Domínio de atração de distribuições α-estáveis sob modelos de mistura finita

Cira E. Guevara Otiniano, **Cátia Regina Gonçalves**,
Depto de Estatística e Matemática, UnB,
70910-900 Brasilia-DF
E-mail: cira@unb.br

Resumo. *Neste trabalho, estudamos a distribuição assintótica da soma normalizada de variáveis aleatórias independentes e identicamente distribuídas, sob modelos de mistura finita. Apresentamos condições necessárias para que uma função de distribuição de uma população mista com k componentes pertença ao domínio de atração de uma distribuição α- estável, assumindo que cada componente da mistura também pertence ao domínio de atração de uma distribuição α-estável. Um exemplo é dado para ilustrar o resultado.*

Palavras-chave: *População mista, Domínio de atração, Índice caudal.*

1. Introdução. Misturas finitas têm seu uso prático em diversas áreas. Os livros [2], [9],[6], e [5] da bibliografia descrevem modelos de mistura finita, propriedades e aplicações. [9] apresenta exemplos de aplicações diretas de modelos de misturas finitas para a pesca , economia, medicina, psicologia, paleontologia, botânica, agricultura, zoologia e confiabilidade, entre outros. Aplicações indiretas incluem *outliers*, mistura de normais para testar robustez, análise de *cluster*, modelos de estrutura latente, inferência bayesiana, estimação não paramétrica de densidades, aproximar modelos de mistura por modelos de não mistura. [7] descreve modelos de sobrevivência multivariados com mistura de distribuições estáveis positivas.

Nas últimas três décadas dados de distribuições com cauda pesada têm sido observados em diversas áreas. Quando as quantias observadas podem ser aproximadas por somas de pequenos termos e as densidades empíricas dos dados exibem cauda pesada e assimetria, o Teorema do Limite Central e as fortes evidências empíricas dos dados indicam que uma modelagem mais realista é obtida com uma distribuição no domínio de atração de uma α- estável.

Dadas $Z_1, \dots Z_n, Z$ variáveis aleatórias (v.a.'s) independentes e identicamente distribuídas (i.i.d.) com distribuição comum G, a v.a. Z ou G é chamada de α-estável se existem constantes $D_n \in \mathbf{R}$ tais que

$$S_n = Z_1 + ... + Z_n \stackrel{d}{=} n^{1/\alpha} Z + D_n. \tag{1}$$

Não existe fórmula fechada para G, porém sua correspondente função característica é conhecida e dada por

$$\varphi(t) = \begin{cases} \exp\left\{i\mu t - \sigma^{\alpha}|t|^{\alpha}[1 - i\beta sign(t) tan(\alpha\frac{\pi}{2})]\right\}, & \text{se} \quad \alpha \neq 1 \\ \exp\left\{i\mu t - \sigma|t|[1 - i\beta \ \frac{2}{\pi} \ sign(t) \ ln(t)]\right\}, & \text{se} \quad \alpha = 1, \end{cases} \tag{2}$$

onde $0 < \alpha \leq 2$ (índice de estabilidade), $-1 \leq \beta \leq 1$ (parâmetro de simetria), $\sigma > 0$ (parâmetro de escala) e $\mu \in \mathbf{R}$ (parâmetro de locação) caracterizam a distribuição G e são únicos, exceto quando $\alpha = 2$. Neste caso denotamos G por $S_{\alpha}(\sigma, \beta, \mu)$.

De (1) temos que a distribuição G de Z é uma distribuição de somas normalizadas; $\frac{S(n)-D_n}{n^{1/\alpha}} \stackrel{d}{=} Z$. O que sugere que G pode ser obtida assintoticamente como distribuição limite de somas normalizadas de v.a.'s i.i.d., tal resultado é conhecido como Teorema do Limite Central Generalizado. Conforme [4] esse resultado em termos de domínio de atração é como segue:

Sejam $X, X_1, X_2, ...$ v.a's i.i.d. com distribuição comum F. A v.a. X ou F pertence ao domínio de atração de uma v.a α-estável, $Z_\alpha \stackrel{d}{=} G_\alpha$ $(F \in D(G_\alpha))$, se existem constantes $b_n \in \mathbf{R}$ tais que

$$\frac{S_n - b_n}{n^{1/\alpha} L(n)} \xrightarrow{d} Z_\alpha,$$

onde "$\xrightarrow{d}$"denota a convergência em distribuição, $S_n = X_1 + X_2 + \ldots + X_n$,

$$b_n = \begin{cases} \mu, & \text{se} \quad 1 < \alpha \leq 2 \\ 0, & \text{se} \quad 0 < \alpha < 1 \\ 0, & \text{se} \quad \alpha = 1 \;\text{ e }\; \text{F simétrica} \end{cases} \tag{3}$$

e L é uma função lentamente variante, ou seja, $L(tx) \sim L(t)$. Usamos a notação $g(y) \sim h(y)$ para indicar $\lim_{y \to \infty} \dfrac{g(y)}{h(y)} = 1$.

Quando $0 < \alpha < 2$, uma caracterização importante de domínio de atração para distribuição α- estável, que pode ser encontrada em [3] ou [1], é a seguinte:

$$F \in D(G_\alpha) \quad \Leftrightarrow \quad F(-x) \sim C^I x^{-\alpha} L(x) \;\; e \;\; \bar{F}(x) = 1 - F(x) \sim C^D x^{-\alpha} L(x), \tag{4}$$

onde $C^I \geq 0$ e $C^D \geq 0$ são tais que $C^I + C^D > 0$ e L é uma função lentamente variante.

Neste trabalho, estenderemos o resultado acima para o caso em que a distribuição original das variáveis básicas é uma mistura finita de k componentes

$$F(x) = \sum_{i=1}^{k} p_i F_i(x), \quad \text{com} \quad p_i \geq 0 \quad \text{e} \quad \sum_{i=1}^{k} p_i = 1. \tag{5}$$

Nesta mistura, a i-ésima componente (função de distribuição da i-ésima sub população) $F_i \in D(G_{\alpha_i})$.

2. Domínio de atração de distribuições α-estáveis baseada em amostras aleatórias de populações mistas

Nesta seção, apresentamos condições necessárias para uma função de distribuição F de (5) estar no domínio de atração de uma distribuição α- estável, usando uma amostra aleatória de uma população mista de k componentes, onde cada componente F_i pertence ao domínio de atração de uma distribuição α- estável. Em outras palavras, o Teorema apresenta condições necessárias para determinar a distribuição limite da soma normalizada de n v.a's i.i.d., onde a distribuição comum é uma mistura de k componentes cujas caudas decrescem conforme a potência $x^{-\alpha_i}$, $i = 1, \ldots, k$.

Para isso, observemos inicialmente que por (4), se $F_i \in D(G_{\alpha_i})$ para $i = 1, \ldots, k$, então

$$F_i(-x) \sim C_i^I x^{-\alpha_i} L_i(x) \quad \text{e} \quad \bar{F}_i(x) = 1 - F_i(x) \sim C_i^D x^{-\alpha_i} L_i(x), \tag{6}$$

onde $C_i^I \geq 0$ e $C_i^D \geq 0$ são tais que $C_i^I + C_i^D > 0$ para $i = 1, ..., k$, e as $L_i's$ são funções lentamente variante.

Teorema. Sejam $X_1, X_2, \ldots, X_n$ v.a's i.i.d. com a função de distribuição comum F dada por (5). Se $F_i \in D(G_{\alpha_i})$ para $i = 1, ..., k$, então
(a) $F \in D(G_{\alpha_m})$ com $\alpha_m = \min\{\alpha_1, \ldots, \alpha_k\}$, se $\alpha_i \neq \alpha_j$ para todo $i \neq j$;
(b) $F \in D(G_\alpha)$ com as constantes em (6) satisfazendo $\frac{C_1^I}{C_1^D} = \frac{C_2^I}{C_2^D} = \ldots = \frac{C_n^I}{C_n^D}$, se $\alpha = \alpha_1 = \ldots = \alpha_k$.
Prova.
(a) De (5) temos que

$$F(-x) = \sum_{i=1}^{k} p_i F_i(-x), \qquad \bar{F}(x) = \sum_{i=1}^{k} p_i \bar{F}_i(x). \tag{7}$$

