k \geq 1$. Se denotarmos a densidade de $S_{k,n}$ por d_k, ou seja,

$$d_k = \rho_1^{k*} * \rho_0^{(n-k)*},$$

então a densidade $r_{k,n}$ de $R_{k,n}$ é dada por

$$r_{k,n}(x) = b_n d_k(a_n b_n + b_n x).$$

Assim, da definição de $\widetilde{p}_n(x)$ em (3.10) e por (3.5) podemos escrever

$$\begin{aligned} \widetilde{p}_n(x) &= \frac{b_n}{1-\varepsilon_n}\rho_{n1}(a_n b_n + b_n x) \\ &= \frac{b_n}{1-\varepsilon_n}\sum_{k=2}^{n}\binom{n}{k}(1-b)^k b^{n-k} d_k(a_n b_n + b_n x) \\ &= \frac{1}{1-\varepsilon_n}\sum_{k=2}^{n}\binom{n}{k}(1-b)^k b^{n-k} r_{k,n}(x). \end{aligned} \tag{3.35}$$

Também, pela definição de $p_{n0}(x)$ em (3.10) e por (3.6), obtemos

$$\begin{aligned} p_{n0}(x) &= \frac{b_n}{\varepsilon_n}\rho_{n0}(a_n b_n + b_n x) \\ &= \frac{b_n}{\varepsilon_n}\sum_{k=0}^{1}\binom{n}{k}(1-b)^k b^{n-k} d_k(a_n b_n + b_n x) \\ &= \frac{1}{\varepsilon_n}\sum_{k=0}^{1}\binom{n}{k}(1-b)^k b^{n-k} r_{k,n}(x). \end{aligned} \tag{3.36}$$

Agora, usando (3.35), (3.36) e a convexidade da função L, obtemos as seguintes cotas superiores para $\widetilde{D}_n$ e D_{n0} em termos da entropia relativa $D(R_{k,n}||Z)$:

$$\begin{aligned}
\widetilde{D}_n &= \int \psi(x) L\left(\frac{\widetilde{p}_n(x)}{\psi(x)}\right) dx \\
&= \int \psi(x) L\left(\frac{1}{1-\varepsilon_n}\sum_{k=2}^{n}\binom{n}{k}(1-b)^k b^{n-k}\frac{r_{k,n}(x)}{\psi(x)}\right) dx \\
&\leq \int \psi(x) \frac{1}{1-\varepsilon_n}\sum_{k=2}^{n}\binom{n}{k}(1-b)^k b^{n-k} L\left(\frac{r_{k,n}(x)}{\psi(x)}\right) dx \\
&= \frac{1}{1-\varepsilon_n}\sum_{k=2}^{n}\binom{n}{k}(1-b)^k b^{n-k}\int \psi(x) L\left(\frac{r_{k,n}(x)}{\psi(x)}\right) dx \\
&= \frac{1}{1-\varepsilon_n}\sum_{k=2}^{n}\binom{n}{k}(1-b)^k b^{n-k} D(R_{k,n}||Z) \qquad (3.37)
\end{aligned}$$

e

$$\begin{aligned}
D_{n0} &= \int \psi(x) L\left(\frac{p_{n0}(x)}{\psi(x)}\right) dx \\
&= \int \psi(x) L\left(\frac{1}{\varepsilon_n}\sum_{k=0}^{1}\binom{n}{k}(1-b)^k b^{n-k}\frac{r_{k,n}(x)}{\psi(x)}\right) dx \\
&\leq \int \psi(x) \frac{1}{\varepsilon_n}\sum_{k=0}^{1}\binom{n}{k}(1-b)^k b^{n-k} L\left(\frac{r_{k,n}(x)}{\psi(x)}\right) dx \\
&= \frac{1}{\varepsilon_n}\sum_{k=0}^{1}\binom{n}{k}(1-b)^k b^{n-k}\int \psi(x) L\left(\frac{r_{k,n}(x)}{\psi(x)}\right) dx \\
&= \frac{1}{\varepsilon_n}\sum_{k=0}^{1}\binom{n}{k}(1-b)^k b^{n-k} D(R_{k,n}||Z). \qquad (3.38)
\end{aligned}$$

Nosso objetivo agora é usar o Corolário 2.2 para obter que

$$D(R_{k,n}||Z) = \mathbf{E}\log\frac{1}{\psi(R_{k,n})} - H(R_{k,n})$$

e assim obter as estimativas finais.

