

DISTRIBUIÇÃO NORMAL GENERALIZADA: PROPRIEDADES, ESTIMAÇÃO E APLICAÇÕES

Fabrcio R. de Carvalho Jnior^{1,2}; Francisco A. M. de Souza¹; Manoel Santos-Neto¹

¹Unidade Acadmica de Estatstica, Universidade Federal de Campina Grande, mn.neco@gmail.com

²Curso de Administrao, Faculdade Maurcio de Nassau, fabrcio.ramalhocj@gmail.com

INTRODUO

A distribuio normal generalizada, tambm conhecida como distribuio exponencial potncia, teve sua origem na dcada de 60 com o trabalho de Box e Tiao (1962). Esse modelo probabilstico tambm conhecido como distribuio de Subottin (SUBOTTIN, 1923). Existem na literatura diversos trabalhos sobre esse modelo e que podem auxiliar no desenvolvimento de novos trabalhos. Alguns destes trabalhos so: Turner (1960), Tiao e Lund (1970), Tadikamalla (1980), Young et al. (1990), Taylor (1992), Kuwana e Kariya (1991), Agro (1995), Achcar e Pereira (1999), Basu et al. (2001), Kotz et al. (2001), Kozubowski (2002), Sanchez-Manzano et al. (2002), Monti (2003), Choy e Walker (2003), Domnguez-Molina et al. (2003), Hazan et al. (2003), DiCiccio e Monti (2004), Tojeiro et al. (2004), Desgagne e Angers (2005), Nadarajah (2005), Pogány e Nadarajah (2010), Salazar et al., (2012), Cintra et al. (2014). Uma referncia sobre o modelo normal generalizado (ou exponencial potncia) tambm pode ser encontrada em Johnson e Kotz (1994).

A escolha pelo modelo normal generalizado para servir como base de estudo deste trabalho, se deve ao fato do mesmo possuir aplicaes em diferentes reas. Esta alta aplicabilidade pode ser justificada pelo fato deste modelo ser uma generalizao das distribuies normal e de Laplace. Algumas das reas, em que o modelo normal generalizado pode ser aplicado so: processamento de sinais, cincia de materiais, finanas quantitativas, medicina, automao e controle remoto, telecomunicao, pesquisa geofísica, inteligncia artificial, cincia e tecnologia de imagem fotogrfica, sistemas de informao, acstica, engenharia de software, ptica, hardware e arquitetura, sensoriamento remoto, engenharia biomdica, radiologia, medicina nuclear e imagiologia mdica, cincia e tecnologia de transportes, engenharia de produo, instrumentos e instrumentao, fsica, qumica analtica, ciberntica, cristalografia, energia e combustveis, engenharia ambiental, biologia computacional, nanociencia e nanotecnologia, neurociencia, pesquisa de operaes e gesto de cincia, confiabilidade e risco e isolamento eltrico e dieltricos (POGÁNY e NADARAJAH, 2010).

Neste trabalho caracterizamos a distribuio normal generalizada. Alm disso, realizamos um estudo das propriedades, via simulaes de Monte Carlo, dos estimadores de mxima verossimilhana dos parmetros que indexam a distribuio normal generalizada. Por fim, utilizamos um conjunto de dados real para mostrar a aplicabilidade deste modelo.

METODOLOGIA

Distribuição Normal Generalizada

A distribuição normal generalizada (NG) é indexada por três parâmetros, μ (de localização), σ (de dispersão), e ϕ (de forma). A função de densidade de probabilidade (f.d.p.) é dada por

$$f(x; \mu, \sigma, \phi) = \frac{\phi}{2 \sigma \Gamma(\frac{1}{\phi})} \exp \left\{ - \left| \frac{x - \mu}{\sigma} \right|^\phi \right\}, \quad (1)$$

em que $-\infty < x, \mu < \infty$ e $0 < \sigma, \phi < \infty$, $\Gamma(\cdot)$ denota a função gama. Usamos a notação $X \sim \text{NG}(\mu, \sigma, \phi)$ para representar uma variável aleatória com distribuição normal generalizada. Na Figura 1 apresentamos diferentes gráficos da f.d.p. sob diferentes combinações dos parâmetros. Podemos observar que o parâmetro ϕ controla a forma da distribuição, μ é um parâmetro de localização e σ controla a variabilidade da distribuição.