Aplicando (6) em (7), obtemos

$$F(-x) \sim \sum_{i=1}^{k} p_i C_i^I x^{-\alpha_i} L_i(x) \quad e \quad \bar{F}(x) \sim \sum_{i=1}^{k} p_i C_i^D x^{-\alpha_i} L_i(x). \tag{8}$$

Como $\alpha_i \neq \alpha_j$ para todo $i \neq j$, seja $\alpha_m = \min\{\alpha_1, \ldots, \alpha_k\}$. Então, como as $L_i's$ são lentamente variante, as expressões de (8) podem ser escritas como

$$\begin{aligned} F(-x) &\sim x^{-\alpha_m} \sum_{i=1}^{k} p_i C_i^I x^{\alpha_m - \alpha_i} L_i(x) \\ &\sim p_m C_m^I x^{-\alpha_m} L_m(x) \end{aligned}$$

e

$$\begin{aligned} \bar{F}(x) &\sim x^{-\alpha_m} \sum_{i=1}^{k} p_i C_i^D x^{\alpha_m - \alpha_i} L_i(x) \\ &\sim p_m C_m^D x^{-\alpha_m} L_m(x), \end{aligned}$$

pois $\lim_{x \to \infty} x^{\alpha_m - \alpha_i} = 0$ para todo $i \neq m$. Assim (4) é satisfeita para a componente com menor índice caudal, e como consequência provamos (a).

(b) Sendo $\alpha = \alpha_1 = \ldots = \alpha_k$, as expressões de (8) podem ser escritas como

$$F(-x) \sim x^{-\alpha} \sum_{i=1}^{k} p_i C_i^I L_i(x) \quad e \quad \bar{F}(x) \sim x^{-\alpha} \sum_{i=1}^{k} p_i C_i^D L_i(x). \tag{9}$$

Usando o fato que soma de funções lentamente variantes é uma função lentamente variante e a condição $\frac{C_1^I}{C_1^D} = \frac{C_2^I}{C_2^D} = \ldots = \frac{C_n^I}{C_n^D}$, temos que

$$F(-x) \sim \frac{A}{C_1^D C_2^D ... C_k^D} x^{-\alpha} \sum_{i=1}^{k} p_i C_i^D L_i(x) \quad e \quad \bar{F}(x) \sim x^{-\alpha} \sum_{i=1}^{k} p_i C_i^D L_i(x), \tag{10}$$

pois

$$\begin{aligned} \sum_{i=1}^{k} p_i C_i^I L_i(x) &= \frac{1}{C_1^D C_2^D ... C_k^D} [C_1^I C_2^D ... C_k^D (p_1 C_1^D L_1(x)) + \ldots + C_k^I C_1^D C_2^D ... C_{k-1}^D (p_k C_k^D L_k(x))] \\ &= \frac{A}{C_1^D C_2^D ... C_k^D} \sum_{i=1}^{k} p_i C_i^D L_i(x), \end{aligned}$$

onde

$$A = C_1^I C_2^D ... C_k^D = C_2^I C_1^D C_3^D ... C_k^D = \ldots = C_k^I C_1^D C_2^D ... C_{k-1}^D.$$

Tomando a função lentamente variante $L(x) = \sum_{i=1}^{k} p_i C_i^D L_i(x)$, $C^I = \frac{A}{C_1^D C_2^D \ldots C_k^D}$ e $C^D = 1$, em (10), a prova de (b) segue de (4).