Para isso, primeiramente da decomposição de $p(x)$, a densidade de X_1, em (3.3) e da Proposição 1.10(a.6) podemos obter

$$\begin{aligned}
D(X_1||Z) &= \int \psi(x) L\left(\frac{(1-b)\rho_1(x) + b\rho_0(x)}{\psi(x)}\right) dx \\
&\geq \int \psi(x) \left[(1-b) L\left(\frac{\rho_1(x)}{\psi(x)}\right) - \frac{1}{e}\frac{\rho_1(x)}{\psi(x)} - \frac{1}{e}\right] dx
\end{aligned}$$

e assim, como U tem densidade ρ_1, temos

$$D(X_1||Z) \geq (1-b)D(U||Z) - \frac{2}{e}.$$

Analogamente, obtemos uma desigualdade similar trocando b por $1-b$:

$$D(X_1||Z) \geq bD(V||Z) - \frac{2}{e},$$

pois V tem densidade ρ_0. Logo, como, por hipótese, $D(X_1||Z) < +\infty$ segue que $D(U||Z)$ e $D(V||Z)$ são finitas. Assim, pelo Corolário 2.2 temos que $H(U)$ e $H(V)$ são finitas, $\mathbf{E}\log^+(\frac{1}{\psi(U)}) < +\infty$ e $\mathbf{E}\log^+(\frac{1}{\psi(V)}) < +\infty$.

Agora, das proposições 2.1 e 2.4, segue que

$$H(R_{k,n}) = -\log b_n + h(S_{k,n}) \geq -\log b_n + H(U), \text{ para } 1 \leq k \leq n$$

e

$$H(R_{0,n}) \geq -\log b_n + H(V).$$

Fazendo $C_0 = \min(H(U), H(V))$, obtemos

$$H(R_{k,n}) \geq -\log b_n + C_0, \tag{3.39}$$

para todo $0 \leq k \leq n$. Assim, $H(R_{k,n}) > -\infty$. Para verificar as condições do Corolário 2.2 aplicado à $D(R_{k,n}||Z)$, vamos separar em dois casos.

Caso 1: $Z \sim N(\mu, \sigma^2)$ para algum $\mu \in \mathbb{R}$ e $\sigma > 0$.
Neste caso, pela Proposição 2.8, como $D(U||Z)$ e $D(V||Z)$ são finitas, então segue que

$\mathbf{E}U^2 < +\infty$ e $\mathbf{E}V^2 < +\infty$. Além disso, usando a Proposição 1.7, obtemos

$$\begin{aligned}
\mathbf{E}R_{k,n}^2 &= \mathbf{E}\left(\frac{S_{k,n}}{b_n} - a_n\right)^2 \\
&= 2\left(\frac{1}{b_n^2}\mathbf{E}S_{k,n}^2 + a_n^2\right) \\
&\leq 2\left(\frac{1}{b_n^2}n(\mathbf{E}U_1^2 + \ldots + \mathbf{E}U_k^2 + \mathbf{E}V_1^2 + \ldots + \mathbf{E}V_{n-k}^2) + a_n^2\right) \\
&= 2\left(\frac{n}{b_n^2}(k\mathbf{E}U^2 + (n-k)\mathbf{E}V^2) + a_n^2\right) \\
&\leq 2\left(\frac{n^2}{b_n^2}\max(\mathbf{E}U^2, \mathbf{E}V^2) + a_n^2\right) < +\infty. \qquad (3.40)
\end{aligned}$$

Logo, da Observação 2.3, $\mathbf{E}\log^+(1/\psi(R_{k,n})) < +\infty$. Como $H(R_{k,n}) > -\infty$, segue da Proposição 2.6 que $H(R_{k,n})$ é finita. Dessa forma, pelo Corolário 2.2, $D(R_{k,n}||Z) < +\infty$ e podemos escrever

$$D(R_{k,n}||Z) = \mathbf{E}\log\frac{1}{\psi(R_{k,n})} - H(R_{k,n}).$$

Novamente, pela Observação 2.3 e da cota inferior para $H(R_{k,n})$ em (3.39), obtemos

$$\begin{aligned}
D(R_{k,n}||Z) &\leq \log(\sqrt{2\pi}\sigma) + \frac{1}{\sigma^2}(\mathbf{E}R_{k,n}^2 + \mu^2) + \log b_n - C_0 \\
&= \log b_n + C_1 + C_2\mathbf{E}|R_{k,n}|^2,
\end{aligned}$$