Segundo Nadarajah (2005) a distribuição normal generalizada possui alguns casos especiais. Quando $\phi = 1$ temos a distribuição de Laplace e para $\phi = 2$ temos a distribuição normal. Além disso, a distribuição normal generalizada tem forma de sino e é unimodal.

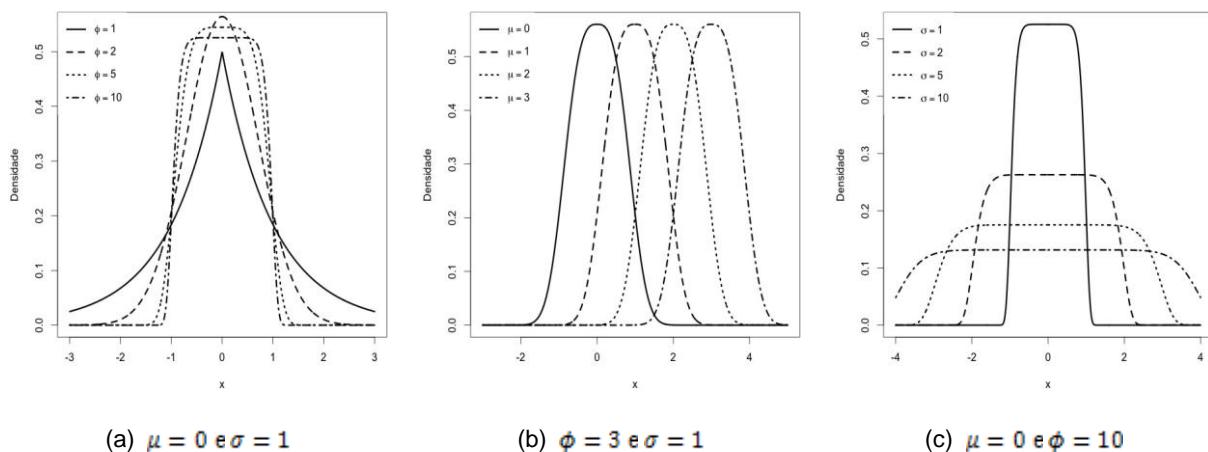


Figura 1 - Função densidade de probabilidade da distribuição normal generalizada para diferentes combinações dos parâmetros μ , σ e ϕ .

O método de estimação por Máxima Verossimilhança

O logaritmo da função de verossimilhança para uma amostra X_1, X_2, \dots, X_n proveniente de uma distribuição NG, com densidade definida em (1), é dado por

$$\ell(\theta; \mathbf{x}) = n \log \left(\frac{\phi}{2 \sigma \Gamma(1/\phi)} \right) - \sum_{i=1}^n \left| \frac{x_i - \mu}{\sigma} \right|^\phi.$$

As derivadas do logaritmo da função de verossimilhança em relação a μ , σ e ϕ , são respectivamente,

$$\begin{aligned}\frac{\partial \ell(\theta; \mathbf{x})}{\partial \mu} &= \frac{\phi}{\sigma^\phi} \left\{ \sum_{x_i \geq \mu} (x_i - \mu)^{\phi-1} - \sum_{x_i < \mu} (x_i - \mu)^{\phi-1} \right\}, \\ \frac{\partial \ell(\theta; \mathbf{x})}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{\phi}{\sigma^{\phi+1}} \sum_{i=1}^n |x_i - \mu|^\phi, \text{ e} \\ \frac{\partial \ell(\theta; \mathbf{x})}{\partial \phi} &= \frac{n}{\sigma} \left\{ \frac{1}{\phi} \Psi\left(\frac{1}{\phi}\right) + 1 \right\} - \sum_{i=1}^n \left| \frac{x_i - \mu}{\sigma} \right| \log \left| \frac{x_i - \mu}{\sigma} \right|,\end{aligned}$$

em que $\Psi(x) = d \log(\Gamma(x))/dx$ é a função psi de Euler.

