



## APRESENTAÇÃO

### A distribuição Birnbaum-Saunders baseada na distribuição Laplace

*Aluno:* Vicenzo Bonasorte Reis Pereira RA225071, e-mail: v225071@dac.unicamp.br

*Orientador:* Filidor Vilca Labra

Depto. de Estatística IMECC-UNICAMP

## 1 Introdução

A distribuição Birnbaum-Saunders, ou BS, (Birnbaum e Saunders, 1969) surgiu no contexto da engenharia, utilizada para descrever processos de degradação. Atualmente, no entanto, por conta desta propriedade, vem sendo utilizada nas mais variadas áreas de aplicação, sendo útil para saúde, passando por economia e até mesmo na previsão de desastres naturais. Nossa proposta é uma extensão da distribuição BS, que é baseada na distribuição assimétrica de Laplace.

### 1.1 A Distribuição BS

Uma variável aleatória positiva  $T$ , que representa o tempo total até que ocorra a falha, é dito ter uma distribuição BS se pode ser representado em termos da distribuição normal padrão por:

$$T = \frac{\beta}{4} \left( \alpha Z + \sqrt{(\alpha Z)^2 + 4} \right), \quad \alpha > 0, \beta > 0, \quad (1)$$

onde  $Z \sim N(0, 1)$ . Esta distribuição é denotada por:  $T \sim BS(\alpha, \beta)$ . A função densidade de probabilidade (fdp) de  $T$ , é dada por:

$$f_T(t) = \phi(a_t(\alpha, \beta)) A_t(\alpha, \beta), \quad t > 0, \quad (2)$$

em que  $\phi(\cdot)$  é a fdp de uma  $N(0, 1)$ ,

$$a_t(\alpha, \beta) = \frac{1}{\alpha} (\sqrt{t/\beta} - \sqrt{\beta/t}) \quad \text{e} \quad A_t(\alpha, \beta) = \frac{d}{dt} a_t(\alpha, \beta). \quad (3)$$

## 1.2 A Distribuição Laplace

Segundo Yu e Moyeed (2001), uma v.a.  $Y$  segue uma distribuição Laplace assimétrica (ALD), com parametro de locação  $\mu$ , parâmetro de escala  $\sigma > 0$  e parâmetro de assimetria  $p \in (0, 1)$ , se sua fdp é dada por:

$$f(y|\mu, \sigma, p) = \frac{p(1-p)}{\sigma} \exp \left\{ -\rho_p \left( \frac{y-\mu}{\sigma} \right) \right\}, y \in R, \quad (4)$$

em que  $\rho_p(\cdot)$  é chamada de função perda definida por  $\rho_p(u) = u(p - \mathbb{I}\{u < 0\})$ .  $Y$  pode ser representada estocasticamente por:

$$Y = \mu + \vartheta_p U + \tau_p \sqrt{\sigma U} Z,$$

em que  $U \sim \exp(\sigma)$ ,  $Z \sim N(0, 1)$ , independentes,  $\vartheta_p = \frac{1-2p}{p(1-p)}$  e  $\tau_p^2 = \frac{2}{p(1-p)}$ .

Esta distribuição será denotada por  $Y \sim ALD(\mu, \sigma, p)$ . Se  $\mu = 0$  e  $\sigma = 1$ , dizemos que  $Y$  segue uma distribuição ALD padrão. Uma das principais aplicações desta distribuição é em regressão quantílica, como pode ser visto em Benites et al. (2015).

## 2 A distribuição BS baseada na distribuição ALD

A função densidade da distribuição BS baseada na distribuição ALD (LBS), é dada por:

$$f(t) = \phi_{ALD}(a_t(\alpha, \beta)) A_t(\alpha, \beta), t > 0,$$

em que  $\phi_{ALD}$  representa a função densidade de uma distribuição ALD padrão, definida em (4),  $a_t$  e  $A_t$  são como definidas em (3)

Para gerar dados  $T \sim LBS(\alpha, \beta, p)$ , os seguintes passos são utilizados:

- 1) Gerar  $Y \sim ALD(0, 1, p)$ ;
- 2) Computar  $T = \frac{\beta}{4}(\alpha Y + \sqrt{(\alpha Y)^2 + 4})^2$ .