onde $C_1 = \log(\sqrt{2\pi}\sigma) + \frac{\mu}{\sigma^2} - C_0$ e $C_2 = \frac{1}{\sigma^2}$. Da desigualdade acima e de (3.40), temos

$$D(R_{k,n}||Z) \leq \log b_n + C_1 + C_3\frac{n^2}{b_n^2} + C_4 a_n^2,$$

onde $C_3 = 2KC_2$ e $C_4 = 2C_2$. Usando as hipóteses iniciais sobre a_n e b_n, segue que $D(R_{k,n}||Z) = o(n^{\gamma'})$ para algum $\gamma' > 0$. De (3.37) e (3.38), obtemos que $\widetilde{D}_n = o(n^{\gamma'})$ e $D_{n0} = o(n^{\gamma'})$. Por fim, de (3.34), concluímos que $|\widetilde{D}_n - D_n| = o(2^{-n})$.

Caso 2: Z possui distribuição estável não-extremal a qual não é normal.
Neste caso, pela Proposição 2.9, como $D(U||Z)$ e $D(V||Z)$ são finitas, segue que

$\mathbf{E}\log(1+|U|) < +\infty$ e $\mathbf{E}\log(1+|V|) < +\infty$. Usando a Proposição 1.11, temos

$$\begin{aligned}
\log(1+|R_{k,n}|) &\leq \log\left(1+\frac{|S_{k,n}|}{b_n}+|a_n|\right) \\
&\leq \log(1+|a_n|)+\log\left(1+\frac{|S_{k,n}|}{b_n}\right) \\
&\leq \log(1+|a_n|)+\frac{1}{b_n}+\log(1+|S_{k,n}|) \\
&\leq \log(1+|a_n|)+\frac{1}{b_n}+\sum_{j=1}^{k}\log(1+|U_j|)+\sum_{j=1}^{n-k}\log(1+|V_j|).
\end{aligned}$$

Logo,

$$\begin{aligned}
\mathbf{E}\log(1+|R_{k,n}|) &\leq \log(1+|a_n|)+\frac{1}{b_n}+k\mathbf{E}\log(1+|U|)+(n-k)\mathbf{E}\log(1+|V|) \\
&\leq \log(1+|a_n|)+\frac{1}{b_n}+nC_5, \qquad (3.41)
\end{aligned}$$

onde $C_5 = \max\left(\mathbf{E}\log(1+|U|), \mathbf{E}\log(1+|V|)\right)$. Da Observação 2.5, existem constantes $A, B > 0$ tais que

$$\mathbf{E}\log^{+}\left(\frac{1}{\psi(R_{k,n})}\right) \leq A + B\mathbf{E}\log(1+|R_{k,n}|). \qquad (3.42)$$

Assim, de (3.41) e (3.42), segue que $\mathbf{E}\log^{+}(1/\psi(R_{k,n})) < +\infty$. Como $H(R_{k,n}) > -\infty$, temos pela Proposição 2.6 que $H(R_{k,n})$ é finita. Portanto, pelo Corolário 2.2, podemos escrever

$$D(R_{k,n}||Z) = \mathbf{E}\log\frac{1}{\psi(R_{k,n})} - h(R_{k,n}).$$

Dessa forma, das desigualdades (3.42), (3.41) e (3.39), temos

$$\begin{aligned}
D(R_{k,n}||Z) &\leq A + B\mathbf{E}\log(1+|R_{k,n}|)+\log b_n - C_0 \\
&\leq \log b_n + C_6 + C_7 n + B(\log(1+|a_n|)+1/b_n),
\end{aligned}$$

onde $C_6 = A - C_0$ e $C_7 = BC_5$. Usando as hipóteses iniciais sobre a_n e b_n, segue que $D(R_{k,n}||Z) = o(n^{\gamma''})$ para algum $\gamma'' > 0$. De (3.37) e (3.38), obtemos que $\widetilde{D}_n = o(n^{\gamma''})$ e $D_{n0} = o(n^{\gamma''})$. Por fim, de (3.34), concluímos que $|\widetilde{D}_n - D_n| = o(2^{-n})$.

$\square$

Observação 3.3 *De acordo com a Observação 3.1, se $Z_n \xrightarrow{\mathfrak{D}} Z$ então as condições do Lema 3.3 sobre as sequências a_n e b_n são automaticamente satisfeitas.*

Finalmente, estamos em condições de demonstrar o teorema principal, de a Bobkov et al. [4].