O Método dos Momentos

Neste trabalho consideramos uma variação dos estimadores de momentos conhecida na literatura como momentos modificados (SANTOS-NETO, et al., 2014). No caso da distribuição normal generalizada esse método resultará nas seguintes equações

$$\begin{cases} \mu = \bar{X}; \\ \frac{\sigma^2 \Gamma\left(\frac{3}{\phi}\right)}{\Gamma\left(\frac{1}{\phi}\right)} = \frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2; \\ \frac{\Gamma\left(\frac{1}{\phi}\right) \Gamma\left(\frac{5}{\phi}\right)}{(\Gamma(3/\phi))^2} = \frac{n \sum_{i=1}^n (X_i - \bar{X})^4}{(\sum_{i=1}^n (X_i - \bar{X})^2)^2}. \end{cases}$$

Daí temos que $\tilde{\mu} = \bar{X}$, $\tilde{\sigma} = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2 \times \left(\frac{\Gamma\left(\frac{1}{\phi}\right)}{\Gamma\left(\frac{3}{\phi}\right)}\right)}$ e $\tilde{\phi}$ é a solução de $\frac{\Gamma\left(\frac{1}{\phi}\right) \Gamma\left(\frac{5}{\phi}\right)}{(\Gamma(3/\phi))^2} - \frac{n \sum_{i=1}^n (X_i - \bar{X})^4}{(\sum_{i=1}^n (X_i - \bar{X})^2)^2} = 0$. Os estimadores de momentos modificados foram utilizados como chutes iniciais no processo de estimação através do método de máxima verossimilhança.

RESULTADOS E DISCUSSÃO

Simulações

Todas as simulações foram realizadas utilizando o *software* R. Para a geração de número pseudo-aleatórios utilizamos o comando `rPE2()` presente no pacote `gamlss`. Para maximizar o logaritmo da função de verossimilhança utilizamos a função `maxLik()` que faz parte do pacote de mesmo nome

Realizamos as simulações considerando 10 000 réplicas de Monte Carlo. Fixamos $\sigma = 1$ e $\mu = 2$ e consideramos os seguintes valores para ϕ : 1.5, 2.5 e 5.0. Quanto aos tamanhos das amostras, consideramos os seguintes valores: 30, 40, 50, 60, 70, 80, 90, 100, 110 e 120. Para estudarmos o comportamento dos estimadores de máxima verossimilhança calculamos o viés normalizado (VNorm) e a raiz quadrada do erro quadrático médio normalizada (REQMNorm). Essas quantidades foram obtidas da seguinte maneira

$$VNorm_i = \frac{Vies_i}{\max\{Vies_{30}, \dots, Vies_{120}\}},$$

e

$$REQMNorm_i = \frac{REQM_i}{\max\{REQM_{30}, \dots, REQM_{120}\}}, \quad i \in \{30, 40, 50, 60, 70, 80, 90, 100, 110, 120\},$$

em que $Vies_i = [E(\hat{\theta}_i) - \theta]$ (viés) e $REQM_i = \sqrt{Var(\hat{\theta}_i) + Vies_i^2}$ (raiz do erro quadrático médio), com $\hat{\theta}_i$ sendo o estimador de máxima verossimilhança de θ para o i -ésimo tamanho amostral.

Na Figura 4 notamos que os vieses dos estimadores estudados tendem, de uma maneira geral, a diminuir com o aumento do tamanho da amostra. Também notamos que o estimador $\hat{\phi}$ mostrou-se mais viesado do que os demais estimadores. Além disso, notamos que para $\phi = 1.5$ os vieses de $\hat{\sigma}$ apresentaram valores relativamente superiores aos valores observados para os valores de $\phi = 2.5$ e 5.0 , ver Figura 4(b).

Na Figura 5, notamos que a REQM dos estimadores estudados diminui com o aumento do tamanho da amostra. Notamos também que o estimador $\hat{\phi}$ apresentou uma REQM maior do que os outros estimadores, ver Figura 5(c). Por fim, notamos que o aumento do valor de ϕ faz com que tenhamos uma redução nos valores do REQM de $\hat{\sigma}$, ver Figura 5(b).