A geração de dados é importante para verificar algumas propriedades do modelo LBS, como por exemplo estudos de simulação sobre a estimativa de máxima verossimilhança dos parâmetros.

**Teorema:** Seja  $T \sim LBS(\alpha, \beta, p)$ . Então:

- i) a função de distribuição acumulada, é da forma:

$$F(t) = \Phi_{ALD}(a_t(\alpha, \beta)),$$

em que  $\Phi_{ALD}$  é a função acumulada da distribuição ALD padrão e  $a_t$  é como definido em (3);

- ii)  $aT \sim LBS(\alpha, a\beta, p)$ , para todo  $a$  positivo;

- iii)  $\frac{1}{T} \sim LBS(\alpha, 1/\beta, 1-p)$ .

### 3 Estimação

Desenvolvemos uma aplicação do algoritmo EM para estimar os parâmetros da distribuição LBS. Esta parte é inspirada nos trabalhos de Yu e Zhang (2005) e Santana (2009).

O algoritmo EM (Dempster et al., 1977) é usado para calcular de forma iterativa estimadores de máxima verossimilhança, útil, principalmente, quando temos dados faltantes.

Seja  $T_1 \dots T_n$  uma amostra aleatória de  $T \sim LBS(\alpha, \beta, p)$ . Usando os teoremas anteriores, podemos obter uma representação hierárquica para a LBS:

$$\begin{aligned} T_i | U_i = u &\sim BSE(\alpha_{u_i,p}, \beta, \sigma = 2, \gamma_{u_i,p}) \\ U_i &\sim exp(1) \end{aligned}$$

Tratando U como dado faltante, podemos implementar o algoritmo EM.

Para  $\boldsymbol{\theta} = (\alpha, \beta, p)^T$ , a função log-verossimilhança para os dados completos é dada por:

$$\ell(\boldsymbol{\theta} | y_c) = \sum_{i=1}^n (\log f_{T|U}(t_i | u_i) + \log f_U(u_i))$$

No passo E, deve-se tomar a esperança da log-verossimilhança condicionada em T e  $\boldsymbol{\theta}^k$ , que denota os valores de  $\boldsymbol{\theta}$  no estágio k do algoritmo, obtendo a função Q.

$$Q(\boldsymbol{\theta} | \boldsymbol{\theta}^k) = \mathbb{E}[\ell(\boldsymbol{\theta} | y_c) | t, \boldsymbol{\theta}^k]$$

No Passo M do algoritmo, deve-se maximizar a função  $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^k)$ . Obtendo o seguinte estimador para  $\alpha$ :

$$\hat{\alpha} = \sqrt{\frac{\sum_{i=1}^n \Delta_{3i} a_{t_i}^2(1, \hat{\beta})}{4((1 - 2\hat{p})A + 2n)}}$$

em que  $A = -\frac{n(1-2\hat{p})}{2\hat{p}(1-\hat{p})} + \frac{1-2\hat{p}}{4\hat{p}^2(1-\hat{p})^2} \sum_{i=1}^n \mathbb{E}[u_i | \boldsymbol{\theta}, t_i]$ .

Para os parâmetros  $p$  e  $\beta$  não temos estimadores com forma analítica fechada, e as estimativas são obtidas, respectivamente, usando:

$$\hat{p} = \arg \max_p \{Q(\hat{\alpha}, \hat{\beta}, p | \boldsymbol{\theta}^k)\} \text{ e } \hat{\beta} = \arg \max_{\beta} \{Q(\hat{\alpha}, \beta, \hat{p} | \boldsymbol{\theta}^k)\}.$$