Teorema 3.2 *Assuma que a sequência de somas normalizadas* Z_n *definidas anteriormente converge em distribuição para uma variável aleatória* Z *estável não-extremal. Então* $D(Z_n||Z) \to 0$ *quando* $n \to \infty$ *se, e somente se,* $D(Z_{n_0}||Z) < +\infty$ *para algum* $n_0 \geq 1$.

Demonstração. Se $D(Z_n||Z) \to 0$ quando $n \to \infty$, então claramente $D(Z_n||Z) < +\infty$ para algum n. Reciprocamente, assuma que $D(Z_{n_0}||Z) < +\infty$ para algum $n_0 \geq 1$. Então, como Z é uma v.a. não-extremal, pela Proposição 2.11, segue que $D(Z_n||Z) < +\infty$, $\forall n \geq n_0$.

Sejam ψ a densidade de Z e p_n a densidade de Z_n, $\widetilde{p}_n(x)$ a densidade modificada construída por (3.11) na Seção 3.2 e considere $\widetilde{Z}_n$ uma v.a. com densidade $\widetilde{p}_n(x)$. Denote $D_n = D(Z_n||Z)$ e $\widetilde{D}_n = D(\widetilde{Z}_n||Z)$, como em (3.30) e (3.31). Então, podemos escrever

$$0 \leq D_n \leq |\widetilde{D}_n - D_n| + \widetilde{D}_n. \tag{3.43}$$

Agora, como por hipótese $Z_n \xrightarrow{\mathfrak{D}} Z$, então conforme as observações 3.1 e 3.2, as sequências normalizadoras a_n e b_n satisfazem as condições dos lemas 3.1 e 3.3 e como $D_n < +\infty$, $\forall n \geq 1$, segue do Lema 3.3 que

$$|\widetilde{D}_n - D_n| < 2^{-n}, \text{ para n suficientemente grande.} \tag{3.44}$$

Assim, $\widetilde{D}_n < +\infty$ para n grande o suficiente e, da definição, segue que $\psi(x) = 0$ implica $\widetilde{p}_n(x) = 0$ q.c..

Por (3.44) temos

$$|\widetilde{D}_n - D_n| \to 0 \text{ quando } n \to \infty \tag{3.45}$$

e então por (3.43) basta mostrarmos que $\widetilde{D}_n \to 0$ quando $n \to \infty$.

Para isso, consideremos a função $L(u) = u \log u$, $u > 0$ e por (3.31) podemos escrever

$$\widetilde{D}_n = D(\widetilde{Z}_n||Z) = \int L\left(\frac{\widetilde{p}_n(x)}{\psi(x)}\right)\psi(x)dx, \tag{3.46}$$

Para obter um limitante superior para $\widetilde{D}_n$ que convergirá para zero, vamos analisar separadamente os casos em que Z é v.a. α-estável não-extremal e não-normal e quando Z é normal.

Caso 1: Z possui distribuição estável não-extremal a qual não é normal ($\alpha \in (0,2)$). Neste caso, utilizaremos a desigualdade elementar da Proposição 1.10(b), ou seja, dado

$\varepsilon \in (0,1]$, existe $C_\varepsilon \geq 1$ tal que

$$L(t) \leq (t-1) + C_\varepsilon |t-1|^{1+\varepsilon}, \quad \forall t \geq 0. \tag{3.47}$$

Isto implica, por (3.46), que

$$\begin{aligned} \widetilde{D}_n &\leq \int \left[\left(\frac{\widetilde{p}_n(x)}{\psi(x)} - 1 \right) + C_\varepsilon \left| \frac{\widetilde{p}_n(x)}{\psi(x)} - 1 \right|^{1+\varepsilon} \right] \psi(x) dx \\ &= C_\varepsilon \int \frac{|\widetilde{p}_n(x) - \psi(x)|^{1+\varepsilon}}{\psi(x)^\varepsilon} dx. \end{aligned}$$

Fazendo $\Delta_n = \sup_x |\widetilde{p}_n(x) - \psi(x)|$, temos pela Proposição 3.1 que $\Delta_n \to 0$. Assim, $\Delta_n < +\infty$ para n suficientemente grande e podemos escrever

$$\widetilde{D}_n \leq C_\varepsilon \Delta_n^\varepsilon \int \frac{1}{\psi(x)^\varepsilon} |\widetilde{p}_n(x) - \psi(x)| dx, \tag{3.48}$$

para n suficientemente grande. Como $\Delta_n \to 0$, então basta provar que a integral em (3.47) é finita. Pela Proposição 1.4, existe $c > 0$ tal que $\psi(x) \geq c(1+|x|)^{-(1+\alpha)}$ para todo $x \in \mathbb{R}$ e pela desigualdade (1.14) da Observação 1.1, temos que $(1+|x|)^s \leq 2^s(1+|x|^s)$, $\forall s > 0$. Então usando estas desigualdades em (3.48) podemos obter

$$\widetilde{D}_n \leq \frac{1}{c^\varepsilon} C_\varepsilon \Delta_n^\varepsilon 2^{\varepsilon(1+\alpha)} \int (1 + |x|^{\varepsilon(1+\alpha)}) |\widetilde{p}_n(x) - \psi(x)| dx.$$