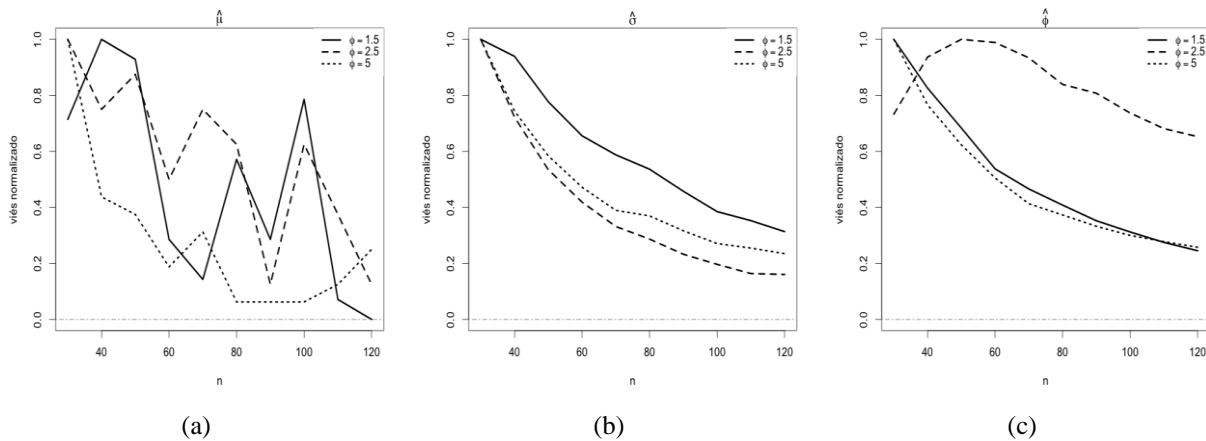


Figura 4 - Viés normalizado dos estimadores de máxima verossimilhança dos parâmetros μ (a), σ (b) e ϕ (c).

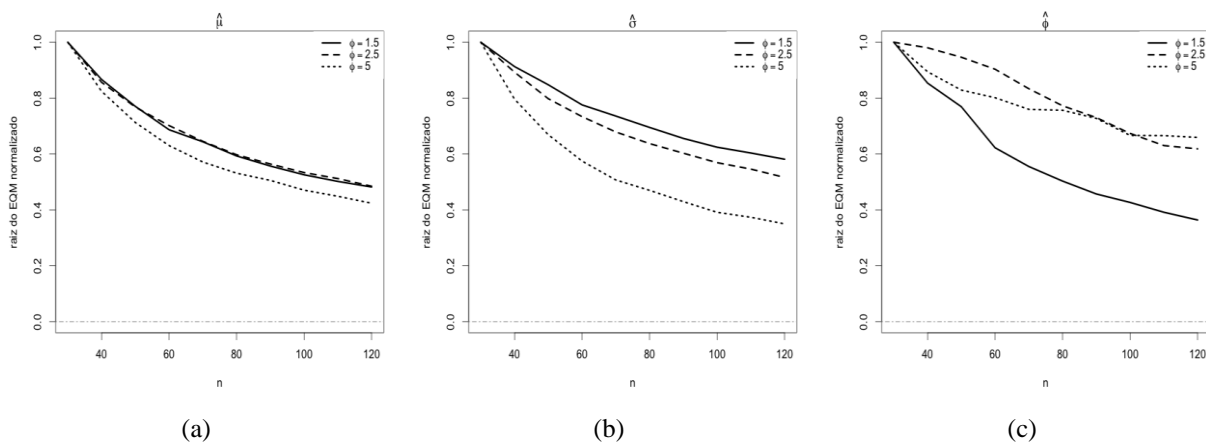


Figura 5 - REQM normalizado dos estimadores de máxima verossimilhança dos parâmetros μ (a), σ (b) e ϕ (c).

Aplicação

Nesta seção apresentamos uma aplicação da distribuição normal generalizada a um conjunto de dados reais. O conjunto de dados refere-se ao comprimento das asas de 86 moscas domésticas. Para

maiores detalhes sobre este conjunto de dados ver, Sokal e Hunter (1955). Além disso, os dados podem ser obtidos em <http://www.seattlecentral.edu/qelp/sets/057/s057web.txt>. Iniciamos nossa aplicação com uma análise descritiva dos dados.

Na Tabela 2 apresentamos as estimativas de máxima verossimilhança dos parâmetros que indexam a distribuição NG. Além disso, apresentamos os erros-padrão dos estimadores de máxima verossimilhança e também as estimativas intervalares dos parâmetros μ , σ e ϕ .

Tabela 2 - Estimativas de máxima verossimilhança, erros-padrão e intervalos de confiança assintóticos.