## 4 Análise de Dados Reais

Por fim, foi feita a análise de um conjunto de dados do Departamento de Minas da universidade do Atacama, no Chile, que mostra medidas de Neodímio (Nd) de 86 amostras de minerais, estes dados já foram explorados em Reyes et al. (2016). O Neodímio é o elemento químico de número atômico 60, não ocorre na natureza como elemento livre, apenas em minérios como monazita, constitui 18% do grupo terras raras de metais com propriedades físico-químicas semelhantes e que ocorrem nos mesmos minérios. É usado para colorir vidros, criando cristais usados por astrônomos para calibrar espectrômetros, como corante de esmaltes e em ímãs poderosos e baratos, comuns em fones de ouvido e discos rígidos de computadores, sendo, assim, de extrema importância encontrar um bom ajuste para modelar quantidade de neodímio em rochas. O objetivo é ajustar a estes dados os modelos Birnbaum-Saunders (BS), Birnbaum-Saunders baseada na Laplace (LBS) e Birnbaum-Saunders baseada na Skew-Normal (BSSN) e compará-los, numa tentativa de justificar a utilidade e eficiência da distribuição apresentada neste trabalho. Nas figuras a seguir vemos comparações entre os ajustes.

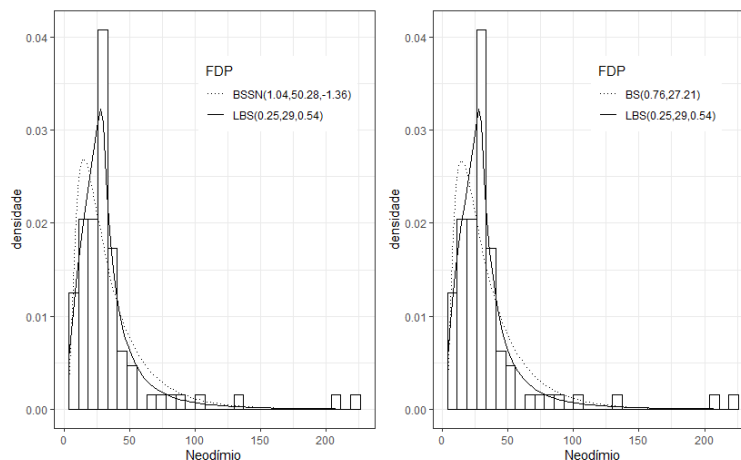


Figura 1: Histograma dos dados com densidades sobrepostas.

## 5 Conclusão

Neste trabalho apresentamos a distribuição BS baseada na distribuição Laplace. Esta distribuição tem propriedades mais flexíveis do que a distribuição BS usual. Para  $0 < p < \frac{1}{2}$  temos uma distribuição assimétrica à direita, e para  $\frac{1}{2} < p < 1$  temos uma distribuição assimétrica à esquerda. Também notamos propriedades leptocúrticas, como pode ser visto na aplicação. A implementação do algoritmo EM resultou ser complexa. Para isso, contamos com a colaboração de um aluno do mestrado em estatística.

## 6 Bibliografia

Birnbaum, Z. W.; Saunders, S. C. (1969) A new family of life distributions. *J. Appl. Prob.*6:637-652.

Yu, K.; Moyeed, R. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54,437-447.

Benites, L. B.; Lachos, V. H. ; Vilca, L. F. (2015). Case-deletion diagnostics for Quantile regression using the asymmetric Laplace distribution. arXiv preprint arXiv, 1509, 05099.

Yu, K.; Zhang, J. (2005). A Three-Parameter Asymmetric Laplace Distribution and Its Extension. *Communications in Statistics—Theory and Methods*, 34:9-10, 1867-1879.

Santana, L. R. (2009). *Estimação e Diagnóstico em Modelos Birnbaum-Saunders Skew-Normal*. Dissertação (Mestrado em Estatística) - Instituto de Matemática, Estatística e Computação Científica, Universidade de Campinas.

Dempster A. P.; Laird, N. M.; Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B*, Vol. 39, No. 1, pp. 1-38.

Leiva, V.; Vilca, F.; Balakrishnan, N.; Sanhueza, A. (2010). A skewed sinh-normal distribution and its properties and application to air pollution. *Comm. Statist. Theor. Meth.* 39, 426–443.

Embrechts, P. (1983). A Property of the Generalized Inverse Gaussian Distribution with Some Applications. *Journal of Applied Probability*, Vol. 20, No. 3, pp. 537-544.

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, 40, pp. 237-260.

Reyes, J.; Vilca, F.; Gallardo, D. I.; Gomez, H. W. (2016). “Modified slash Birnbaum-Saunders distribution,” *Hacettepe Journal of Mathematics and Statistics*, vol. 46, no. 112, pp. 1–1.