Escolhendo $\varepsilon < \frac{\alpha}{\alpha+1}$ e fazendo $s = \varepsilon(1+\alpha) < \alpha$, a última integral pode ser limitada por

$$\int (1+|x|^s)|\widetilde{p}_n(x) - p_n(x)| dx + \int (1+|x|^s) p_n(x) dx + \int (1+|x|^s)\psi(x) dx.$$

Pelo Teorema 1.9, temos $\sup_n \mathbf{E}|Z_n|^s < +\infty$ e como $\mathbf{E}|Z|^s < +\infty$, segue do Lema 3.1 que todas as integrais acima são limitadas por uma constante. Dessa forma, temos $\widetilde{D}_n \to 0$ quando $n \to \infty$, o que implica, por (3.43) e (3.45) que $D_n \to 0$ e o teorema está provado para o caso em que Z não é normal.

Caso 2: Z possui distribuição normal.

Suponha, sem perda de generalidade, que Z tem distribuição normal padrão. Para uma sequência T_n, dividiremos a integral em (3.46) em duas regiões $\{|x| \leq T_n\}$ e $\{|x| > T_n\}$, ou seja,

$$\widetilde{D}_n = \int_{|x| \leq T_n} L\left(\frac{\widetilde{p}_n(x)}{\psi(x)} \right) \psi(x) dx + \int_{|x| > T_n} L\left(\frac{\widetilde{p}_n(x)}{\psi(x)} \right) \psi(x) dx.$$

Para a primeira integral na soma acima, usaremos a desigualdade (3.47) com $\varepsilon = 1$ e $C_\varepsilon = 1$ e fazendo como no caso anterior limitaremos o integrando por Δ_n. Dessa forma, temos

$$\begin{aligned}\widetilde{D}_n &\leq \int_{|x|\leq T_n} \frac{|\widetilde{p}_n(x)-\psi(x)|^2}{\psi(x)}dx + \int_{|x|>T_n} \widetilde{p}_n(x)\log\frac{\widetilde{p}_n(x)}{\psi(x)}dx \\ &\leq \Delta_n^2 \int_{|x|\leq T_n} \frac{1}{\psi(x)}dx + \int_{|x|>T_n} \widetilde{p}_n(x)\log\frac{\widetilde{p}_n(x)}{\psi(x)}dx.\end{aligned}$$

Agora, como ψ é a densidade Normal padrão, obtemos

$$\widetilde{D}_n \leq 2T_n\sqrt{2\pi}e^{T_n^2/2}\Delta_n^2 + \int_{|x|>T_n} \widetilde{p}_n(x)\left[\log\widetilde{p}_n(x) + \frac{x^2}{2} + \log\sqrt{2\pi}\right]dx.$$

Mas, pela Proposição 3.1, podemos concluir que $\widetilde{p}_n(x) \leq 1$ para n suficientemente grande, o que implica $\log\widetilde{p}_n(x) \leq 0$ para n suficientemente grande. Então escolhendo T_n, $T_n \to \infty$ de forma suficientemente lenta tal que $2T_n\sqrt{2\pi}e^{T_n^2/2}\Delta_n^2 \to 0$, podemos obter

$$\begin{aligned}\widetilde{D}_n &\leq \int_{|x|>T_n} \widetilde{p}_n(x)\left[\frac{x^2}{2} + \log\sqrt{2\pi}\right]dx + o(1) \\ &= \frac{1}{2}\int_{|x|>T_n} x^2\widetilde{p}_n(x)dx + \log\sqrt{2\pi}\int_{|x|>T_n}\widetilde{p}_n(x)dx + o(1).\end{aligned}$$

Como $\widetilde{p}_n(x)$ é integrável, segue que $\int_{|x|>T_n}\widetilde{p}_n(x)dx = o(1)$. Logo

$$\widetilde{D}_n \leq \frac{1}{2}\int_{|x|>T_n} x^2\widetilde{p}_n(x)dx + o(1).$$