Exemplo. Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma população mista cuja função de distribuição é $F(x) = pF_1(x) + (1-p)F_2(x)$, onde $0 < p < 1$ e F_1, F_2 são as distribuições estáveis

$$F_1 = S_{\alpha_1}(\sigma_1, \beta_1, \mu_1) \;\; e \;\; F_2 = S_{\alpha_2}(\sigma_2, \beta_2, \mu_2),$$

com $0 < \alpha_1, \alpha_1 \leq 2$. Temos que

$$F_1 \in D(\alpha_1) \;\; e \;\; F_2 \in D(\alpha_2),$$

pois, conforme [10], [8] entre outros, as caudas das distribuições satisfazem

$$F_i(-x) \sim C_i^I x^{-\alpha_i} \quad \text{e} \quad \bar{F}_i(x) \sim C_i^D x^{-\alpha_i}, \quad i = 1, 2,$$

onde

$$C_i^I = \frac{(1-\alpha_i)(1-\beta_i)\sigma_i^{\alpha_i}}{2\Gamma(2-\alpha_i)\cos(\frac{\pi\alpha_i}{2})}, \qquad C_i^D = \frac{(1-\alpha_i)(1+\beta_i)\sigma_i^{\alpha_i}}{2\Gamma(2-\alpha_i)\cos(\frac{\pi\alpha_i}{2})} \quad \text{para} \quad \alpha_i \neq 1$$

e

$$C_i^I = \frac{(1-\beta_i)\sigma_i^{\alpha_i}}{\pi}, \qquad C_i^D = \frac{(1+\beta_i)\sigma_i^{\alpha_i}}{\pi} \quad \text{para} \quad \alpha_i = 1.$$

Se $\alpha_1 < \alpha_2$ então $F(-x) \sim pC_1^I x^{-\alpha_1}$ e $\bar{F}(x) \sim pC_1^D x^{-\alpha_1}$. Daí $F_1 \in D(\alpha_1)$, o que ilustra (a) do Teorema.

Se $\alpha = \alpha_1 = \alpha_2$ e $\beta_1\beta_2 + 2(\beta_1\beta_2) = 0$, temos que

$$F(-x) \sim \frac{1-\beta_1}{1+\beta_1}[pC_1^I + (1-p)C_2^I]x^{-\alpha} \quad \text{e} \quad \bar{F}(x) \sim [pC_1^I + (1-p)C_2^I]x^{-\alpha}.$$

Então $F \in D(\alpha)$, o que ilustra (b) do Teorema.

3. Conclusão. As caudas da distribuição assintótica da soma normalizada de v.a's i.i.d., cuja distribuição comum é uma mistura de k componentes com caudas decrescendo conforme a potência $x^{-\alpha_i}$, $i = 1 \ldots, n$, decrescem com a potência do menor índice caudal das k componentes, quando os índices caudais são todos distintos. No caso dos índices caudais das componentes serem iguais,uma condição adicional entre o peso das caudas das componentes é necessária para a distribuição assintótica da soma normalizada decrescer com a mesma potência das componentes.

4. Agradecimentos

Agradecemos à CAPES - PROCAD e FINATEC - UnB, pelo financiamento.

Referências

[1] P. Embrechts, T. Mikosch, C. Kluppelberg, Modelling Extremal Events for Insurance and Finance. Springer - Verlag, berlin, 1997.

[2] B.S. Everitt, D.J. Hand, Finite Mixture Distributions. Chapman and Hall, London, 1981.

[3] W. Feller, An Introduction to Probability Theory and Its Applications II. 3rd edition, Wiley, New York, 1968.

[4] I.A. Ibragimov, Yu. V. Linnik, Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen, 74-85, 1971.

[5] B. G. Lindsay, Mixture Models : Theory, Geometry and Aplications. The Institute of Mathematic and Statistics, Hayward, CA, 1995.

[6] G.J. McLachlan, K.E.Basford, Mixture Models: Aplications to Clustering. Marcel dekker, New York, 1988.

[7] N. Ravishanker, D.K. Dey, Multivariate Survival Models with a Mixture of Positive Stable Frailties, *Journal Methodology and Computing in Applied Probability*, 2 (2004) 293-308.

[8] G. Samorodnitsky, M. Taqqu, Stable Non-Gaussian Random Process, Chapman and Hall, New York, 1994.

[9] D.M. Titterington, A.F.M. Smith, U.E. Makov, Statistical Analysisof Finite Mxture Distributions.wiley, New York, 1985.

[10] V. M.Zolotarev, One Dimensional Stable Distributions. American Mathematical Society, Providence, 1986.