Parâmetro	Estimativa	Erro-Padrão	Intervalo de Confiança (95%)
μ	45.5000	0.3861	[44,7432; 46,2568]
σ	5.9447	0.7164	[4,5406; 7,3488]
ϕ	2.3657	0.5934	[1,4470; 3,8678]

Nas Figuras 6 (a)-(b), podemos notar que a distribuição normal generalizada apresenta um bom ajuste ao conjunto de dados estudado.

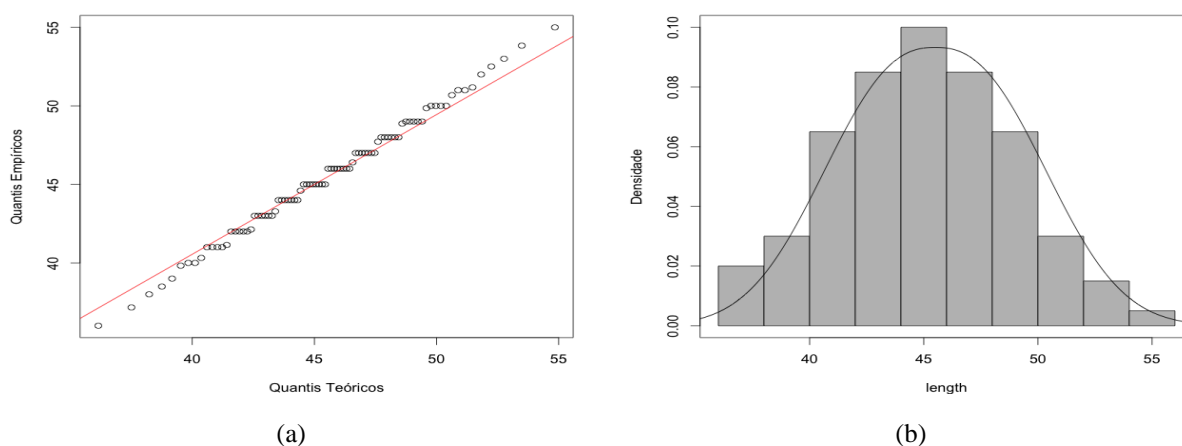


Figura 6 - QQplot (a) e histograma com a densidade estimada (b).

CONCLUSÕES

Neste trabalho estudamos e apresentamos as demonstrações de algumas propriedades da distribuição normal generalizada. Além disso, realizamos simulações de Monte Carlo para estudarmos o comportamento dos estimadores de máxima verossimilhança dos parâmetros desta distribuição. Conseguimos observar, por exemplo, que o estimador de máxima verossimilhança de ϕ , $\hat{\phi}$, é mais viesado do que os estimadores dos demais parâmetros (μ e σ). Por fim, apresentamos uma aplicação usando dados reais onde mostramos a aplicabilidade do modelo.

REFERÊNCIAS BIBLIOGRÁFICAS

ACHCAR, J. A.; PEREIRA, G. A. Use of Exponential Power Distributions for Mixture Models in the