Somando e subtraindo $p_n(x)$, segue que

$$\widetilde{D}_n \leq \frac{1}{2}\int_{|x|>T_n} x^2|\widetilde{p}_n(x)-p_n(x)|dx + \frac{1}{2}\int_{|x|>T_n} x^2 p_n(x)dx + o(1). \tag{3.49}$$

Note que, pela Proposição 2.8, como Z é normal e $D(Z_{n_0}||Z) < +\infty$ temos que $\mathbf{E}Z_{n_0}^2 < +\infty$. Dessa forma,

$$\mathbf{E}S_{n_0}^2 = b_{n_0}^2\mathbf{E}(Z_{n_0}+a_{n_0})^2 \leq 2b_{n_0}^2(\mathbf{E}Z_{n_0}^2 + a_{n_0}^2) < +\infty$$

e, como $X_1,\ldots,X_n$ são i.i.d., segue que

$$\mathbf{E}X_1^2 \leq n_0\mathbf{E}X_1^2 + n_0(n_0-1)(\mathbf{E}X_1)^2 = \mathbf{E}S_{n_0}^2 < +\infty.$$

Então podemos usar o Lema 3.1 para concluir que a primeira integral em (3.49) tende a zero. Como $\mathbf{E}X_1^2 < +\infty$, segue de forma análoga ao Lema 1.1 que a sequência Z_n^2 é uniformemente integrável e assim a segunda integral em (3.49) também converge para zero. Portanto $\widetilde{D}_n \to 0$ e por (3.43) e (3.45) segue que $D_n \to 0$. Concluímos então a prova do teorema.

$\square$

Referências Bibliográficas

[1] ASH, R.B. and DOLÉANS-DADE. C. *Probability and Measure Theory.* Academic Press, 2000.

[2] BARRON, A.R. *Entropy and the central limit theorem.* Ann. Probab. 14(1), 336-342 (1986).

[3] BINGHAM, N.H., GOLDIE, C.M. and TEUGELS, J.L. *Regular Variation.* Cambridge University Press. UK. 1987

[4] BOBKOV, S.G., CHISTYAKOV, G.P. and GÖTZE, F. *Convergence to stable laws in relative entropy.* J. Theoret. Probab, 2013, 803-818.

[5] BOBKOV, S.G., CHISTYAKOV, G.P. and GÖTZE, F. *Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem.* Ann. Probab. Volume 41, Number 4 (2013), 2479-2512.

[6] BROWN, L. *A proof of the Central Limit Theorem motivated by the Cramér-Rao inequality.* In Kallianpur, G., Krishnaiah, P., and Ghosh, J., editors, Statistics and Probability: Essays in Honour of C.R. Rao, North-Holland, 1982, 141-148.

[7] COMMENGS D. *Information Theory and Statistics: and overview.* arXiv:1511.0086v1 [math.ST] 3Nov2015.

[8] FELLER, W. *An introduction to probability theory and its applications II.* Willey, New York, 1971.

[9] HOYOS, M. L. C. *Convergência em informação de fisher relativa de somas parciais para variáveis aleatórias estáveis.* 2016. 72 p. Dissertação (Mestrado em Matemática) - Universidade de Brasília, Brasília.

[10] IBRAGIMOV, I.A. and LINNIK J.V. *Independent Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen, 1971.

[11] JOHNSON, O. *Information Theory and The Central Limit Theorem*. Imperial College Press. London, 2004.

[12] KULLBACK, S. and LEIBLER, R. *On information and sufficiency*. Ann. Math-Statist., 1951, 79-86.

[13] LAMPERTI, J.W. *Probability. A Survey of the Mathematical Theory*. Wiley Series in Probability and Statistics, Wiley, New York, 1996.

[14] LINNIK, J.V. *An information-theoretic proof of the central limit theorem with Lindeberg conditions*. Theory Probab. Appl, 1959, 288-299.

[15] NOLAN, J. *Stable Distributions Models for Heavy Tailed Data*. Birkhauser, 2007.

[16] PETROV, V.V. *Sums of Independent Random Variables*. Springer, New York, 1975.

[17] SENETA, E. *Regulary varying functions*. Springer, 1976.

[18] SHANNON, C.E. *A mathematical theory of communication*. The Bell System Technical Journal, 1948, 379-423, 623-656.

[19] ZOLOTAREV, V.M. *One-dimensional stable distributions*. Probability Theory and mathematical statistics. Moscow: Nauka, 1983.