- Presence of Covariates. **Journal of Applied Statistics**, v. 26, n. 6, p. 669–679, 1999.
- AGRO, G. Maximum likelihood estimation for the exponential power function parameters, **Communication in Statistics Simulation and Computation**, v. 24, n. 2, p. 523–536, 1995.
- BASU, S.; MICCHELLI, C. A.; OLSEN, P. Power exponential densities for the training and classification of acoustic feature vectors in speech recognition, **Journal of Computational Graphical Statistics**, v. 10, n. 1, p. 158–184, 2001.
- BOX, G. E. P.; TIAO, G. C. A further look at robustness via Bayes's theorem. **Biometrika**, v. 49, n. 3/4, p. 419–432, 1962.
- CINTRA, R. J.; RÊGO, L. C.; CORDEIRO, G. M.; NASCIMENTO, A. D. C. Beta generalized normal distribution with an application for SAR image processing, **Statistics**, v. 48, n. 2, p. 279–294, 2014.
- CHOY, S. T. B.; WALKER, S. G. The extended exponential power distribution and Bayesian robustness, **Statistics and Probability Letters**, v. 65, n. 3, p. 227–232, 2003.
- DESGAGNE, A.; ANGERS, J. F. Importance sampling with the generalized exponential power density, **Statistics and Computing**, v. 15, n. 3, p. 189–196, 2005.
- DICICCIO, T. J.; MONTI, A. C. Inferential aspects of the skew exponential power distribution, **Journal of the American Statistical Association**, v. 99, n. 466, p. 439–450, 2004.
- DOMÍNGUEZ-MOLINA, J. A.; GONZÁLEZ-FARIÁS, G.; RODRÍGUEZ-DAGNINO, R. M. A practical procedure to estimate the shape parameter in the generalized Gaussian distribution, p. 1–37, 2003. URL: <http://goo.gl/nPN677>.
- HAZAN, A.; LANDSMAN, Z.; MAKOV, U. E. Robustness via a mixture of exponential power distributions, **Computational Statistics and Data Analysis**, v. 42, n. 1/2, p. 111–121, 2003.
- JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Univariate Distributions**. 2ª ed. New York: Wiley, vol I, 1994.
- KOTZ, S., KOZUBOWSKI, T. J.; PODGORSKI, K. **The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering and Finance**. Boston, MA: Birkhauser Boston, 2001.
- KOZUBOWSKI, T. J. On the vertical density of the multivariate exponential power distribution, **Statistics**, v. 36, n. 3, p. 219–221, 2002.
- KUWANA, Y.; KARIYA, T. LBI tests for multivariate normality in exponential power distributions, **Journal of Multivariate Analysis**, v. 39, n. 1, p. 117–134, 1991.
- MONTI, A. C. A note on estimation of the skew normal and the skew exponential power distributions, **Metron**, v. 61, n. 2, p. 205–219, 2003.
- NADARAJAH, S. A Generalized Normal Distribution, **Journal of Applied Statistics**, v. 32, n. 7, p. 685–694, 2005.
- POGÁNY, T. K.; NADARAJAH, S. On the characteristic function of the generalized normal distribution, **C. R. Acad. Sci. Paris, Ser. I**, v. 348, n. 3/4, p. 203–206, 2010.
- SALAZAR, E.; FERREIRA, M. A. R.; MIGON, H. S. Objective Bayesian analysis for exponential power regression models, **Sankhya B**, v. 74, n. 1, p. 107–125, 2012.
- SANCHEZ-MANZANO, E. G.; GOMEZ-VILLEGAS, M. A.; MARIN-DIAZARAQUE, J. M. A matrix variate generalization of the power exponential family of distributions, **Communications in Statistics, Theory and Methods**, v. 31, n. 12, p. 2167–2182, 2002.
- SANTOS-NETO, M.; CYSNEIROS, F. J. A.; LEIVA, V.; BARROS, M. On a reparameterized Birnbaum-Saunders distribution and its moments, estimation and application. **Revstat Statistical Journal**, v. 12, n. 3, p. 247–272, 2014.
- SOKAL, R. R.; HUNTER, P. E. A morphometric analysis of DDT-resistant and non-resistant housefly strains **Ann. Entomol. Soc. Amer.** v. 48, n. 6, p. 499-507, 1955.
- SUBOTTIN, M. T. On the law of frequency of error, **Mathematicheskii Sbornik**, v. 31, n. 2, p. 296–301, 1923.
- TADIKAMALLA, P. R. Random sampling from the exponential power distribution, **Journal of the American Statistical Association**, v. 75, n. 371, p. 683–686, 1980.
- TAYLOR, J. M. G. Properties of modelling the error distribution with an extra shape parameter, **Computational Statistics and Data Analysis**, v. 13, n. 1, p. 33–46, 1992.
- TIAO, G. C.; LUND, D. R. The use of OLUMV estimators in inference robustness studies of the location

parameter of a class of symmetric distributions, **Journal of the America Statistical Association**, v. 65, n. 329, p. 370–386, 1970.

TOJEIRO, C. A. V.; LOUZADA-NETO, F.; BOLFARINE, H. A Bayesian analysis for accelerated lifetime tests under an exponential power law model with threshold stress, **Journal of Applied Statistics**, v. 31, n. 6, p. 685–691, 2004.

TURNER, M. E. On heuristic estimation methods, **Biometrika**, v. 16, n. 2, p. 299–301, 1960.

YOUNG, D. M.; SEAMAN, J. W.; TURNER, D. W. A ratio-of-uniforms method for generating exponential power variates, **Journal of Statistical Computation and Simulation**, v. 35, n. 1/2, p. 13–18, 199